Archiving South African digital research data

Diverse diets of Mio-Pliocene carnivorans of Langebaanweg

Australopithecus sediba

- landscape and diet
- simulation of cranial development
- osteogenic tumour

JULY/AUGUST 2016

eISSN: 1996-7489

Applying scientific thinking in the service of society

20 YEARS

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

SOUTH AFRICAN Journal of Science

volume 112
number 7/8

JULY/AUGUST

eISSN: 1996-7489

SOUTH AFRICAN
Journal of Science

volume 112

number 7/8

Roseanne Diab
Academy of Science of South Africa

Hassina Mouri
Department of Geology,
University of Johannesburg

Johann Mouton
Centre for Research on Science and Technology, Stellenbosch University

Maano Ramutsindela
Department of Environmental & Geographical Science, University of Cape Town

Published by
the Academy of Science of South Africa (www.assaf.org.za) with financial assistance from the Department of Science & Technology.

Design and layout
SUN MeDIA Bloemfontein
T: 051 444 2552
E: admin@sunbloem.co.za

Correspondence and enquiries
sajs@assaf.org.za

Submissions
Submissions should be made at http://mc.manuscriptcentral.com/sajs

Cover caption
The African bear (*Agriotherium*) – arguably the most impressive carnivoran to have ever inhabited the continent of Africa and the iconic species of Langebaanweg – defending its *Sivatherium* prey (image: Adam Hartstone-Rose). In an article on page 61, Hartstone-Rose and colleagues describe the diets of this extinct African bear and the other Mio-Pliocene carnivorans of Langebaanweg, South Africa.

Review Article

Research Article

Research Letter

More scientific thinking needed to feed society: The NSTF tackles hunger

Food shortages, hunger, starvation and serious problems with long-term food security are far from new issues in Africa, nor in the rest of the developing world. The most recent figures for Africa suggest that 20% of the population (some 233 million, mostly young, people) face dire hunger on a daily basis. This number also represents more than a quarter of the global figure for people who live with hunger bordering on starvation.

The geographical distribution of the crisis in sub-Saharan Africa is uneven, with those countries whose environmental conditions are least suitable for agriculture, and/or those which have been subject to prolonged periods of warfare or internal social conflict and poor governance, being the most severely at risk. In more recent years, the broad effects of global climate change and the prolonged effect of the current El Niño phenomenon have intensified the problem.

Ethiopia, South Sudan, Zimbabwe, Malawi, Angola and Mozambique are all at risk, while South Africa will, this year, possibly cull some 800 000 head of cattle, and import at least 300 000 tonnes of maize (and pay with a weakened rand). As a result, this 'old' topic has become increasingly important in current research and development. The South African Department of Science and Technology created, through the National Research Foundation, a Centre of Excellence and Chair in Food Security, shared by the Universities of Pretoria and the Western Cape in 2015. A recent article in *The Conversation* makes the following points:

> *Economic growth and wealth is necessary to make progress in reducing poverty and hunger, especially in the face of an expanding population. But governments need to do more than pursue economic growth. The key factor in ensuring food security is inclusive growth – growth that promotes access for everyone to food, assets and resources. Governments need to adopt an integrated approach to effectively reduce hunger, food insecurity and malnutrition in sub-Saharan Africa. They also need to implement a mix of complementary and comprehensive food security and nutrition policies and programmes. ... In South Africa, a national policy on food security and nutrition seeks to provide an overarching guiding framework for the different strategies and programmes of government and civil society.*

However, underlying such broad social and economic development, and government policies, is the need for substantial and reliable scientific knowledge and the means for its implementation. It is this essential set of conditions needed to address hunger and food security that South Africa's National Science and Technology Forum (NSTF) has addressed through a 'Discussion Forum on Pulses and Food Security', held in June this year (the UN's International Year of Pulses). Partners in this NSTF meeting included the Department of Science and Technology (the sponsors), the Department of Agriculture, Forestry and Fisheries, and AGT Foods /Advanced Seed, Africa.

The importance of the Forum lies not just in its theme and its public and private sector partners, but in the issues raised and recommendations made by those attending the event. These are also placed in the wider context of an ongoing parliamentary debate as to what constitutes an adequate social security payment to feed a South African family in an adequate manner.

The preliminary findings of the NSTF meeting make the critical point that, while food is a human necessity, food products are also commodities, which gives rise to 'competing interests' – those of the consumers, for adequate food is a basic human right, and those of the players along the food value chain who feel entitled to compensation for their role in growing agricultural food and moving it into the consumers' market place. This challenge is not, of course, unique to South Africa, but is part of the worldwide problem: the world produces more food than paying consumers really need, resulting in substantial waste, while hunger and starvation remain endemic, primarily in developing regions of the world.

In South Arica, the case is intensified by two conditions: a highly unequal distribution of land ownership and, more importantly when it comes to food production, the general suitability of agricultural land in the country for production. Of this land surface, 12% has high agricultural potential, 22% is suitable for production, while 66% is only marginally suitable for production – in short, of the total agricultural land area, only 34% is effectively usable. Yet mining activities erode a fair proportion of that 34%. As a result, fewer and fewer people are engaged in food production while, at the same time, the demand for adequate food is increasing as the population grows – which means that the most vulnerable sector of the 'at risk' population in relation to hunger is that of young people.

The preliminary report of the NSTF meeting provides four recommendations:

1. The need for better communication among producers, distributors and consumers in the food supply chain so that expectations are aligned and the food supply chain is less vulnerable to competing interests.
2. The need for collaboration – not only across the food supply chain but between government and the private sector and between governmental departments.
3. The recognition of indigenous knowledge systems around food security.
4. The urgent need for investment in research and development – in other words, in relevant science and technology and the implementation of the outcome of scientific and developmental work.

The significance of the first and fourth recommendations is underscored by the debate around what might be considered to be the minimum level of funding that will secure sufficient nutritious food for a low-income family. It is the view of the South African Minister for Social Development that ZAR753 a month is sufficient to feed a low-income family of five people – a figure that was hotly debated and contested in Parliament this June. Best estimates, which take account of the impact of drought and the effects of El Niño on food prices, suggest that the minimum figure for the most basic basket of food for a family of five is, in fact, closer to ZAR1147 a month – 52% more than the figure cited by the Minister.

It would seem that the food supply chain is subject not only to competitors, but also to inaccurate, and possibly opportunistic, political assessments.

All the more reason for more rigorous science, better insights, and fewer opportunities for guess work at the expense of low-income, hungry families.

HOW TO CITE: Butler-Adam J. More scientific thinking needed to feed society: The NSTF tackles hunger. S Afr J Sci. 2016;112(7/8), Art. #a0168, 1 page. http://dx.doi.org/10.17159/sajs.2016/a0168

ASSAf 1996–2016
Commemorative article

Science advisory role of national science academies

AUTHORS:
Roseanne Diab[1]
Susan Veldsman[1]

AFFILIATION:
[1]Academy of Science of South Africa, Pretoria, South Africa

CORRESPONDENCE TO:
Roseanne Diab

EMAIL:
roseanne@assaf.org.za

POSTAL ADDRESS:
Academy of Science of South Africa, PO Box 72135, Lynnwood Ridge 0040, South Africa

KEYWORDS:
science advisory ecosystem; evidence-based advice; consensus studies; evidence-informed policy

HOW TO CITE:
Diab R, Veldsman S. Science advisory role of national science academies. S Afr J Sci. 2016;112(7/8), Art. #a0169, 3 pages. http://dx.doi.org/10.17159/sajs.2016/a0169

Rising global interest in science advice

The celebration of the Academy of Science of South Africa's (ASSAf's) 20th anniversary presents an opportune time to examine ASSAf's role in the science advisory space in South Africa and to reflect, more generally, on the nature and value of academy advice. Although very young in comparison with other global science academies, some of which are more than 360 years old, ASSAf is fortunate to have benefitted from the experience of others and since its inception has had its science advisory role clearly articulated in its statutes. This is not true of many of the older science academies that were founded with the intention of fulfilling a largely honorific role and as such have had to grapple with the transition to a working academy model.

The topic of science advice is assuming a rapidly evolving and increasingly prominent role. The importance of science advice is commonly attributed to the scientific nature of the challenges confronting modern society – examples include climate change, disaster risk management, food security, and the bulk of the 17 Sustainable Development Goals. However, a defining moment in this newfound prominence must be attributed to the workshop on science advice hosted by Professor Sir Peter Gluckman, Chief Science Advisor (CSA) to the Prime Minister of New Zealand, under the auspices of the International Council for Science (ICSU) General Assembly in Auckland, New Zealand in August 2014. Further interest has been stimulated by the creation of the International Network of Government Science Advice (INGSA), led by Gluckman.

Science advice has also been the topic of international meetings hosted by ASSAf. These meetings include the InterAcademy Partnership (IAP) General Assembly and conference hosted by ASSAf in February 2016 in Hermanus; the capacity building workshop on science advice for African scientists that was held as a pre-event to the IAP meeting; and a South African Young Academy of Science (SAYAS) event on youth perspectives on science advice to governments that was held in March 2016 alongside the IAP meeting. All these events have placed the spotlight on science advice and raised awareness amongst scientists and policymakers.

Science advisory ecosystem

A useful concept is that of a science advisory ecosystem, which accommodates a range of co-existing science advisory modalities, with individual models assuming varying importance in different countries, giving rise to ecosystems that may have quite different geographical expressions. Some of these science advisory modalities may include individual scientists, industry and/or business groupings, non-governmental organisations, science and technology committees, statutory bodies mandated to provide advice, government scientists, national academies and CSAs.[1]

A variety of advisory structures is in place in South Africa, which include those listed above, as well as others such as advisors in individual ministries, sector-specific advisory bodies and early warning advisory bodies.[2] It is therefore not difficult to understand why in South Africa, the term 'crowded advisory space' is often used.

Here we focus on the role of science academies in this ecosystem, highlight the strengths of academy advice, give some examples from ASSAf's experience and reflect on how academy advice might evolve in the future in the South African context.

The academy's niche in the science advisory space is clearly carved out. Its strength lies in long-term, in-depth, evidence-based studies known in academy parlance as 'consensus studies'. Consensus studies are executed by a panel of volunteer members (not necessarily academy members/fellows). The panel members are selected for their knowledge and excellence in the field, their willingness to serve in a volunteer capacity and in the South African context, with due consideration given to demographic diversity. Consensus studies are a unique academy methodology, providing a multi-perspective, evidence-based view on a particular topic. Findings and recommendations are synthesised and published in a peer-reviewed report that is made available in the public domain.

ASSAf's experience in implementing consensus studies was gained from the United States National Academies during the African Science Academy Development Initiative (ASADI), which was a long-term capacity development initiative that introduced ASSAf to many such activities. It is acknowledged that the term 'consensus study' is troubling to some, particularly scholars in the humanities. However, the term should not be perceived as a notion that may stifle or constrain rigorous debate and argument, but rather as a means to distill core points of agreement that can provide policymakers with some degree of certainty about the complex world in which we live.

Addressing uncertainty in a complex, post-normal world

It is acknowledged that the complexity of the space in which science advisory bodies are operating can be overwhelming. This state was pointed out by Gluckman[3] when referring to a post-normal world, characterised by uncertainty and stochastic phenomena. The need to reach consensus on key findings and recommendations in such complex situations is critical when giving advice to policymakers. At no time in our history has the need to identify the core been more urgent. Faced with an overwhelming amount of information and complexity, the role of academies in distilling the complexity into a finite number of evidence-based recommendations agreed upon by a panel of experts, is critical.

Challenges facing academies in giving science advice

Some of the challenges faced by academies in giving advice are those of relevance, timeliness and receptivity. If an academy-initiated consensus study does not align with the current needs of policymakers, it may be relegated to dusty bookshelves. This does not necessarily imply that only studies requested by government should be undertaken. On the contrary, it is expected that academicians will be sufficiently alert and practised in their fields that they will be able to identify emerging topics and embark on studies that will be useful to policymakers regardless of the genesis. Relevance also relates to the scope and focus of the study. The findings and recommendations must be precise, concise, targeted and useful. A related challenge is that of timeliness. Ideally, if the timing of a study is right, its relevance will be enhanced.

One of the biggest challenges is that of receptivity or country readiness for science advice. A report will have limited impact on policy if government is either unwilling or unprepared to receive the advice. Perhaps one of the best examples of this scenario is ASSAf's 2007 report *HIV/AIDS, TB and Nutrition*, which was published when the government at the time held alternative views about the cause of AIDS. The ASSAf report gained no traction in South Africa at the time, despite being lauded abroad as an example of a bold academy report seeking to present the scientific evidence. It is interesting to note that the full impact of this report was only experienced many years after its publication.

A second example is that of ASSAf's 2015 report *Diversity in Human Sexuality: Implications for Policy in Africa*, which drew some bold conclusions based on recent scientific evidence, and in so doing dispelled many myths surrounding human sexuality. The reception of this report in South Africa, which is known to uphold human rights of all persons regardless of sexual orientation, contrasted dramatically with its uptake in other African countries, 60% of which have legislation criminalising same-sex sexual conduct and some even have laws by which such conduct is punishable by death. Publishing an evidence-based report that challenges widely held belief systems underpinned by legislation takes courage. In this respect, the Ugandan National Academy of Sciences (UNAS) is to be commended for their courageous stance in the face of severe government opposition to the findings in the report. For ASSAf it was less of a reputational risk as the findings concurred with generally held views in the country. Exceptionally disconcerting, however, was the large number of science academies in Africa that refused to endorse the evidence-based report, in all probability because the evidence challenged official government positions or belief systems in those countries. Such a situation in which so few are willing to tackle a controversial topic and uphold the evidence base, does not bode well for academy advice in Africa.

Impact of ASSAf consensus study reports

To date ASSAf has conducted 19 consensus studies, all of which are available on the ASSAf website. A question that is often asked relates to the impact and uptake of these reports. Recently, ASSAf has introduced a monitoring and evaluation framework and has begun to track impact, although it is acknowledged that this is a complex undertaking that is still in its infancy. Some examples of a direct policy influence of ASSAf's reports are listed:

- The 2006 *Report on a Strategic Approach to Research Publishing in South Africa* and the 2009 report *Scholarly Books: Their Production, Use and Evaluation in South Africa Today* had a major influence on the Research Output Policy of the Department of Higher Education and Training (DHET) that was published in 2015 and implemented with effect from January 2016. Recommendations from each of these reports have been incorporated into the DHET policy, specifically those relating to quality, peer-review practices, subsidy units allocated for books, and the inclusion of Scientific Electronic Library Online (SciELO)-SA as an index for automatic accreditation of South African scholarly journals.
- The findings and recommendations of the 2010 report, entitled *The PhD Study*, were used by the Council for Higher Education as a basis for discussion when revising the Higher Education Qualifications Framework. The report recommendations were also implemented by the Department of Science and Technology, DHET and the National Research Foundation.
- The 2009 report *Revitalising Clinical Research in South Africa* led to announcements by the Minister of Health regarding increased health research funding and a website has been developed to raise awareness about and promote clinical research in South Africa.

Limitations of science advice in the policymaking process

Policy is rarely determined by scientific evidence alone. It is contended that to some extent the provision of advice may be viewed as separate from the policymaking process. The advice that is given should be based on the best available information which underpins objective conclusions and recommendations to policymakers. That a policymaker may elect on occasions not to follow the advice given must be acknowledged. There may be many competing and compelling considerations that have little to do with the scientific evidence, such as financial constraints, public opinion and political obligations. Gluckman[3] therefore prefers the term evidence-informed policy as opposed to evidence-based policy.

Advice versus advocacy

Advocacy differs from objective, impartial science advice in that it reflects the interests and/or value systems of the party providing the information. Ideally, the body providing advice in this instance should be transparent about their bias and how it may influence the conclusions reached; non-governmental organisations typically fall into this category. Such responsible, science-based advocacy differs from science advice but may still have a place.[1]

Straying into the advocacy terrain is dangerous for academies. At the heart of academy advice is the reputation for honest, objective, unbiased advice. Damage to these core attributes would place the advisory function at risk and render academy advice questionable.

Confidentiality of science advice

There is certainly a need for confidential science advice, but this is not a role that is or should be fulfilled by academies. Academy advice is valued for its transparency; all ASSAf's science advisory reports are published and available in the public domain. Arguably there are other bodies that are more suited to this type of science advice and the distinction between such bodies and academies in respect of their science advisory roles should remain – further strengthening the notion of a science advisory ecosystem and a distinct but synergistic role for all the ecosystem components.

ASSAf's role in science advice in the future

The two main bodies mandated to provide science advice in South Africa – ASSAf and the National Advisory Council on Innovation (NACI) – have each defined their niche and form part of the science advisory ecosystem in the country. The former focuses on long-term, in-depth studies as described above and the latter on shorter timescale studies of the order of a few months and produces concise briefs for the Minister of Science and Technology. There is a gap at the very short end of the temporal spectrum, when advice is required on a timescale of the order of days or weeks.

It is generally acknowledged that academies are not good at providing rapid response advice. The question might then be posed as to who is best positioned to give rapid response advice. Certainly, a CSA can play a critical role and in times of disasters or emergencies may even become part of the decision-making team as noted by Gluckman[3]. However, even a CSA is unlikely to possess all the expertise required at a time that rapid response advice is required and would likely depend on advice from a network of experts in such instances. It is contended that in the absence of a CSA in South Africa, ASSAf should assume this role. Academies

are renowned for their convening capabilities, largely expressed through the hosting of workshops and symposia. Yet, there is nothing preventing this experience and proficiency from being exercised in a shorter time frame; the same network of national and international experts could be convened electronically or telephonically and either collectively or individually, to seek advice in such situations. Currently, in South Africa, neither ASSAf nor NACI is active in this advisory space. A role for ASSAf should be explored.

Final reflection

Notwithstanding the fact that ASSAf's role as a science advisory body is enshrined in the *ASSAf Act*, it is incumbent on ASSAf to mould and establish this role such that its advice is highly respected and sought after. The fundamental tenets of academy advice, essentially objective, evidence based, free of vested interests, based on volunteerism and multi-perspective, must be cherished and protected as ASSAf matures in its role in the science advisory ecosystem in South Africa.

References

1. Hutchings JA, Stenseth NC. Communication of science advice to government. Trends Ecol Evol. 2016;31(1):7–11. http://dx.doi.org/10.1016/j.tree.2015.10.008
2. Netshiluvhi T. Overview: Africa's science advice landscape. Proceedings of the Science Advice for African Scientists' Workshop; 2016 February 26–27; Hermanus, South Africa. Pretoria: ASSAf; 2016. p. 5–7. In press 2016.
3. Gluckman P. Keynote address. Proceedings of the InterAcademy Partnership (IAP) Conference on Science Advice; 2016 February 28 – March 01; Hermanus, South Africa. Pretoria: ASSAf; 2016. p. 6–8. In press 2016.

Performance in Chemistry of students who started in the University Preparation Programme: The ripple effect

AUTHORS:
Charlene Marais[1]
Rina Meintjes[1]
C. Robert Dennis[1]
Jannie C. Swarts[1]

AFFILIATION:
[1]Department of Chemistry, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
C. Robert Dennis

EMAIL:
crobbie.dennis@gmail.com

POSTAL ADDRESS:
Department of Chemistry, University of the Free State, PO Box 339, Bloemfontein 9300, South Africa

KEYWORDS:
natural sciences; graduates; postgraduates; foundation programmes

HOW TO CITE:
Marais C, Meintjes R, Dennis CR, Swarts JC. Performance in Chemistry of students who started in the University Preparation Programme: The ripple effect. S Afr J Sci. 2016;112(7/8), Art. #a0162, 3 pages. http://dx.doi.org/10.17159/sajs.2016/a0162

Chemistry has been offered in the University Preparation Programme of the University of the Free State for more than 20 years and has been the only university-accredited module in the Natural Sciences Stream for many years. The positive impact of recurriculation during this time on Chemistry pass rates is demonstrated. Although only a few students from the University Preparation Programme graduated with Chemistry as a major and pursued postgraduate studies in Chemistry, a total of 284 degrees has been awarded to students who started in the Natural Sciences Stream with Chemistry as a core subject, since inception in 1994.

In South Africa, efforts have been made over the past 20 years to address the needs of school-leavers with a matric (Senior Certificate or National Senior Certificate) who did not meet the entrance requirements of higher education institutions. Most of these institutions have addressed these needs by implementing bridging or foundation programmes. Typically, bridging programmes do not earn subsidy, thus forcing higher education institutions to rely on donor funding to implement such programmes.[1]

The University of the Free State (UFS) and nine other training institutions in the region came together in a consortium in 1993 to implement the Need for Education and Elevation Programme. The name changed to the Career Preparation Programme in 2004 and since 2010 has been known as the University Preparation Programme (UPP). The primary aim of the programme is to offer school-leavers an opportunity to enter general-formative and vocationally directed studies at various higher and further education institutions in the region.[1,2] Apart from registering for certain prescribed university-accredited modules, students at UFS also have to register for compulsory development modules. These modules include Introduction to Computer Usage, Skills and Competencies in Life-long Learning and Academic Literacy.[1-3] At the different colleges at which the programme is presented, students have to register for the Further Education and Training modules in Computer Practice and Communication. Through the additional compulsory modules, an attempt is therefore made to not only offer students academic information but also address the wider needs of students in higher education.

Initially, the general entrance requirements for the UPP were a Senior Certificate and a matriculation score (M-score) of a minimum of 12 points calculated from a student's Grade-12 subject symbols.[1] In 2009, these entrance requirements changed to a National Senior Certificate and a minimum admission point score of 17 points calculated from the achievement levels of the Grade-12 school subjects. From 2011, the entrance requirements have been:

- a National Senior Certificate;
- a minimum admission point score of 20;
- four subjects with a minimum achievement level of 3 (40–49%); and
- language of instruction (English or Afrikaans) on a minimum achievement level of 3 (40–49%).

Students interested in the UPP have a choice of streams in Humanities and Social Sciences, Economic Sciences, and Natural Sciences.[1,2] When the UPP was initially launched, a pass in Grade-12 Mathematics on a standard-grade level was a prerequisite for the Economic Sciences and Natural Sciences Streams.[1] Currently, additional criteria must be met for enrolment in the Natural Sciences Stream:

- Grade-12 Mathematics on a minimum achievement level of 3 (40–49%) and
- Grade-12 Life Sciences *or* Physical Sciences on a minimum achievement level of 3 (40–49%).

The objective of this discussion is the performance of UPP students in Chemistry, and so the focus will be on the Natural Sciences Stream of the UPP. At the inception of the UPP, students in the Natural Sciences Stream had to register for university-accredited modules in Mathematics (WIS104) and Chemistry (CEM108). The success rate for the Mathematics module was very poor, as can be seen from Figure 1. Marks higher than 70% were mainly achieved by students who did qualify for mainstream study but registered for the UPP as it gave them the opportunity to complete some of the first-year university modules at a venue closer to home. The pass rate of UPP students for the Introductory Mathematics module, over the period 1994 to 2004, was a dismal 29%. Because of this high failure rate, the mathematics modules for the Natural Sciences Stream of the UPP were downgraded to foundation mathematics modules (WTV154 and WTV164) with a level between school (Grades 11 and 12) and first-year university mathematics.

The Chemistry module started as a 32-credit General Chemistry year module (CEM108). It was changed in 2006 to two 16-credit year modules (CHE104 and CHE194) which run in parallel. Since 2010, four 8-credit Chemistry modules have been offered (CHE112 and CHE132 in the first semester, CHE142 and CHE122 in the second semester). The first-semester Chemistry modules are a prerequisite for the second-semester Chemistry modules as is a pass in Mathematics (WTV154) or an achievement level 4 for National Senior Certificate or National Certificate Vocational Mathematics.

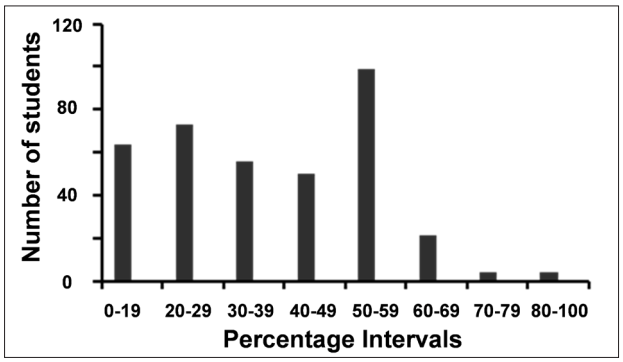


Figure 1: Performance of students in the University Preparation Programme in the first-year Mathematics module (WIS104) over the period 1994–2004.

The UPP students performed relatively well in the Chemistry modules (Figure 2). The pass rate in the 32-credit module during the period 1994–2005 was 47.0%, while the pass rates in the two 16-credit year modules (CHE104: Inorganic and Analytical Chemistry and CHE194: Physical and Organic Chemistry) were 49.7% and 57.7%, respectively. Since the introduction of the four 8-credit Chemistry modules in 2010, the pass rates for Chemistry have increased dramatically (Figure 3). The average pass rates for the first-semester modules (CHE112: General Foundation Chemistry and CHE132: Organic Chemistry) for the period 2010–2014 were 84.0% and 77.7%, respectively, while pass rates for the second-semester modules (Physical Chemistry and Inorganic and Analytical Chemistry) were 91.1% and 87.9%, respectively.

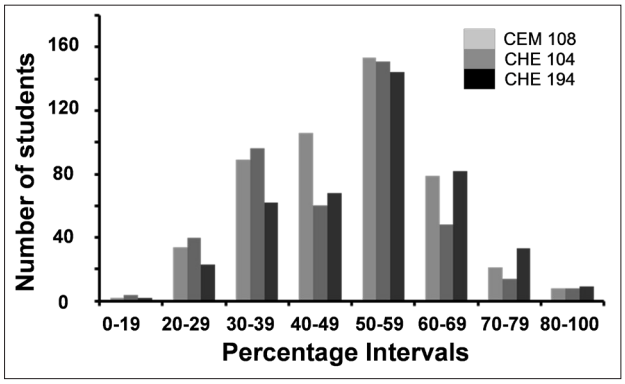


Figure 2: Performance of students in the University Preparation Programme in the first-year Chemistry modules (CEM108 over the period 1994–2005, CHE104 and CHE194 over the period 2006–2009).

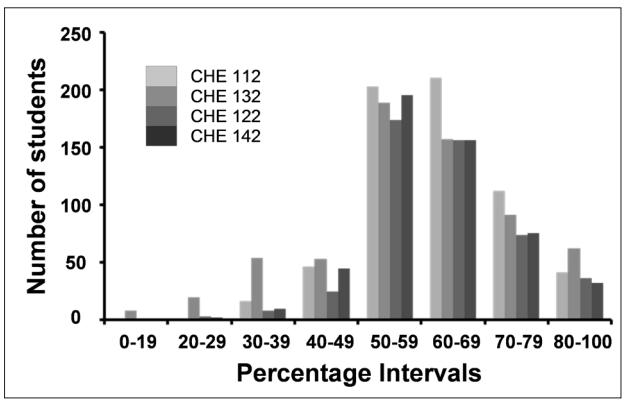


Figure 3: Performance of students in the University Preparation Programme in the first-year Chemistry modules (CHE112, CHE132, CHE122 and CHE142 over the period 2010-2014). The skewness to the right is an indication of a high pass rate.

Considering these statistics, prospects for graduate students in Chemistry look bright. However, the interest of the UPP students to continue with Chemistry as a major is disappointing, and furthermore, the academic performance of UPP students continuing with second- and third-year Chemistry at UFS shows the contrary.

Since the start of the UPP until 2014, 1784 students have enrolled for the Natural Sciences Stream of the programme. In total, only 191 students who qualified, enrolled for a second-year Chemistry module over this period. Of these 191 students, 127 students passed at least one of the four second-year Chemistry modules and 69 students passed all four second-year Chemistry modules and were thus eligible to enrol for third-year Chemistry modules. Only 59 out of a possible 69 students enrolled for third-year Chemistry modules. Of these 59 students, 20 of the students (1.1%) passed all four of the third-year Chemistry modules and obtained a BSc Chemistry degree. Of the 20 Chemistry graduates, 6 eventually also obtained a BSc Honours degree in Chemistry followed by an MSc (Chemistry) qualification. One of these Chemistry graduates completed a PhD in Chemistry. (Although outside our sample period, we can report that a second candidate from the above-mentioned six UPP students completed a PhD study in Chemistry during 2015.)

Table 1: Bachelor degree majors and number of bachelor degrees conferred on students in the Natural Sciences Stream of the University Preparation Programme over the period 1994–2014

Degree majors	Number of degrees
Agricultural Sciences	22
Biological Sciences	88
Building Sciences	2
Chemistry	20
Geography	10
Geology	15
Information Technology	9
Mathematics and Statistics	9
Physics	4
Total	**179**

Although graduates in Chemistry coming through the UPP are few, the success of the UPP Natural Sciences Stream can clearly be seen from the graduates in other BSc programmes (Table 1) as well as the graduates in other faculties who started in the Natural Sciences Stream and changed to other faculties after the first year of study (Figure 4).

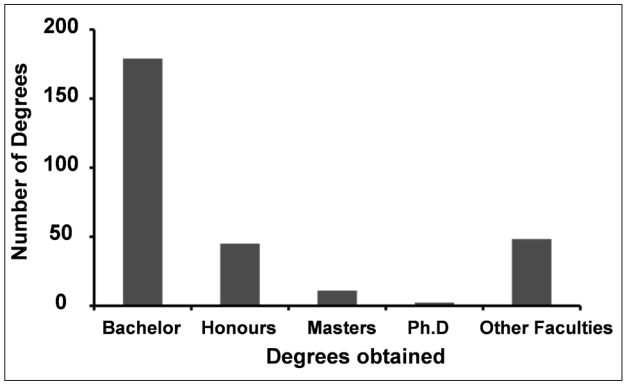


Figure 4: Degrees obtained at the University of the Free State by students who started in the Natural Sciences Stream of the University Preparation Programme over the period 1994–2014.

The number of graduates in other BSc programmes (i.e. those not majoring in Chemistry) was 159 (8.9%). The number of graduates in other faculties was 48 (2.7%). Postgraduate degrees obtained in the Faculty of Natural and Agricultural Sciences by UPP students from the Natural Sciences Stream over this period was 57 (Figure 4).

The Chemistry modules, and from 2014 also Biology modules, are the only university-accredited modules in the first year of the UPP in the Natural Sciences Stream. The graduation rate with Chemistry as a major is therefore a concern (only 20 BSc Chemistry degrees (1.1%)). From our perspective, the reason for this low rate could be mainly because of two factors:

1. First-year, first-semester university mathematics is a prerequisite for second-year Chemistry modules at the UFS. UPP students register for foundation mathematics in the first year of study. The required mathematics module is therefore taken in their second year of study and consequently the UPP students can only start with second-year Chemistry in their third year of study after a break of at least one year from Chemistry.
2. The second-year Chemistry modules are popular with students. At the UFS an average of 60% in first-year Chemistry modules is required to qualify for second-year Chemistry modules. UPP students have to compete with mainstream students for a place in the second-year class, as the Chemistry Department can accept only a limited number of second-year students because of facility constraints.

In general, the Faculty of Natural and Agricultural Sciences of the UFS is proud of what has been achieved within the UPP. The fact that 284 degrees have been awarded to students who started in the Natural Sciences Stream since 1994, underlines the success of the UPP. The UPP provides deserving students who are not able to meet the admission requirements at universities because of imbalances in the school system, an opportunity to build themselves a career in the competitive work environment.

Acknowledgement

This study would not have been possible without the help of Annemari Eksteen and Jacques Botha of the Directorate for Institutional Research and Academic Planning at the UFS who compiled the data.

References

1. Hay HR, Marais F. Bridging programmes: Gain, pain or all in vain. S Afr J High Educ. 2004;18(2):59–75. http://dx.doi.org/10.4314/sajhe.v18i2.25454
2. Rabie NE. Access to higher education: The case of the Career Preparation Programme at the University of the Free State [mini-thesis]. Cape Town: University of the Western Cape; 2008.
3. Massyn L. Open learning with success: Practical implementations for facilitators and learners in rural areas. Paper presented at: Second Pan-Commonwealth Forum on Open Learning; 2002 July 28 – August 02; Durban, South Africa.

Harry Oppenheimer Fellowship Awards, 2016

AUTHOR:
John Butler-Adam[1]

AFFILIATION:
[1]Academy of Science of South Africa, Pretoria, South Africa

CORRESPONDENCE TO:
John Butler-Adam

EMAIL:
j.butleradam@gmail.com

POSTAL ADDRESS:
Academy of Science of South Africa, PO Box 72135, Lynnwood Ridge 0040, Pretoria, South Africa

KEYWORDS:
fungal genomics; ascomycetes; tree pathogens; Nelson Mandela; biography; Romantic hero; tragic hero

HOW TO CITE:
Butler-Adam J. Harry Oppenheimer Fellowship Awards, 2016. S Afr J Sci. 2016;112(7/8), Art. #a0164, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0164

The Harry Oppenheimer Fellowship Awards were initiated in 2001 to commemorate the Oppenheimer Memorial Trust's founder and all he stood for. Especially important were his efforts to support human and intellectual development, advance scholarship and encourage innovative ideas. The Trust has a long tradition of investing in education and many beneficiaries have gone on to make important contributions to South African public life. The fellowships build on this tradition and, with a monetary value of R1.5 million, rank as the Trust's premier annual awards. These are special investments to encourage excellence in scholarship in all disciplines and in all its forms. The awards serve to acknowledge cutting-edge, internationally significant work that has particular application to the advancement of knowledge, teaching, research and development in South Africa.

The 2016 award was unusual in that two awards were made instead of a single recipient being selected, as in most years. One award was made to Professor Brenda Wingfield of the University of Pretoria and another to Professor Xolela Mangcu of the University of Cape Town.

Brenda Wingfield

Brenda Wingfield is a professor of genetics at the University of Pretoria. She holds the South African Research Chairs Initiative (SARChI) Chair in Fungal Genomics, and is internationally recognised as a world leader in her field. She is a past chair of the National Science and Technology Forum, and currently serves in leadership roles in national and international structures. She is vice-president of the Academy of Science of South Africa; convener of the National Research Foundation (NRF) rating specialist committee for microbiology and plant pathology; secretary-general of the International Society of Plant Pathology; and a project leader at the Centre of Excellence in Tree Health Biotechnology (run by the Department of Science and Technology together with the NRF). She is Deputy Dean of Research in the University's Faculty of Natural and Agricultural Sciences.

Brenda Wingfield was born in Zambia, completed her schooling in Zimbabwe, and calls South Africa her home. She obtained her BSc and BSc Hons degrees at the Universities of Natal and Cape Town respectively, a Master's in biochemistry from the University of Minnesota, and her Ph.D at the University of Stellenbosch. Prof. Wingfield and her husband Prof. Mike Wingfield took up posts at the University of Pretoria in the late 1990s, and together with other experts established the Forestry and Agricultural Biotechnology Institute.

Brenda Wingfield's contributions to science have been widely recognised. Over the years she has received the following awards:

- Women in Water, Sanitation, and Forestry Research Award: Department of Water Affairs and Forestry (2007)
- Distinguished Women in Science Award: Department of Science and Technology (2008)
- Women in Science (Southern Region) Award: African Union (2009)
- BHP Billiton Outstanding Contribution to SET Research Capacity Building: National Science and Technology Forum (2014)
- Christiaan Hendrik Persoon Medal for Outstanding Achievement in Plant Pathology: Southern African Society for Plant Pathology (2015).

In 2013, Prof. Wingfield received an A-rating from the NRF. Her research over the past 20 years has focused on the global movement and evolution of fungal pathogens, particularly those which cause tree disease. She was instrumental in developing the first DNA-based phylogenes for a number of important tree pathogens. She also helped to develop molecular tools to study the population diversity, origins and movements of many tree pathogens around the world. She sequenced the first fungal genome in Africa.

Prof. Wingfield's current project aims at achieving better understanding and characterisation of the mating type locus in a group of fungi known as ascomycetes. The vast majority of tree pathogens are fungal and belong to this group. Genes are central to sexual reproduction in fungi, as elsewhere, and recombination between mating genomes is one of the driving forces in their evolution and genetic change. Studies such as Prof. Wingfield's are important for managing plant and tree pathogens because of the 'arms race' between pathogens and their hosts. The capacity of a pathogen to adapt and change in response to the defences of its host defines the pathogen's ability to survive. The international community of mycologists has embarked on 'The 1000 Fungal Genomes Project', and Prof. Wingfield with her collaborators and graduate and postdoctoral students at the University of Pretoria are enthusiastic members of this initiative. The Harry Oppenheimer Fellowship Award will make the project possible on an ambitious scale.

Xolela Mangcu

Xolela Mangcu is a professor in the Department of Sociology at the University of Cape Town. He has held fellowships at the prestigious Brookings Institution, the W.E.B. Du Bois Institute, and the Hutchins Center for African and African–American Studies at Harvard University. He is a leading columnist and political commentator, and has published nine books including *Biko: A Biography,* for which he won the University of Cape Town book award.

Xolela Mangcu was born to a family of educators in Ginsberg Township in King William's Town, which was also the hometown of Steve Biko (his childhood political inspiration). Prof. Mangcu studied at local schools, and later attended the University of the Witwatersrand in 1984 under a quota system for then designated white universities.

He soon became a prominent member of the Black Consciousness student movement. After completing a degree in law and sociology, he obtained a master's degree in Development Planning (Wits, 1988) and soon after was admitted as a Fellow in the Special Program in Urban and Regional Studies at Massachusetts Institute of Technology. Prof. Mangcu completed his PhD in city and regional planning at Cornell University in 1997. Other fellowships followed – at the Rockefeller Foundation and Harvard's John F. Kennedy School of Government – before he returned to South Africa in 1999. He launched the Steve Biko Foundation in 2000, in partnership with the Biko family and local youth. From 2006 to 2011 he divided his time between the University of the Witwatersrand and the University of Johannesburg, and he joined the University of Cape Town in 2012.

Prof. Mangcu will use the Harry Oppenheimer Award to write a new and highly contextualised biography of Nelson Mandela. One of the qualities Mangcu greatly admired in Mandela was that he was embraced by the great man, even though he was essentially Mandela's critic. The idea of someone being Mandela's critic was jarring in a world where Mandela has been treated as a god-like figure. Prof. Mangcu believes that because of this 'hero-worship', many of Mandela's biographers have abstracted Mandela from the political debates and controversies that have animated the world of black South Africans since their earliest encounters with colonialism. The result is an abstracted morality tale instead of engagement with Mandela as a contested figure, both within the African National Congress and various political movements. (The latter include the All-Africa Convention, the New Unity Movement, the Pan Africanist Congress and the Black Consciousness movement.)

Prof. Mangcu thus intends to address broader questions of history. Among his aims are to examine the activities of organisations such as the South African Native Association (Imbumba) in Tembuland; the political and electoral campaigns for African representation in the Cape Parliament; the rise of the African Christian church; and the political and intellectual writing of African greats such as Tiyo Soga, Walter Rubusana, D.D.T. Jabavu, John Tengo Jabavu, David Malasi, Richard Kawa, Meshack, James Pelem and many others.

Writing the new book will involve conducting empirical research in South Africa and spending time at Harvard to write up the findings and engage with leading scholars. Prof. Mangcu has titled his work *Paradoxical Mandela: Romantic Hero, Tragic Hero* to highlight the difference between the predominantly Romantic representation of Mandela as an individual hero of the liberation struggle and the Classic idea of tragedy as communal action. Whereas liberal Romantic tragedy focuses on what happens *to* the hero, Classic tragedy focuses on what happens *through* the hero. (This formulation was proposed by Raymond Williams.) Thus, in Classic Greek tragedy, tragic action continues even after the hero has died or stepped off the stage. The action continues through the chorus, the audience's response to the tragedy and the energies that are released in the action. The hero, chorus and audience keep returning to the scene despite its dangers, driven by the desire to give to the world what Steve Biko called 'a more human face'. This is another way to describe tragic hope as the spur for human action. These insights are profound for present-day South Africa, and I look forward to the completion of this important task.

Previous recipients of the Oppenheimer Fellowship Award are:

2014: Helen Rees, Obstetrics and Gynaecology, Wits Reproductive Health & HIV Institute, University of the Witwatersrand

2013: Keertan Dheda, Respiratory Medicine, University of Cape Town

2012: Robin Crewe, Zoology and Entomology, University of Pretoria

2011: Leslie Underhill, Zoology, University of Cape Town

2010: Duncan Mitchell, Physiology, University of the Witwatersrand

2009: Jill Farrant, Molecular Physiology, University of Cape Town

2008: Philippe-Joseph Salazar, Rhetoric Studies, University of Cape Town

2007: Jeff Guy, History, University of KwaZulu-Natal, Winston Hide, Bioinformatics and Cancer Research, University of the Western Cape

2006: Paul Cilliers, Philosophy and Ethics, University of Stellenbosch

2005: Norman Owen-Smith, Environmental Sciences, University of the Witwatersrand

2004: Igor Barashenkov, Mathematics and Applied Mathematics, University of Cape Town Frank Brombacher, Immunology and Infectious Diseases, University of Cape Town

2002: Jan-Hendrik Hofmeyr, Biochemistry, University of Stellenbosch

2001: David Glasser, Chemical Engineering, University of the Witwatersrand

The possibility of lichen growth on bones of *Homo naledi*: Were they exposed to light?

AUTHOR:
J. Francis Thackeray[1]

AFFILIATION:
[1] Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Francis Thackeray

EMAIL:
Francis.thackeray@wits.ac.za

POSTAL ADDRESS:
Evolutionary Studies Institute, University of the Witwatersrand, Private Bag 3, Wits 2050, South Africa

KEYWORDS:
Homo naledi; manganese; lichen; taphonomy

HOW TO CITE:
Thackeray JF. The possibility of lichen growth on bones of *Homo naledi*: Were they exposed to light? S Afr J Sci. 2016;112(7/8), Art. #a0167, 5 pages. http://dx.doi.org/10.17159/sajs.2016/a0167

Dirks et al[1] report on the taphonomy of bones from the Dinaledi Chamber in the Rising Star cave complex, situated in the Cradle of Humankind in South Africa. Fossils of *Homo naledi* have been recovered from this chamber.[2] An age estimate for the *Homo naledi* fossils, based on morphometric analyses of crania, is in the order of 2 million years,[3] although it is recognised that a younger date may apply if morphological patterns persisted for an unknown period of time.

Dirks et al.[1] state that 'some bones and teeth are dotted with black iron-manganese oxy-hydroxide deposits and coatings'. The question arises as to what factors contributed to the 'dotted' or spotty distribution of manganese oxy-hydroxide on bone surfaces. Abiotic geological factors may account for continuous (matted) deposition of manganese oxy-hydroxide coatings in some cases, but why should these be *dotted* or *spotted* in the instance of many bones from the Dinaledi Chamber – or from other nearby sites, such as Sterkfontein? To try to answer this question, it is appropriate to refer to observations by Thackeray et al.[4] in which we noted, firstly, that in the Cradle of Humankind, lichen can grow on certain substrates (including bone or rock) with a dotted or spotted distribution. Secondly, the spotted distribution of lichen is sometimes associated with dotted distributions of manganese oxy-hydroxide on the same surfaces.[4] The source of manganese would include dolomite and chert, 2 billion years old, relating to the shallow saline sea that existed at that time.

Lichen are symbionts[5] that include a fungus (a mycobiont) and a photosynthetic partner (a photobiont). This partner could be a cyanobacterium or a green alga. The growth of the symbiont results in a lichen body called a thallus. On substrates such as chert or bone, under suitable micro-environments the spottiness of lichen may relate to the dispersal of fungal spores. An example of lichen growth is shown on chert at Kromdraai in the Cradle of Humankind (Figure 1). In this case, small thalli of lichen – possibly *Rhizocarpon* – grow adjacent to older central thalli in which the lichens form continuous mats. The size of lichen thalli generally decreases outwards from the central thalli. Very small spots at the edge of lichen growth are relatively young growths of lichen.

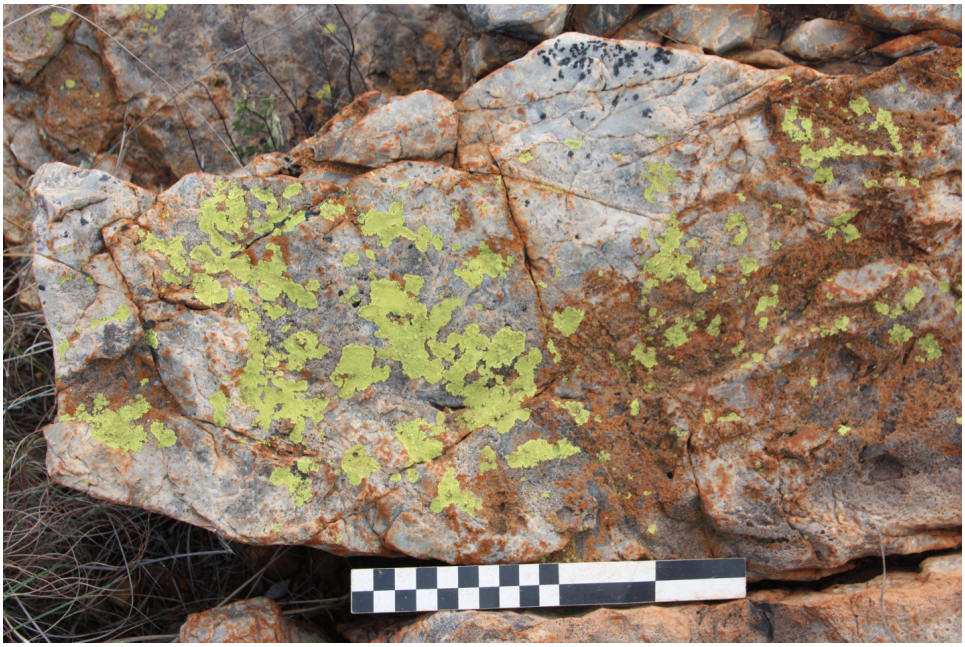


Figure 1: A chert rock at Kromdraai, showing the distribution of actively growing green lichen which is dispersing in small thalli from larger, denser central mats. The size of the thalli generally decreases outwards from the central mats. The small lichen spots to the right are relatively young growths. Lichen is a fungal–algal–bacterial symbiont, and dispersal from central thalli is facilitated by fungal spores. Note the presence of adjacent black spots of manganese dioxide (upper margin of photograph). Scale: centimetre squares are shown in black and white. Photograph: Benjamin Lans.

In the Cradle of Humankind, lichen can grow not only on chert but also on dolomite. However, weathering of the latter substrate may inhibit extensive lichen growth, which certainly does occur on more durable chert as well as on bone.

As mentioned above, lichens are algal–fungal–bacterial symbionts. Most bacteria include a manganese-containing superoxide dismutase or Mn-SOD[6], and the association with manganese is of special interest in the study of lichen in the Cradle of Humankind[4]. Elsewhere, Pentecost et al.[7] reported lichen on a substrate rich in manganese. Geomicrobial transformation of manganese compounds by fungi can also occur. Hansen et al.[8] demonstrate that ascomycetes are capable of oxidising manganese. Beckett et al.[9] indicate the existence of redox cycling in lichens, and refer to a process which 'will increase the photosynthetic capacity of thalli during winter when light intensities are low'. This may apply also to micro-environments in dolomitic caves under limited light conditions.

Figure 2: A chert rock at Kromdraai, showing the distribution of spots of manganese oxy-hydroxide. These spots may relate to former growth of lichen which dispersed in spots from central thalli. The size of spots of manganese oxy-hydroxide generally decreases from a central mat. The smaller spots of manganese oxy-hydroxide to the right may relate to relatively young growth of lichen (compare with lichen growth in Figure 1). Scale: centimetre squares are shown in black and white. Photograph: Benjamin Lans.

In Figure 1, lichen is shown in close association with spots of manganese oxy-hydroxide on chert in the Cradle of Humankind. Figure 2 shows a chert rock at the same site (Kromdraai), which is more extensively coated by spots of manganese oxy-hydroxide, apparently dispersing (with diminishing size) from a central area. The central area is potentially analogous to a continuous mat of a central thallus of lichen, as shown in Figure 1.

In the case of durable rocks such as chert at Kromdraai, lichen may have absorbed manganese from substrates.[4] Because magnesium is potentially toxic, the lichen may then have subsequently deposited and fixed the spots or dots of manganese dioxide on surfaces such as rock or bone.[4] Episodes of lichen growth could have occurred over long periods of time, such that the same rock might show relatively old spots of manganese oxy-hydroxide as well as relatively younger lichen thalli. Good examples of the association between modern lichen and manganese oxy-hydroxide at Kromdraai are shown in Figure 3 and Figure 4. These images record the spotty distribution of modern lichen, and the spotty or dotted distribution of black manganese oxy-hydroxide that may relate to former (spotty) growth of lichen.

Hominid cranial specimens from Sterkfontein, such as Stw 505 (*Australopithecus africanus*) and Sts 71 (attributed to *A. prometheus*), have small dots or spots of manganese oxy-hydroxide on bone surfaces, even within the inner cranial wall of these skulls. These spots of manganese oxy-hydroxide may represent areas where lichens were able to grow in a partially sunlit micro-environment and for a relatively short time, before the deposition of sediments that covered the crania, thereby halting lichen growth through the absence of light.[4] Figure 5 shows such spots on specimen Stw 505. Because this relates partly to a biological process, and manganese has a radioactive isotope that decays with a half-life in the order of 1 million years, there is exciting potential to use this theory or process to date fossils in the Cradle of Humankind.[4]

Modern dolomitic cave systems in the Cradle of Humankind include the Wonder Cave, which is situated within 10 km of the Rising Star cave complex. One encounters the following set of circumstances as one walks from the entrance of the cavern into the darkness of the dolomitic solution cavity:

1. Where there is intense light and heat on dry exposed surfaces outside the cave, there is little or no lichen growth.
2. As one walks into moister and slightly darker regions of the cave, there is an area where light and moisture appear to be optimal for the present lichen colonists, and lichen growth is extensive on cave walls if not also on bone.
3. As one enters the darkness of the cave where there is little light, there is a decrease in the mean size of lichen thalli until there is no lichen growth at all; this is the case even where there may be some moisture as a result of water dripping through the phreatic maze of the dolomitic cave system.

As mentioned above, the growth of lichen includes a photobiont (the photosynthetic partner to the fungal mycobiont), and this photobiont is dependent on at least some degree of exposure to light for growth to occur. It has been proposed that the Dinaledi Chamber in prehistory could be accessed only by means of very narrow and circuitous passages, in complete darkness.[1] The question arises whether the Dinaledi Chamber had an additional entrance. In response to this question, I hypothesise that in the case of the Dinaledi Chamber, bones such as tibia shaft specimen UW 101-1070 (Figure 6) may show spotty or dotted coatings of manganese oxy-hydroxide, of the kind reported by Dirks et al.[1], at least in part because of the growth of lichen.

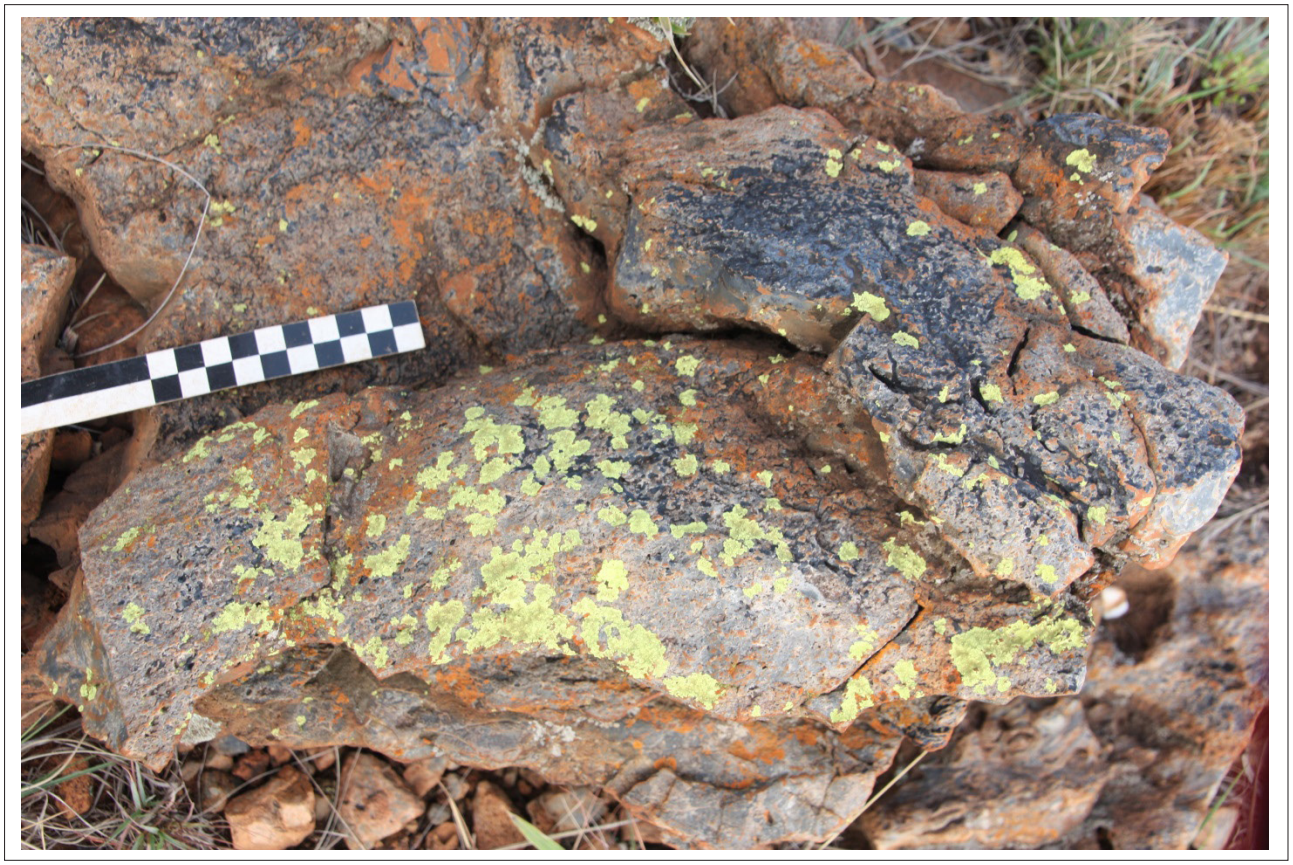


Figure 3: A chert rock at Kromdraai, showing black manganese oxy-hydroxide in association with green lichen. The manganese oxy-hydroxide may relate to earlier episodes of lichen growth on the same rock. Scale: centimetre squares are shown in black and white. Photograph: Benjamin Lans.

Figure 4: A chert rock at Kromdraai, showing the distribution of spots of manganese oxy-hydroxide, which may relate to former growth of lichen which dispersed in spots. Note the presence of actively growing green lichen, which in this instance may be very young compared with earlier phases of lichen growth and deposition of manganese oxy-hydroxide. Photograph: Jean-Baptiste Fourvel.

Figure 5: Manganese oxy-hydroxide spots on the inner cranial bone of Stw 505 (*Australopithecus africanus*) from Sterkfontein. It is suggested that the spots relate to the temporary growth of lichen before the cranium was covered by sediment, when lichen growth would have ceased in the absence of light. Photograph: Francis Thackeray (University of the Witwatersrand).

Figure 6: Tibia shaft specimen UW 101-1070, *H. naledi* from the Dinaledi Chamber, with dotted coatings of manganese oxy-hydroxide. It is suggested that the black dots result, at least in part, from the growth of lichen as a bacterial–algal–fungal symbiont that includes a photobiont. The growth of lichen on such bone surfaces, even for a limited time, may have occurred in subdued but essential lighting. Note the distribution of manganese oxy-hydroxide, extending from a continuous mat to more dotted occurrences; this pattern is potentially analogous to the dispersal of lichen from a central thallus. Photograph: F Thackeray (University of the Witwatersrand).

As mentioned, lichen are bacterial–fungal–algal symbionts, including a photobiont. The growth of lichen on bone surfaces, even for a limited period of time, may have occurred in subdued but essential lighting. If correct, this theory may mean there was at some time a second entrance to the Dinaledi Chamber in the phreatic maze cave system. Such an entrance might have allowed penetration of at least some light into the cave, facilitating the growth of lichen and subsequent deposition of manganese oxy-hydroxide on bones. It is further hypothesised that such an entrance, if it existed, was temporary. A rock fall in the phreatic maze cave system may subsequently have sealed the entrance at some stage in the dolomitic solution cavity, whereupon lichen growth would have terminated, in darkness.

If there was more than one entrance into the Dinaledi Chamber, as suggested here and by Val[11], the 'intentional depositional model' of Dirks et al.[1] would need to be re-assessed. Such re-assessment would need to account for the occurrence of 15 individuals of *Homo naledi* in the Dinaledi Chamber.[1] More recently, Dirks et al.[12] have responded to comments by Val[11], and those comments are relevant to this discussion. However, it should be noted that the age distribution of the 15 individuals seems to correspond to a family group. Perhaps this group died as a result of a crisis near an entrance to the Dinaledi Chamber, and/or a roof collapse in a phreatic maze, such that the group was trapped.

It is certainly accepted that abiotic factors may account for deposition of manganese oxy-hydroxide on bone surfaces. However, the question is why manganese oxy-hydroxide should be dotted or spotted, as in the case of bones from Dinaledi Chamber[1] or other nearby sites such as Sterkfontein[4]. The answer may lie in a biological process associated with lichen requiring light, rather than an entirely abiotic (geological) process whereby manganese oxy-hydroxide may have been distributed as a continuous mat.

Dominguez-Rodrigo and Barba[12] conducted experimental work on modern bone to investigate taphonomic processes associated with fungi and bacteria. They reported that 'bone surfaces are affected by dark circular marks caused by mycelial fungi, with bacteria playing a decidedly smaller role'.[12] They also noted that the excretions deposited on bone included 'organic and inorganic acids, enzymes, pigments, and toxins'.[12] However, their investigation did not relate to lichen and manganese in the context of a bacterial–fungal–algal symbiotic relationship as discussed here. Long-term experimental work on lichen on bone substrates in the Cradle of Humankind is planned for a 10-year period, in and around cave environments.

The concepts advocated here can be assessed in the context of empirical data already available. The mean size of manganese oxy-hydroxide spots on a selection of bones from the Dinaledi Chamber is 1.11 ± 0.31 mm, for $n = 72$ measurements (on mandibular and cranial bone as well as tibiae, a femur and a humerus). This mean size does not differ significantly from that of manganese oxy-hydroxide spots on the inner cranial bone of Stw 505 (1.32 ± 0.42 mm; $n = 25$ measurements). The smaller spot size for *H. naledi* specimens may perhaps relate to a lower degree of light exposure that allowed at least temporary growth of lichen on bone surfaces.

Apart from the possibility of lichen growth, it is pertinent to mention that Dirks et al.[1] reported taphonomic alteration to bones from the Dinaledi Chamber in the form of damage by snails and beetles. Snails and beetles may be dependent to some extent on the presence of leaf litter, and leaves in turn depend on light. Therefore, such damage may also indicate there was at least a temporary opening into the Dinaledi Chamber, in addition to the opening accessed by modern explorers. The Rising Star cave complex is a phreatic maze where episodic openings and rock falls can be common within long periods of time.

Acknowledgements

This research was supported by the National Research Foundation and the Andrew W. Mellon Foundation and the Centre of Excellence for the Palaeosciences. I am grateful to Benjamin Lans and Jean-Baptiste Fourvel for the photographs; to Bernhard Zipfel and the Access Committee for access to *H. naledi* specimens to take measurements of manganese oxy-hydroxide spots, based on photographs with a scale; to Don Cowan, Richard Beckett, Mike Wingfield, Sarah Elton and anonymous commentators for helpful recommendations on an earlier version of this manuscript; and to Laurent Bruxelles for useful discussions on phreatic maze cave systems.

References

1. Dirks PHGM, Berger LR, Roberts EM, Kramers JD, Hawks J, Randolph-Quinney PS, et al. Geological and taphonomic context for the new hominin species *Homo naledi* from the Dinaledi Chamber, South Africa. eLife 2015;4:e09561 http://dx.doi.org/10.7554/eLife.09561
2. Berger LR, Hawks J, de Ruiter DJ, Churchill SE, Schmid P, Delezene LK, et al. *Homo naledi*, a new species of the genus *Homo* from the Dinaledi Chamber, South Africa. eLife 2015;4:e09560 http://dx.doi.org/10.7554/eLife.09560
3. Thackeray JF. Estimating the age and affinities of *Homo naledi*. SA J Sci. 2015;111:3–4. http://dx.doi.org/10.17159/sajs.2015/a0124
4. Thackeray JF, Sénégas F, Laden G. Manganese dioxide staining on hominid crania from Sterkfontein and Swartkrans: Possible associations with lichen. SA J Sci. 2005;101:28. http://reference.sabinet.co.za/webx/access/electronic_journals/sajsci/sajsci_v101_n1_a4.pdf
5. Purvis W. Lichens. The Natural History Museum, London; 2000.
6. Law N, Caudle M, Pecoraro V. Manganese redox enzymes and model systems: Properties, structures, and reactivity. Adv Inorg Chem 1998;46:305–440. http://dx.doi.org/10.1016/S0898-8838(08)60152-X
7. Pentecost A, Spiro B, Williamson B. A note on the relationship between some saxicolous lichens and manganese ore in North Wales UK. Geomicrobiol J. 2010;27:349–352. http://dx.doi.org/10.1080/01490451003707668
8. Hansel CM, Zeiner CA, Santelli CM, Webb SM. Mn(II) oxidation by an ascomycete fungus is linked to superoxide production during asexual reproduction. Proc Natl Acad Sci USA. 2012;109(31):12621–12625. http://dx.doi.org/10.1073/pnas.1203885109
9. Beckett RP, Ntombela N, Scott E, Gurjanov OP, Minibeyeva FV, Liers C. Role of laccases and peroxidases in saprotrophic activities in the lichen *Usnea undulata*. Fungal Ecol. 2015;14,71–78. http://dx.doi.org/10.1016/j.funeco.2014.12.004
10. Val A. Deliberate body disposal by hominins in the Dinaledi Chamber, Cradle of Humankind, South Africa? J Hum Evol. 2016;96:145–148. http://dx.doi.org/10.1016/j.jhevol.2016.02.004
11. Dirks PHGM, Berger LR, Hawks J, Randolph-Quinney PS, Backwell LR, Roberts EM. Comment on 'Deliberate body disposal by hominins in the Dinaledi Chamber, Cradle of Humankind, South Africa?' J Hum Evol. 2016;96:149–153. http://dx.doi.org/10.1016/j.jhevol.2016.04.007
12. Dominguez-Rodrigo M, Barba R. Five more arguments to invalidate the passive scavenging version of the carnivore-hominid-carnivore model: A reply to Blumenschine et al. (2007a). J Hum Evol. 2007;53:427–433. http://dx.doi.org/10.1016/j.jhevol.2007.05.010

A temporally constrained re-evaluation of temperature inferences from Boomplaas and isotope records from Cango Caves: Comments on Thackeray (2016)

AUTHOR:
Jennifer M. Fitchett[1]

AFFILIATION:
[1]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Jennifer Fitchett

EMAIL:
jennifer.m.fitchett@gmail.com

POSTAL ADDRESS:
Evolutionary Studies Institute, University of the Witwatersrand, Private Bag 3, Wits 2050, South Africa

KEYWORDS:
Holocene; climate fluctuation; palaeoclimate reconstruction

HOW TO CITE:
Fitchett JM. A temporally constrained re-evaluation of temperature inferences from Boomplaas and isotope records from Cango Caves: Comments on Thackeray (2016). S Afr J Sci. 2016;112(7/8), Art. #a0161, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0161

The Holocene has been characterised by rapid climatic fluctuations.[1,2] The climatic and biogeographical variations across southern Africa have presented challenges in accurately reconstructing the palaeoenvironmental and palaeoclimatic record for this period.[3] Difficulties have included the integration of palaeoclimatic reconstructions from multiple locations and the identification of contemporaneous climatic events.[4] The improvement of dating methods has facilitated temporally well-constrained comparisons and reduced the margin of errors in the ascription of ages to environmental and climatic anomalies inferred from proxy evidence.[5]

Thackeray[6] provides a comparison of his previous temperature record inferred from mammalian and insectivore records from Boomplaas[7] with Talma and Vogel's[8] speleothem isotope data from the nearby site Cango Cave. Notable are similarities in peaks and troughs demonstrated in the two records, argued to represent contemporaneous climatic events.[6] These records are not, however, presented on comparable axes, with the Cango Cave record plotted against calibrated years before present (cal yr BP), while the Boomplaas inferences are plotted against sample number. The sample numbers from Boomplaas have been demonstrated to be largely sequential, yet referring to Thackeray's Table 1[6], sample 57 is dated at 6400 years before present. If the existing radiocarbon ages, as presented by Thackeray[6,7], are to be used to argue temporally sequential samples, these ages should also be used in comparison between sites. Importantly, as the Cango Cave record presented terminates at 6000 cal yr BP, the radiocarbon date presented for sample 57 indicates that this and subsequent sample numbers plotted in Thackeray's Figure 1 precede the Cango Cave record.[6] The alignment of peaks and troughs for the two records is therefore not temporally consistent.

To facilitate direct comparison between these two records, the radiocarbon dates for Boomplaas require calibration, and it is necessary to interpolate ages for the Boomplaas samples for which there are not yet radiocarbon or other dates. The previously published radiocarbon dates[7] are therefore calibrated using ShCal13, with dates for the remaining samples interpolated by depth using Bacon, a Bayesian age-depth model using Markov Chain Monte Carlo simulation[9] (Figure 1).

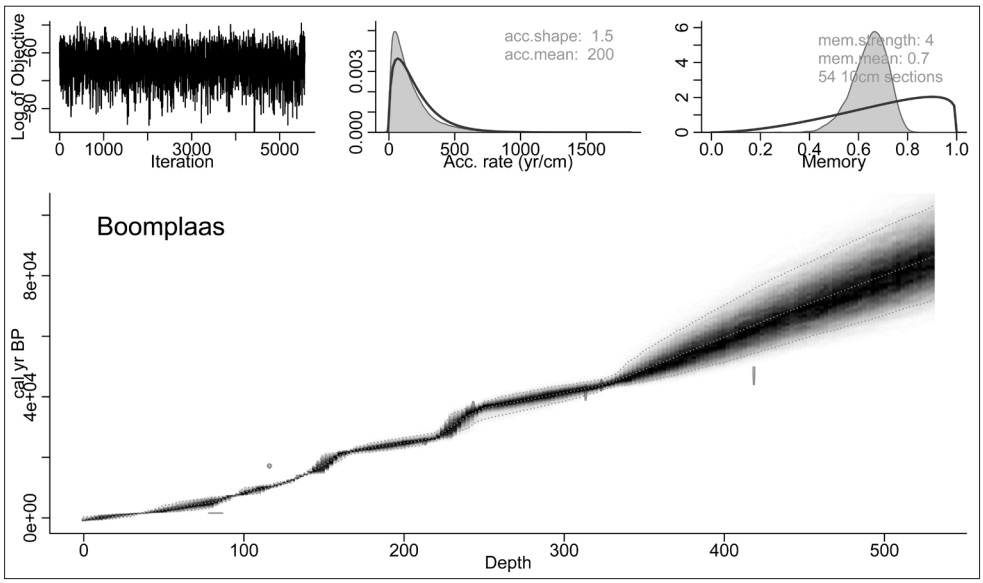


Figure 1: Bacon age-depth interpolation for Boomplaas Cave, based on the published radiocarbon ages[7].

Plotting the Boomplaas record against the calibrated ages, the significantly coarser temporal resolution of the Boomplaas record becomes apparent; much of the detail in the Cango Cave record is not captured in the Boomplaas sequence, for which only a more smoothed trend can be produced (Figure 2). This is not uncommon when making comparisons with speleothem isotope data, for which a very high resolution of analysis is permitted.[10] Only one of Thackeray's identified peaks remains, yet notable similarities exist between the records (Figure 2). The peak in the Boomplaas sequence corresponds very closely with the largest peak in the Cango Cave records, and most likely represents the height of the Holocene Altithermal (Figure 2). There is a progressive decrease in the inferred temperatures and isotope records with time (Figure 2). This pattern is consistent with the transition from the Holocene Altithermal to the slightly cooler contemporary conditions.[4] Thackeray[6] is therefore correct in his argument that there is a notable correlation between the two records, but the coincidence of events appears to be at a coarser temporal resolution than he presents.

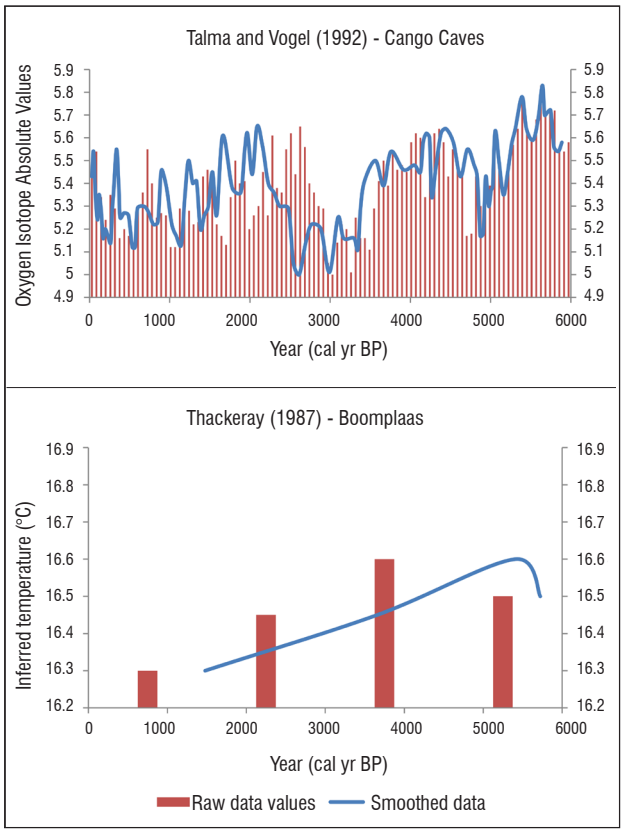


Figure 2: Comparison of the Boomplaas Cave and Cango Cave records, with the Boomplaas sequence constrained by calibrated radiocarbon ages.

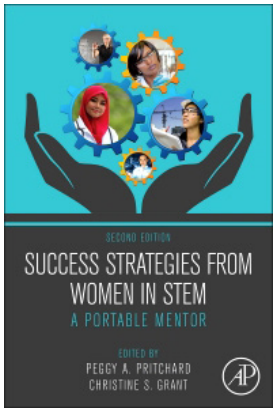
With increased access to dating facilities, it is surprising for contemporary publications to plot results against sample number; an abstract measure to any reader unfamiliar with the study site in question. It would be of interest to understand the reason for the omission of existing radiocarbon dates in the plot, and for the inclusion of samples 57–59 which, by the age data presented in the paper, pre-date the Cango Cave sequence. Although methods of 'wiggle matching' are increasingly being accepted, these usually require that such adjustments occur within the error range of the measured dates, and are performed through statistical analyses of the fit of each date to the calibration curve.[11] Thackeray presents a very interesting comparison between two sites of relatively close proximity, and argues for the simplification of temperature inferences.[6] To derive the maximum benefit from these comparisons it is important to understand the temporal chronologies of the two sites, and the logic involved in making visual over statistical comparisons. Once these concerns have been resolved, Thackeray's arguments regarding the potential for reducing temperature inferences from isotope records to more simplistic regression models[6] can be explored with greater confidence.

References

1. Mayeweski PA, Rohling EE, Stager JC, Karlén W, Maasch KA, Meeker LD, et al. Holocene climate variability. Quaternary Res. 2004;62(3):243–255. http://dx.doi.org/10.1016/j.yqres.2004.07.001
2. Wanner H, Mercolli L, Grosjean M, Ritz SP. Holocene climate variability and change: A data-based review. J Geol Soc London. 2015;172(2):254–263. http://dx.doi.org/10.1144/jgs2013-101
3. Meadows ME. Seven decades of Quaternary palynological studies in southern Africa: A historical perspective. T Roy Soc S Afr. 2015;70(2):103–108. http://dx.doi.org/10.1080/0035919X.2015.1004139
4. Scott L, Neumann FH, Brook GA, Bousan CB, Norström E, Metwally AA. Terrestrial fossil-pollen evidence of climate change during the last 26 thousand years in southern Africa. Quaternary Sci Rev. 2012;32:100–118. http://dx.doi.org/10.1016/j.quascirev.2011.11.010
5. Meadows ME. Recent methodological advances in Quaternary palaeoecological proxies. Prog Phys Geog. 2014;38(6):807–817. http://dx.doi.org/10.1177/0309133314540690
6. Thackeray JF. Comparison of Holocene temperature data (Boomplaas Cave) and oxygen isotope data (Cango Caves). S Afr J Sci. 2016;112(5/6), Art. #a0156, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0156
7. Thackeray JF. Late Quaternary environmental changes inferred from small mammalian fauna, southern Africa. Climatic Change. 1997;10:285–305. http://dx.doi.org/10.1007/BF00143907
8. Talma AS, Vogel JC. Late Quaternary paleotemperatures derived from a speleothem from Cango Caves, Cape Province, South Africa. Quaternary Res. 1992;37:203–213. http://dx.doi.org/10.1016/0033-5894(92)90082-T
9. Blaauw M, Christen JA. Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian Anal. 2011;6(3):457–474. http://dx.doi.org/10.1214/ba/1339616472
10. Holmgren K, Karlén W, Lauritzen SE, Lee-Thorp JA, Partridge TC, Piketh S, et al. A 3000-year high-resolution stalagmite-based record of palaeoclimate for northeastern South Africa. Holocene. 1999;9(3):295–309. http://dx.doi.org/10.1191/095968399672625464
11. Ramsey CB, Van der Plicht J, Weninger B. "Wiggle matching" radiocarbon dates. Radiocarbon. 2001;43(2):381–390.

A primer for success in science

BOOK TITLE:
Success strategies from women in STEM: A portable mentor (2nd ed)



EDITORS:
Peggy A. Pritchard,
Christine S. Grant

ISBN:
9780123977755 (ebook)

PUBLISHER:
Academic Press, San Diego;
USD44.95

PUBLISHED:
2015

REVIEWER:
Brenda Wingfield[1]

EMAIL:
Brenda.Wingfield@up.ac.za

AFFILIATION:
[1]Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

POSTAL ADDRESS:
Faculty of Natural and Agricultural Sciences, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

HOW TO CITE:
Wingfield B. A primer for success in science. S Afr J Sci. 2016;112(7/8), Art. #a0165, 1 page. http://dx.doi.org/10.17159/sajs.2016/a0165

This book is extremely useful and is strongly recommended. However, the language is dense to read; it is not the kind of book that can be read from cover to cover in one or two sittings – or even three. I suspect that the authors never meant it to be read this way, as each chapter is really a stand-alone entity.

My overriding comment is that this book is not for only women. The strategies, in most cases, are as useful for men as they are for women. Indeed, I would recommend this book as essential reading matter for *all* young academics in the field of science, technology, engineering and mathematics (STEM). That is not to say more mature academic readers could not learn from the book, but many of the strategies discussed are perhaps most useful for those who are starting out or are still doing their graduate studies. I certainly would have liked to read many of the chapters while much younger. Of course, the question is whether I would have really understood what I was reading and whether I would have acted on the information.

The book starts with a very interesting chapter on career management, followed by a chapter on networking and then a chapter on mentoring. These three chapters overlap, but the messages are very clear. With the hindsight of a long career in STEM, I support much of what is covered and suggested. Chapter 4, on mental toughness, is in a different vein but equally important. It is one of the chapters that focus on women and it highlights some of the common mistakes women make. Chapter 5, titled 'Time stress', highlights some important issues – including one of my personal favourites, the issue of juggling 'important' versus 'urgent' tasks. The chapter includes a very useful table which explains the differences. I would be tempted to say that everyone should have this table displayed on a wall where they can see it daily.

Chapter 6 rather nicely talks about personal style. Often young women – and men – do not fully believe the idiom 'clothes make the man'. First impressions do count, and you need to think about what these impressions say about you. Clothes are not everything; the chapter also discusses action and voice. For example, women's voices and speech patterns are often softer than men's, and may be less likely to command respect and attention. Chapter 7 follows up on this topic rather well. It provides sound advice for how scientists should communicate, and includes a list of dos and don'ts – which I thought was particularly appropriate.

Chapter 8 is an obvious addition to the book, as it talks about social media. This is something I have less experience with, and I will go back to the chapter to learn more. This chapter is followed by a chapter on negotiation, which again should be read by both men and women. The most useful 'take home' message is that negotiations are a process – something our younger generations are less likely to understand. Our young scientists are part of Generation Y and expect things to happen instantaneously, but negotiations do not happen this way.

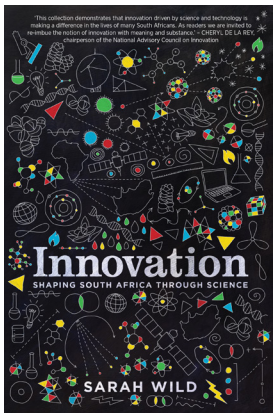
The chapter on leadership is very well written. It too contains some useful tables, but is extremely dense. I think the most important message of the chapter is that one needs to define one's own style. However, this message gets a bit lost in the detail, despite the many tables being important and interesting. I found the chapter on 'climbing the ladder' less compelling, and the penultimate chapter on balancing one's professional and personal life was unsatisfying. Some of the statistics given in this chapter are quite alarming, and suggest that the most common reason for women not to continue in full-fledged academic careers relates to child bearing and rearing. I am not convinced there is anything balanced about reaching the top rungs of an academic career. One of the problems is that we need to be clearer, early on in mentoring young academics – particularly young women – about what it takes to succeed in academia. This is not to frighten people off but to enable them to deal with the reality and plan accordingly. Such mentorship should include giving them the tools to cope. Somehow this is not done, and women end up having to make difficult decisions in a vacuum.

The book ends with a chapter on transitions. This chapter is another that contains messages for men as well as women. It is also applicable not only to STEM but to all careers. The examples used are for STEM, and tend to focus on the fact that people in STEM are likely to move around geographically, and this is likely to make transitions from one stage to the next – such as school to university – quite complex.

This book is a very useful summary of many different aspects that STEM professionals are likely to encounter as they climb the professional ladder. The book includes many different stories and in this way gives examples of situations STEM professionals are likely to encounter. All examples are drawn from women's lives and this makes the book perhaps more accessible to women. However, all the examples come from North America, making the book perhaps a bit biased. Despite this, I found that many of the examples resonated strongly. I suggest that all young academics and STEM professionals should have this book as a reference to use as they embark on their careers. I will be recommending particular chapters to some of the academics I mentor.

From the ocean to outer space – and almost everything in between

BOOK TITLE:
Innovation: Shaping South Africa through science



AUTHOR:
Sarah Wild

ISBN:
9781770104389 (softcover)

PUBLISHER:
Pan Macmillan (in partnership with Gordon Institute of Business Science), Johannesburg; ZAR215

PUBLISHED:
2015

REVIEWER:
Tim G.B. Hart[1,2]

EMAIL:
thart@hsrc.ac.za

AFFILIATIONS:
[1]Economic Performance and Development, Human Sciences Research Council, Pretoria, South Africa

[2]Department of Sociology and Social Anthropology, Stellenbosch University, Stellenbosch, South Africa

POSTAL ADDRESS:
Economic Performance and Development, Human Sciences Research Council, Private Bag X41, Pretoria 0001, South Africa

HOW TO CITE:
Hart TGB. From the ocean to outer space – and almost everything in between. S Afr J Sci. 2016;112(7/8), Art. #a0160, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0160

If you have ever wondered how the innovation sector in South Africa has transformed since 1994, *Innovation,* by Sarah Wild is an extremely informative and interesting publication on the topic. The book is not an academic monograph nor is it a textbook for students. Rather it presents a diversity of the innovations underway in South Africa, along with their challenges and their potential to benefit humankind, in a very easy to read style that avoids unnecessary jargon. One striking value of Wild's style is that in many cases a chapter (describing a single innovation) can be read in under 10 minutes.

Pre-1990, science, technology and innovation in South Africa primarily focused on the military-industrial complex and labour-intensive commodity production. Wild's collection of 30 concise articles, based on personal interviews and encounters over the last decade, suggests that this situation has changed significantly. She highlights five areas of South African society in which transformation is taking place in the application of science, technology and innovation: environment, energy, health, industry and education.

Following a brief and rather sombre introduction, in which Wild summarises the historical and the contemporary local and global challenges dogging the innovation sector in South Africa, the remaining chapters provide captivating accounts of many of the research and development (R&D) innovations in various stages: development, completion, application and commercialisation. The multitude of innovations extend from unmanned ocean-going vessels that monitor weather and carbon data (Chapter 7), to monitoring solar storms to ensure the safety of regional aviation and maritime services (Chapter 3), and satellites in outer space that inform us of fire threats across the continent (Chapter 1) or monitor the water quality in our dams (Chapter 2). In between there are numerous other innovations: potential life-saving genetic research (Chapter 16); mitigation of fresh produce wastage on supermarket shelves (Chapter 4); the possibility of robots in our mines to reduce the human cost of underground mining (Chapter 23); deriving energy from the sun (Chapter 9), to power homes in informal settlements (Chapters 13 and 14), and energy from burning unmined underground coal seams (Chapter 10); cameras to predict power line interruptions (Chapter 12); ultrasonic transducers to ensure railway line safety (Chapter 22); and diamond fingerprinting (Chapter 27). For many, the journey has been exciting and challenging; for others, the journey is far from complete (Chapters 5, 9, 11, 18 and 24) and they need to circumnavigate financial and legal obstacles. A few more recent innovations have been purposefully designed to address specific local societal challenges identified post-1994 (Chapters 13, 14, 28 and 30).

Some of the innovations described are South African in origin (Chapters 6, 9, 20 and 22), or have had a long history in South Africa (Chapters 11, 12 and 19). Others are the improvement of external ideas and inventions through their local adaption to the particular needs of South Africa and similar developing countries, including several countries in sub-Saharan Africa (Chapters 1, 2, 3, 7, 8, 15, 26 and 29).

Most innovations described involve local partners from universities, science councils and private research laboratories. A few include international partnerships, which in some cases simply provide access to costly infrastructure or information that is shared with many countries (Chapters 1, 2, 3 and 11). For other innovations, there is close collaboration in design and research with international partners (Chapters 5, 8, 16, 18, 21 and 27). What is strongly evident throughout most of the book is the diversity of the scientists, engineers and technicians who comprise the innovation sector in South Africa. Innovations arise not from individuals but through the hard work of cosmopolitan, multi-racial, interethnic, gender inclusive and intergenerational teams, all striving to create or expand on new ideas to improve society in the broadest sense, including the environment, industry and the health and contentment of the population. Exclusion from the National System of Innovation (NSI) is no longer the norm, while government remains a key partner in innovation funding, although amounts are less (p. 2–3). Regrettably, social scientists do not feature in these teams, which are predominantly composed of information technologists, engineers and natural scientists, indicating that social scientists are still far from readily acknowledged by the prevailing actors in South Africa's NSI. Another group that appears to take a backseat in most examples are the actual beneficiaries of innovation. In most cases they are unidentified and supposedly indirect beneficiaries. When they are identified, it is often as test subjects in pilot studies (Chapters 6, 13, 14 and 28).

While balanced and very positive in its outlook, the book does not place much emphasis on how innovation in South Africa can tackle the three primary socio-economic challenges of inequality, poverty and unemployment – all of which have increased since 1994. Clearly the knowledge economy is growing. It is diverse in its make-up and skills base, and the innovation sector is steadily moving forward. But as Prof. De la Rey asks in the Foreword (p. vii), how can the objectives of development, social inclusion and innovation be aligned? When Wild considers several examples of improvement of the quality of life of the poor and social inclusion, the focus is heavily towards high-tech scientific solutions, led by the NSI actors rather than the poor, who remain treated largely as passive recipients. Wild acknowledges their significance but confines it to their uptake or adoption of innovations and scientific solutions rather than indicating their ability to innovate in the broader sense of inventing and adapting (Chapters 6, 13, 14 and 28). In reality, it is more about matching the right solutions to specific problems and many of these solutions may not be of a scientific or technical nature. As some scholars note, we should not repeat the regular mistake of trying to interpret social problems as technical and thus suitable to technical interventions, when they are not.[1,2] Rather the solutions might be simple innovations developed out of need by people who are still completely separated from South Africa's NSI, but who encounter such problems on a daily basis.

While Wild reports on some successes and challenges in these examples, many are pilot studies situated on the urban periphery of major centres (housing, toilets) that provide the poor or previously marginalised actors with close access to the developing and implementing partners (Chapters 13 and 14). Others involve pilot studies located further afield in distant rural areas (Chapters 6 and 28). Ultimately many of these scientific solutions are only available to a select few and are unlikely to be rolled out to reach all those in need. The ICT4RED in the iNciba locality surrounding Cofimvaba in the Eastern Cape is a pilot project which is generously funded by the Department of Science and Technology. The concern, as Wild rightly points out, is that the Department of Science and Technology cannot continually (or sustainably) fund such an endeavour and the National and Provincial Department of Basic Education does not have the money to replicate the project across the province or to maintain the existing ITC infrastructure after the current implementing agencies exit. No long-term funders are volunteering to keep the high-tech solutions going forward.

Despite their usefulness and appropriateness to many walks of life, high-tech scientific solutions may be inappropriate in some South African contexts, as they create expectations that they seem unlikely to fulfil because of the challenges, which include infrastructure requirements and the high costs involved in implementation and maintenance.

References

1. Ferguson J. The anti-politics machine: Development, depoliticization and bureaucratic power in Lesotho. Cambridge: Cambridge University Press; 1990.
2. Li TM. The will to improve: Governmentality, development, and the practice of politics. Durham, NC: Duke University Press; 2007. http://dx.doi.org/10.1215/9780822389781

Citizens and cities: A South African review

BOOK TITLE:
Popular politics in South African cities: Unpacking community participation

EDITOR:
Claire Bénit-Gbaffou

ISBN:
9780796924643 (softcover)

PUBLISHER:
HSRC Press, Cape Town; ZAR270

PUBLISHED:
2015

REVIEWER:
Alan Mabin[1]

EMAIL:
alan@alanmabin.org

AFFILIATION:
[1]Wits City Institute, University of the Witwatersrand, Johannesburg, South Africa

POSTAL ADDRESS:
Wits City Institute, University of the Witwatersrand, Room 150, Central Block, East Campus, Braamfontein, Johannesburg, South Africa

HOW TO CITE:
Mabin A. Citizens and cities: A South African review. S Afr J Sci. 2016;112(7/8), Art. #a0166, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0166

This book is a special achievement in that it presents substantial original material, ultimately based on a rather small research grant under the CORUS framework (administered by the French International Development Research Agency). The book shows what can be achieved through dedicated construction of groups of diverse scholars working together over sustained periods – in this case 4 years. Such collaboration should include interaction between graduate students, younger scholars and a few more established figures. In this case, it also included collaboration between geographers based in France and diverse South African academics. Such projects require determined leadership. The editor of this collection clearly played that role over the life of the project – which apart from this book has resulted in several special issues of journals and other individual papers. As a reviewer, I had a minor role in initiating this project but I did not play an active role in research, or in other ways, once the project was under way. I take no personal credit for its success.

The collection is about collective citizen participation in processes intended to achieve significant change in urban politics and local government practices. Although such participation generally does not result in dramatic alterations, many people continue to engage in such processes – and this is the underlying theme of the contributions. The book is divided into two sections. The first section contains five studies of participation in which party politics play a central role in the dynamics of governance. The second section offers six studies, each of which is informed by seeking to go beyond the typical distinction between 'invited' (usually official, institutional) and 'invented' forms of collective expression related to city problems, plans and prospects. The 'invited' forms are usually official and institutional, whereas the 'invented' forms imply something ranging from unruly to revolutionary.

The studies presented are mostly drawn from Johannesburg, with Cape Town being a second focal city. Not quite in the former category is a study of public meetings in Vosloorus, a large township in Ekurhuleni, not far from Johannesburg. A couple of chapters take a more general view of participatory democracy in South Africa – and indeed beyond. The book is introduced by the editor in a dozen compact pages, describing the nature of the process, and setting out key themes and arguments. The closing piece, titled a postscript, tries to 'view South African urban governance from an Indian perspective' – an apparently odd fit, but in this short text Glyn Williams offers a commentary on South African experience in a novel way.

The book presents a wide range of collective citizen activities, which sometimes include 'foreigner' activities – a dangerous distinction indeed in recent months. These activities attempt to redirect public investment, alter governmental direction, or fulfill demands for both small and large changes in urban space and priorities. By no means does a single political organisation dominate in all ways at all times in almost any area presented here.

Luke Sinwell offers a fascinating account of shifting alliances between the nascent Landless People's Movement and the well-known African National Congress (ANC) and Democratic Alliance (DA) in a relatively new area in the Soweto complex. Sinwell's account highlights the lack of stasis that many narratives of politics in South Africa might lead readers to expect. Laurence Piper shows something similar in his study of Imizamo Yethu, one of the classic cases of a generally poor informal area wedged within a high-income, formal space in Cape Town. The editor, writing with two of her students, portrays similar complexity and contradictory loyalties in an inner-city area of Johannesburg (co-authored with Eulenda Mkwanazi) and in a part of Soweto (with Boitumelo Matlala). Obvious Katsaura draws on his doctoral work in Yeoville to explore party attempts to shape direction in a local official forum, again with far more mixed results than a simplistic understanding of political allegiance might suggest. These studies in the first part of the book present important narratives, and raise intriguing questions that political developments alone can answer.

The second part of the book turns towards suggesting newer ways of thinking through citizen collective action in South Africa's young democracy. Daryl Glaser reflects on the unruliness of the citizenry, drawing on both the 'people's power' phase of the township struggle against apartheid in the 1980s, and the violent xenophobia in recent months and years. Questions of political theory, political power, recourse to jurisprudence, and actions beyond the purview of the state intersect in these pages. The mix of legal and courtroom strategy with street action is perhaps the most globally intriguing phenomenon here, with Laila Smith and Margot Rubin explicitly addressing the issues in two Johannesburg cases. Alex Wafer and Sophie Oldfield similarly offer a deeply informed study of a changing neighbourhood in Cape Town.

The nature of 'citizen voice', the expression merely of grievances as well as demands for radically different action, and the apparent simplicity of the 'public meeting' all come under scrutiny in Marianne Morange's work with Cape Town street traders. Philippe Gervais-Lamobony's engagement with Vosloorus residents offers a similar analysis. Chloe Buire provides insights into the relatively unusual Cape Town attempt to 'bring government closer to the people' through its more than 20 subcouncils, in a situation in which all large cities in South Africa have but one elected council structure. The results, as Buire illustrates, have been very mixed.

In general, the methods applied to the cases are mixed and diverse. Some studies draw on archival material, mostly participant observation and interviews, and there is strong use of media and other secondary sources. Most chapters pay specific attention, in limited space, to the nature of the research. More attention could have been given to the limitations of the work done, and to future research possibilities.

A few themes link this volume together more successfully than many an edited collection, perhaps because of the process through which the authors connected over several years. One theme is how various forms of dominance are reproduced in local political contexts. These contexts may take the form of political party dominance through

patronage networks; a 'dominance of disempowerment' through almost endless and sometimes meaningless activities; or even participation in the life of institutions that perpetuate domination, by the very people who are dominated – as Morange argues in the case of Cape Town street traders. Another example is the ways in which local participation both opens and closes creative and positive relationships between the state and citizens. Open exploration of these themes is not as common as one might have wished for in South Africa (or indeed many other countries). For this reason, the volume is well worth the attention of all readers who are interested in the direction of change within South African cities.

Naturally, there is room for doubt and dispute around some of the material in the book. Not all claims made are supported by evidence. For example, Sinwell suggests that 'where the DA is in power in the Western Cape, it has had arguably similar exclusionary effects to the ANC nationally' (p. 79). This statement might or might not be correct; either way, the support cited for this statement in a footnote has to rank as weak. The claim seems unnecessary in the midst of Sinwell's rare account of complexities of party and other political formations. Similarly, in Buire's chapter, it seems a slightly limited view – even in the light of her own evidence – to go beyond the important study she presents and claim that Cape Town's subcouncils are merely about 'Christmas trees' and not 'substantive palpable' matters (p. 222).

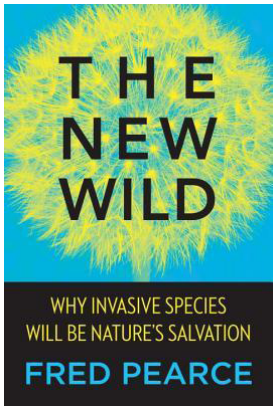
Some readers might find they are left with a nagging feeling that the studies in the book are mostly based on strong, but not always stated, views on the state and 'popular politics'. The phrase 'popular politics' could suggest that collective views framed outside formal processes can somehow be more legitimate than the processes of 'the state'. Other readers might feel the work is based on sources just a little too narrow to justify some of the strong conclusions put forward.

One is reminded, by the overwhelming presence of Johannesburg and Cape Town in the pages of this volume, that studies on other South African cities are sorely lacking. Although Ethekwini (Durban) receives more attention than most cities, the book focuses mainly on just these two of South Africa's many cities. Most of the other cities and towns – whether small, medium or large – are effectively sidelined. A great deal of the 'celebrated civic activism' (p. 289) associated with South Africa has historically occurred in places such as Cradock, Emalahleni (Witbank) and Nelson Mandela Metro (Port Elizabeth). It was in Nelson Mandela Metro that the first powerful 'black township civic' arose, after all, in the 1970s.

In coming years, one would expect to see views that confound – or at least complicate – some of the arguments in the present volume, as more work takes place across a broader variety of city spaces in South Africa. In the same way, many other kinds of citizen activity (some but by no means all addressed in other publications emanating from the same project) could usefully receive more attention in future. The opportunity is indeed wide open and will surely be stimulated by reading this provocative volume. The subject will remain of great interest, as political contest among various forces – particularly at local level – is likely to accelerate through 2016 and beyond. This contest will have profound implications for the future of the country and for its citizens both individually and collectively.

The new wild: The uncomfortable truth

BOOK TITLE:
The new wild: Why invasive species will be nature's salvation



AUTHOR:
Fred Pearce

ISBN:
9780807033685 (hardcover)

PUBLISHER:
Beacon Press, Boston;
USD26.95

PUBLISHED:
2015

REVIEWER:
Roy A. Lubke[1]

EMAIL:
r.lubke@ru.ac.za

AFFILIATION:
[1]Department of Botany, Rhodes University, Grahamstown, South Africa

POSTAL ADDRESS:
Department of Botany, Rhodes University, Grahamstown 6140, South Africa

HOW TO CITE:
Lubke RA. The new wild: The uncomfortable truth. S Afr J Sci. 2016;112(7/8), Art. #a0163, 2 pages. http://dx.doi.org/10.17159/sajs.2016/a0163

Fred Pearce, an environmental journalist from London, has produced a book that will set the proverbial cat amongst the pigeons of restoration ecology. For some decades, ecologists have been set on eradicating alien invasive species and restoring pristine environments. In this book, Pearce exposes their efforts as fruitless, because there is no such thing as a pristine environment on earth. All environments have been disturbed by humans. Pearce is also of the opinion that many of these so-called harmful introduced species may in fact be beneficial in cleaning up degraded or polluted environments.

Pearce has extensive experience in scientific reporting – he has written five books on similar topics. These topics include the demise of earth's water resources, the human population crisis and climate change. Now he attempts to provide an optimistic outcome to human 'interference' in nature, by accepting that invasive species are part of 'the new wild'. This view directly contrasts that of Paddy Woodworth[1] who, in an analysis of restoration projects around the world, champions the work of restoration ecologists by describing their successful projects. Pearce quotes many examples in which alien species have simply fitted into the new habitats, often benefitting the changed ecosystems. Is there a middle road where both approaches are compatible? A more detailed analysis of *The New Wild* might answer this question.

In Part 1 of the book, titled 'Alien empires', Pearce analyses several alien invasions and the 'myths' behind these invasions, many of which are commonly reported in the popular press. Pearce contends that alien species are, with human help, moving into disturbed places where they are often 'helping nature's recovery'. The first chapter (On Green Mountain) is about islands, where some of the greatest damage to ecosystems has occurred. Pearce states that the 'malign invaders were simply taking advantage of ecosystems that had already been wrecked by humans' and the invaders 'were often doing jobs that natives could not accomplish' as they were regenerates of the disturbed environment. For example, on Ascension Island in the South Atlantic, 90% of the biota now comprises alien sources, and it is unlikely that restoration is possible. However, the loss of rare and endangered species, such as birds from other islands that are sparsely populated, is hardly comparable, and there is an inconsistency in Pearce's theory. Should we simply accept that on most islands, 'alien species add to local diversity and enrich species-poor ecosystems' – even when rare species are being decimated on remote islands? Ecologists working to save colonies of seabirds that have been ransacked by invasive rats or cats are unlikely to accept this view.

Chapter 2 (New Worlds) is on the accidental introduction of weeds, pests and diseases. Pearce attempts to lessen the blow caused by these 'villains', explaining that their presence is the result of people's pollution or other disturbances of habitats, and contending that such species fit in with the 'new ecosystem'. Ecologists and conservationists who witness disturbances to aquatic systems by invading water weeds or thickets of introduced alien woody plants displacing grasslands or fynbos, would hardly find this idea acceptable. Pearce also examines the invasion of marine habitats by alien species (Chapter 3) and the alien invaders of America (Chapter 4) and Britain (Chapter 5). Top of the British hate list for weeds is Japanese knotweed. Pearce gives an almost comical account of the ridiculousness of being overly obsessed with alien invasive species. He analyses the costs of clearing the weed, citing claims of a '$250-million-a-year problem – or a pint of beer for every person in the land', which is disputed by some ecologists working on the weed.

In Part 2 of the book, titled 'Myths and demons', Pearce refers to 'our misplaced notions about how aliens affect the real world and how we do conservation'. This section of the book is likely to rile conservationists, restoration ecologists and biological control scientists, especially in South Africa.

Chapter 6 (Ecological Cleansing) reports on the mega-project of Working for Water in eradicating alien woody species that are annihilating natural ecosystems and sapping South African water supplies. Pearce rather frivolously considers this a 'rarely successful' process of ecological cleansing. His opinion is based largely on the high cost of the operation for very little return of land restored, and the fact that alien seeds remain in the soil – meaning alien plants are likely to return. He argues that the tree species being eradicated in South Africa are encouraged for use in arid northern areas in Africa, to increase 'the water-holding capacity of soils and generate income for farmers'. There is no mention of the aesthetics of an environment of alien thicket species compared with restored mountain fynbos. Pearce also does not mention the loss of grazing land to alien invasions. He describes many other studies of eradication, and the chapter ends thus: 'Conservationists, it seems, are dedicated to protecting the weak and vulnerable, the endangered and the abused. Nature generally promotes the strong and the wily, the resilient and versatile.' The reader is left to wonder whether this is an argument to allow invasive species to take over.

Chapter 7 (Myths of the Aliens) charts the development of 'invasion biology'. Pearce begins by discussing the work of Elton, who wrote *The Ecology of Invasions by Animals and Plants* (1958) and promoted the study of 'ecological explosions' of alien biota. Pearce cites many alleged facts and figures about alien species and the contradictions made by ecologists. He concludes that despite all the inconsistencies and 'myths' about invasive species, there is still a strong call for academics to support action against invasive species.

Most of Chapter 8 (Myths of the Pristine) describes so-called pristine environments such as tropical forests, nearly all of which are in fact disturbed or influenced by people. There is overkill in Pearce's argument here. Although many ecologists agree that the pristine environment concept is overstated, numerous sites do still exist on the globe that have been little affected by humans.

Chapter 9 (Nativism in the Garden of Eden) explores Clements' idea of collective associations of plants, and contrasts it with Gleason's concept of individualistic action of species. Gleason's concept fits Pearce's view that alien species introduced into new habitats act individualistically and can fit in with the ecosystem. The ecological

concept of associations is derided as it does not suit Pearce's ideas on the new wild. Pearce cites succession on the newly formed volcanic island of Surtsey, near Iceland, as an example of the haphazard formation of a new ecosystem – such as in his concept of the new wild. Many ecologists will question the shallowness of his arguments to advance his concepts.

Having described the myths, in the final part of his book (The New Wild), Pearce attempts to offer some solutions. He believes we need to change our ideas on conservation and 'lose our dread of the alien and the novel'. Alien species and novel ecosystems will be the heart of the new wild.

In Chapter 10 (Novel Ecosystems), Pearce never really defines what restoration ecologists mean by 'novel ecosystems'. Restoration ecologists have discovered that in attempting to recreate the original system, an ecological threshold is often encountered and is very difficult to overcome. This 'new' unoriginal state is the so-called novel ecosystem. By contrast, Pearce's concept of a novel ecosystem is one in which humans have introduced alien species, often in a disturbed environment, and 'the system itself does not depend on humans to keep it going'. He cites examples such as Puerto Rico, where disturbed secondary forests are invaded by and contain many introduced species. He quotes Lugo et al.[2] as calling this the 'new wild' where transformed sites are 'beautifully functioning new forests, with greater biodiversity than the old forests'.

In Chapter 11 Pearce promotes railway sidings, landfill sites, graveyards, market gardens and brownfield sites as unique habitats, where many rare and endangered plant species find refuge and protection. Such sites are part of the new wild. He also describes how many animals find cities safer than the countryside. The ugliness of these environments and how they would mesh with graceful parks and gardens is not considered.

In the concluding chapter (Call of the New Wild), Pearce explores the effects of climate change and hybridisation of species, and how these trends will contribute to changes in community assemblages in creating the new wild. He feels that in this century, biologists should not fight a losing battle to protect 'pristine nature' but should rather encourage the rebirth of nature, and accept invasive alien species.

Is there a middle road between alien eradication on the one hand, and, on the other, restoration of environments and acceptance of the new wild, such that both approaches are compatible? The answer is 'No'. If biologists were to accept these concepts, they would have to give up all they have attempted to achieve in conserving endangered species and restoring 'pristine' environments – often regarded as 'wilderness' areas. Pearce disregards aesthetics and the beauty of the natural environment, which ecotourists travel to view every day. His arguments also make no mention of how to use the new wild. Environmentalists are closely involved in trying to maintain or restore desirable sites for appropriate use by people. Land use is not considered in Pearce's book, and I also see no tranquillity as a 'sense of place' in the new wild.

Despite the lack of depth of research and the shortcomings mentioned above, there is a need for a book like this to be widely read by the scientific community. Firstly, Pearce has noted that he is 'far from alone' in his concern 'that we have bought into some dangerous mythology about how nature works'. We need to re-examine the literature in which there is opposition to the eradication of invasive aliens and ecosystem restoration. Perhaps we have overlooked, as Pearce claims, figures regarding the danger of aliens and costs of their eradication. Secondly, experts must interrogate Pearce's arguments and pose counter-arguments to their students and in their publications. Thirdly, experts working on projects of alien eradication or ecosystem restoration should read this book carefully to enforce their requests for funding, because of the question Pearce raises about the possible waste of such funding. Fourthly, environmentalists, ecologists and conservationists involved in planning and environmental impact assessments might find his arguments (such as 'there is no such thing as a pristine ecosystem') used by developers who wish to push their projects through the environmental impact assessment or planning phases. On such grounds, scientists and the concerned public should retain the useful ideas in Pearce's book – and reject the rest.

Finally, what is the uncomfortable truth in this book? However scientists feel about its concepts, what Pearce proposes as the 'new wild' will certainly occur sometime in the future. On Ascension Island, it is happening already. Ecologists cannot indefinitely eradicate all aliens from 'pristine' or even slightly disturbed environments. We will have to limit our resources to some areas, while allowing new systems – the new wild – to develop in others. Moreover, humans might not be present on this planet forever. Then, as Pearce claims in his final paragraph, 'nature never goes back; it always moves on'. The vagabonds or aliens will be the pioneers and colonists in this constant renewal, and the new wild will become a reality across the planet.

References

1. Woodworth P. Our once and future planet: Restoring the world in the climate change century. Chicago, IL: University of Chicago Press; 2013. p.515.
2. Lugo AE, Carlo TA, Wunderle JM. Natural mixing of species: Novel plant–animal communities on Caribbean Islands. Anim Conserv. 2012;15(3):233–241. http://dx.doi.org/10.1111/j.1469-1795.2012.00523.x

A multi-disciplinary review of late Quaternary palaeoclimates and environments for Lesotho

AUTHORS:
Jennifer M. Fitchett[1,2]
Stefan W. Grab[1]
Marion K. Bamford[2]
Anson W. Mackay[3]

AFFILIATIONS:
[1]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa
[2]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa
[3]Environmental Change Research Centre, Department of Geography, University College London, London, United Kingdom

CORRESPONDENCE TO:
Jennifer Fitchett

EMAIL:
jennifer.m.fitchett@gmail.com

POSTAL ADDRESS:
Evolutionary Studies Institute, University of the Witwatersrand, Private Bag 3 Johannesburg, Johannesburg, Gauteng, WITS 2050, South Africa

DATES:
Received: 04 Feb. 2016
Revised: 09 Mar. 2016
Accepted: 09 Mar. 2016

KEYWORDS:
Lesotho; archaeology; geomorphology; palaeobotany

HOW TO CITE:
Fitchett JM, Grab SW, Bamford MK, Mackay AW. A multi-disciplinary review of late Quaternary palaeoclimates and environments for Lesotho. S Afr J Sci. 2016;112(7/8), Art. #2016-0045, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20160045

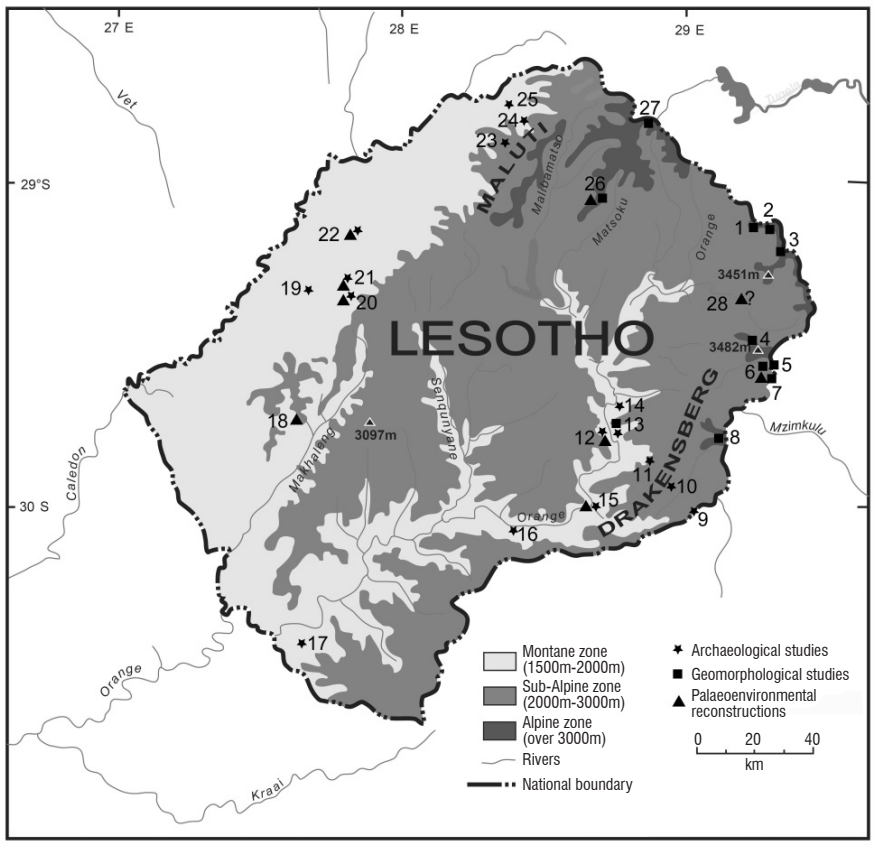
Lesotho provides a unique context for palaeoclimatic research. The small country is entirely landlocked by South Africa, yet has considerable variation in topography, climate, and associated vegetation over an approximate east–west transect. The region has been of archaeological interest for over a century, and hosts many Early to Late Stone Age sites with occupation preceding 80 000 years before present. The eastern Lesotho highlands are of interest to periglacial and glacial geomorphologists because of their well-preserved relict landforms and contentious evidence for permafrost and niche glaciation during the late Quaternary. However, continuous proxy records for palaeoenvironmental reconstructions for Lesotho are scarce and hampered by a range of methodological shortfalls. These challenges include uncertain ages, poor sampling resolution, and proxies extracted from archaeological excavations for which there may be bias in selection. Inferences on palaeoclimates are thus based predominantly on archaeological and palaeogeomorphological evidence for discrete periods during the late Quaternary. This review paper presents a more detailed multidisciplinary synthesis of late Quaternary conditions in Lesotho. We simultaneously considered the varying data that contribute to the under-studied palaeoenvironmental record for southern Africa. The collective palaeoenvironmental data for eastern Lesotho were shown to be relatively contradictory, with considerable variations in contemporaneous palaeoclimatic conditions within the study area. We argue that although methodological challenges may contribute to this variation, the marked changes in topography result in contrasting late Quaternary palaeoenvironments. Such environments are characterised by similar contrasting microclimates and niche ecologies as are witnessed in the contemporary landscape. These spatial variations within a relatively small landlocked country are of importance in understanding broader southern African palaeoenvironmental change.

Introduction

Southern African palaeoenvironmental science has developed rapidly over recent decades, clarifying many key debates on environmental and climatic boundaries, and relating key climatic events with those of the northern hemisphere.[1] Whilst this work spans an increasing number of study sites, these are predominantly determined based on their accessibility and the availability of well-preserved proxies.[1,2] To date, site selection for palaeoenvironmental research in southern Africa has been predominantly based on ease of access. Therefore, sites with highly vulnerable niche ecologies, heightened spatial ecological variation owing to rapid topographic changes, or which have particularly well established archaeological records are often neglected. Lesotho provides such a study region, yet relative to the surrounding southern African countries, the late Quaternary climatic and environmental record remains uncertain and at best undefined.[3,4] Existing literature for Lesotho is based almost entirely on archaeological and geomorphological evidence that spans discrete periods of human occupation or glacial and periglacial activity (Figure 1). Such past research has offered numerous inferences on possible palaeoclimatic changes throughout the late Quaternary in Lesotho[5,6], but the chronological continuity and quantification of past climates lack both detail and objective confirmation[7-9].

A variety of Quaternary periglacial and glacial studies on Lesotho's eastern high mountain region have emphasised colder conditions during the late Pleistocene, including but not limited to the Last Glacial Maximum (LGM)[6,10-13] and Holocene neoglacial episodes. The latter events are presumed to have occurred at ~4500 years before present (yr BP), 4000 yr BP, and 3000–2000 yr BP respectively.[14] However, the absence of ages for most geomorphic phenomena studied, the lack of consistent chronologies, and the often disputed interpretation of landforms have limited their palaeoenvironmental value.[4,8,13,15] The paucity of palaeoenvironmental work in the Lesotho region (Figure 1) stems largely from the difficulty in accessing sites in this mountainous country, and the considerable logistical challenges in extracting material[16]. However, these challenges provide tremendous impetus for such work in the region because of the unique high-altitude setting, with resultant vulnerable niche ecosystems[3,15].

We explore the current understanding of Lesotho's late Quaternary palaeoclimatic and palaeoenvironmental history, critically analysing the evidence published to date from archaeological, palaeogeomorphological and palaeoecological records. Our results highlight the importance of this region as a key site for Quaternary Science on the African sub-continent. The aim here is to provide a review of the three disciplines, followed by a synthesis of the palaeoclimatic data recorded to date. In so doing, we offer a robust framework for future research. This framework can guide palaeoenvironmental and palaeoclimatic research in the region, interrogate valuable proxies, extend studies into under-studied regions, and address outstanding questions and uncertainties.



1. Mafadi; 2. Njesuthi; 3. Popple Peak; 4. Thabana Ntlenyana; 5. Tsatsa-la-Manguaung; 6. Sani Top;
7. Sekhokong; 8. Leqooa; 9. Belleview; 10. Ha Soloja; 11. Moshebi's Shelter; 12. Likoaeng; 13. Sehonghong;
14. Mashai; 15. Melikane; 16. Bohlahla; 17. Masitise; 18. Tsoaing River; 19. Leqhetsoana; 20. Tloutle;
21. Ha Makotoko; 22. Ntloana Tsoana; 23. Liphofung; 24. Muela; 25. Hololo Crossing; 26. Tlaeeng;
27. West of Mont Aux Sources; 28. Eastern Lesotho Wetland

Figure 1: The location of archaeological, geomorphological and palaeoenvironmental study sites in Lesotho from which published palaeoclimatic inferences have been made.

Lesotho: A key site for palaeoenvironmental research

Considerable work has been undertaken to reconstruct palaeo-environments for southern Africa, and in recent years reviews of such work have been published[1,17]. However, there exists little synthesis of the research (albeit sparse) undertaken in Lesotho. Not only does the neglect of this country form a notable research gap, but as the highest altitude region in southern Africa, eastern Lesotho reflects a unique environment which has been significantly influenced by snow and ice throughout the late Quaternary[18]. Western Lesotho, by contrast, is situated at low elevation, consisting of lowlands and foothills.[19] The contrast in elevation, topography, and resultant ecosystems across the country presents a valuable context for examining environmental lapse rates and topographic influences on the environment. Furthermore, synoptically this region is important as much of the moisture derives from the Indian Ocean, yet winter snow is a function of the passage of mid-latitude cyclones originating from the southwest of southern Africa[18]. A critical review of the existing palaeoenvironmental literature for this region is thus of importance in directing future research to more accurately address key research uncertainties.

Lesotho is located between the coordinates 28°30' S to 30°40' S and 27°00' E to 29°30' E, and covers a terrestrial area of 30 355 km^2.[20] The country is landlocked, entirely surrounded by provinces of the Republic of South Africa, namely Free State, KwaZulu-Natal and Eastern Cape. Lesotho is classified into four main physiographic zones: the western lowlands, ranging in elevation from 1500 masl to 1800 masl; the foothills, ranging from 2000 masl to 2500 masl; the Senqu River valley; and the eastern highlands, which include elevations greater than 2500 masl.[20] The eastern highlands comprise 75% to 80% of the total land area, but the majority of the population lives in the lowlands, where the climate, soils and topography are most suitable for agriculture[21]. The high altitude and steep slopes contribute to a set of flora and fauna markedly different from those observed for much of southern Africa.

Eastern Lesotho accounts for a substantial portion of the Drakensberg Alpine Centre (DAC), a biome listed as a centre of endemism. Increasing efforts have been made to protect and understand the biodiversity of the region.[22] The eastern Lesotho Drakensberg is a deeply dissected plateau, with elevations ranging from 2290 masl to 3482 masl[23], and all exposed and underlying rock in the eastern Lesotho Drakensberg forms part of the Karoo Supergroup[16]. This includes the Drakensberg Group flood basalts of Upper Triassic to Jurassic age, intruded by fine-grained dolerite dykes, underlain by the Stormberg Group of sandstone – including the Clarens Formation, which is exposed to an altitude of ~2500 masl, and the underlying Elliot and Molteno Formations[24]. Climate data for Lesotho are insufficient relative to the varied climate across the region, and historical data represent an even more sparse set of locations of uncertain accuracy[8]. However, the contemporary climate of eastern Lesotho can broadly be described as a distinct seasonal alpine climate with cool wet summers and cold dry winters.[25] Spanning a trajectory from the eastern Lesotho highlands to the western lowlands, temperatures increase while rainfall and humidity decrease. The climate of eastern Lesotho, as well as the broader country, is influenced predominantly by altitude and distance from the rain-shadow of the Drakensberg Mountains, but there is considerable local variation arising from aspect and temperature inversions[4,16].

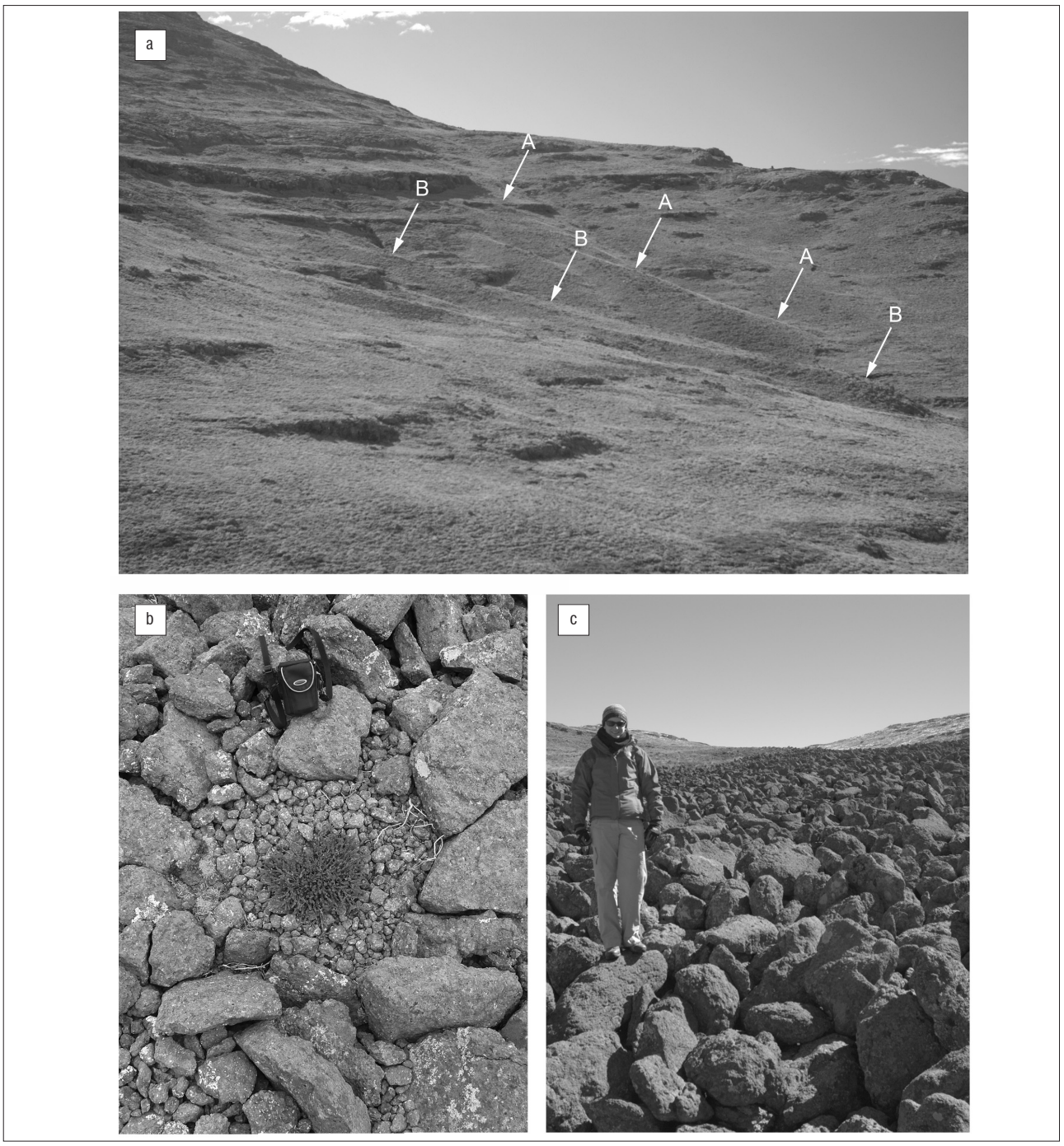


Figure 2: Examples of relict glacial and periglacial geomorphic features in eastern Lesotho: (a) glacial moraines on the south-facing slope of Leqooa Valley (ridges are marked A and B), (b) large sorted patterned ground at Mafadi summit (~40 cm in diameter), and (c) a blockstream near Sani Pass.

The palaeogeomorphological record

Many inferences on past climates for Lesotho have been made on the basis of periglacial and glacial palaeogeomorphic evidence. The eastern Lesotho highlands host an impressive array of relict glacial and periglacial geomorphic features (Figure 2), many of which are presumed to have developed during the LGM, and some periglacial features may have been reactivated during Holocene cold periods as recent as the Little Ice Age (LIA)[8,11,14,26]. Owing to high altitudes, relatively cold temperatures persisted during interglacial periods. Palaeogeomorphic phenomena are thus relatively well preserved and provide evidence for discrete past cold events.[27] Given that these landforms are relatively unique to southern Africa, they have attracted considerable research attention in recent decades[8,15]. However, many phenomena identified, and their associated palaeoenvironmental and palaeoclimatic inferences, have been contentious.[8,13,28]

The most contentious palaeogeomorphic features in eastern Lesotho are those interpreted to be of glacial origin.[8,10,29] Many critiques of such glacial interpretations arise from the predominantly qualitative earlier approaches used, and difficulties in positive feature identification and age determinations[8,15]. However, more recent quantitative approaches, including detailed mapping, macro- and micro-sedimentology, and obtaining ages for small moraines (Figure 2a), have provided more substantive evidence for Quaternary glaciation in eastern Lesotho. Such glaciation is dated to ~19 500–17 350 calibrated years before present (cal yr BP)[6,12,13], which coincides with the terminal stage of the LGM in southern Africa[2]. While the accuracy of the term 'niche glaciation'

has been queried[29], evidence based on these moraines for small-scale, spatially restricted glaciation during the LGM is undisputed to date[18].

Evidence for contemporary and relict periglacial features in the eastern Lesotho highlands (Figure 2) is more widely accepted[8,15,30]. The contemporary alpine region of Lesotho is considered a marginal periglacial environment.[8,31,32] Active features, including earth hummocks, sorted patterned ground (Figure 2b) and stone-banked lobes still develop as a result of diurnal and seasonal frost action[14]. Inactive periglacial landforms dating back to colder periods are generally larger in scale than their contemporary counterparts. In the case of stone-banked lobes, the older lobes are between 1.4 m and 4 m long and were formed in association with prolonged periods of sub-zero air and ground temperatures[11,15]. Similar to the contested glacial evidence, greater insolation on north-facing slopes requires that the majority of such features are situated on south-facing slopes[14,28,33-35]. Larger relict periglacial features (e.g. sorted patterned ground) could be indicative of sporadic permafrost during their genesis, most likely during the LGM[11]. Such permafrost might have been highly localised and possibly of relatively short duration, spanning a few centuries at most.[11]

The palaeo–periglacial phenomena indicate colder conditions than at present, with the size of landforms and their likely genesis providing some indication of relative temperature departures[11,15,18]. However, palaeoenvironmental interpretations of associated precipitation levels during these cold periods are conflicting.[6,8,15,31] Boelhouwers and Meiklejohn[8] suggested that some periglacial features, including large patterned ground, solifluction mantles and blockstreams (Figure 2c) were indicative of dry, cold conditions at their time of formation, whereas typically erosional features, including nivation hollows, cirques and cryoplanation terraces are likely indicative of wetter conditions during cold periods[36]. Other scholars[11,26] have argued that similar periglacial features, including large sorted patterned ground near Mafadi Summit (Figure 2b) and at Thabana Ntlenyana, and pronival ramparts at Thabana Ntlenyana, imply relatively high moisture availability during their formation. In addition, small glaciers would have required sufficient snow cover during the LGM, implying a shift in the seasonality of precipitation, relative to contemporary conditions[6,12,13,18].

Glacier reconstructions from five sites in eastern Lesotho have enabled the calculation of equilibrium line altitudes and glacier viability under specific climatic conditions.[18] From this, using the high-resolution HadAM3h climate model, Mills et al.[18] demonstrate a likely decrease in summer precipitation during the LGM. Assuming temperatures of 6 °C during the LGM, this would be associated with a 28% increase in solid precipitation during the autumn, winter and spring months[18]. These models support a growing number of recent studies that posit relatively wet LGM conditions over parts of south-eastern Africa[2,37,38].

Relict periglacial features provide evidence for single discrete cold events. The restricted climatic requirements for such landform genesis can facilitate improved quantification of climatic change during these cold periods.[11,18] Continued debates concerning these cold region landforms highlight many unresolved uncertainties as to their identification, processes and environmental drivers, oftentimes with an over-reliance on northern hemisphere analogues[28,36,39,40]. The identification of features may become increasingly difficult to address as landscape disturbance increases in response to greater human and livestock pressure in the region[27]. Temporal constraint for the period of landform formation is critical if such debates are ever to be resolved adequately. However, age-dating remains a challenge because of difficulties in precisely dating relict periglacial and glacial phenomena.[8,15] Ages obtained for pronival ramparts and moraines have confirmed the timing during which these features developed, and interestingly coincide with a Holocene neoglacial event and the LGM respectively. This knowledge has improved the potential for palaeoenvironmental inferences to be drawn from such features.[6,12,13] Accurate inferences could be achieved through the provision of corroborating environmental evidence, and by filling in the large temporal gaps between cold periods represented by these features[6,8,18,36].

The archaeological record

Archaeological work in Lesotho has been relatively slow in forthcoming, resulting from problems endemic to field-based research in the region: difficulty in accessing sites because of their remoteness in eastern Lesotho; the deeply dissected landscape; and a lack of institutional knowledge[16,41]. The first significant contribution to understanding Stone Age occupation in the eastern Lesotho highlands was made by Carter[42,43], with the specific aim of understanding occupation patterns resulting from changes in the natural environment, driven by climatic changes[5,16,43]. While some aspects of Carter's methodology have been critiqued, this work importantly confirmed that occupation dates back to the Early Stone Age at >80 000 yr BP, and that the highlands do not represent an enclave only recently inhabited[7,16,44].

From the late 1980s onwards, the development of the Southern Perimeter Road and the Lesotho Highlands Water Scheme initiated a more intensive developer-funded archaeological effort in the regions[41,45]. This included further work on the timing and seasonal patterns of human occupation in the region, based on excavations at predominantly shelter sites[4,5,7,16,19,43-48] and later work on open-air sites[9,50-53]. These studies have largely supported Carter's initial findings and interpretations, and have improved the resolution of detail. The environmental inferences from these studies are explored in this section of our paper, beginning with long-term patterns of occupation and absence at these sites, followed by evidence for the seasonal occupation of sites, and finally evidence of human influence on the environment. A small body of archaeological work undertaken in the region has analysed palaeoecological proxies for the periods of settlement[4,9,19,53]. This will be discussed predominantly in the palaeoecological section which follows.

The selection of locations for settlement by Early through to Late Stone Age communities in southern Africa necessarily involved considerations of the climate, natural environment and resource availability. These factors presented varied opportunities and threats at each location, particularly due to the reduced habitat adaptation measures available[43,47,49,50]. It has been posited that Stone Age settlement in the eastern Lesotho highlands would have occurred under two dominant environmental conditions. The first is relatively arid conditions which would have forced people into the moist highlands in search of water, plants and animals.[4,48,53,54] The second is a shift to temperatures warm enough to support varied vegetation, with a consequent range of animals, and with limited snow cover[5,41,48,49]. Given the low capacity for improved warmth, settlement under more arid cold conditions could not have occurred at the highest altitudes, as temperatures would have been intolerably cold and would have prevented plant growth and deterred animals.[7,41,47,49] Settlements could thus develop only during sufficiently warm periods and with enough moisture to support a varied ecology. Pulsed occupation throughout the Pleistocene and Holocene has commonly been observed for archaeological sites across Lesotho (Table 1). For many sites, periods of absence are greater than those of occupation.[7,41,47,48] For the highest altitude sites, such as Belleview, Moshebi's Shelter and Ha Soloja (Figure 1), occupation ceases during known periods of very cold conditions, most recently the LGM[16,47,50]. Occupation resumes during warm periods, as is most clearly noted for Likoaeng[16,47,50,52]. For lower altitude sites, including Sehonghong and Melikane (Figure 1), there remains evidence of occupation throughout very cold periods including the LGM, although some scholars suggest this evidence of occupation reflects visits of short duration for hunting purposes[46,53].

Although it would appear that occupation has been more intensive and prolonged since the mid-Holocene[49], notably, there are no clear overlaps in periods of occupation or absence between the Lesotho sites (Table 1), nor for the broader southern African sites[55]. Therefore, microclimate and microtopography appear to have been important factors determining the selection of locations for occupation.[47] However, a far greater set of influencing factors is likely to be responsible for the selection of sites and timing of occupation than climate and associated environment alone.[49,53] These pulses of occupation may critically involve climatically-driven seasonal movements between sites, both within Lesotho and from farther afield.[43,47,48]

Table 1: Known periods of occupation at Stone Age sites in Lesotho

Author	Year	Site	Occupation Periods
Mitchell et al.	1994	Hololo	1660 AD (cal) 1640 AD (cal)
Mitchell et al.	1994	Bolahla	1885–1914 AD (cal) 1179 AD (cal)
Mitchell et al.	1994	Tloutle	1272–1305 AD (cal) 1446–1638 AD (cal)
Esterhuysen & Mitchell	1996	Ha Makotoko	10 000–8370 yr BP
Esterhuysen & Mitchell	1996	Ntloana Tsoana	12 100–8780 yr BP
Esterhuysen & Mitchell	1996	Tloutle	9000–5000 yr BP
Mitchell	1996a	Sehonghong	~6000 yr BP ~7000 yr BP ~10 000 yr BP ~11 000 yr BP ~12 000–12 500 yr BP ~20 000 yr BP ~25 000 yr BP
Mitchell	1996b	Sehonghong	850–1870 AD (cal)
Mitchell et al.	1998	Sehonghong	12 500–12 000 yr BP ~11 000 yr BP 9800–9300 yr BP
Mitchell et al.	2011	Likoaeng	1700–1000 BC (cal) 1000–200 BC (cal) 100 BC–250 AD (cal) 800–899 AD (cal)
Parker	2011	Likoaeng	>3380–2860 cal yr BP 2860–2440 cal yr BP 2510–2150 cal yr BP 2070–1600 cal yr BP 1260–1070 cal yr BP
Stewart et al.	2012	Melikane	~80 000 cal yr BP ~60 000 cal yr BP ~50 000 cal yr BP ~46 000–38 000 cal yr BP ~24 000 cal yr BP ~9000 cal yr BP ~3000 cal yr BP ~1800 cal yr BP
Roberts et al.	2013	Ntloana Tsoana	~61 000 cal yr BP 57 000 cal yr BP 14 200–9600 cal yr BP
Loftus et al.	2015	Sehonghong	35 000–32 000 cal yr BP 30 000–9000 cal yr BP 25 000–21 000 cal yr BP 19 000–18 500 cal yr BP 16 000–12 500 cal yr BP 11 500–10 000 cal yr BP 8000–7500 cal yr BP 7000–6500 cal yr BP 1700–1000 cal yr BP

Note: All dates are cited directly from the literature and have not been re-calibrated (owing to insufficient data).

Key: cal yr BP – calibrated years before present, yr BP – years before present

Carter[42] originally suggested that evidence of occupation at sites in Lesotho may reflect periods of seasonal duration, with movement between two or more sites over an annual period. This hypothesis was supported by observations of seasonal movements between eastern Lesotho and low-altitude sites in KwaZulu-Natal in the late 19th century, with the highest altitude or coldest site being used in spring and summer and a warmer site used in winter each year[47,56]. This in turn is supported by documentary records for the 19th century from eastern Lesotho, which record particularly cold conditions[57]. There are three main examples of seasonal occupation shifts in eastern Lesotho: Sehonghong and Melikane[43,47]; Belleview and lower-altitude sites in KwaZulu-Natal[42]; and Likoaeng[51,52]. Sehonghong and Melikane are both situated on the Senqu/Orange River, but Sehonghong has a considerably colder climate as it is west-facing, compared with the north-facing Melikane[16,52].

The synchronous dates of occupation at the two sites during climatically adverse periods are argued to be explained by seasonal migration between the two sites, using Sehonghong in summer and Melikane in the colder winter months during interglacial periods[47], whereas during the LGM it is likely that both sites were used only briefly in summer[7]. This seasonal pattern of use is supported by evidence of fish, flower-heads and grasses found at Sehonghong but not Melikane, indicating preferential occupation of Sehonghong in spring[16,47,56]. At the higher altitude sites on the eastern escarpment, including Belleview, Moshebi's Shelter and Ha Soloja, a different system of seasonal movement appears to have occurred, across far greater distances, covering a transect from the eastern Lesotho highlands to the KwaZulu-Natal lowlands[16,42]. High-altitude Lesotho sites would have been used only briefly in summer, with movement through the mist belt and tall-grassveld of KwaZulu-Natal, and coastwards to the thornveld during winter months[16,42,43]. The larger distances travelled and more extreme cold temperatures at the escarpment sites explain the complete abandonment of high-altitude sites during the LGM.[5,7,43]

Likoaeng is one of the few open-air sites in Lesotho for which archaeological records have been reported, with remains of mammals, birds, reptiles and molluscs, demonstrating a shift from hunting to fishing at the site at ~4000 cal yr BP[9,50]. Higher resolution seasonal occupation is evident from Likoaeng, where fish bones provide evidence of deliberate occupation of a site timed by recurrent biological events[9,50-52]. The late Holocene deposits at this site are dominated by very large numbers of fish bones (1.3 million counted), suggesting that the site was occupied specifically to intercept the spawning runs[52]. The species of fish from these deposits demonstrate a marked shift in dominance from smallmouth yellowfish (*Labeobarbus aeneus*) to Orange River mudfish (*Labeo capensis*) over time, with Orange River mudfish accounting for the majority of fish bones after 560 cal yr BP[51]. As these fish have different spawning times, centred in mid-summer for smallmouth yellowfish and early spring for Orange River mudfish, the timing of occupation of Likoaeng demonstrates preferences for different spawning groups – based most likely on the availability of food elsewhere[51,52]. Over longer time periods, occupation at Likoaeng appears to have been interrupted primarily by periods of flooding rather than changes in temperature or snowfall.[52]

The palaeoenvironmental significance of these occupation pulses identified in the archaeological record, together with evidence of seasonal movement between sites, implies that an ameliorating climate would have been necessary for these Stone Age occupants of high-altitude sites to survive[7]. Although it seems reasonable that these occupation pulses and seasonal movements were driven predominantly by climatic changes in the eastern Lesotho highlands, it is impossible to quantify the extent of any climatic influence or to account for all non-environmental drivers[7,55]. Social interactions between different communities, and changes in food and landscape preferences, could alter settlement patterns and occupation pulses entirely, independent of changes in climate and vegetation[4,49,51]. Where settlement location and consequent occupation pulses were driven by climate changes, it is difficult to determine whether these were shifts in temperature, precipitation, or both.[4,43]

It is likely that occupation in the eastern Lesotho highlands occurred either during warmer periods, or during periods of lowland drought

when moisture available in the highlands was a valuable and necessary resource – which would surpass the detriments of having to endure the cold[4,43,47]. For most sites there is insufficient evidence to interpret independently which of these drivers was dominant at any given period. Thus, many of the climate inferences based on occupation cycles are dependent on palaeoclimatic information from the site or from elsewhere in southern Africa.[43,52] Consequently, there have been continued calls, and recent efforts made, to extract palaeobotanical climate proxies – including charcoal, faunal remains, phytoliths and stable isotopes – from material recovered at archaeological excavations in the region.[4,7,9,19,48,59]

Palaeoenvironmental reconstructions

Detailed continuous palaeoenvironmental records from Lesotho are limited.[3,5,7] For example, the only pollen reconstruction for the eastern Lesotho highlands[60] was sampled at poor resolution and lacks chronology, and no information on site location is provided. Fluctuations in Holocene climate have been reconstructed on the basis of sedimentary characteristics from gully exposures in eastern Lesotho[61-63], but without an analysis of biological proxies the palaeoenvironmental inferences made from such studies remain relatively generalised[3]. The majority of studies that have explored palaeoenvironmental changes using biological proxies have been conducted in the western Lesotho lowlands, and many have relied on archaeological evidence and hence provide environmental records only for periods of occupation[3,4,19]. Archaeological material is further compromised by human bias, because material that was brought to shelters often did not directly reflect the broader spectrum of dominant regional vegetation but rather preferences for fuel, food and bedding material.[9,19,53]

As part of a broader preliminary study of the wetlands in eastern Lesotho, van Zinderen Bakker[60] presents the results of pollen analysis from a sediment profile measuring 90–100 cm in length, extracted from an unidentified wetland. The pollen results reflect a shift from Cyperaceae-dominated samples at the bottom of the profile to a proportion of Poaceae (Gramineae) near the top. This pattern is argued to represent a regional decrease in moisture over the period represented by the profile.[60] The period during which this regional drying would have occurred remains unclear because of the absence of age-dating technology at the time of the study. In a later and more detailed vegetation analysis of eastern Lesotho wetlands, van Zinderen Bakker and Werger[64] present a bottom date for a peat bog in eastern Lesotho of 8020 ± 80 yr BP, which could potentially be related to this profile. However, the lack of site details for either study makes the inference impossible to confirm. The initial study was preliminary work in the region, which most likely was intended to lay a foundation for future studies given the availability of well-preserved pollen and diatoms in the many highland wetlands[60].

Sedimentary sequences with alternating layers of organic rich clays, peat and gravels, exposed along deep gully systems at Sani Top and Tlaeeng (Figure 1), were investigated to develop a regional late Quaternary palaeoclimatic and palaeoenvironmental history[61-63]. Interpreting the sequence to represent dry periods through layers of orange-coloured gravels, and wet periods through organic rich clay and peat layers, a chronology of moisture fluctuations for eastern Lesotho was developed, constrained by radiocarbon dates[61-63]. Results suggest cold conditions prior to 13 500 yr BP, a warm and wet period from 13 500–9000 yr BP, drier and potentially colder conditions from 9000–5000 yr BP, and warmer wet conditions from 5000–1000 yr BP.[63] The methodology requires critical review, as the samples were not pre-dried, and organic material was determined by burning the samples at 105 °C rather than the standard 550 °C.[64] The low burning temperature accounts for percentage organic content measurements that were, by the researcher's own admission, exceptionally high.[61] These measures of percentage organic content removed therefore rather reflect the percentage of moisture loss. Nonetheless, distinct differences in colour and texture between gravel and organic-rich sediments are recorded, and the temporal constraint of these shifts provides some understanding of Holocene precipitation changes in the highlands[61,62].

A further study[3] focused on reconstructing palaeoenvironments of the western Lesotho lowlands, by examining a sedimentary sequence exposed along a 13-metre deep gully sidewall in the Tsoaing River Basin (Figure 1). Pollen and phytoliths were extracted for analysis from nine visually distinct sedimentary layers within the sequence, with dates ranging from 12 000 to 4000 cal yr BP.[3] Very little pollen was available for counting, with representative samples for only the bottom two layers of the sequence, possibly indicating oxidation under seasonal dry conditions for the remaining layers[3]. The environmental reconstruction was thus made primarily from the phytolith data, which indicate a rapid transition to dry conditions from 8600 to 8450 cal yr BP, with wetter conditions by 7000 cal yr BP and again after 4500 cal yr BP.[3] The absence of pollen for much of the sequence, limited information available from the phytoliths, and low sampling resolution all restrict the palaeoclimatic value of this work. However, the study does indicate considerable potential for detailed future analysis.

Work undertaken on biological proxies from excavated sediments at archaeological sites provides some of the most valuable palaeoenvironmental information for Lesotho.[4,9,19,53,59] Holocene charcoal assemblages were analysed from three shelters (Tloutle, Ha Makotoko and Ntloana Tsoana) along a tributary of the Caledon River in western Lesotho (Figure 1).[19,66] Considerable differences in species composition were noted between Tloutle, which is closer to the river, and Ha Makotoko and Ntloana Tsoana throughout the periods of occupation; these findings highlight the influence of microclimates within small geographic areas in Lesotho[19]. Ha Makotoko and Ntloana Tsoana had charcoal evidence of species well adapted to cold and drought, including Protea *sp.*, *Leucosidea sericea*, *Rhamnus prinoides*, Rhus *sp*, *Passerina montana* and Erica *sp.*, for the period 12 110–8370 yr BP. The species composition from Tloutle for the period 8680–5080 yr BP was indicative of more mesic woodland, including high percentages of *Olea africana, Celtis africana*, *Maytenus heterophylla*, Rhus *sp*. and small percentages of *Podocarpus*.[19]

Examining changes throughout the periods of occupation, two mesic periods were identified by the appearance of *Podocarpus* charcoal. The first was centred at 8700 yr BP, possibly the result of a continued increase in moisture from 10 000 yr BP; the second period, 6900–5000 cal yr BP, reflected an even wetter period[19,66]. These wet periods were separated by dry conditions, characterised by the appearance of *Euphorbia sp*. and increases in charcoal from *Buddleia sp*. and *Passerina montana*[19].

Analyses of stable carbon and oxygen isotopes from ungulates, and stable carbon isotopes from organic material, provide further palaeoenvironmental evidence for these three sites.[4,59] The isotopic values for herbivore grazers reflect the proportion of C_3:C_4 vegetation consumed, from which the vegetation balance at the site can be inferred[59]. Evidence for the three Caledon River sites in Lesotho indicate a progressive warming from 16 000 to 6000 cal yr BP, marked by considerable fluctuations.[59] The fluctuations notably include periods that were sufficiently cool to encourage the growth of C_3 plants during the periods 16 000–14 000 cal yr BP, 10 200–9600 cal yr BP and 8400–8000 cal yr BP.[59]

A more recent study using stable carbon isotopes provides further environmental proxy data for Ntloana Tsoana and Tloutle for the period spanning the Pleistocene-Holocene transition.[4] The $\delta^{13}C$ values are distinct for C_3 and C_4 plants, and the ratio of C_3:C_4 plants is related to temperature and water availability. In Lesotho, temperature is altitudinally constrained, with C_3 plants found largely in the cold high-altitude regions, and C_4 plants dominating the vegetation in the warmer lowlands.[4] Isotope analysis confirms predominantly C_3 vegetation during the LGM, followed by a progressive increase in the proportion of C_4 plants.[4] Within this period of overall warming, the Pleistocene-Holocene transition was marked by rapid temperature fluctuations of as much as 4 °C from 11 200 cal yr BP to 9500 cal yr BP, with at least three cycles of warming and cooling reflected in the carbon isotope record[4]. More stable, warm conditions were attained by 9500 cal yr BP.[4] In addition to providing valuable proxy evidence for the period preceding the charcoal record reported by Esterhuysen and Mitchell[19], this record notably demonstrates the existence of very cold periods and extreme climatic fluctuations during longer periods of occupation, which, without proxy data, would have been assumed to reflect continuous moderate conditions[4].

Caution must be used when interpreting environmental proxy data from archaeological sites, particularly in the case of shelters, as the species

composition depends on what people selectively brought into the shelter[19]. Depending on the intended use of the plant material, biases may exist, and species that were common in the area may be absent in the record.[19]

Evidence of environmental and climatic change from a 3-metre stratified sequence at Likoaeng open air shelter is derived from phytoliths and stable carbon isotopes, spanning periods of both occupation and absence over the period 3400–1070 cal yr BP.[9] Results suggest a warm period dominated by C_4 panicoid and chloridoid grass types from the beginning of the sequence to 2960 cal yr BP.[9] This is followed by an abrupt neoglacial cooling period from 2960 to 2160 cal yr BP, dominated by C_3 pooid grassland, with the presence of *Erica* and *Euryops sp.*[9] Thereafter, the environment is characterised by a mixture of C_3 and C_4 species until 1600 cal yr BP, followed by a return to the original environment dominated by C_4 panicoid and chloridoid.[9] This proxy evidence complements findings on occupation pulses at the site based on mammal and fish remains.[50-52] Although occupational hiatuses at the site resulted from flooding rather than unsuitable temperatures, unlike most archaeologically derived environmental reconstructions, the record spans both occupation and absence[9].

Most recently, the analysis of stable isotopes derived from soil organic matter and faunal enamel from Sehonghong[53] greatly improves the temporal resolution of palaeoclimatic reconstructions for eastern Lesotho. Notably, that study provides evidence for the absence of C_4 plants at Sehonghong during the LGM.[53] With the isotope record reflecting the re-emergence of C_4 species in the early Holocene, and supporting evidence of C_4 species occurrence during the LGM at the lower-altitude western Lesotho sites[9], a temperature depression of at least 5 °C is confirmed[53]. The Sehonghong isotope records further indicate a sudden cooling at ~11 000 cal yr BP, argued to provide possible evidence for southern African cooling during the Younger Dryas[53].

Conclusions and research questions

Combining evidence from archaeological, geomorphological and palaeobiological archives highlights considerable variability in temperature and precipitation, biosystems, geomorphology and human occupation throughout the late Quaternary in Lesotho (Figure 3). Specific variables of interest include temperature and precipitation, biosystems, geomorphology and human occupation. Although there are some

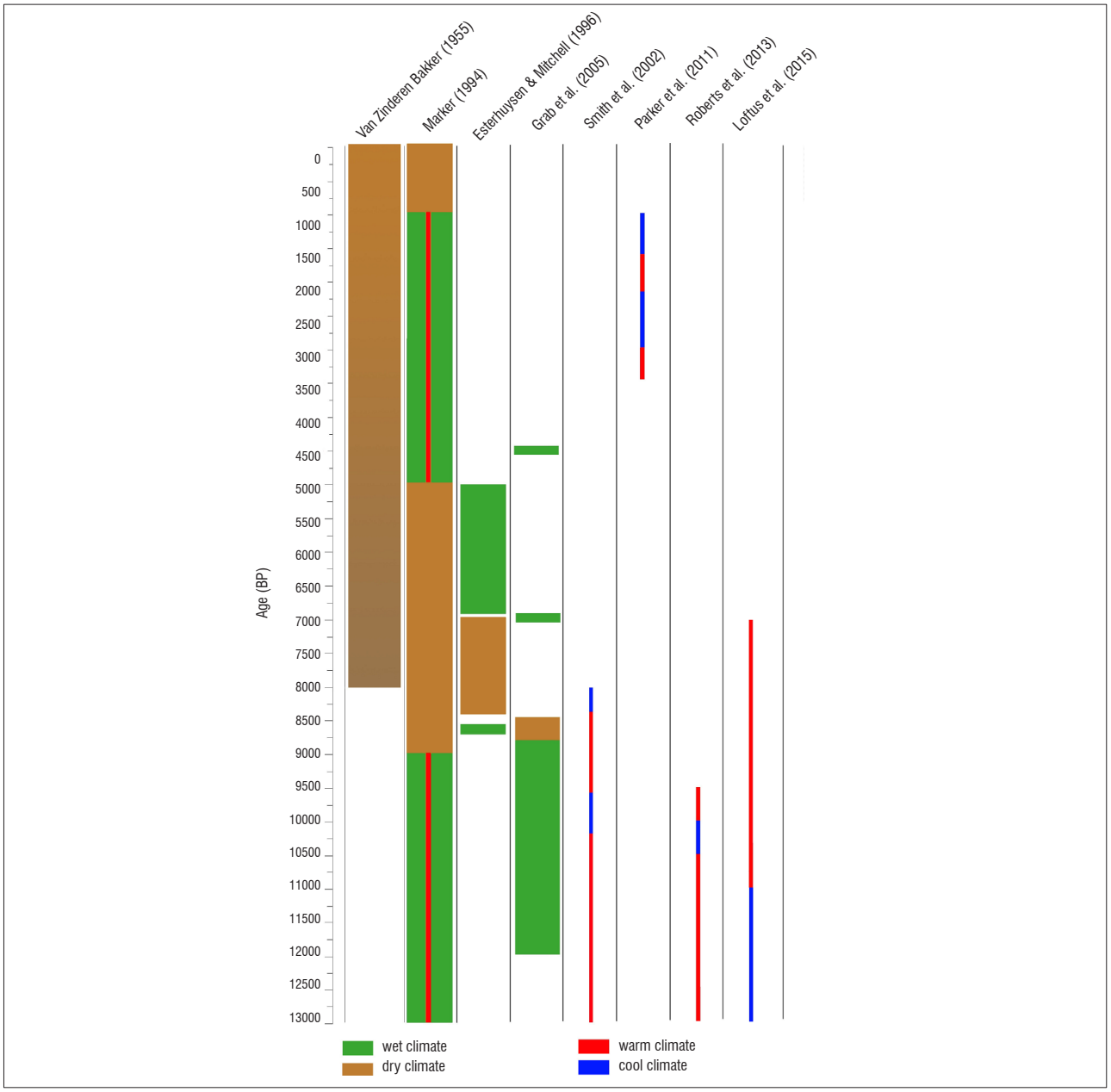


Figure 3: Palaeoclimatic inferences based on geomorphological, archaeological and palaeobiological studies in Lesotho from 13 000 cal yr BP to present. Graduated colour represents progressive drying.

conflicting interpretations of the available records, these often occur where coarse-resolution long-term records do not register the intricacies of higher-resolution records that span shorter time periods (Figure 3). Most notable is the absence of any high-resolution long-term record that can bridge the gap and resolve these apparent conflicts. Long-term temporally continuous palaeoenvironmental reconstructions, based on proxy data, would greatly improve the palaeoclimatic inferences made from such geomorphic features.[8,18] Comparisons with higher resolution and more temporally continuous records from Braamhoek Wetland[67,68] and Mahwaqa Mountain[37] in the adjacent South African Drakensberg indicate lags in temperature depressions and changes in moisture availability. These findings support the inference of topographical controls on lapse rates and orographic rainfall.

The reconstructed palaeoclimatic and palaeoenvironmental history of Lesotho relies predominantly on archaeological and geomorphic evidence. Both sources of evidence are produced through pulsed events. In the case of archaeology, these pulses are recurrent periods of usually warmer temperatures, whereas for cryogenic landforms, phenomena reflect development during what is usually a single pulse of cold conditions[4,6,13]. Although such periglacial landforms may become reactivated (active) again after warmer climatic episodes terminate, it has thus far not been possible to ascertain their age and periods of past activity. When combined, palaeogeomorphological and archaeological evidence can be used to infer the timing of both cold and warm events. However, such data are collectively unable to provide a complete and continuous climatic and environmental history for the region, as the climatic extremes they represent are often short-lived and seldom overlap, and are influenced significantly by conditions leading into and out of the events they document[4,9,26]. Thus, while potentially valuable for defining more extreme palaeoclimatic periods, and useful for reconstructing broad environmental conditions where no other data exist, these records are relatively limited in providing more precise palaeoclimatic and environmental information on the late Quaternary[3].

Continuous, well dated palaeoecological records are thus essential for determining the nature of climatic and associated environmental changes, during which these extreme events took place, during the periods leading up to these events, and the smaller temporal fluctuations between them[3,6]. Because of the complex relationships between temperature and precipitation in Lesotho, it would be beneficial to employ multiple proxies to quantify the relative effects of these two climate variables[4]. Finally, given the large differences in palaeoclimatic and palaeoenvironmental information gained from geomorphic features, human occupation, and present-day environments at sites in very close proximity to each other, these features highlight distinct microclimatic differences. Only once detailed analyses across many sites are forthcoming will it be possible to develop a high-resolution understanding of past climates at specific sites in Lesotho, and to integrate the findings of geomorphological and archaeological records.

Authors' contributions

J.F. was the team leader, collated the existing literature, and wrote the first draft. S.G., M.B. and A.M. supervised this process, contributed to the literature database, and assisted in the argumentation throughout the paper. S.G. provided photographs of geomorphological features.

References

1. Meadows ME. Seven decades of Quaternary palynological studies in southern Africa: a historical perspective. T Roy Soc S Afri. 2015;70(2):103–108. http://dx.doi.org/10.1080/0035919X.2015.1004139
2. Chase BM, Meadows ME. Late Quaternary dynamics of southern Africa's winter rainfall zone. Earth-Science Reviews. 2007;84:103–138. http://dx.doi.org/10.1016/j.earscirev.2007.06.002
3. Grab S, Scott L, Rossouw L, Meyer S. Holocene palaeoenvironments inferred from a sedimentary sequence in the Tsoaing River Basin, western Lesotho. Catena. 2005;61:49–62. http://dx.doi.org/10.1016/j.catena.2005.01.002
4. Roberts P, Lee-Thorp JA, Mitchell PJ, Arthur C. Stable carbon isotopic evidence for climate change across the late Pleistocene to early Holocene from Lesotho, southern Africa. J Quaternary Sci. 2013;28(4):360–369. http://dx.doi.org/10.1002/jqs.2624
5. Mitchell P, Parkington J, Wadley L. A tale from three regions: The archaeology of the Pleistocene/Holocene transition in the Western Cape, the Caledon Valley and the Lesotho Highlands, Southern Africa. Quatern Int. 1998;50:105–115. http://dx.doi.org/10.1016/S1040-6182(97)00056-6
6. Mills SC, Grab SW, Carr SJ. Recognition and palaeoclimatic implications of late Quaternary niche glaciation in eastern Lesotho. J Quaternary Sci. 2009;24(7):647–663. http://dx.doi.org/10.1002/jqs.1247
7. Mitchell PJ. The late Quaternary of the Lesotho Highlands, southern Africa: Preliminary results and future potential of ongoing research at Sehonghong shelter. Quatern Int. 1996;33:35–43. http://dx.doi.org/10.1016/1040-6182(95)00097-6
8. Boelhouwers JC, Meiklejohn KI. Quaternary periglacial and glacial geomorphology of Southern Africa: Review and synthesis. S Afr J Sci. 2002;98:47–55.
9. Parker AG, Lee-Thorp J, Mitchell PJ. Late Holocene neoglacial conditions from the Lesotho highlands, southern Africa: Phytolith and stable carbon isotope evidence from the archaeological site of Likoaeng. P Geologist Assoc. 2011;122:201–211. http://dx.doi.org/10.1016/j.pgeola.2010.09.005
10. Harper G. Periglacial evidence in South Africa during the Pleistocene epoch. Palaeoeco A. 1969;4:71–91.
11. Grab S. Characteristics and palaeoenvironmental significance of relict sorted patterned ground, Drakensberg plateau, southern Africa. Quat Sci Rev. 2002;21:1729–1744. http://dx.doi.org/10.1016/S0277-3791(01)00149-4
12. Mills SC, Grab SW. Debris ridges along the southern Drakensberg escarpment as evidence for Quaternary glaciation in southern Africa. Quatern Int. 2005;129:61–73. http://dx.doi.org/10.1016/j.quaint.2004.04.007
13. Mills SC, Grab SW, Carr SJ. Late Quaternary moraines along the Sekhokong Range, eastern Lesotho: Contrasting the geomorphic history of north- and south-facing slopes. Geog Ann. 2009;91(2):121–140. http://dx.doi.org/10.1111/j.1468-0459.2009.00359.x
14. Grab S. Stone-banked lobes and environmental implications, High Drakensberg, Southern Africa. Permafrost Periglac. 2000;11:177–187. http://dx.doi.org/10.1002/1099-1530(200007/09)11:3<177::AID-PPP349>3.0.CO;2-R
15. Grab SW, Mills SC, Carr SJ. Periglacial and glacial geomorphology. In: Holmes P, Meadows M, editors. Southern African Geomorphology: Recent Trends and New Directions. Bloemfontein: Sun Media; 2012. p. 231–266.
16. Mitchell PJ. Archaeological research in Lesotho: A review of 120 years. Afr Archaeol Rev. 1992;10:3–34. http://dx.doi.org/10.1007/BF01117694
17. Scott L. Grassland development under glacial and interglacial conditions in southern Africa: Review of pollen, phytollith and isotope evidence. Palaeogeography, Palaeoclimatolgy, Palaeoecology. 2012; 177:47–57. http://dx.doi.org/10.1016/S0031-0182(01)00351-0
18. Mills S, Grab S, Rea B, Carr S, Farrow A. Shifting westerlies and precipitation patterns during the Late Pleistocene in southern Africa determined using glacier reconstruction and mass balance modelling. Quat Sci Rev. 2012;55:145–159. http://dx.doi.org/10.1016/j.quascirev.2012.08.012
19. Esterhuysen A, Mitchell P. Palaeoenvironmental and archaeological implications of charcoal assemblages from Holocene sites in western Lesotho, Southern Africa. Palaeoeco A. 1996;24:203–232. http://dx.doi.org/10.1007/s11069-013-0964-8
20. Grab SW, Linde J. Mapping exposure to snow in a developing African context: Implications for human and livestock vulnerability in Lesotho. Nat Hazards 2014;71:1537–1560. http://dx.doi.org/10.1111/j.0016-7398.2004.05002.x
21. Ziervogel G. Targeting seasonal climate forecasts for integration into household level decisions: The case of smallholder farmers in Lesotho. Geogr J 2004;170(1):6–21.
22. Carbutt C, Edwards TJ. The flora of the Drakensberg Alpine Centre. Edinburgh Journal of Botany 2004;60(3):581–607.
23. Nüsser M, Grab S. Land degradation and soil erosion in the eastern Highlands of Lesotho, Southern Africa. Beiträge sur Physischen Geographie 2002;133:291–311.
24. Schmitz G, Rooyani F. Lesotho geology, geomorphology, soils. 1987; National University of Lesotho: Maseru.
25. Grab SW. Alpine turf exfoliation pans in Lesotho, southern Africa: Climate–process–morphological linkages. Geomorphology 2010;114:261–275. http://dx.doi.org/10.1016/j.geomorph.2009.07.007

26. Grab SW, Mills SC. Quaternary slope processes and morphologies in the upper Sehonghong Valley, eastern Lesotho. P Geologist Assoc. 2011;122:179–186. http://dx.doi.org/10.1016/j.pgeola.2010.02.001

27. Grab SW. A note on the morphology of miniature sorted stripes at Mafadi Summit, High Drakensberg. S Afr Geogr J. 1996;78(2):59–63. http://dx.doi.org/10.1080/03736245.1996.9713609

28. Grab, S. Block and debris deposits in the High Drakensberg, Lesotho, southern Africa: Implications for high altitude slope processes. Geog Ann. 1999;81(1):1–16. http://dx.doi.org/10.1111/j.0435-3676.1999.00045.x

29. Hall K. The shape of glacial valleys and implications for southern African glaciation. S Afr Geogr J. 2010;92:35–44. http://dx.doi.org/10.1080/03736245.2010.485360

30. Butzer KA. Pleistocene 'periglacial' phenomena in southern Africa. Boreas. 1973;2:1–11. http://dx.doi.org/10.1111/j.1502-3885.1973.tb00245.x

31. Hanvey PM, Marker ME. Present-day periglacial microforms in the Lesotho Highlands: Implications for present and past climatic conditions. Permafrost Periglac. 1992;3:353–361. http://dx.doi.org/10.1002/ppp.3430030409

32. Grab S. A note on needle-ice mound formation in the High Drakensberg, southern Africa. Permafrost Periglac. 2002;13:315–318. http://dx.doi.org/10.1002/ppp.427

33. Boelhouwers J. Quaternary slope development in the Lesotho highlands: review and alternative model. S Afr J Sci. 2003;99:44–46.

34. Sumner P. Relict Sorted Patterned Ground in Lesotho. Permafrost Periglac. 2004;15:89–93. http://dx.doi.org/10.1002/ppp.459

35. Sumner PD. Geomorphic and climate implications of relict openwork block accumulations near Thabana-Ntlenyaya, Lesotho. Geograf Ann. 2004;86(3):289–302. http://dx.doi.org/10.1111/j.0435-3676.2004.00232.x

36. Boelhouwers J, Holness S, Meiklejohn I, Sumner P. Observations on a blockstream in the vicinity of Sani Pass, Lesotho Highlands, Southern Africa. Permafrost Periglac. 2002;13:251–257. http://dx.doi.org/10.1002/ppp.428

37. Neumann FH, Botha GA, Scott L. 18,000 years of grassland evolution in the summer rainfall region of South Africa: Evidence from Mahwaqa Mountain, KwaZulu-Natal. Veg Hist Archaeobot. 2014;23(6):665–681. http://dx.doi.org/10.1007/s00334-014-0445-3

38. Scott L, Neumann FH, Brook GA, Bousan CB, Norström E, Metwally AA. Terrestrial fossil-pollen evidence of climate change during the last 26 thousand years in Southern Africa. Quat Sci Rev 2012;32:100–118. http://dx.doi.org/10.1016/j.quascirev.2011.11.010

39. Grab S, Hall K. North-facing hollows in the Lesotho/Drakensberg mountains: Hypothetical palaeoenvironmental reconstructions? S Afr J Sci. 1996;92:183–184.

40. Knight J. The shape of glacial valleys: Comment on Hall (2010). S Afr Geogr J. 2012;94(1):1–3. http://dx.doi.org/10.1080/03736245.2012.663953

41. Arthur C, Mohapi M, Mitchell P. Archaeology and dam projects in Lesotho. Conservation and Management of Archaeological Sites. 2011;13(3):231–252. http://dx.doi.org/10.1179/175355211X13179154166196

42. Carter PL. Late Stone Age exploitation patterns in southern Natal. S Afr Archaeol Bull. 1970;25:55–58. http://dx.doi.org/10.2307/3887948

43. Carter P. The effects of climatic change on settlement in eastern Lesotho during the middle and later Stone Age. World Archaeol. 1976;8:197–206. http://dx.doi.org/10.1080/00438243.1976.9979664

44. Cain C. Cultural heritage survey of Lesotho for the Maloti-Drakensberg Transfrontier Project, 2005–2006: Palaeontology, archaeology, history and heritage management. S Afr Archaeol Bull. 2009;64(189):33–44.

45. Mitchell PJ, Parkington JE, Yates R. Recent Holocene archaeology in western and southern Lesotho. S Afr Archaeol Bull. 1994;49(159):33–52. http://dx.doi.org/10.2307/3889171

46. Carter PL. Moshebi's Shelter. Lesotho. 1969;8:13–23.

47. Mitchell PJ. Revisiting the Robberg: New results and a revision of old ideas at Sehonghong Rock Shelter, Lesotho. S Afr Archaeol Bull. 1995;50(161):28–38. http://dx.doi.org/10.2307/3889272

48. Stewart B, Dewar G, Morley M, Inglis R, Wheeler M, Jacobs Z, et al. Afromontane foragers of the Late Pleistocene: Site formation, chronology and occupational pulsing at Melikane Rockshelter, Lesotho. Quatern Int. 2012;270:40–60. http://dx.doi.org/10.1016/j.quaint.2011.11.028

49. Plug I. Late Pleistocene and Holocene Hunter–Gatherers in the Eastern Highlands of South Africa and Lesotho: A faunal interpretation. J Archaeol Sci. 1997;24:715–727. http://dx.doi.org/10.1006/jasc.1996.0153

50. Plug I, Mitchell P, Bailey G. Animal remains from Likoaeng, an open-air river site, and its place in the post-classic Wilton of Lesotho and eastern Free State, South Africa. S Afr J Sci. 2003;99:143–152.

51. Plug I, Mitchell P, Bailey G. Late Holocene fishing strategies in southern Africa as seen from Likoaeng, highland Lesotho. J Archaeol Sci. 2010;27:3111–3123. http://dx.doi.org/10.1016/j.jas.2010.07.012

52. Mitchell P, Plug I, Bailey G, Charles R, Esterhuysen A, Thorp J, et al. Beyond the drip-line: A high-resolution open-air Holocene hunter-gatherer sequence from highland Lesotho. Antiquity. 2011;85(330):1225–1242 http://dx.doi.org/10.1017/S0003598X00062025

53. Loftus E, Stewart BA, Dewar G, Lee-Thorp J. Stable isotope evidence of late MIS 3 to middle Holocene palaeoenvironments from Sehonghong Rockshelter, eastern Lesotho. J Quaternary Sci. 2015;30(8):805-816. http://dx.doi.org/10.1002/jqs.2817

54. Plug I, Mitchell PJ. Fishing in the Lesotho highlands: 26,000 years of fish exploitation, with special reference to Sehonghong Shelter. J Afr Archaeol. 2008;6:33–55. http://dx.doi.org/10.3213/1612-1651-10102

55. Mitchell PJ. The Late Holocene assemblages with pottery. S Afr Archaeol Bull. 1996;51(163):17–25. http://dx.doi.org/10.2307/3888928

56. Carter PL, Mitchell PJ, Vinnicombe P. Sehonghong: The middle and later Stone Age industrial sequence at a Lesotho rock-shelter. Oxford: British Archaeological Reports; 1988.

57. Nash DJ, Grab SW. A sky of brass and burning winds: Documentary evidence of rainfall variability in the Kingdom of Lesotho, southern Africa, 1824–1900. Climatic Change 2010;101:617–653. http://dx.doi.org/10.1007/s10584-009-9707-y

58. Humphreys AJB. Prehistoric seasonal mobility: What are we really achieving? S Afr Archaeol Bull. 1987;42:34–38. http://dx.doi.org/10.2307/3887771

59. Smith JM, Lee-Thorpe JA, Sealy JC. Stable carbon and oxygen isotopic evidence for late Pleistocene to middle Holocene climatic fluctuations in the interior of southern Africa. J Quaternary Sci. 2002;17(7):683–695. http://dx.doi.org/10.1002/jqs.687

60. Van Zinderen Bakker EM. A preliminary survey of the peat bogs of the alpine belt of northern Basotholand. Acta Geogr. 1955;14:413–422

61. Marker ME. Sedimentary sequences at Sani Top, Lesotho highlands, southern Africa. Holocene. 1994;4(4):406–412. http://dx.doi.org/10.1177/095968369400400408

62. Marker ME. Late Quaternary environmental implications from sedimentary sequences at two high altitude Lesotho sites. S Afr J Sci. 1995;91:294–298.

63. Marker ME. New radiocarbon dates from Lesotho. S Afr J Sci. 1998;94:239–240.

64. Van Zinderen Bakker EM, Werger MJA. Environment, vegetation and phytogeography of the high-altitude bogs of Lesotho. Vegetatio. 1974;29:37–49. http://dx.doi.org/10.1007/BF02390894

65. Heiri O, Lotter AF, Lemcke G. Loss on ignition as a method for estimating organic and carbonate content in sediments: Reproducibility and comparability of results. J Paleolimnol. 2001;25:101–110. http://dx.doi.org/10.1023/A:1008119611481

66. Esterhuysen AB, Mitchell PJ, Thackeray JF. Climatic change across the Pleistocene/Holocene boundary in the Caledon River southern Africa: Results of a factor analysis of charcoal assemblages. Southern African Field Archaeology. 1999;8:28–34.

67. Finné M, Norström E, Risberg J, Scott L. Siliceous microfossils as late-Quaternary paleo-environmental indicators at Braamhoek wetlands, South Africa. Holocene 2010;20(5):747–760. http://dx.doi.org/10.1177/0959683610362810

68. Norström E, Neumann FH, Scott L, Smittenberg RH, Holmstrand H, Lundqvist S, et al. Late Quaternary vegetation dynamics and hydro-climate in the Drakensberg, South Africa. Quat Sci Rev 2014;105:48–65. http://dx.doi.org/10.1016/j.quascirev.2014.09.016

The stable isotope setting of *Australopithecus sediba* at Malapa, South Africa

AUTHORS:
Emily Holt[1]
Paul Dirks[1,2,3]
Christa Placzek[1]
Lee Berger[2]

AFFILIATIONS:
[1]Department of Geosciences, College of Science and Engineering, James Cook University, Queensland, Australia
[2]Evolutionary Studies Institute, DST/NRF Centre of Excellence in Palaeosciences, University of the Witwatersrand, Johannesburg, South Africa
[3]School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Paul Dirks

EMAIL:
paul.dirks@jcu.edu.au

POSTAL ADDRESS:
James Cook University: College of Science and Engineering, Townsville Campus, Townsville QLD 4811, Australia

DATES:
Received: 19 Oct. 2015
Revised: 10 Feb. 2016
Accepted: 12 Feb. 2016

KEYWORDS:
palaeovegetation; carbonate-cemented; cave sediments; palaeoclimate

HOW TO CITE:
Holt E, Dirks P, Placzek C, Berger L. The stable isotope setting of *Australopithecus sediba* at Malapa, South Africa. S Afr J Sci. 2016;112(7/8), Art. #2015-0351, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20150351

We report $\delta^{13}C$ and $\delta^{18}O$ results from carbonate-cemented cave sediments at Malapa in South Africa. The sediments were deposited during a short-period magnetic reversal at 1.977±0.003 Ma, immediately preceding deposition of Facies D sediments that contain the type fossils of *Australopithecus sediba*. Values of $\delta^{13}C$ range between -5.65 and -2.09 with an average of -4.58±0.54‰ (Vienna Pee Dee Belemnite, VPDB) and values of $\delta^{18}O$ range between -6.14 and -3.84 with an average of -4.93±0.44‰ (VPDB). Despite signs of diagenetic alteration from metastable aragonite to calcite, the Malapa isotope values are similar to those obtained in two previous studies in South Africa for the same relative time period. Broadly, the Malapa $\delta^{13}C$ values provide constraints on the palaeovegetation at Malapa. Because of the complex nature of the carbonate cements and mixed mineralogy in the samples, our estimates of vegetation type (C_4-dominant) must be regarded as preliminary only. However, the indication of a mainly C_4 landscape is in contrast to the reported diet of *A. sediba*, and suggests a diverse environment involving both grassland and riparian woodland.

Introduction

Palaeoclimate and palaeoenvironment studies are important for identifying the drivers of hominin evolution.[1,2] Long-term shifts and variability in climate are linked to changes in floral and faunal assemblages, which have been matched with significant points in hominin development.[1-3] Studies that investigate the link between climate and hominin evolution have focused on Africa because of its relatively extensive hominin fossil record, which extends back millions of years and coincides with significant palaeoenvironmental changes.[2]

Caves are prevalent in the Cradle of Humankind (CoH) area of South Africa. These caves are important because they contain records of changes in climate[4] and they host hominin fossils[5]. The caves act as natural traps for floral and faunal remains, capture wind- and water-borne sediments, and are host to carbonate cave deposits.[6,7] In a number of cave settings, fossils and sediments have been preserved in a definable stratigraphic sequence, and owing to the relatively stable nature of the cave these layers have remained undisturbed over long periods of time.[6] Thus, CoH caves are invaluable for analysing the links between changes in climate and terrestrial environments with hominin evolution.[8]

Stable isotope proxies in carbonate cave deposits have been used to investigate late Pliocene and early Pleistocene climate and environmental conditions in South Africa.[9,10] In the Makapansgat Valley, an area rich in hominin fossils, Hopley et al.[9] used oxygen isotope values to time-constrain cave carbonate (flowstone) deposition at Buffalo Cave, and to determine the major orbital cycles influencing climate and vegetation patterns. In addition, the flowstone carbon isotope values and organic matter contained therein were used to ascertain changes in the dominant vegetation type (i.e. C_3 versus C_4) during the early Pleistocene. A major finding by Hopley et al.[9] was an increase in C_4 vegetation in samples less than 2 million years (Ma) old, with the most significant increase in C_4 vegetation after 1.7 Ma.[9] Similarly, Pickering et al.[10] used carbon isotope values from cave carbonate deposits at Gladysvale Cave in the CoH. They reported a mixed C_4-dominant CoH landscape and a cool dry environment during the early Pleistocene (~1.8±0.7 Ma).

The ability of cave carbonates to reveal information about Plio-Pleistocene climate and vegetation has been established. However, studies from the modern summer rainfall region of South Africa are lacking. Limestone mining has displaced key stratigraphic sequences, which has led to dislocation of cave deposits and the proxies they contain, from important faunal fossils.[9,11] Challenges in defining stratigraphy are compounded by the difficulty in forming time-depth series because of the uneven rates of sedimentation and mineral growth, and the shortage of appropriate techniques for absolute dating.[7,12] Furthermore, in dolomitic areas such as the CoH, interpretations of oxygen and carbon isotope values from carbonates are affected by post-depositional diagenetic alteration of metastable aragonite to calcite.[13]

Despite the difficulties presented by the setting, cave carbonate studies in the CoH remain important for linking climate and environmental change with hominin evolution. Expansion of stable isotope studies to carbonate-cemented cave sediments associated with hominin fossils in the CoH provide a valuable source of information for palaeoclimate studies. In this paper, we present stable oxygen and carbon isotope data from Malapa in the CoH. The isotope data were obtained from a thin flowstone drape and carbonate cements, deposited within a cave setting in close proximity to the type fossils of *A. sediba*.[7,14] Based on a combination of uranium-series dating and palaeomagnetic methods,[7,14] the depositional age of the tested flowstone drape coincides with the age obtained for the *A. sediba* fossils at 1.977±0.003 Ma.

We compared the raw stable isotope results with the results from studies at Gladysvale[10] and Buffalo Cave[9]. Modelling was applied to the carbon isotope values in an effort to place constraints on the dominant vegetation type and past environment overlying the cave. Our results provide an initial assessment of possible vegetation conditions during the time when *A. sediba* was alive. In addition, important textural and mineralogical constraints are defined, providing guidance for future studies.

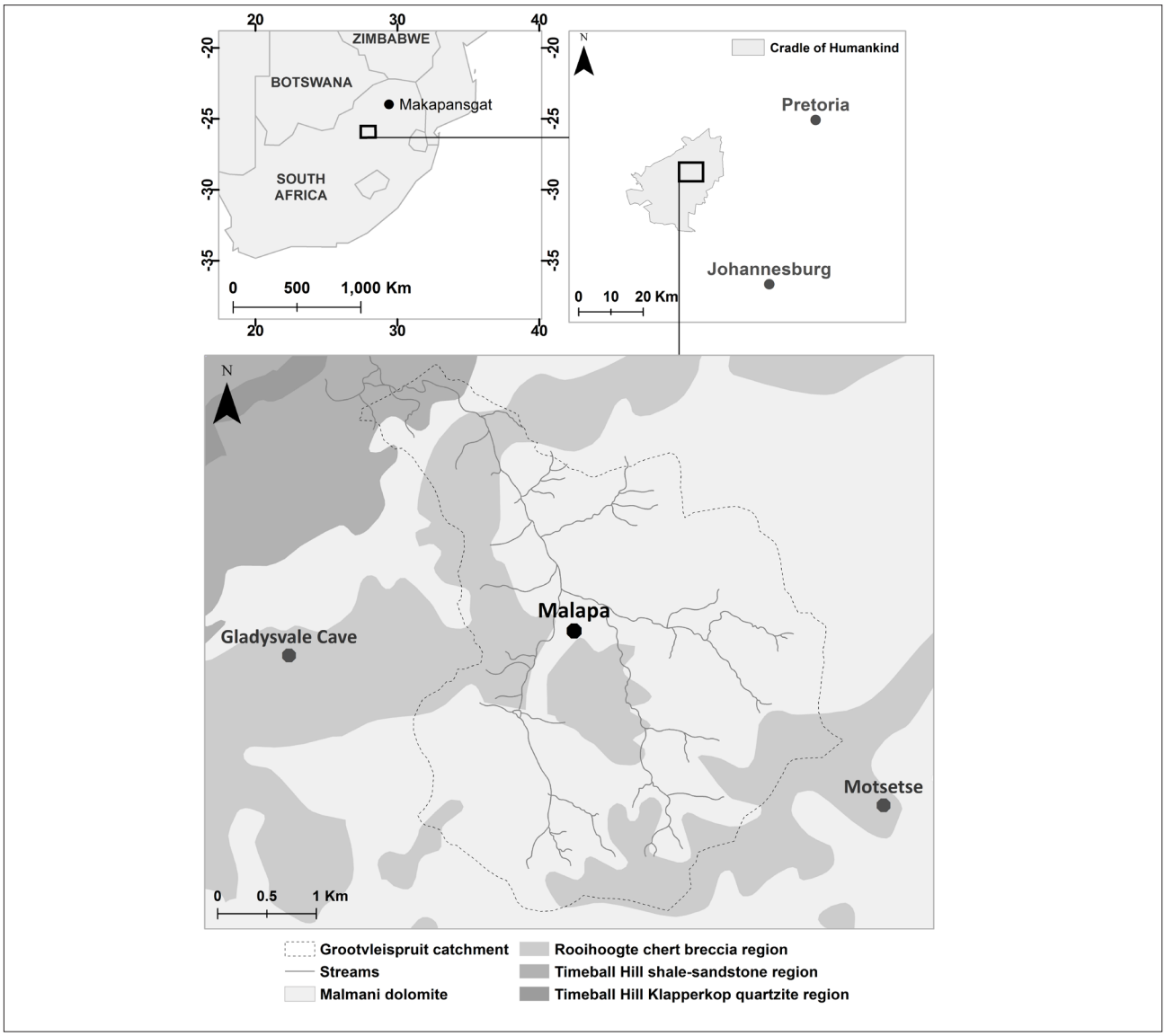


Figure 1: Location and geological setting of Malapa. Also shown is the location of another important hominin fossil site in South Africa, the Makapansgat Valley in Limpopo Province.

Site location

Malapa is located approximately 40 km northwest of Johannesburg at 25°52'S, 27°48'E (elevation 1440 masl) in the Grootvleispruit catchment, and is hosted by chert-free dolomite of the Palaeoproterozoic Lyttelton Formation in the Malmani Group (Figure 1).[7] The site comprises two small pits excavated by miners in the early part of the 20th century.[14] The larger pit, Pit 1, measures approximately 20 m^2 on the ground and 4 m deep, and hosts the type fossils of *A. sediba*.[15] Pit 1 was the source of samples obtained in our study. A second smaller pit, Pit 2, which is roughly 12 m^2 in size and 1 m deep, is situated nearby and also hosts hominin remains.[7,14]

Current climate and vegetation

Pretoria (elevation 1330 masl) is the closest monitored weather station to Malapa, and demonstrates a clear seasonal pattern in temperature and monthly rainfall. The highest temperatures occur from December to February (local summer), with February recording an average daytime temperature of 23.5 °C (1999–2013).[16] July (local mid-winter) is the coldest month, with an average daytime temperature of 12.6 °C.[16] The mean annual temperature (MAT) is 19.3 °C and the average annual rainfall is 700 mm, with most rain occurring in the summer months.[16]

Vegetation cover in the CoH is a consequence of precipitation, geology, MAT, and location within the landscape.[17-19] Malapa is situated within the Carleton Dolomite Grassland, which is dominated by a single layer of C_4 photosynthetic process grasses.[19] However, this area also has a high level of species richness because of the heterogeneous nature of the dolomitic landscape.[19] A range of woodland communities occurs in specific areas,[18,19] such as near cave entrances[18] and at spring sites or along streams. In addition, C_3-type grasses are common at relatively high and cool elevations, and in areas that experience winter rainfall.[17,19]

Depositional setting and stratigraphy

The depositional history and stratigraphy of the Malapa site has been well described.[7,14] The type fossils of *A. sediba* (MH1 and MH2)[15] were found within poorly-sorted sandstone of Facies D, which has been interpreted as a mass flow deposit.[7] A flowstone unit from Pit 1, measuring 5 cm to 20 cm in thickness (Flowstone 1), has been dated to 2.026 ± 0.021 Ma[7] and presents an important chronostratigraphic marker for Facies D. Flowstone 1 consists of a single sheet in the southeast corner of the pit, and splits first into two and then into three separate sheets towards the centre of the pit, where the top sheet directly underlies the *A. sediba* fossils contained in Facies D.

The three sheets of Flowstone 1 are separated by layers of fossil-bearing sandstone and siltstone, with thickness increasing from southeast to northwest.[7,14] The basal layer of Flowstone 1 correlates with a flowstone layer in the northwest wall of Pit 1, from which the U-Pb date was obtained. This correlated flowstone records normal magnetic polarity, assigned to the Huckleberry Ridge event (~2.05–2.03 Ma) at its base, followed by intermediate polarity, and then reversed polarity. The middle and upper flowstone layers in Flowstone 1 record stable reverse polarity of the Matuyama Chron (2.03–1.95 Ma). These layers cannot be correlated with flowstone in the west wall of the pit, but thin out and disappear.[7]

Along the northwest wall of Pit 1, a finely laminated unit occurs in a lateral up-dip position from the flowstone. This laminated unit is composed of mm-scale sandstone–siltstone layers, with angular clastic fragments (mainly dolomite, chert and speleothem) intercalated with regular drapes of flowstone (0.1 mm – 1 mm thick). Stratigraphically, the laminated unit referred to in this study as 'Facies Da' accumulates on top of Flowstone 1 by as much as 30 cm. The top of the unit (composed of Facies Da) constitutes the base of palaeomagnetic sample UW88-PM04. Facies Da also comprises most of palaeomagnetic sample UW88-PM09.[7,14] Thin flowstone layers in UW88-PM04 and UW-PM09 record normal polarity.[7,14]

Overlying Facies Da sediments are other sediments that belong to the finer-grained, weakly layered topmost section of Facies D (which were originally described as Facies C).[7] The sediments of Facies D also record intermediate and normal polarity, indicating that Facies Da and Facies D were both deposited during a brief magnetic reversal (the Pre-Olduvai event) at 1.977±0.003 Ma.[14] Facies D sediments are overlain by horizontally laminated, poorly sorted muddy sandstone of Facies E, which is also fossiliferous.[7] Along the northwest wall of Pit 1, the sediments of Facies Da, D and E overlap an erosion remnant of peloidal fine-grained muds, which remain undated and belong to Facies C.[7,14] Textural evidence indicates that the sedimentary units composed of Facies C, Da, D and E were deposited in a water-saturated environment.[7]

Methods

Samples

As shown in Figure 2, four samples from Pit 1 were studied for stable isotope analysis: UW88-PM09 (PM09), PM09-2 (a duplicate of PM09), UW88-PM04 (PM04) and PM04-2 (a duplicate of PM04). These samples have a well-constrained age (1.977±0.003 Ma), and contain thin carbonate drapes as part of Facies Da sediment, which were deposited during the same magnetic reversal recorded in Facies D immediately before deposition of the *A. sediba* fossils. Sample PM09 comprises a base of Facies C overlain by sediment of Facies Da and Facies D, and contains a thin laminated flowstone drape between Facies C and Facies Da. Sample PM09-2 is derived from the same sample block as PM09 but was set further back in the pit wall. Samples PM04 and PM04-2 represent two separate slices of the same original rock sample collected in the north wall of Pit 1,[7] and are largely composed of Facies D sediment.

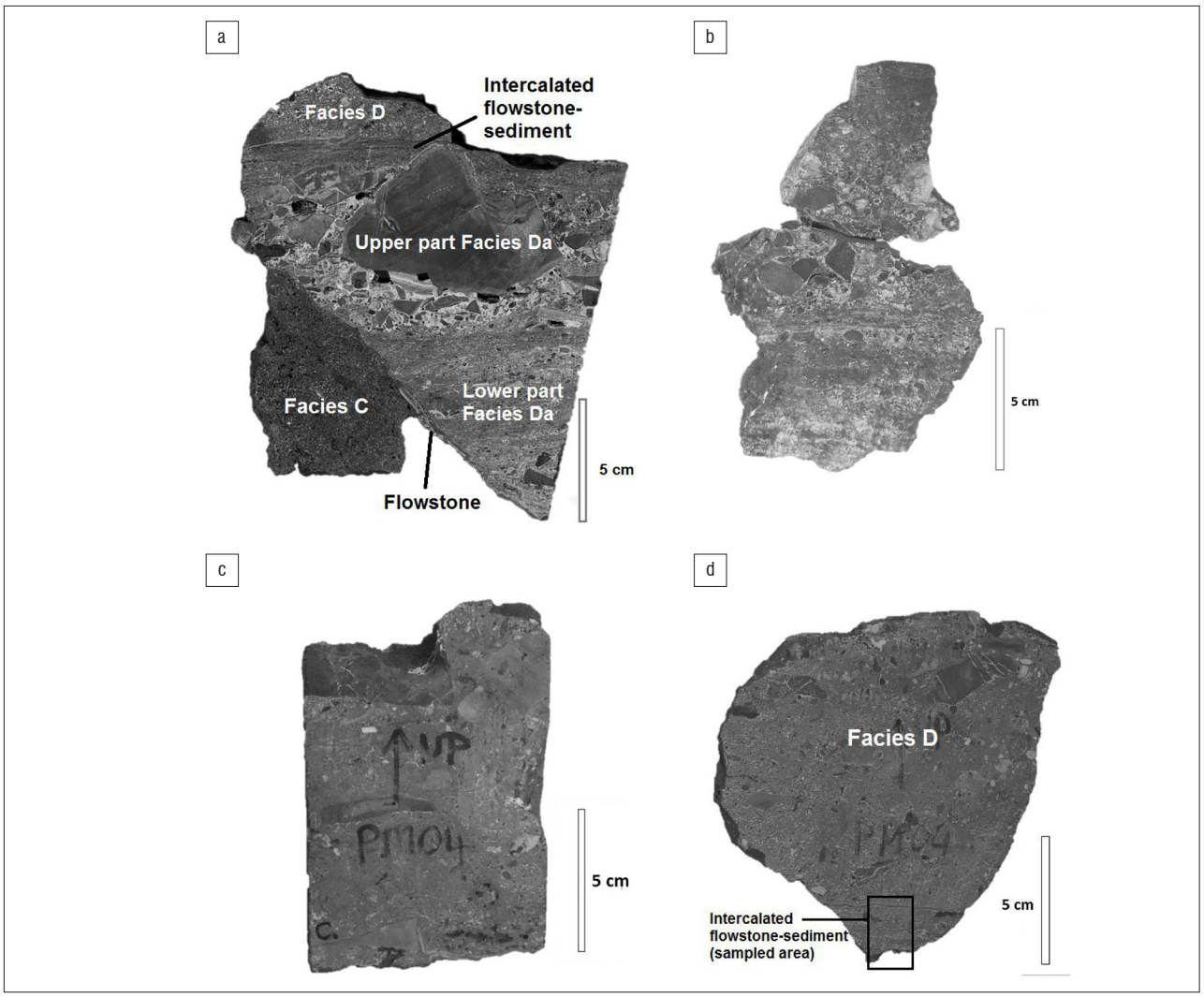


Figure 2: Malapa samples used in this study for carbonate stable isotope analysis: (a) UW88-PM09 (PM09), (b) PM09-2, (c) UW88-PM04 (PM04), (d) PM04-2. The PM09-2 is largely composed of Facies Da, whereas PM04 is largely Facies D.

The studied samples exhibit a number of relatable features, including an intercalated flowstone-sediment unit, which represents the top of Facies Da and the abrupt transition to Facies D. The inferred stratigraphic relationships between the samples indicate that they comprise a full profile of deposition that directly preceded (Facies Da) or coincided (Facies D) with the burial of *A. sediba*.

Feigl's staining of thin sections of sample PM09 revealed mixed carbonate mineralogy, evidence of post-depositional diagenesis, and a varying depositional environment (Figure 3). Evidence of primary acicular aragonite ($CaCO_3$) is common in all samples; however, the aragonite has largely been recrystallised to or replaced by sparry columnar calcite ($CaCO_3$). Thin carbonate layers in Facies Da, void fillings (laminated) in Facies Da and Facies D, and the flowstone drape separating Facies C and Facies Da show calcite-preserving relics of aragonite crystals. These crystals take various forms, including botryoidal, acicular or micrite crystals. Where aragonite relics are preserved, the diagenetic process is interpreted to be the result of calcitisation by thin water films.[13] However, complete dissolution of primary aragonite and replacement by secondary calcite[13] has also occurred, as evidenced by the lack of aragonite relics in parts of all facies within the samples. Crystal form confirms that Malapa was subject to fluctuating water flow and sediment input, resulting in a mixed phreatic-vadose depositional environment that was conducive to both carbonate deposition and post-depositional changes.

Sample preparation for stable isotope analysis

A total of 99 carbonate sub-samples were collected for stable isotope analysis from flowstone layers and carbonate cement within the samples described above. Samples were obtained by hand, using a standard engraving drill with removal bits measuring between 1 mm and 3 mm. Cross-contamination between sites was reduced by discarding surface material and rinsing the drill bits in 5% nitric acid, deionized water and ethanol, respectively. Carbonate powders were collected in PCR tubes, then weighed and transferred into clean glass sampling tubes.

Stable isotope analyses were performed at the Advanced Analytical Centre at the James Cook University in Cairns, Australia. The equipment used was the Thermo Scientific Delta V gas source isotope ratio mass spectrometer (IRMS) together with GasBench III and Conflo IV interfaces. Carbonate samples were digested in 99% phosphoric acid at 25 °C, and the resulting CO_2 gas was analysed after equilibrating for 18 h. Results were normalised to Vienna Pee Dee Belemnite (VPDB) using the calibrated reference materials of NBS 19 and NBS 18, and were reported in parts per mil (‰) with delta (δ) notation. Mean analytical precision of repeat reference materials is ±0.1‰ for both $\delta^{18}O$ and $\delta^{13}C$.

Replication

Flowstones and carbonate-cemented sediments are not normally subject to Hendy criteria tests. This is because of the variable nature of growth, including difficulty in defining vertical versus lateral advances over time.[20] For these types of samples, replication has been suggested as a more purposeful and valuable method in stable isotope studies.[21] In our study, sample PM09-2 was chosen for replicate analysis of the primary sample UW-PM09. The methodology for sample collection and data analysis of the replicate was the same as for the original samples.

Modelling palaeovegetation at Malapa

The values of $\delta^{18}O$ and $\delta^{13}C$ in cave deposits are governed by a complex set of variables and processes.[22,23] Under ideal equilibrium conditions, speleothem $\delta^{18}O$ values reflect cave temperature and source water $\delta^{18}O$ values.[22,24] The speleothem $\delta^{13}C$ values are indicative of overlying vegetation, atmospheric CO_2 and carbonate host rock-source water interaction.[24] However, stable isotope analysis of cave carbonates, in the context of palaeoclimate and palaeoenvironmental determinations, relies on several constraints and assumptions.[22,24-27]

Fractionation refers to the relationship between stable isotope values in the source water (or gas) and the resulting carbonate (Equation 1).[23,28-30]

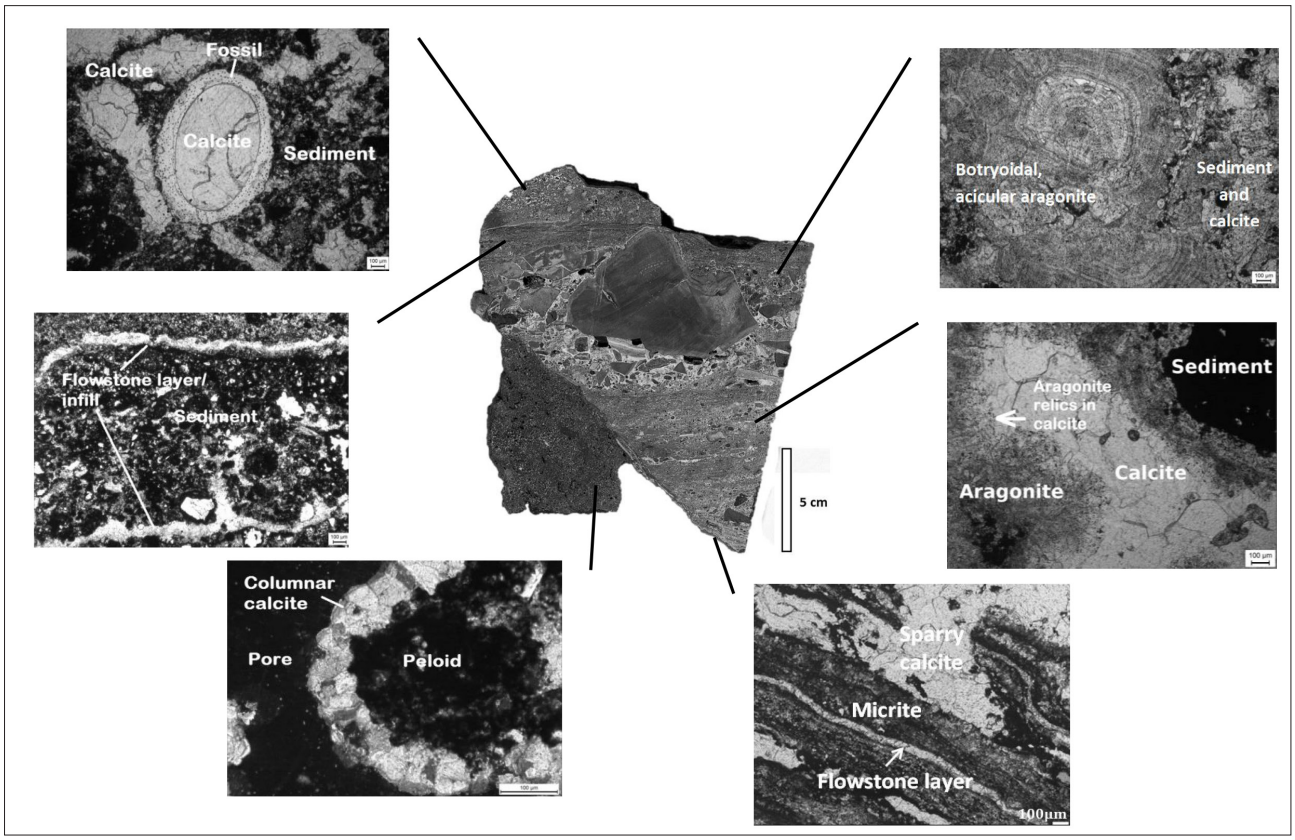


Figure 3: Thin section views from Malapa sample UW-PM09. The thin sections are shown in plane-polarised light, except Facies C which is viewed in cross-polarised light. The views illustrate a mix of carbonate mineralogy, including aragonite relics in calcite, as well as fossils and sediments.

Fractionation leads to the enrichment or depletion of the heavier isotope as phase changes (gas–liquid–solid) and reactions occur from the source to the resulting carbonate. Temperature-dependent equilibrium fractionation factors are available for carbon isotopes within cave carbonates (Equation 1).[28] However, the accurate application of equilibrium fractionation factors requires that either the source water or soil-gas stable isotope value, or the temperature at time of deposition must be known.[26,31] The fractionation equation is as follows:

$$\alpha_{\text{carbonate-source}} = \frac{1000 + \delta^{18}O(\delta^{13}C)_{\text{carbonate}}}{1000 + \delta^{18}O(\delta^{13}C)_{\text{source}}} \qquad \text{Equation 1}$$

where α (alpha) = fractionation factor, which is temperature-dependent; and the values for stable isotopes are relevant to the same standard.

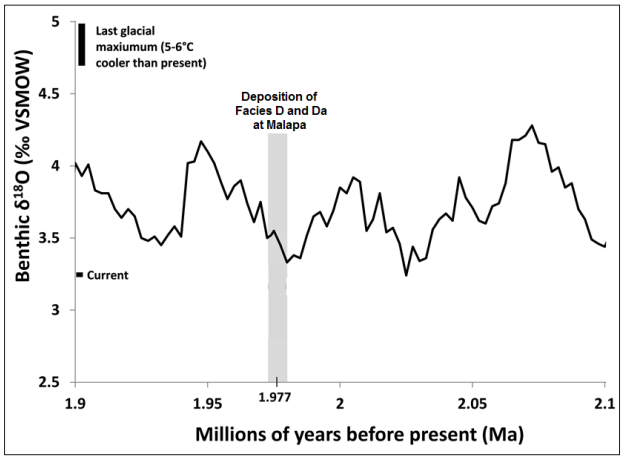
Constraining all parameters that influence stable isotope fractionation in palaeocave settings such as Malapa is not possible, if the cave environment and source-water chemistry are unknown.[26] Although a number of these factors, including the depositional environment, can be determined using petrographic studies and statistical analysis of stable isotope results (i.e. Hendy criteria tests for kinetic fractionation), others must be estimated. Examples of estimable factors include source-water $\delta^{13}C$ values and cave temperature. Here, we outline the estimable variables and fractionation factors used to calculate source carbon $\delta^{13}C$ values from the Malapa stable isotope values.

Fractionation factors

The primary equilibrium fractionation factors chosen for carbon source calculations in our study were those of Romanek et al.[32] These factors, which include calcite-$CO_{2(g)}$ and aragonite-$CO_{2(g)}$ equilibrium reactions, were obtained experimentally[32] and have been used in travertine studies to determine dissolved organic and inorganic carbon components.[33] In addition, we used re-evaluated factors for stepwise calcite-$CO_{2(g)}$ fractionation[34] and back-calculated factors for step-wise aragonite-$CO_{2(g)}$ fractionation.[32,34]

Cave temperature

Cave interiors have minimal fluctuation in temperature, with cave temperature equalling surface MAT.[21,35] Estimates of temperature at the time of cave carbonate deposition are based on known changes in MAT during glacial and interglacial periods. The minimum temperature in southern Africa during the Last Glacial Maximum (LGM) is consistent with the global temperature decrease, and has been quantified as MAT minus 6 °C.[36] This estimate has been obtained by analysing noble gases in groundwater.[36]



VSMOW – Vienna Standard Mean Ocean Water

Figure 4: Change in benthic $\delta^{18}O$ during the dated Malapa time period. The benthic $\delta^{18}O$ stack of 57 globally distributed sites by Lisiecki and Raymo[37] shows that changes in average ocean temperature are cyclical, and could indicate global oceanic or atmospheric cooling during the Malapa sample time period.

The samples in our study were deposited within a short period of 6000 years, at 1.977±0.003 Ma. During this period global temperature might have been relatively cool, as evidenced by changes in benthic $\delta^{18}O$ values (Figure 4).[37] Tectonic uplift and local insolation also have a bearing on MAT; however, the quantification of these effects at 1.977±0.003 Ma is currently unattainable. We estimated the minimum temperature at Malapa for the studied period as having been 12.1 °C, using the current MAT (calculated as 18.1 °C)[16,38] and the quantified minimum at the LGM. In addition, a maximum temperature of MAT minus 0.5 °C can be assumed, based on the southern African response to the LGM.[36]

Host rock contribution to $\delta^{13}C$ values

The contribution of host rock $\delta^{13}C$ to the dissolved inorganic carbon (DIC) pool in source water can be as high as 50%, depending on various factors (e.g. source-water pH, residence time, and host rock mineralogy).[25,30] A greater contribution of host rock $\delta^{13}C$ to the DIC pool leads to more strongly positive $\delta^{13}C$ values in the tested carbonate,[26] which skews the results in favour of a C_4 vegetation-dominated source. The dissolution and fractionation of dolomite, such as that which occurs at Malapa, is not well-defined. However, calcite and low magnesium calcite levels are considered to be influential components in the $\delta^{13}C$ value of source water as affected by dolomite host rock dissolution.[27,39] For this reason, we used the calcite fractionation factor of Mook[34] to calculate the host rock contribution to DIC in source water, in addition to an average $\delta^{13}C$ value of -0.74‰ that was obtained from two dolomite samples at Malapa. The relative contribution of host rock dissolution to the final $\delta^{13}C$ source-water value was calculated using a simple percent contribution calculation (i.e. 10% host rock + 90% soil CO_2 = source-water $\delta^{13}C$ value).

Results

Stable isotope values

The stable isotope values for the four samples are presented in Table 1. These values demonstrated a relatively small range for $\delta^{18}O$ and a larger range for $\delta^{13}C$. For all 99 values, the mean (± standard deviation) $\delta^{18}O$ value was -4.93±0.44‰, and the mean $\delta^{13}C$ value was -4.58±0.68‰. Trends moving up the growth axis profile (i.e. from older to younger) show a statistically significant (α=0.05) decrease in $\delta^{18}O$ values in PM09-2 (Facies Da) and the lower part of Facies Da in PM09. Replicate stable isotope results of Facies Da in UW-PM09 and PM09-2 were examined using the Welch modified 2-sample *t*-test.[40] The results confirmed that the mean values for $\delta^{18}O$ and $\delta^{13}C$ in the two samples differed significantly (α=0.05).

Source carbon values

The calculated values of source $\delta^{13}C$ varied 1‰ according to the different estimated temperatures at deposition and the fractionation factors we used. The minimum temperature estimate (12.1 °C) yielded slightly more negative values of $\delta^{13}C$ compared with the maximum temperature estimate (17.6 °C) results. In addition, the re-evaluated fractionation factors[34] resulted in more strongly positive calculated $\delta^{13}C$ values compared with the experimentally obtained factors.[32]

The greatest effect on calculated source $\delta^{13}C$ values was the contribution of host rock dissolution to the DIC pool. The modelling demonstrates that as the estimated host rock contribution increases from 0% to 50%, so the calculated source $\delta^{13}C$ values move towards the C_3 vegetation range (i.e. becoming more negative). This result can be explained by the model's accounting for the effect of host rock DIC on the $\delta^{13}C$ values of the tested carbonate. For a host rock contribution to the DIC pool of up to 50%, the calculated source $\delta^{13}C$ values for Malapa fit predominantly within the C_4 vegetation range (Figure 5).

Discussion

When carbonate samples are unaltered and are deposited in isotopic equilibrium with the cave environment, variations in the $\delta^{18}O$ values are linked to changes in palaeotemperature and palaeohydrology.[30,41] The $\delta^{13}C$ values reflect the vegetation type and interaction with the host rock.[30,41] However, the interpretation of stable isotope results from the carbonate-bearing samples of Malapa deposited at 1.977±0.003 Ma is challenging.

Table 1: Stable isotope data for analysed Malapa samples

Sample and unit/facies	*n*	$d^{18}O‰$ (VPDB)		$d^{13}C‰$ (VPDB)	
		range	mean ± s.d.	range	mean ± s.d.
PM04 Facies D	7	-5.42 to -4.77	-5.13 ± 0.23	-4.17 to -3.89	-4.35 ± 0.26
PM04-2 Facies D Intercalated unit (top Facies Da)	1 9	N/A -6.13 to -5.17	-5.23 -5.37 ± 0.30	N/A -5.10 to -4.08	-4.74 -4.73 ± 0.36
PM09-2 Facies Da	25	-5.70 to -5.10	-5.31 ± 0.14	-4.88 to -3.45	-4.26 ± 0.48
PM09 Facies D Intercalated unit (top Facies Da) Upper part of Facies Da Lower part of Facies Da All Facies Da Flowstone drape separating Facies C[a] and Facies Da	 1 5 13 33 51 5	 N/A -6.14 to -4.33 -5.25 to -4.25 -5.33 to -4.05 -6.14 to -4.05 -5.08 to -3.84	 -4.25 -4.89 ± 0.75 -4.75 ± 0.27 -4.65 ± 0.29 -4.76 ± 0.27 -4.43 ± 0.53	 N/A -5.46 to -4.62 -5.02 to -2.09 -5.65 to -4.30 -5.65 to -2.09 -5.29 to -3.55	 -4.80 -5.04 ± 0.33 -4.29 ± 0.86 -4.90 ± 0.26 -4.75 ± 0.32 -4.35 ± 0.72

[a]*Facies C is undated; therefore, results for this facies are not presented in this study*

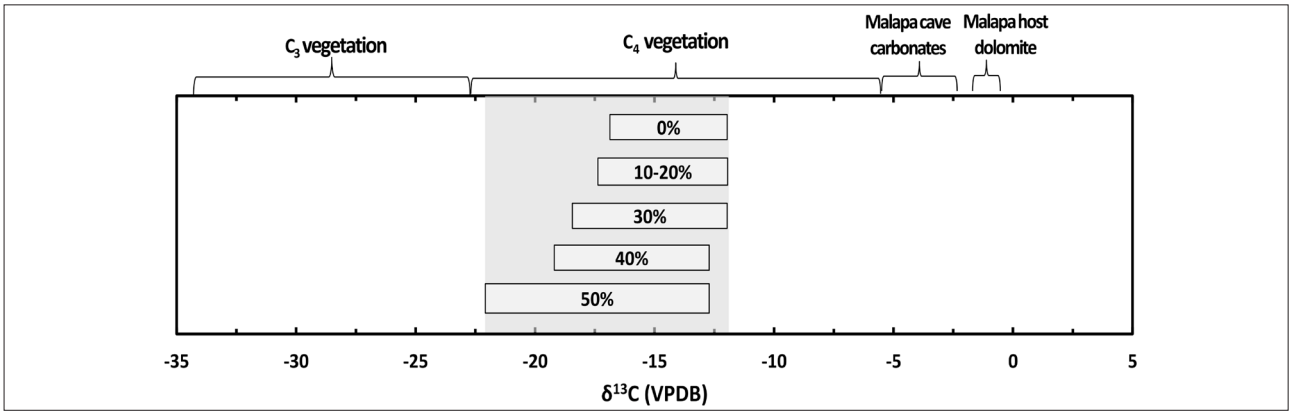


Figure 5: Calculated source $\delta^{13}C$ values for Malapa according to host rock contribution. The percentage (0%–50%) of host rock contribution to the DIC pool determines the final value of source $\delta^{13}C$. The source of carbon from vegetation overlying Malapa at 1.977 Ma is dominated by C_4 vegetation when the host rock contribution to the DIC pool is 50% or less. If a higher contribution by the host rock is assumed, results indicate a greater contribution by C_3 vegetation ($\delta^{13}C$ values for vegetation type as reported by Faure and Mensing[41]).

The sampled carbonates have been altered and are of mixed mineralogy, and the effect of the host rock contribution to the DIC pool is difficult to estimate. In particular, the $\delta^{18}O$ values can be markedly affected by alteration, with timing of the diagenetic event(s) difficult to define.[42,43]

Climate change and cyclicity in Plio-Pleistocene southern Africa have been linked to, and primarily attributed to, North Atlantic sea surface temperatures and high-latitude ice volumes.[44] The Earth's orbital cycles also played a role.[9] In particular, the influence of orbital precession on monsoonal patterns during that period is considered to have been a major driver of climate,[9] especially prior to 2.8 Ma when 23 000-year cycles dominated.[45] The vegetation type fluctuated in response to glacial (stadial) and interglacial periods and the associated changes in rainfall and MAT, with shifts in dominant vegetation occurring rapidly.[46] Hopley[9] suggests that after 1.7 Ma, a shift occurred towards grassland (C_4 vegetation) and away from forested landscapes (C_3 vegetation), with increasing aridity. This theory is in keeping with the change towards higher amplitude 40 000-year cycle glacial periods after 1.7 Ma.[45]

The suggestion that the landscape in the region of Malapa at 1.977±0.03 Ma was dominated by C_4 vegetation aligns with estimates for palaeovegetation patterns at Gladysvale Cave, albeit for the later time of ~1.8 Ma (Figure 6).[10] In addition, the Malapa vegetation might have been similar to the mixed C_3-dominant vegetation in the Makapansgat Valley for the reported period, although Malapa carbonates have a more positive average $\delta^{13}C$ value.[9] Despite difficulties in precisely correlating the timing of carbonate deposition between sites, and the effects of host rock contribution, the Malapa results could indicate that the CoH experienced a shift towards a C_4-dominant vegetation before 1.7 Ma.

Deposition of the Malapa carbonates occurred at a time when benthic $\delta^{18}O$ values were increasing (Figure 4). An increase in the benthic $\delta^{18}O$ value indicates an increase in global ice volume, because O-16 accumulates in ice and snow[47]; or it might signal cooler ocean temperatures that precede global atmospheric changes[48]. Thus, a simple interpretation of benthic values during the period under study suggests that the MAT at Malapa was decreasing. However, changes in source-water $\delta^{18}O$ values, including an increase in precipitation amount, or a change in the factors affecting $\delta^{18}O$ values in precipitation (e.g. continentality, wind direction, insolation and orbital cycles) would better explain this trend. Considering the large number of variables, the data from our study were not sufficiently constrained to provide accurate palaeotemperature estimates. Nevertheless, $\delta^{18}O$ values might reflect a change in palaeohydrology.

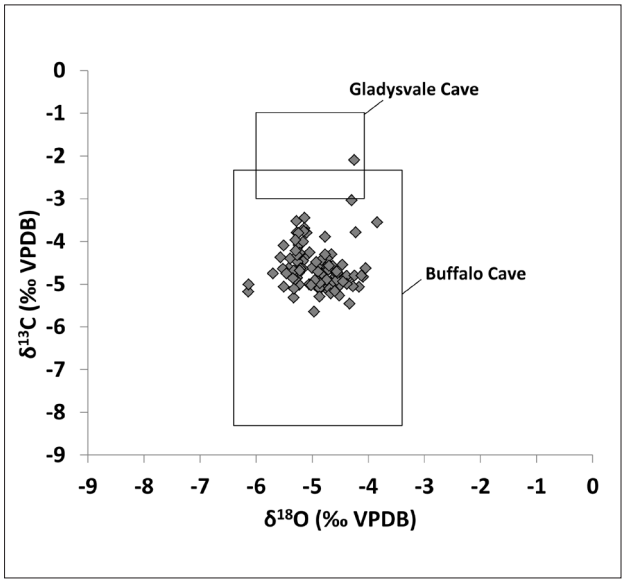


Figure 6: Comparison of stable isotope results for early Pleistocene cave carbonates in South Africa. Stable isotope results from Malapa (rhombus) are shown with reported ranges of results from Gladysvale Cave[10] in the CoH and Buffalo Cave[9] in Makapansgat Valley.

Interpretations of relative changes in aridity using $\delta^{18}O$ values in cave carbonates within the summer rainfall region of South Africa differ.[46,49] A decrease in the $\delta^{18}O$ values of cave carbonates has been interpreted as indicating a drier climate.[9] However, deep vertical convection, ascribed to movement of the intertropical convergence zone and subtropical highs throughout the seasonal year[50], also lead to a decrease of $\delta^{18}O$ values in precipitation and therefore also in cave carbonates. Modern data from the Global Network of Isotopes in Precipitation for Pretoria show that as rainfall and temperatures increase, the $\delta^{18}O$ value in precipitation decreases – known as the rainfall 'amount effect' (Figure 7).[51-53] Furthermore, there is a strong correlation between increased rainfall in East Africa as a consequence of a positive Indian Ocean dipole, and significant depletions in $\delta^{18}O$ values in precipitation.[52]

Implications for palaeovegetation–evolution linkages

The suggested palaeovegetation at Malapa, as reconstructed in this paper, provides an important insight into the ecological niche that *A. sediba* occupied. In studying pollen and phytolith remains recovered from plaque on teeth of *A. sediba* fossils, Henry et al.[54] found that their diet consisted wholly of leaves, fruits, the bark of trees, and herbaceous plants (C_3 vegetation). The diet of *A. sediba* is at odds with the diet of other hominin species (such as *A. africanus*)[55] in the CoH, including those from the same time period.[54] Therefore the location of Malapa, in a relatively sheltered valley near the confluence of two streams, might have provided the ideal riparian environment for *A. sediba* in an otherwise C_4-dominated grassland.

Conclusion

Stable isotope values of cave carbonates at Malapa during the early Pleistocene are similar to those from two previous studies conducted in the summer rainfall region of South Africa. The $\delta^{13}C$ values have been used to reconstruct preliminary palaeovegetation estimates of a mixed, but possibly C_4-dominant, landscape at 1.977 ± 0.003 Ma. Additionally, current relationships between $\delta^{18}O$ values in rainfall and rainfall amounts, and the slight decrease in the $\delta^{18}O$ values as the samples become younger, may indicate that there was an increase in rainfall and/or a change in source-water provenance during the tested time period. However, the current data are not sufficiently well constrained to derive accurate temperature estimates using $\delta^{18}O$ values, or to quantify the effect of diagenesis on $\delta^{13}C$ values. Nevertheless, carbonate-cemented cave sediments from hominin sites in the CoH are an important resource of proxies for future studies of palaeoclimate and palaeoenvironments.

Acknowledgements

We thank the South African Heritage Research Authority (SAHRA) and the Nash family for allowing us generous access to the study sites. Samples were collected under a permit granted by SAHRA (permit number 80/08/09/001/51). Funding was received by P.D. from the Australian Research Council (DP140104282), the University of the Witwatersrand, and James Cook University. We also thank the School of Geosciences and the Evolutionary Studies Institute at the University of the Witwatersrand, and the South African National Centre of Excellence in Palaeosciences, for their logistical support. Special thanks to staff and colleagues at James Cook University, especially Dr Liz Tynan for her invaluable comments on the draft manuscript.

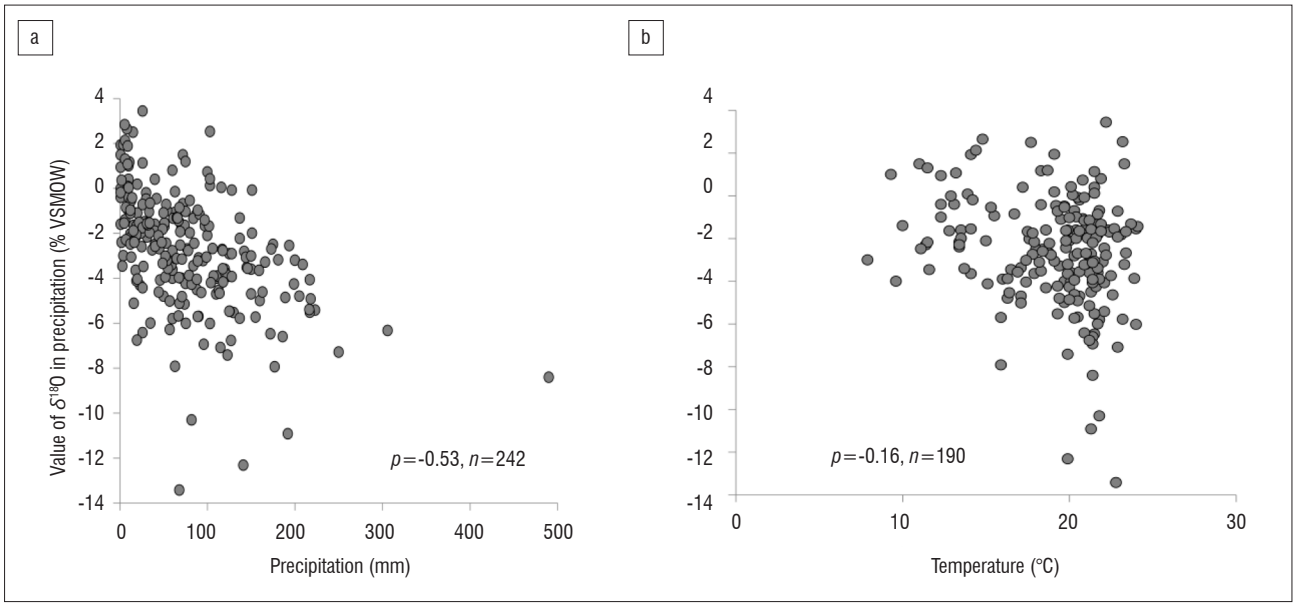


Figure 7: Oxygen isotopes in precipitation, measured for Pretoria. Monotonic relationship (Spearman's rank correlation ρ) between reported monthly $\delta^{18}O$ values in precipitation and (a) total monthly precipitation, (b) average monthly temperatures. Both results were significant at the 95% level.

VSMOW = Vienna Standard Mean Ocean Water

Data source: IAEA/WMO[48]

Authors' contributions

P.D. initiated the project, and supplied the samples. E.H. prepared the samples, performed the isotope analyses and results analysis, completed calculations for modelling, and wrote the initial version of the manuscript. P.D. contributed knowledge of the study area, commented on the interpretation of results, named 'Facies Da', and reviewed the manuscript. C.P. assisted with the modelling, provided direction for the research, and reviewed the manuscript. L.B. provided logistical support during the field component of the project, and facilitated access to the site.

References

1. deMenocal P. Climate and human evolution. Science. 2011;331:540–2. http://dx.doi.org/10.1126/science.1190683
2. Potts R. Hominin evolution in settings of strong environmental variability. Quat Sci Rev. 2013;73:1–13. http://dx.doi.org/10.1016/j.quascirev.2013.04.003
3. Vrba E. On the connections between paleoclimate and evolution. In: Vrba E, Denton G, Partridge T, Burckle L, editors. Paleoclimate and evolution, with emphasis on human origins. New Haven, USA and London, UK: Yale University; 1995. p. 24–45.
4. White WB. Cave sediments and paleoclimate. J Cave Karst Stud. 2007;69:76–93.
5. Dirks PHGM, Berger LR. Hominin-bearing caves and landscape dynamics in the Cradle of Humankind, South Africa. J Afr Earth Sci. 2013;78:109–31. http://dx.doi.org/10.1016/j.jafrearsci.2012.09.012
6. Lundelius Jr EL. Cave site contributions to vertebrate history. Alcheringa. 2006;30:195–210. http://dx.doi.org/10.1080/03115510609506863
7. Dirks PHGM, Kibii JM, Kuhn BF, Steininger C, Churchill SE, Kramers JD, et al. Geological setting and age of *Australopithecus sediba* from Southern Africa. Science. 2010;328:205–208. http://dx.doi.org/10.1126/science.1184950
8. Hopley PJ, Maslin MA. Climate-averaging of terrestrial faunas: An example from the Plio-Pleistocene of South Africa. Paleobiology. 2010;36:32–50. http://dx.doi.org/10.1666/0094-8373-36.1.32
9. Hopley PJ, Marshall JD, Weedon GP, Latham AG, Herries AIR, Kuykendall KL. Orbital forcing and the spread of C4 grasses in the late Neogene: Stable isotope evidence from South African speleothems. J Hum Evol. 2007;53:620–634. http://dx.doi.org/10.1016/j.jhevol.2007.03.007
10. Pickering R, Hancox PJ, Lee-Thorp JA, Grün R, Mortimer GE, McCulloch M, et al. Stratigraphy, U-Th chronology, and paleoenvironments at Gladysvale Cave: Insights into the climatic control of South African hominin-bearing cave deposits. J Hum Evol. 2007;53:602–619. http://dx.doi.org/10.1016/j.jhevol.2007.02.005
11. Latham AG, Herries A, Quinney P, Sinclair A, Kuykendall K. The Makapansgat Australopithecine site from a speleological perspective. Geol Soc Sp. 1999;165:61–77. http://dx.doi.org/10.1144/GSL.SP.1999.165.01.05
12. Woodhead J, Pickering R. Beyond 500 ka: progress and prospects in the U-Pb chronology of speleothems, and their application to studies in palaeoclimate, human evolution, biodiversity and tectonics. Chem Geol. 2012;322–323:290–299. http://dx.doi.org/10.1016/j.chemgeo.2012.06.017
13. Hopley PJ, Marshall JD, Latham AG. Speleothem preservation and diagenesis in South African hominin sites: Implications for paleoenvironments and geochronology. Geoarchaeology. 2009;24:519–547. http://dx.doi.org/10.1002/gea.20282
14. Pickering R, Dirks P, Jinnah Z, de Ruiter D, Churchil S, Herries A, et al. *Australopithecus sediba* at 1.977 Ma and implications for the origins of the Genus *Homo*. Science. 2011;333:1421–1423. http://dx.doi.org/10.1126/science.1203697
15. Berger LR, De Ruiter DJ, Churchill SE, Schmid P, Carlson KJ, Dirks PHGM, et al. *Australopithecus sediba*: a new species of Homo-like Australopith from South Africa. Science. 2010;328:195–204. http://dx.doi.org/10.1126/science.1184944
16. South African Weather Service. Weather data for Pretoria, South Africa. 2013.
17. SANBI(a). South African National Biodiversity Institute. Savanna biome [website page on the Internet]. [Cited 2015 May 30]. Available from: http://www.plantzafrica.com/vegetation/savanna.htm.
18. Eloff G. The phytosociology of the natural vegetation occurring in the Cradle of Humankind world heritage site, Gauteng, South Africa. [Thesis]. Pretoria: University of South Africa; 2010.
19. SANBI(b). South African National Biodiversity Institute. Grassland biome [website page on the Internet]. [Cited 2015 May 30]. Available from:http://www.plantzafrica.com/vegetation/grassland.htm.
20. Li ZH, Driese SG, Cheng H. A multiple cave deposit assessment of suitability of speleothem isotopes for reconstructing palaeo-vegetation and palaeo-temperature. Sedimentology. 2014;61:749–766. http://dx.doi.org/10.1111/sed.12078
21. Dorale J, Liu Z. Limitations of Hendy test criteria in judging the paleoclimate suitability of speleothems and the need for replication. J Cave Karst Stud. 2009;71:73–80.
22. Lachniet MS. Climatic and environmental controls on speleothem oxygen-isotope values. Quat Sci Rev. 2009;28:412–432. http://dx.doi.org/10.1016/j.quascirev.2008.10.021
23. McDermott F. Palaeo-climate reconstruction from stable isotope variations in speleothems: A review. Quat Sci Rev. 2004;23:901–918. http://dx.doi.org/10.1016/j.quascirev.2003.06.021
24. Mickler PJ, Banner JL, Stern L, Asmerom Y, Edwards RL, Ito E. Stable isotope variations in modern tropical speleothems: Evaluating equilibrium vs. kinetic isotope effects. Geochim Cosmochim Ac. 2004;68:4381–4393. http://dx.doi.org/10.1016/j.gca.2004.02.012
25. Clark I, Fritz P. Environmental isotopes in hydrogeology. USA: CRC Press; 1997.
26. Fairchild IJ, Baker A. Speleothem science: From process to past environments. Chichester, West Sussex UK: John Wiley & Sons Ltd; 2012. http://dx.doi.org/10.1002/9781444361094
27. Ford D, Williams P. Karst hydrogeology and geomorphology. Chichester, UK: John Wiley & Sons Ltd; 2007. http://dx.doi.org/10.1002/9781118684986
28. Chacko T, Deines P. Theoretical calculation of oxygen isotope fractionation factors in carbonate systems. Geochim Cosmochim Ac. 2008;72:3642–3660. http://dx.doi.org/10.1016/j.gca.2008.06.001
29. Chacko T, Mayeda TK, Clayton RN, Goldsmith JR. Oxygen and carbon isotope fractionations between CO_2 and calcite. Geochim Cosmochim Ac. 1991;55:2867–2882. http://dx.doi.org/10.1016/0016-7037(91)90452-B
30. Hendy CH. The isotopic geochemistry of speleothems–I. The calculation of the effects of different modes of formation on the isotopic composition of speleothems and their applicability as palaeoclimatic indicators. Geochim Cosmochim Ac. 1971;35:801–824. http://dx.doi.org/10.1016/0016-7037(71)90127-X
31. Daëron M, Guo W, Eiler JM, Genty D, Blamart D, Boch R, et al. $^{13}C^{18}O$ clumping in speleothems: Observations from natural caves and precipitation experiments. Geochim Cosmochim Ac. 2011;75:3303–3317. http://dx.doi.org/10.1016/j.gca.2010.10.032
32. Romanek CS, Grossman EL, Morse JW. Carbon isotopic fractionation in synthetic aragonite and calcite: Effects of temperature and precipitation rate. Geochim Cosmochim Ac. 1992;56:419–430. http://dx.doi.org/10.1016/0016-7037(92)90142-6
33. Prado-Pérez AJ, Huertas AD, Crespo MT, Sánchez AM, Pérez Del Villar L. Late Pleistocene and Holocene mid-latitude palaeoclimatic and palaeoenvironmental reconstruction: An approach based on the isotopic record from a travertine formation in the Guadix-Baza basin, Spain. Geol Mag. 2013;150:602–625. http://dx.doi.org/10.1017/S0016756812000726
34. Mook W. Introduction to isotope hydrology: Stable and radioactive isotopes of hydrogen, oxygen and carbon. London: Taylor & Francis; 2006.
35. Talma AS, Vogel JC. Late Quaternary paleotemperatures derived from a speleothem from Cango Caves, Cape Province, South Africa. Quat Res. 1992;37:203–213. http://dx.doi.org/10.1016/0033-5894(92)90082-T
36. Stute M, Schlosser P. Atmospheric noble gases. In: Cook P, Herczeg A, editors. Environmental tracers in subsurface hydrology. Massachusetts, USA: Kluwer Academic Publishers; 2000. p. 349–377. http://dx.doi.org/10.1007/978-1-4615-4557-6_11

37. Lisiecki LE. A benthic δ^{13}C-based proxy for atmospheric pCO_2 over the last 1.5 Myr. Geophys Res Lett. 2010;37:1–5. http://dx.doi.org/10.1029/2010GL045109

38. Schaefli B, Hingray B, Picouet C, Musy A. Snow hydrology in mountainous regions. In: Hingray B, Picouet C, Musy A, editors. Hydrology: A science for engineers. USA: CRC Press; 2105. p. 274–320.

39. Johnson KR, Hu C, Belshaw NS, Henderson GM. Seasonal trace-element and stable-isotope variations in a Chinese speleothem: The potential for high-resolution paleomonsoon reconstruction. Earth Planet Sci Lett. 2006;244:394–407. ttp://dx.doi.org/10.1016/j.epsl.2006.01.064

40. Jones R, Gilliver R, Robson S, Edwards W. S-Plus for the analysis of biological data. Aitkenvale, Queensland: James Cook Univerity Press; 2013.

41. Bar-Matthews M, Ayalon A, Kaufman A. Timing and hydrological conditions of Sapropel events in the Eastern Mediterranean, as evident from speleothems, Soreq cave, Israel. Chem Geol. 2000;169:145–156. http://dx.doi.org/10.1016/S0009-2541(99)00232-6

42. Perrin C, Prestimonaco L, Servelle G, Tilhac R, Maury M, Cabrol P. Aragonite-calcite speleothems: Identifying original and diagenetic features. J Sediment Res. 2014;84:245–269. http://dx.doi.org/10.2110/jsr.2014.17

43. Zhang H, Cai Y, Tan L, Qin S, An Z. Stable isotope composition alteration produced by the aragonite-to-calcite transformation in speleothems and implications for paleoclimate reconstructions. Sediment Geol. 2014;309:1–14.

44. deMenocal P, Bloemendal J. Plio-Pleistocene climatic variability in subtropical Africa and the paleoenvironment of hominid evolution: A combined data-model approach. In: Vrba E, Denton G, Partridge T, Burckle L, editors. Paleoclimate and evolution, with emphasis on human origins. New Haven, USA and London, UK: Yale University Press; 1995. p. 262–288.

45. deMenocal P. African climate change and faunal evolution during the Pliocence-Pleistocene. Earth Planet Sci Lett. 2004;220:3–24. http://dx.doi.org/10.1016/S0012-821X(04)00003-2

46. Hopley PJ, Weedon GP, Marshall JD, Herries AIR, Latham AG, Kuykendall KL. High- and low-latitude orbital forcing of early hominin habitats in South Africa. Earth Planet Sci Lett. 2007;256:419–432. http://dx.doi.org/10.1016/j.epsl.2007.01.031

47. McClymont EL, Sosdian SM, Rosell-Melé A, Rosenthal Y. Pleistocene sea-surface temperature evolution: Early cooling, delayed glacial intensification, and implications for the mid-Pleistocene climate transition. Earth-Science Reviews. 2013;123:173–93. http://dx.doi.org/10.1016/j.earscirev.2013.04.006

48. Lisiecki LE, Raymo ME. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18}O records. Paleoceanography. 2005;20(2), Art. #PA1003, 17 pages. http://dx.doi.org/10.1029/2004PA001071

49. Stager JC, Ryves DB, King C, Madson J, Hazzard M, Neumann FH, et al. Late Holocene precipitation variability in the summer rainfall region of South Africa. Quat Sci Rev. 2013;67:105–20. http://dx.doi.org/10.1016/j.quascirev.2013.01.022

50. Feng X, Faiia AM, Posmentier ES. Seasonality of isotopes in precipitation: A global perspective. J Geophys Res Atmos. 2009;114, Art. #D08116, 12 pages. http://dx.doi.org/10.1029/2008JD011279

51. IAEA/WMO. Global Network of Isotopes in Precipitation: The GNIP database. 2014. [Cited 2014, December 31]. Available from: http://www.iaea.org/water.

52. Vuille M, Werner M, Bradley RS, Chan RY, Keimig F. Stable isotopes in East African precipitation record Indian Ocean zonal mode. Geophys Res Lett. 2005;32:1–5. http://dx.doi.org/10.1029/2005GL023876

53. Gat, J. R., Mook, W. G. and Meijer, H.: Environmental isotopes in the hydrological cycle, in: Vol. II Atmospheric Water, edited by: Mook, W., International Atomic Energy Agency, Groningen, 2000. 73 pages.

54. Henry AG, Ungar PS, Passey BH, Sponheimer M, Rossouw L, Bamford M, et al. The diet of Australopithecus sediba. Nature. 2012;486:90–93. http://dx.doi.org/10.1038/nature11185

55. Klein R. Stable carbon isotopes and human evolution. Proceedings of the National Academy of Sciences. 2013;110(26):10470–10472. http://dx.doi.org/10.1073/pnas.1307308110

Note: This article is accompanied by supplementary material.

Archiving South African digital research data: How ready are we?

AUTHORS:
Margaret M. Koopman[1]
Karin de Jager[2]

AFFILIATIONS:
[1]South African Environmental Observation Network – Fynbos Node, Cape Town, South Africa
[2]Library and Information Studies Centre, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Margaret Koopman

EMAIL:
margaret@saeon.ac.za

POSTAL ADDRESS:
SAEON – Fynbos Node, Private Bag X07, Claremont 7735, South Africa

DATES:
Received: 19 Aug. 2015
Revised: 08 Feb. 2016
Accepted: 13 Feb. 2016

KEYWORDS:
research data management; long-term ecological data; data repositories; data sharing; data preservation

HOW TO CITE:
Koopman MM, De Jager K. Archiving South African digital research data: How ready are we? S Afr J Sci. 2016;112(7/8), Art. #2015-0316, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20150316

Digital data archiving and research data management have become increasingly important for institutions in South Africa, particularly after the announcement by the National Research Foundation, one of the principal South African academic research funders, recommending these actions for the research that they fund. A case study undertaken during the latter half of 2014, among the biological sciences researchers at a South African university, explored the state of data management and archiving at this institution and the readiness of researchers to engage with sharing their digital research data through repositories. It was found that while some researchers were already engaged with digital data archiving in repositories, neither researchers nor the university had implemented systematic research data management.

Introduction

A number of articles published in this journal are pertinent to the topic of digital data archiving[1,2], in particular the need for the preservation of long-term ecological data sets, which are crucial for understanding the management of the South African environment[3]. Research data have not traditionally had a home in university libraries or university archives, and have instead remained the responsibility of research units and researchers, or, in some cases, have been archived in special collections associated with a particular research unit and its specialised focus.[4] Data are the currency of research; but analogue and digital research data generated within academia have largely been an invisible resource utilised within the research unit and shared with a select group of trusted colleagues, and consequently their management is poorly understood. Digital data may have various states – raw data, which probably contain errors, require verification and, without metadata, only have meaning within a research discipline, and, at the other end of the spectrum, analysed data with metadata that can be downloaded from a repository and understood more broadly across disciplines. Each research discipline produces unique data, which require a range of specialised metadata languages and ontologies as well as subject-focused management and archiving solutions.[5]

The international focus on research data makes it important for South African researchers and policymakers to engage with the imperatives of ensuring that data are managed in a way that enables long-term security and accessibility. Data have commercial and intrinsic value, and in both cases it is important that they are archived for future use, particularly because re-collecting data is costly, in both time and money.[6]

The international context

The Advanced Research Projects Agency Network (ARPANET) was established in 1969 specifically to enable researchers to share data between laboratories in geographically distant locations.[7] ARPANET was the template upon which the Internet was subsequently built. The ubiquity of the Internet was the cornerstone of the open access initiative[8] which raised the question of universal access to research, particularly publicly funded research. There are, however, fundamental underlying factors that have led to the current preoccupation with research data archiving:

- Global climate change research has alerted governments and researchers to the value of long-term ecological studies.[9]
- Garnering funding has become an extremely competitive exercise and major funders want evidence that the research has not previously been undertaken, that the data collected will be preserved, and that the research will be open to scrutiny.[10]
- Providing underlying data is regarded as a way to prevent fraud in research, as the findings in the publications are expected to have robust scientific data underlying the research.[11]
- There is global awareness that digital records are in danger of being lost, or have already been lost because of inadequate management and preservation initiatives.[12]

Concern about the accessibility of digital data is universal and a plethora of published articles on the topic can be identified in the literature. Numerous case studies have been published which report on surveys conducted among the researchers who generate the data to establish the fate of research data.[13-15] In each case the findings were similar: lack of institutional support for research data management, lack of suitable data repositories to archive data for the long term and no incentives or mandates in place to encourage systematic data archiving, resulting in researchers keeping their data within the research unit.

Compounding this situation are attitudes towards sharing data. On the one hand, there are defensible reasons for data sharing:

- creating opportunities for further integrated research[16]
- contributing to global research initiatives, e.g. natural resource use decision-making[17]
- preventing expensive duplication of research[18]

- verifying research findings[19]
- sharing data to make research more efficient[20] and to ensure continuation of research
- making research transparent[21]
- improving researchers' international profiles[22]

On the other hand, there are the cautious and often negative attitudes of the researchers who produce the data and who are slow to archive or make their data available.[23] Ecological researchers do not have a tradition of sharing research data, other than with trusted colleagues and collaborators. In his interview for a Data Matters blog from *Scientific Data*, Gavin Simpson, a Canadian environmental scientist, succinctly presented the point of view of ecologists: 'If you've toiled in the field for years to collect data then you're not going to be very easily convinced to make the data available. It's not part of our culture'.[24] It would appear that the only way to resolve the concerns around archiving and sharing research data in a formal repository is to make data archiving mandatory, to formalise data management and to ensure that data generators benefit from sharing their data. Digital Object Identifiers (DOIs) for data enable data users to acknowledge data generators in the same way that the authors of articles and books are acknowledged. Ensuring that data are available for long-term reuse, and that they can be acknowledged through DOIs will enable data generators to use data citations, in addition to article citations, when preparing funding proposals for further research.

A number of mainstream academic journals have made data archiving mandatory – *American Naturalist*, *Molecular Ecology*, *Nature*, the Public Library of Science (PLoS) journals, Royal Society of London journals and *Science*, to name a few in the field of ecology. Funder mandates are seen as the most reliable method for making data management and archiving a part of the research data life cycle[25] – the process whereby a researcher plans and documents the various steps in data creation, processing, and analysing as part of the research design. A data management plan includes the preservation of the data and a process whereby data can be shared and reused along with the detailed data description, or metadata, that must be archived with the data. A recent editorial in *Nature*[26] pertinent to open access publishing reveals that the Research Council UK, with oversight of seven public funders in the UK, has found that mandatory open access publishing continues to be problematic, with considerably less than 100% compliance. It is not surprising that archiving research data to make them openly accessible is in a far less developed state.

International initiatives that stand out in their response to digital data archiving initiatives include:

- The Digital Curation Centre (DCC) at Edinburgh University, established in 2004[27]
- The Data Archiving and Networked Services (DANS) in the Netherlands, established in 2004/2005[28]
- The Long-Term Ecological Research Network of the National Science Foundation in the USA, established in 1982[29]
- GenBank, the genetic sequence database provided by the US National Center for Biotechnology Information, established in 1982[30]

Numerous international solutions can be used by researchers to archive their data and by policymakers and institutional managers as examples of best practice for a range of research disciplines.[31] The growth in digital data repositories has resulted in the establishment of an international, peer-reviewed process – 'The Data Seal of Approval' – initiated at DANS, which enables institutions to evaluate the reliability of their repository.[32] A repository carrying the Data Seal of Approval is immediately recognisable to researchers and policymakers as a reliable source of data and a reliable site on which to deposit data.

The South African context

Several initiatives for archiving ecological data have been in operation in South Africa, such as the Southern African Data Centre for Oceanography (SADCO) that has been in existence since the 1960s; AfrOBIS, the African component of the international Ocean Biogeographic Information System (OBIS) that was set up in 1997 as a project of the International Oceanographic Commission; and the South African Bird Ringing Unit (SAFRING) which has been contributing to knowledge about bird migration since 1948. The main digital data archiving platform focusing on environmental data is that of the South African Environmental Observation Network (SAEON), established in 2002.

A survey was undertaken to investigate the state of data management and archiving within the Department of Biological Sciences at the University of Cape Town (UCT) and the readiness of researchers to engage with sharing their digital research data in repositories. It will be seen from the results of the survey reported below that these repositories are among those utilised by the academic researchers who were surveyed.

Survey of data archiving expertise and initiatives

Researchers from the Department of Biological Sciences at UCT participated in an online multiple-choice survey, designed to be both interrogative and informative, about their data management and archiving initiatives. The survey was a variation of the computerised self-administered questionnaire[33] – an anonymous web-based survey in which the respondents linked to an identified site and completed the questionnaire online without assistance. The survey was designed using Google Forms and consisted of 32 multiple-choice questions. The research was undertaken after ethical clearance from UCT (reference number UCTLIS201408-01).

Face-to-face interviews were conducted with a small group of research technicians and emeritus/retired researchers using the questions from the self-administered survey.

Out of an estimated target population of 318, a total of 163 researchers completed the survey. The survey was conducted over a 5-week period with weekly email reminders sent out to the target population.

To enable an understanding of the Department's researchers' data management issues and activities, the questions were divided into different categories:

- researcher characteristics
- researcher funding streams
- publishing characteristics
- data characteristics
- data ownership, intellectual property and copyright
- housekeeping routines and responsibilities
- long-term data potential, archiving and metadata
- institutional engagement and data management education possibilities

Researcher characteristics

The respondents were divided into seven categories of researchers: Emeritus/Retired (10 individuals); Academic (24 individuals); Research Associate (14 individuals); Postdoctoral (21 individuals); PhD (39 individuals); Master's (32 individuals); and Honours (9 individuals). An additional category of respondents – Other/Technical – consisted of 14 individuals made up of research technicians and research support staff.

The respondents were highly qualified, with 71% having either a master's degree or a doctorate. Among the Other/Technical category there were a number of PhD and master's graduates.

Researcher funding streams

Biological research is generally expensive to fund, particularly marine and Antarctic research that require ocean-going vessels that are not available through the university. Such research requires international collaboration and involvement in government initiatives which are publicly funded programmes. The proportion of public funding of the respondents' research is high: 73% of research is at least partially funded through public funds (Table 1). Such funding renders researchers accountable to the public to make their research openly available and to ensure that their data are available for future research.

Table 1: Percentage of public funding of respondents' research

Researcher category	Percentage of public funding					
	100%	75%	50%	25%	0%	Don't know
Emeritus/ Retired	5	1	3	1	0	0
Academic	14	5	0	1	4	0
Research Associate	4	4	0	2	3	1
Postdoctoral	15	0	3	0	1	2
PhD	18	6	6	0	5	4
MSc	12	0	3	1	7	9
Honours	3	0	1	2	2	1
Total	**71**	**16**	**16**	**7**	**22**	**17**

Information on co-funding through international collaboration was extracted by an examination of published research output during 2007, 2010 and 2014. This examination demonstrated that private and overseas co-funding matched public funding in 2007, and exceeded public funding in 2010 and 2014. The authors collaborating on one paper may have been co-funded by more than one party, resulting in more co-funding categories than total articles (Figure 1).

Publishing characteristics

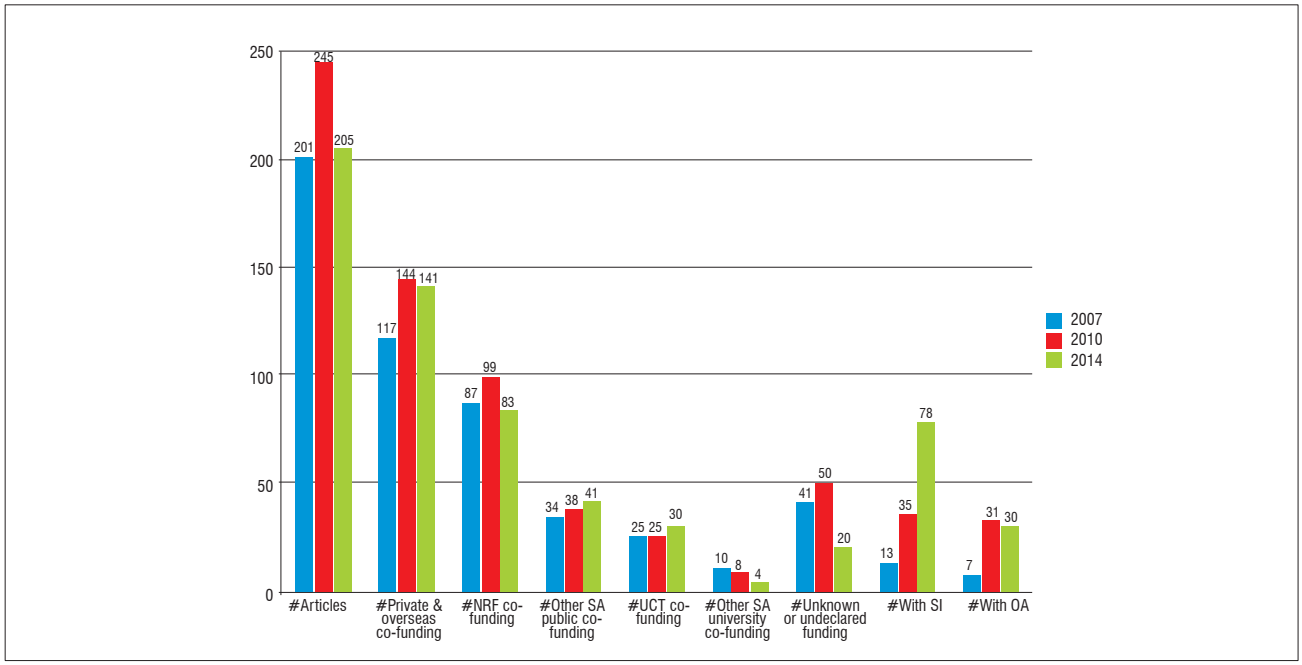
Academic research findings were made available in the past through the publications of learned societies. When learned societies ceased their publications, the task of publishing findings was taken on by discipline-specific journals published by commercial or not-for-profit scholarly publishers. In both cases research was largely hidden from the general public.

The trend for researchers to make their publications openly accessible has grown because of funding and collaboration mandates and/or to ensure that the research receives the widest audience possible. During 2014, open-access articles amounted to 15% of the total published output of the Department of Biological Sciences (Figure 1). In order to comply with future public funding mandates, the percentage of articles – with accompanying underlying data – would be expected to at least match the percentage of public funding.

An investigation was undertaken into the publication output in scientific journals of researchers in the Department of Biological Sciences, in parallel with the survey, to establish how many articles were published with supplementary information – for example data, code, images or extended bibliographies – as a way of sharing other products relating to the published research. In 2007, only 6% of the published papers were accompanied by supplementary information. This percentage climbed to 14% in 2010 and jumped to 38% in 2014 (Figure 1).

Data characteristics

It was found that researchers in the Department of Biological Sciences have worked with a range of long-term data sets, ranging from over 10 years to over 50 years in extent. Past research has generated digital data in many different formats, which have been archived on various media such as zip drives and 8-, 5¼- or 3½-inch floppy disks. Many digital data sets were in proprietary formats such as Lotus, dBase, Quattro Pro and other Corel products, or early versions of Microsoft, creating problems for long-term data accessibility. Emeritus, retired and senior academics reported data lost because of incompatibility with contemporary computer hardware, operating systems or software programs.



NRF, National Research Foundation; SI, supplementary information; OA, open access

Figure 1: Number of articles by researchers in the Department of Biological Sciences funded through private and overseas funding compared to those funded through public funding.

Other researchers had retained field or experimental notebooks which contained their raw data in analogue format. The majority (91%) of digital data formats generated by younger researchers were spreadsheets (in XML or CSV format). Improved management of digital data would prevent obsolescence and loss of data from this cohort of researchers.

Data were reused and shared within a controlled group of collaborating researchers. Very few researchers allowed open use of the data sets under their control, with the exception of the researchers in the Animal Demography Unit of the Department of Biological Sciences who managed large sets of 'citizen science' data that carried an open data mandate. Analogue data sets also existed within the department, but were largely invisible through lack of description and archiving. Many past students' data sets accompanied dissertations and remained as appendices in analogue theses, also lacking indexing and description. A number of retired research staff reported that their data either had been thrown away ($n=3$) or had nowhere to go ($n=5$) because of a lack of interest among colleagues and the institution. Early digital data had also been lost through lack of institutional support, foresight and responsible management. Instances of old digital data still in existence but inaccessible on contemporary computer platforms were common among senior academics. At the time of the investigation, there were no institutional plans in place to rescue these digital data sets.

Data ownership, intellectual property and copyright

Data ownership was found to be a key inhibitor to data sharing. Opinion about ownership varied: does the funder, the institution, the research unit, the supervisor, or the student own the data? Or is the owner a combination of all these potential data owners?

Researchers' responses varied according to the category in which they were placed, with 'Researcher ownership' scoring the highest by 'Emeritus/Retired', 'Academic' and 'Postdoctoral' respondents; 'UCT ownership' was scored highest by the 'PhD' category; and 'Supervisor ownership' of data was scored highest by the 'MSc' category. The highest score allocated by both 'Research Associate' and 'Honours' categories was for the response 'Don't know'. Other data owners were reported as 'the organisation I work for', 'citizen science observers', 'South African government', and 'open data from a repository'. Overall responses to data ownership can be seen in Figure 2.

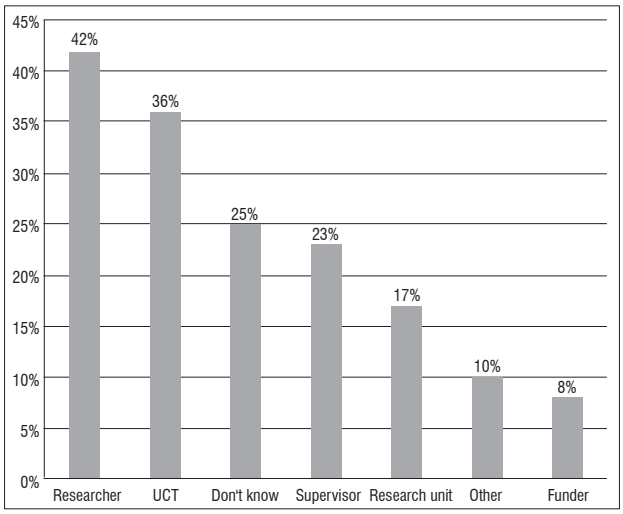


Figure 2: Percentage of respondents who attributed data ownership to each possible response.

When asked if their data should be made available for future research, 88% of researchers responded positively. But there were caveats to this response, which are reported in Table 2. The raw data revealed cascading requirements for making data available. For example, respondents were willing to share data only after publication and only if the data generator was offered co-authorship, or, after publication and only on request so that the data generator could evaluate the researcher and project wishing to use the data. Being able to trust the person with whom data would be shared was an important consideration.

Table 2: Conditions for sharing research data as reported by respondents

Condition for sharing data	% Respondents
Only after publication	62
Only on request so that I can discriminate	37
Only if open access with acknowledgement	30
Only if I am offered co-authorship	28
Only to a trusted researcher	25
Only if my data sets have DOIs	7
Only if data sets have a Creative Commons licence	3

The raw data demonstrated that the respondents who were prepared to share data through acknowledgement, inclusion of DOIs or through publishing under Creative Commons licences were those contributing to or utilising data sets such as 'citizen science' data that already had an open mandate.

In some cases, researchers indicated that there were copyright restrictions on the data they were using and that they were not permitted to share these data sets. In another case, a research group reported that their data had been misappropriated by another research group on campus, because no memorandum of understanding had been in place to specify agreed terms of data use.

Researchers were sharing data, but in the majority of cases this sharing was not through data repositories. Respondents to the survey could select multiple answers and reported sharing of data through the following methods:

- by email on request, 70%
- within published papers, 38%
- in the public domain, 17%
- through a collaborative initiative, 15%
- through a repository, 12%
- through the research unit's server, 3%

Until data ownership is resolved, through funding or collaborative agreements – and incentives such as the acknowledgement of data generators through the use of data DOIs are commonplace – data sharing will remain a contested issue.

Housekeeping routines and responsibilities

The survey interrogated data management and preservation activities such as storage, back-up routines and data migration routines. Questions on responsibility for data storage revealed a range of perceptions (Table 3) that focused data responsibility on the research unit rather than institutional IT departments, the university library or repositories. At the time of the survey, the institution took no responsibility for research data, although researchers could avail themselves of storage space on an IT server at a cost. Departmental IT personnel interviewed for the investigation (included in the Other/Technical category) indicated that they would give advice to researchers for the storage of data but that they were not responsible for researchers' data. Researchers and research units took responsibility for their data, as this was considered to be the status quo.

Table 3: The number of respondents who attributed responsibility for storage of data sets to each possible response

Data storage location	Number of respondents
Researcher/supervisor	95
Research unit	68
University library	37
National repository	33
Departmental IT personnel	28
International repository	18
Don't know	13
University IT department	12
Other	3

Researchers were diligent about back-up routines. The type and location of data back-ups is of interest as it demonstrates changing trends in data storage (Table 4). CD/DVDs are falling out of fashion and cloud storage is becoming more popular, although some researchers expressed reservations about data privacy on cloud storage. There appeared to be no consensus at the time of the survey, although external hard drives were the favoured medium, and keeping a back-up at home was the favoured location. One cannot predict how research data will be stored and backed up in the future. But, the move to cloud storage as a more accessible format which does not require the researcher to purchase or carry around additional hardware appears to be gaining popularity. Institutional commitment to research data management through the provision of staging repositories for active research data could improve the security of research data.

Table 4: Percentage of respondents who back up data at each location

Back-up location	% Respondents
Hard drive	83
PC/laptop	55
At home	42
Cloud storage	39
Office	26
Flash drive	25
Server	15
CD/DVD	4
Other	4
UCT ICTS	0.6

Long-term data potential, archiving and metadata

A range of data repositories was utilised by the researchers who were obliged to archive their data, either through collaborations, funding or publisher mandates, or through disciplinary mandates, such as for genome data. Repositories were also utilised by researchers to access data to use in their research. The repositories reportedly utilised by respondents are listed in Table 5.

Table 5: Repositories utilised by researchers in the Department of Biological Sciences

Repository	% Respondents
GenBank	23
SAEON	17
UCT Libraries Digital Repository	9
GBIF/SANBIF	8
Other†	8
EMBL	4
Dryad	3
JStor Global Plants	3
Movebank	3
AfrObis/Obis	2
SADCO	2

†Other repositories mentioned were a suite of Animal Demography Unit (UCT) databases, SANBI data archives, The British Library Sound Archive, and those of The National Marine Linefish System, The BirdLife Seabird Tracking Database, UvA-BiTS (University of Amsterdam Bird Tracking System) and Iziko Museum.

It was found that only 12% of the respondents had used a repository as a means of sharing data, although responses shown in Table 3 indicated that a higher percentage considered a repository to be the appropriate place for responsible data archiving. This apparent contradiction is understandable as routine data archiving in repositories was unknown to many researchers.

Table 6: Types of metadata considered important to describe research data as reported by respondents

Metadata field	Number of respondents
Name of creator/ research unit's name	91
Title of the data set	90
Geographical coordinates	87
Description of the data set	86
Contact details of creator/ research unit	78
Date of data creation	78
Collection methods	78
Taxonomic names	68
Beginning and end dates of project	57
Equipment used to gather data	57
Data format/s	54
Keywords	45
Names of funders	38
Copyright provisions	37
Title of umbrella project	31
Contact details of umbrella project	14
Don't assign metadata	14
Contact details of funders	9
Don't know	7

As metadata are a fundamental component of data sharing, the survey was devised to include a question for which there was a range of mandatory metadata fields as possible answers; respondents could select multiple answers. The possible answers and percentage of respondents who gave each answer are shown in Table 6.

The question that elicited the responses in Table 6 was also intended to sensitise researchers to metadata fields that could be used to describe their research, as the assignment of metadata was a new concept for many researchers at the time. Maintaining detailed descriptions about their data through the use of metadata did not appear to be a routine activity and a number of researchers indicated that they did not assign metadata. The fields shown in Table 6 represent those required by the Ecological Metadata Language (EML) standard.

Researchers were asked what they thought was the purpose of data curation (Table 7). Migrating data into formats that could be used by current software and operating systems received the lowest response. This was a neglected aspect of data management among senior academics that had resulted in data obsolescence instead of data remaining viable for long-term research.

Table 7: Purposes of data curation as reported by respondents

Purpose of data curation	% Respondents
Storing data for access and use	83
Ensuring that data are secure and backed up and available	79
Making sure data are available for future use	77
Maintaining research data in the long term to enable reuse	68
Ensuring that data are organised and indexed	54
Migrating data to new platforms/software	35

In order to build up long-term data sets for long-term ecological research such as land-use or climate change, data management will need to become an integrated part of the research life cycle.

Institutional engagement and data management education possibilities

The survey contained three questions posed in order to gauge the appetite of researchers for data management education. The questions and percentage responses can be seen in Table 8.

Table 8: Potential for data management education as indicated by researchers in the Department of Biological Sciences

Question	Answer		
	Yes	No	Other
Would you attend a workshop to discuss metadata generation?	50%	40%	Metadata are not applicable (10%)
Would you attend a workshop to discuss data management?	58%	28%	Would prefer an online resource (14%)
Do you require data management assistance?	51%	45%	Hire students to assist with data management (4%)

Various aspects of data management, such as information about metadata languages and standards, are an opportunity for librarians to develop online resources in support of data-generating disciplines. Tools such as LibGuides[34] are ideally suited for providing such online support for researchers. Although some data generators were using students to assist in data management, 51% of the respondents said that they 'would like more information about managing data efficiently'.

For UCT libraries to give appropriate support to researchers, librarians with specialised backgrounds or experience and the ability to interact with researchers would be required. Much of the advice needed at undergraduate and postgraduate levels is generic, such as file naming conventions, data back-up habits, keeping records of the what, how, when, where and why data were gathered (metadata), and types of metadata protocols required for archiving specific data types. The UCT libraries also have a role to play in directing researchers to other divisions on campus where information on topics – such as research funding, ethics support, IP support and temporary data storage – can be found.

During 2014 an eResearch Centre was established at UCT[35] which initiated a number of activities such as workshops and conferences to support research data generators. Collaborators in this initiative were UCT libraries, UCT ICTS (Information and Communication Technology Services) and the UCT Research Office who were in the process of developing a research data management policy for the university. The UCT libraries established a web presence for research data management (RDM) to advertise 'resources and training on research data management'[36].

Institutional managers have a role to play in ensuring that data management and curation are accounted for in research budgets. Whereas 88% of respondents indicated that their research should be made available for future research, only 18% budgeted for data management and data curation and only 26% had a data preservation plan.

Conclusion

The survey demonstrated that, even within the Department of Biological Sciences, research was varied and data collection and interpretation required a range of specialist skills, equipment and tools. Any discussion of metadata should include the standards and metadata languages appropriate for all types of research data in order for researchers to successfully describe their data for long-term preservation. The link between metadata and sharing has to be made in order for researchers to see the importance of comprehensive data descriptions, as without metadata, their data have no long-term value.

There had been no systematic interventions at UCT for supporting researchers with data management or data storage facilities, and an ad-hoc situation with varying success in the preservation of research data had been the status quo. Research data archiving for long-term preservation requires secure funding streams as well as training in RDM. Assistance with the development of RDM plans, soon to be required by South African research funders, is one of the ways in which the institution can assist researchers to apportion funding for data preservation.

Systematic RDM and archiving will only come about when proposed policies have been established in consultation with researchers. RDM education of the new cohort of researchers is a prerequisite for establishing systematic data archiving, and initiatives should be introduced at senior undergraduate level. Because RDM is a relatively new concept to South African researchers, support should also be offered to senior- and mid-level academic researchers so that they are sufficiently informed to ensure that student data are properly managed and archived. New research projects should include a data archiving and sharing plan as part of the overall project plan.

At the time of the survey, there was no strategy in place for the management or archiving of pre-digital or early digital research data, and some of these data were still in the hands of the retired and emeritus staff of the Department of Biological Sciences who were interviewed. Ensuring that long-term data sets are preserved is urgent and important, as it is not possible to recreate long-term ecological data because the impacts of human population expansion and resource usage change ecological systems over time.

Acknowledgements

We thank the Percy FitzPatrick Institute of African Ornithology for co-funding; the survey respondents for their participation; and the anonymous reviewers for their input in improving the manuscript.

Authors' contributions

M.M.K. was the project leader; M.M.K. and K.d.J. were responsible for project design; M.M.K. undertook the survey; K.d.J. made conceptual contributions and supervised the dissertation on which the article is based; M.M.K. prepared, analysed and documented the results of the survey; and M.M.K. and K.d.J. wrote the manuscript.

References

1. Grundlingh ML, Von St Ange UB, Bolton JJ, Bursey M, Compagno L, Cooper R, et al. AfrOBIS: A marine biogeographic information system for sub-saharan Africa. S Afr J Sci. 2007;103:91–93.
2. Van Jaarsveld AS, Pauw JC, Mundree S, Mecenero S, Coetzee BWT, Alard GF. South African Environmental Observation Network: Vision, design and status. S Afr J Sci. 2007;103:289–294.
3. Henschel J, Pauw J, Banyikwa F, Brito R, Chabwela H, Palmer T, et al. Developing the Environmental Long-Term Observatories Network of Southern Africa (ELTOSA). S Afr J Sci. 2003;99:100–108.
4. South African Bird Ringing Unit [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://safring.adu.org.za/content.php?id=1
5. Digital Curation Centre. Disciplinary metadata [homepage on the Internet]. No date [cited 2016 Jun 21] Available from: http://www.dcc.ac.uk/resources/metadata-standards
6. Kenall A, Edmonds S, Goodman L, Bal L, Flintoft L, Shanahan DR, et al. Better reporting for better research: A checklist for reproducibility. BMC Neurosci. 2015;16, Art. #44, 3 pages. http://dx.doi.org/10.1186/s12868-015-0177-z
7. Dasgupta S. Arpanet. In: Encyclopedia of virtual communities and technologies. London: Idea Group Reference; 2006. p. 173. http://dx.doi.org/10.4018/978-1-59140-563-4
8. Budapest Open Access Initiative [homepage on the Internet]. c2002 [cited 2016 Jun 21]. Available from: http://www.soros.org/openaccess/read.shtml
9. Walther GR, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, et al. Ecological responses to recent climate change. Nature. 2002;416:389–395. http://dx.doi.org/10.1038/416389a
10. The Editors. Editorial: Dr No Money: The broken science funding system. Sci Am. 2011 April 19; Mind. Available from: http://www.scientificamerican.com/article/dr-no-money/
11. Doorn P, Dillo I, Van Horik R. Lies, damned lies and research data: Can data sharing prevent data fraud? Int J Digit Curation. 2013;8:229–243. http://dx.doi.org/10.2218/ijdc.v8i1.256
12. Computer History Museum. Digital Dark Age: Revolution preview [video on the Internet]. c2011 [cited 2016 Jun 21]. Available from: https://www.youtube.com/watch?v=PSIMzirvsFc
13. Diekmann F. Data practices of agricultural scientists: Results from an exploratory study. J Agr Food Inform. 2012;13:14–34. http://dx.doi.org/10.1080/10496505.2012.636005
14. Elliot G. Otago biodiversity data management project report. Part 1. Questionnaire report. Dunedin: University of Otago Library; 2008. Available from: http://otago.ourarchive.ac.nz/handle/10523/198
15. Scaramozzino JM, Ramírez ML, McGaughey KJ. A study of faculty data curation behaviors and attitudes at a teaching-centered university. Coll Res Libr. 2012;75:349–365. http://dx.doi.org/10.5860/crl-255
16. Costello ML. Motivating online publication of data. BioScience. 2009;59:418–427. http://dx.doi.org/10.1525/bio.2009.59.5.9
17. Huang X, Hawkins BA, Qiao G. Biodiversity data sharing: Will peer-reviewed data papers work? BioScience. 2013;63(1):5–6. http://dx.doi.org/10.1525/bio.2013.63.1.2
18. Fry J, Lockyer S, Oppenheim C. Identifying benefits arising from the duration and open sharing of research data produced by UK higher education and research institutes [programme/project deposit]. c2009 [cited 2016 Jun 21]. Available from: http://repository.jisc.ac.uk/279/
19. Borgman CL. The conundrum of sharing research data. J Am Soc Inform Sci Technol. 2012;63:1059–1078. http://dx.doi.org/10.1002/asi.22634
20. Piwowar HA. Who shares? Who doesn't? Factors associated with openly archiving raw research data. PLoS ONE. 2011;6(7), Art. #e18657, 13 pages. http://dx.plos.org/10.1371/journal.pone.0018657
21. Molloy JC. The Open Knowledge Foundation: Open data means better science. PLoS Biol. 2011;9(12), Art. #e1001195, 4 pages. http://dx.plos.org/10.1371/journal.pbio.1001195
22. Piwowar HA, Day RS, Fridsma DB, Ionnidis J. Sharing detailed research data is associated with increased citation rate. PLoS ONE. 2007;2(3), Art. #e308, 5 pages. http://dx.plos.org/10.1371/journal.pone.0000308.
23. Van Noorden R. Confusion over open-data rules. Nature. 2014;515:478. http://dx.doi.org/10.1038/515478a
24. Hufton A. Data matters: Interview with Gavin Simpson. Scientific Data Updates. 2014 August 26. Available from: http://blogs.nature.com/scientificdata/2014/08/26/data-matters-interview-with- gavin-simpson/
25. Vines TH, Andrew RL, Bock DG, Franklin MT, Gilbert KJ, Kane NC, et al. Mandated data archiving greatly improves access to research data. FASEB J. 2013;27:1304–1308. http://dx.doi.org/10.1096/fj.12-218164
26. Editorial: All that glitters. Nature. 2015;520:131. http://dx.doi.org/10.1038/520131a
27. Digital Curation Centre [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.dcc.ac.uk/
28. Data Archiving and Networked Services (DANS) [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.dans.knaw.nl/en
29. The Longterm Ecological Research Network (LTER) [homepage on the Internet]. c2013 [cited 2016 Jun 21]. Available from: http://www.lternet.edu
30. GenBank overview [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.ncbi.nlm.nih.gov/genbank/
31. Registry of Research Data Repositories [homepage on the Internet]. c2012 [cited 2016 Jun 21]. Available from: http://www.re3data.org/
32. Data Seal of Approval [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://datasealofapproval.org/en/information/about/
33. Babbie E, Mouton J. The practice of social research. Cape Town: Oxford University Press; 2001.
34. SpringShare. LibGuides [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://springshare.com/libguides/
35. University of Cape Town. eResearch Centre [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.eresearch.uct.ac.za
36. University of Cape Town Libraries. Research data management [homepage on the Internet]. No date [cited 2016 Jun 21]. Available from: http://www.digitalservices.lib.uct.ac.za/dls/rdm

Development of census output areas with AZTool in South Africa

AUTHORS:
Tholang Mokhele[1,2]
Onisimo Mutanga[2]
Fethi Ahmed[3]

AFFILIATIONS:
[1]Population Health, Health Systems and Innovation, Human Sciences Research Council, Pretoria, South Africa
[2]School of Agriculture, Earth and Environmental Sciences, University of KwaZulu-Natal, South Africa
[3]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Tholang Mokhele

EMAIL:
TAMokhele@hsrc.ac.za

POSTAL ADDRESS:
Population Health, Health Systems and Innovation, Human Sciences Research Council, Private Bag X41, Pretoria 0001, South Africa

DATES:
Received: 12 Jan. 2015
Revised: 03 Aug. 2015
Accepted: 13 Feb. 2016

KEYWORDS:
automated zone design; enumeration areas; output areas

HOW TO CITE:
Mokhele T, Mutanga O, Ahmed F. Development of census output areas with AZTool in South Africa. S Afr J Sci. 2016;112(7/8), Art. #2015-0010, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20150010

The use of a single geographical unit to both collect and disseminate census data is common in many countries across the world, especially in developing countries. In South Africa this approach poses some challenges, as the design of small geographical units called enumeration areas to facilitate data collection differs considerably from the design of units that aid data analysis and interpretation. We aimed to create optimised census output areas using the Automated Zone-design Tool (AZTool) program, using the 2001 census enumeration areas as building blocks at various spatial levels, for both rural and urban settings in two South African provinces. The results were consistent and stable. The primary criterion of the confidentiality limit of 500 people was respected at all geographical levels or regions, in both urban and rural settings, for newly created optimised output areas. For the second criterion, lower intra-area correlation values at lower geographical levels for both rural and urban areas showed that higher geographical levels produced more homogeneous output areas than did lower geographical levels or regions. Our obtained intra-area correlation of 0.62 for the two provinces combined indicated that the selected homogeneity variables were good indicators of social homogeneity for creating optimised output areas in South Africa. We conclude that the AZTool software can be used to effectively and objectively create optimised output areas for South African data. Further research on the comparison of the newly created output areas with existing output areas in South Africa should be explored.

Introduction

Many countries use a single geographical layer for both census data collection and dissemination. This was the case in South Africa prior to the 2001 census. However, this practice has caused challenges for census data users. Firstly, the ideal characteristics of an area to facilitate efficient data collection are not the same as those which aid data analysis.[1,2] Secondly, in some enumeration areas (EAs) the population falls below the confidentiality limits, resulting in those EAs being merged with nearby EAs.[1,3,4] Thirdly, the design of EAs prior to census data collection did not consider social homogeneity.[4,5] Lastly, the shape compactness was also not directly considered.[1,3] Certain exceptions exist, such as the United Kingdom, where output areas have been used for census disseminations.[5-10]

In South Africa, EAs typically contain between 100 and 250 households. The most important criteria for the design of EAs are that they should not overlap, they should have boundaries that can be identified on the ground, and they should be of approximately equal population size to enable an enumerator to cover each unit within the census period.

The fact that census data are collected at household level but are disseminated at higher geographical levels, such as EAs, raises concerns. One problem is the modifiable areal unit problem (MAUP), a term first used by Openshaw[11] but originally coined by Gehlke and Biehl[12]. The MAUP has two components: (1) *scale problem* – namely the variation in results caused by the progressive aggregation of smaller areas into larger areas; and (2) *zoning problem* – the variation in results caused by different arrangements of a set of zones.[11-20] Openshaw[11] developed the automated zoning procedure (AZP) in an attempt to solve the MAUP problem. Briefly, the AZP algorithm works by iteratively combining and recombining sets of building blocks to create output areas that optimise a set of pre-specified design criteria.[10,21,22] The AZP model was further enhanced by Openshaw and Rao[23]. It was then further reviewed and extended to automated zone matching (AZM) software by Martin in 1998 and 2003, to permit its application to the intersection of two zonal geographical systems.[21] In 2006, Cockings, Martin and Harfoot at the University of Southampton developed the AZTool software from AZM. This tool was further enhanced to the current version (AZTool 1.0.3), which does not require ArcInfo for preparing .pat and .aat files.

Among many studies on the automated zone design applications, in 2002 Martin and the Office for National Statistics created output areas for the 2001 census for England and Wales using automated zone design methods[24]. These output areas were designed to respect minimum population and household threshold sizes of 100 and 40 respectively, as well as a compact shape and with a degree of homogeneity in terms of housing tenure and type. In addition, these output areas had to be nested within higher geographical regions. This project was seen as a success even though there were some concerns about the resulting abstract nature of output area boundaries.

The applications of automated zone design techniques were further employed in the health research environment by Cockings and Martin[25] and Flowerdew et al.[26] Flowerdew et al.[26] used the 1991 limiting long-term illness data in Great Britain, with enumeration districts as building blocks to construct alternative zonal systems with the AZTool zone design algorithm, to determine if neighbourhoods defined in various ways would have similar implications for health. Their results showed that, for sets of pseudo-wards that made sense in terms of population equality and shape, the zonation effect was real. Hence they concluded that it did matter where boundaries are drawn.

Haynes et al.[27] compared automated zone design program zones – the A2Z developed by Daras[28] – with areal units identified subjectively by local government officers as being communities in the city of Bristol, United Kingdom. They found that automated zone design came close to replicating the subjective communities when the balance

of objectives and boundary constraints was adjusted. In 2009, Ralphs and Ang[20] developed new geographies in New Zealand using the AZTool. Their results indicated that the newly created geographies substantially outperformed the current geographies across almost all their optimisation criteria. Ralphs and Ang[20] argued that the algorithm they used was stable and consistent, hence it could repeatedly generate high-quality solutions in a timely manner. In France, Sabel et al.[22] used the AZTool program (using 250 m x 250 m cells as building blocks) to create new zones to explore relationships between asthma and deprivation in Strasbourg. Their newly-produced synthetic neighbourhood solution performed better than the then-existing IRIS census areas, as shown by stronger statistical relationships between asthma and deprivation.

In South Africa, for the 1991 and 1996 censuses, the same EAs were used for both census enumeration and dissemination. For the 2001 census, it was decided that census data must be released for an area larger than an EA to enhance confidentiality.[4] For that purpose, two names were attached to each EA, and a spatial layer was created from the name attributes (SubPlaces and MainPlaces). In many instances the areas created were too large for most census data users. In 2005, a non-automated zone design approach was employed to create a small area layer (SAL) for dissemination of the 2001 census in an effort to meet user needs. A similar non-automated zone design approach was employed in the creation of SALs for the 2011 census data. This was mainly to create a spatial area layer that corresponded as closely as possible to the EA layer while adhering to the confidentiality limit of 500 people.[4] The following criteria were set and adhered to as far as possible for the creation of the SAL. Firstly, EAs can only be merged if they are within the same SubPlace; secondly, EAs can only be merged if they have the same EA geography type; thirdly, an EA can only be merged if its population is less than 500; and lastly, the resulting small area polygons must have a population total of 500 or more.[4] These new criteria resulted in 56 255 SALs from the previous 80 787 EAs, as shown in Table 1. Table 1 highlights the South African geographical levels or regions used for the 2001 census.

Table 1: South African geographical levels or regions for census 2001

Regions	Number	Population mean
Provinces	9	4 979 997
District municipalities[a]	52	861 923
Local municipalities[b]	257	174 397
MainPlaces	3109	14 416
SubPlaces	21 243	2110
Small area layers	56 255	797
Enumeration areas	80 787	555

Source: Stats SA

[a] *Includes 6 metropolitans that are both district and local municipalities*

[b] *Includes 20 district management areas*

The maintenance and update of provincial and municipality boundaries is the responsibility of the Municipal Demarcation Board, while the National Statistics Office (Statistics South Africa, Stats SA) is responsible for the creation and maintenance of MainPlaces, SubPlaces, SALs and EAs. In the creation of the SAL, only the census confidentiality limits and population size were addressed; social homogeneity and output shape were not. It is worth noting that out of 56 255 SALs, 13.5% of the SALs breached the confidentiality limit.[4] Although the issue of census output areas being too large for most South African census data users had been addressed by the creation of the SAL, the issue of confidentiality remains a concern. Policies for census output areas vary from country to country, but confidentiality requirements are strictly enforced in almost all countries. Therefore, the advanced techniques of automated zone design methods such as the AZTool are worth exploring for the creation of optimal output areas in South Africa. Our study attempted to address this by creating census output areas with AZTool software, using the 2001 census EAs as building blocks and prioritising the confidentiality limit (minimum population threshold of 500) as well as homogeneity, population mean target and shape compactness. In addition, we examined the performance of the AZTool program for both urban and rural areas in South Africa at different geographical levels or regions. This was to give a general picture as to how the program was likely to perform when the entire country was analysed.

Methods

Study area

The study area comprised two of the nine provinces of South Africa. These were Free State, representing rural settings, and Gauteng, the most populated province but the smallest in area, representing urban areas. Free State province includes the former homelands of Phuthaditjhaba and Botshabelo, and is one of only two provinces that did not experience any provincial boundary change for the 2001 and 2011 censuses. It experienced less than 1.5% population increase within that period. Gauteng, by contrast, is the most populated and developed province in South Africa, and had the highest population growth between 2001 and 2011 (2.7%). The analysis of both provinces provides examples of rural and urban settings in South Africa.

The provincial, district, municipality and MainPlace levels were analysed for each province, to gain a better understanding of the performance of the AZTool at each geographical level in both rural and urban settings. In the Free State, Thabo Mofutsanyane district and Maluti-a-Phofung municipality were selected. In addition, Phuthaditjhaba MainPlace (a former homeland) was analysed to gain a full understanding of the behaviour of the AZTool at lower geographical levels in a rural setting. For Gauteng, City of Tshwane Metropolitan (which is both a district and a metropolitan municipality) was analysed, and Pretoria MainPlace was selected from this district or metro to explore the potential challenges that might occur in urbanised settings at lower geographical levels or regions.

AZTool software

The AZTool software version 1.0.3 (available from www.geodata.soton.ac.uk[10]) was used for the creation of optimal output areas in our study. As mentioned earlier, this software was derived from the AZP developed by Openshaw[11]. The AZP-based algorithms normally take input building blocks and iteratively aggregate them into larger output areas from an initial random aggregation, by checking the effect of swapping individual building blocks between output areas based on criteria set by the user. Such criteria include mean population target, minimum population threshold, homogeneity and compactness of the shape. The user may also set various options for how the AZTool will operate; for example, how many iterations and swaps the AZTool should run, whether donuts are allowed or not (that is, one output area surrounding another), setting minimum boundary lengths, and allowing output areas to be wholly contained within higher geographical levels or regions.[20]

Data preparation

Our study employed EAs from 2001 census estimates[29] as building blocks to create new census output areas in South Africa. We were not able to access data at household and EA levels from Stats SA. The ESRI ArcGIS 10.2 software was used to prepare data for analysis by the AZTool and to display output results. The variables employed were total population, dwelling type and geotype, and higher geographical levels or regions. The geotype variable was the geography type of the EA, which was divided into the following categories: Geotype1=Formal Urban, Geotype2=Informal Urban, Geotype3=Informal Rural (tribal areas), and Geotype4=Formal Rural (farms). The AZTool expects the intra-area correlation variables to be provided as counts. Therefore, the geotype variable – which contained the above four categories – was expanded into four attributes, one for each category (each with a count of 0 or 1).

The AZTImporter, which is part of the AZTool software download, was used to convert the building block shapefile (geospatial vector data format) to polygon attribute table (.pat) and arc attribute table (.aat) files. These files have the format required by the AZTool software.

Zone design criteria

The criteria or rules for the AZTool runs were set in the .xml parameter file. This file specifies the location of .aat and .pat files as well as defining the parameters, rules, constraints, criteria and column position of variables in the .pat file to be used in the AZTool run. The following criteria were considered for developing optimised output areas:

- minimum threshold population size: 500 (minimum used by Stats SA[4])
- homogeneity: IAC measure of dwelling type and geotype variables
- shape compactness: perimeter squared per area (P2A)
- mean target population: 1000

The minimum threshold population size is a hard constraint, as are the higher geographical regions. The others are soft constraints, which are traded off in the objective criteria as in previous studies – which also indicated that it is not possible to satisfy all four criteria[20,25,30]. The weights for population target, homogeneity (IAC score) and shape compactness were left at the default weight of 100%, indicating that all were weighted equally. The same design criteria were applied to all geographical levels in both rural and urban settings.

Confidentiality limit

The population variable from the 2001 census was used for respecting the confidentiality limit, with a minimum population of 500 set for output areas.[4] Generally, statistical spatial data analysis requires the aggregation of basic spatial areas into larger areas to preserve confidentiality, to minimise population differences and to reduce inaccuracies in the data.[18] Therefore, the population target mean was also set to 1000 in this study in order to minimise population differences.

Degree of homogeneity

To measure the degree of homogeneity within the created output areas, IAC was employed. The IAC is a direct measure of within-area homogeneity, which is the correlation for a given variable between different people living in the same areal unit.[7,31-33] Higher values indicate a higher degree of homogeneity within areas, coupled with a higher degree of heterogeneity between areas.[7,33,34] The homogeneity variables that were selected from the 2001 census data included dwelling type and geotype. The dwelling type or housing type is a variable commonly used as a proxy for the social built environment homogeneity measure, as it has been identified as one of the variables that tends to possess a high degree of homogeneity.[7,20] It was therefore applied in our study. The EA geographical type (geotype) was also used as one of the rules to create the SAL that was used to disseminate 2001 census data in South Africa.[4]

Shape compactness

Shape compactness, adapted from Cockings and Martin[25] and Haynes et al.[27], was used in an effort to produce more compact (circular rather than linear) output areas. The overall perimeter squared per area (P2A) was used as a measure of shape compactness. Lower P2A mean values indicated that output shapes were more compact, whereas higher P2A mean values indicated that output areas were less compact.

Results

This section presents results for both rural and urban settings at all geographical levels. Figure 1a shows the boundaries of original EAs of Phuthaditjhaba MainPlace, and indicates an EA that is widely spread on the northern part of the study area. This is typical especially for rural areas in South Africa. In most cases, EAs that are large in size in the rural areas are sparsely populated. Figure 1b shows the newly created output areas for the same area of Phuthaditjhaba. Donut EAs or building blocks, which are areas that completely surround other areas (such as the ones on the north-eastern part of our study area), no longer show in the new output areas. These have been combined with other building blocks to form the largest output area in terms of size. However, the largest EA in terms of coverage or area does not possess the greatest total population. This indicates that some building blocks which formed the new output area were not as populated as some of their counterparts in the same northern part, or in the southern part of the study area.

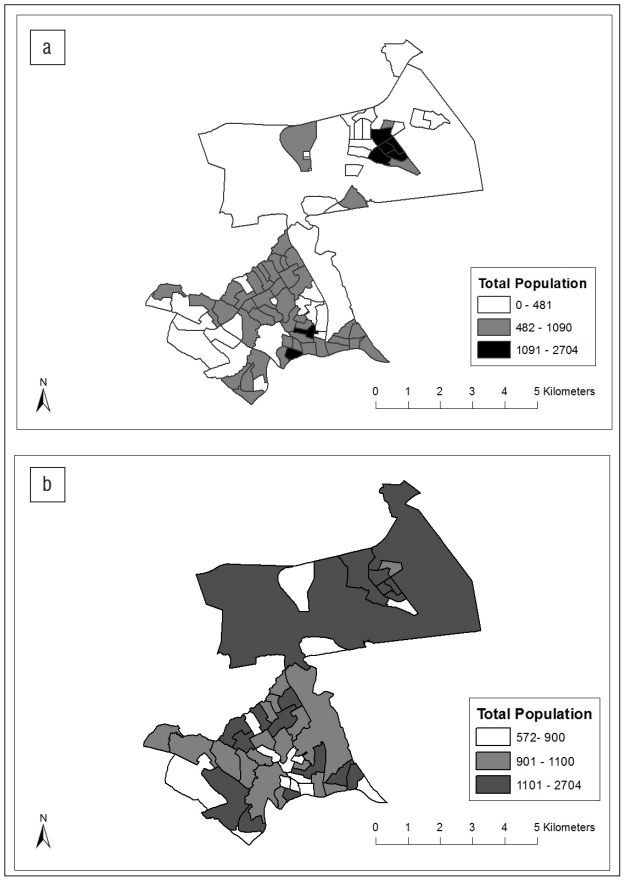


Figure 1: Phuthaditjhaba MainPlace: (a) original building block EAs, (b) newly created output areas.

Table 2 highlights the statistical characteristics of the original EAs and the newly created output areas for the rural areas in all four geographical regions. It is important to note that the original EAs were slightly more homogeneous and compact than the newly created output areas, at all geographical levels. The confidentiality threshold of 500 was not breached at any of the four geographical levels (MainPlace, municipality, district and provincial). The results show a steady increase in IAC from the lower geographical level (0.22) to the higher geographical level (0.59), meaning that the degree of homogeneity within–area increased as the geographical level increased. The mean population sizes were also close to the targeted mean, with reasonable standard deviations, but the MainPlace level had a higher mean value and standard deviation compared with the municipal level. The mean shape tended to increase from lower to higher geographical levels, indicating that the output areas at higher geographical levels were much less compact in shape than those of lower geographical levels (which had lower means and lower standard deviations).

For urban areas, a similar trend was noticed. The EAs were slightly more homogeneous and compact than the newly created output areas at all geographical levels. Table 3 shows that the IAC increased dramatically from 0.09 at MainPlace level to 0.46 for the district or metro level. The provincial level experienced a slight decrease to 0.45. The mean population limit was adhered to at all geographical levels, as it was for the rural areas. For both rural and urban areas, IAC values at the lower geographical levels were smaller than IAC values at higher geographical levels.

Table 2: Statistical characteristics of enumeration areas and output areas at various geographical levels for rural settings

	Number of zones	Population				Shape		Homogeneity
		min	max	mean	s.d.	mean	s.d.	Intra-area correlation
Enumeration areas								
Phuthaditjhaba MainPlace	86	0	2704	621	451	25	9	0.25
Maluti-a-Phofung Municipality	747	0	2704	480	313	26	9	0.58
Thabo Mofutsanyane District	1412	0	6196	518	410	26	9	0.66
Free State Province	5182	0	9269	519	454	26	10	0.65
Output areas								
Phuthaditjhaba MainPlace	49	572	2704	1090	341	27	10	0.22
Maluti-a-Phofung Municipality	349	610	2704	1027	232	32	13	0.50
Thabo Mofutsanyane District	667	581	5292	1087	403	33	13	0.56
Free State Province	2440	547	9269	1101	489	31	12	0.59

Table 3: Statistical characteristics of enumeration areas and output areas at various geographical levels for urban settings

	Number of zones	Population				Shape		Homogeneity
		min	max	mean	s.d.	mean	s.d.	Intra-area correlation
Enumeration areas								
Pretoria MainPlace	865	0	4625	610	358	24	9	0.11
City of Tshwane District	2115	0	8802	726	538	24	9	0.50
Gauteng Province	13 200	0	9627	667	563	24	8	0.50
Output Areas								
Pretoria MainPlace	500	621	5026	1056	320	28	11	0.09
City of Tshwane District	1276	502	8802	1203	514	27	10	0.46
Gauteng Province	7253	501	9627	1214	520	27	9	0.45

This means that higher geographical levels produced more homogeneous output areas than did the lower levels. This finding might be due to the fact that at higher geographical levels there are many building blocks that output areas can be constructed from, whereas at lower geographical levels there are fewer building blocks and hence the AZTool has limited options with regard to improving the IAC or any other constraints. With regard to the compactness of the shape of the output areas, a contradiction to what happened in rural areas was noted. The lower geographical levels' output areas were less compact compared with those of higher geographical areas.

Table 4 shows the comparison of rural and urban provinces as well as their combined results. The mean population threshold was not breached for either province or when the two were combined. The urban province, Gauteng, seemed to be outperformed by the rural province with regard to the degree of homogeneity, but it outperformed the rural province with regard to compactness of the output shapes. Similar trends were noted at other geographical levels. The IAC for two provinces combined was higher than that of the two provinces run separately, whereas the shape of output areas for the combined provinces was more compact than that of Free State province. The higher degree of homogeneity (IAC of 0.62) for both provinces combined suggests that the selected variables can be used as good indicators of social homogeneity, to create homogeneous output areas across the entire country.

Optimal number of AZTool runs

Several AZTool runs were performed for Phuthaditjhaba and Pretoria MainPlaces to see if increasing the number of iteration runs would improve the results. For Phuthaditjhaba, there was only a slight improvement in the results when the runs were increased up to 1000. For Pretoria MainPlace, IAC values remained constant throughout the runs (10 to 1000) at 0.09, while the mean shape compactness declined only slightly from 28 to 27 after 500 runs. It is worth noting that the higher number of runs came at a price of increased processing time. Therefore, if no tangible improvement with regard to output areas is achieved with a higher number of runs, it may be wise to retain a low number of runs; hence, 10 runs were kept in this case. On average, it took approximately three to four hours of processing time when Gauteng and Free State provinces were combined, with 10 runs. Free State province on its own, and lower geographical regions in both rural and urban areas, took a much shorter time to complete. We anticipate that if the creation of census output areas using the AZTool program is considered for the entire country, processing might take between 10 and 18 hours. But with an increased number of runs, such as 1000, processing would take even longer.

Table 4: Statistical characteristics of Free State and Gauteng Province output areas and the two provinces combined

Region	Output areas	Population				Shape		Homogeneity
		min	max	mean	s.d.	mean	s.d.	Intra-area correlation
Gauteng	7253	501	9627	1214	520	27	9	0.45
Free State	2440	547	9269	1101	489	31	12	0.59
All combined	9773	502	9627	1176	515	28	10	0.62

Table 5: Statistical outputs of merged districts against Free State Province (single run)

Region	Output areas	Population				Shape		Homogeneity
		min	max	mean	s.d.	mean	s.d.	Intra-area correlation
Lejweleputswa	558	541	9269	1143	580	30	12	0.40
Motheo	669	517	6252	1093	425	31	12	0.60
Northern Free State	409	573	7116	1116	551	30	11	0.44
Thabo Mofutsanyane	667	581	5292	1087	403	33	13	0.56
Xhariep	123	578	5183	1105	525	33	11	0.34
Merged districts	**2426**	**517**	**9269**	**1108**	**487**	**N/A**	**N/A**	**0.47**
Free State	**2440**	**547**	**9269**	**1101**	**489**	**31**	**12**	**0.59**

Respecting higher geographical levels or regions

In an effort to make sure that output areas were nested within higher geographical regions such as MainPlace, municipality and district, the 'Respect Regions' rule was set to 'True' in the .xml parameter file for the AZTool program. The AZTool could not successfully produce any solutions when any of the higher geographical regions were respected. To overcome this, higher geographical regions were analysed separately and merged at the end to produce an overall output for Free State province (Table 5). The results show that an average IAC score (0.46) for the five districts was below the exact IAC score (0.59) for Free State. The importance of census output areas nesting within higher geographical levels is to enable exact statistics to be compiled for geographical areas used for applications such as elections or public resource allocation. However, these higher geographical levels change regularly as the population grows, which makes it difficult to keep census output areas nested within them. Hence, some countries – including Australia, England and Wales – have removed the requirement for census output areas to be nested within certain higher geographical levels.[34]

Boundary length constraint

Boundary length is the length of the perimeter of boundaries that is shared between adjacent building blocks. When minimum boundary length was set to 5%, the shape of the output areas for Pretoria MainPlace improved slightly compared with the shape of the original EA building block. At the same time, the population mean increased away from the population target mean. The IAC measure did change compared to when the minimum boundary length was ignored. Care should be taken when using this option, as many building blocks might become isolated due to boundary length restrictions. This was the case for Phuthaditjhaba MainPlace.

Donut constraint

Overall, donut areas were allowed in all analyses. Figure 2a shows some donut output areas (shaded on the map) in the western part of Pretoria MainPlace. Further experiments were undertaken with donuts not allowed in the final output areas for MainPlace, municipality and district levels for the two provinces. When comparing 'donuts allowed' against

Figure 2: Newly created output areas for Pretoria MainPlace: (a) with donuts allowed, (b) with donuts not allowed.

'donuts not allowed', the results showed that not allowing donuts had little or no effect on the shape compactness of the output or on the IAC score's degree of homogeneity. For example, the western part of Pretoria MainPlace no longer contained donuts, as illustrated in Figure 2b. In general the inclusion of a donut constraint made no real difference in this experiment. The donut criterion might be of importance in a broader application for avoiding disjointed census output areas, especially when output areas are created for mapping and analysis.

Discussion

This paper illustrates the potential of automated zone design techniques, and possible challenges that might occur when applying such techniques in the creation of optimised output areas in South Africa. Generally, we noted that the original building blocks were slightly more homogeneous and compact than the newly created output areas, at all geographical levels and for both rural and urban settings. The IAC values were smaller at lower geographical levels than those of any higher geographical levels, for both rural and urban areas. This indicates that higher geographical levels produced more homogeneous output areas than did lower geographical levels. One reason might be that at a higher geographical level there are many building blocks from which output areas can be constructed, whereas at lower geographical levels there are fewer building blocks. Hence, the AZTool has limited options with regard to improving IAC and other constraints. Similarly, in New Zealand, Ralphs and Ang[20] found that larger areas seemed to be more homogenous with each other compared with smaller areas (i.e. evidence of a scale effect of MAUP). Lower IAC scores for lower geographical levels – MainPlace levels – were also noted in previous studies at detailed levels.[9,10,33]

When comparing the performance of the AZTool at the two spatial settings, urban and rural areas, the newly created output areas from rural areas had higher degrees of homogeneity than their urban counterparts. However, the urban areas were more compact than the rural areas. Overall, the relatively high degree of homogeneity for all provinces combined (urban and rural provinces), with an IAC of 0.62, suggests that the selected variables can be used as good indicators of social homogeneity in creating homogeneous output areas across South Africa. Generally, an IAC of 0.5 is regarded as a very reasonable degree of homogeneity.[22] It is also important to mention that in all our experiments that were performed for urban and rural areas at all geographical levels, the confidentiality limit was adhered to.

Ideally, increasing the number of runs should improve the AZTool's solution, as it enables finding better optimal output areas. On the contrary, results from this study did not show reasonable improvement of optimal output areas when different numbers of runs were explored. Our results concur with previous studies that were conducted with the AZTool, such as those of Ralphs and Ang[20] and Sabel et al.[22] When increasing the number of runs, they found that increasing numbers of iteration runs achieved little real improvement in the quality of outputs, while significantly increasing computing time. Therefore, we are confident that setting the number of runs to 10 can still produce quality output areas even when expanding the analysis to the entire country.

The donut constraint had minimal effect on the quality of output areas with regard to shape and degree of homogeneity. Therefore, we did not restrict output areas to exclude donuts in the final output areas. That means donuts were allowed as in the initial run. To make sure that output areas were nested within a higher geographical level or region, the AZTool was set to respect higher geographical regions. Unfortunately the program did not produce any solutions when any higher geographical levels were respected. Cockings et al.[10] argued that having to respect a higher geographical region constraint is particularly restrictive and often prevents solutions being found at all. Further investigations need to be performed to examine the cause of this in the context of South African geographical areas. To overcome this challenge, higher geographical regions can be analysed separately and merged at the end to produce an overall output, even though this might be time-consuming for larger samples.

The uniqueness of our approach in this study was that we considered the performance of the AZTool program for both urban and rural areas at different geographical levels or regions. This provided a clear indication as to how the program is likely to perform when the whole of South Africa is analysed. In addition, the current version of the AZTool has promising potential for application in developing countries, including South Africa, as it does not require an ArcInfo licence for preparing the contiguity files (.aat and .pat files). However, further consultations with relevant stakeholders should be undertaken before output areas from this study can be considered for possible use for any census data dissemination, as each set of output areas is the product of a set of criteria determined by the authors.

From a policy and practice point of view, it is important to note that this research was a stand-alone project with the aim of influencing policies and practice of government stakeholders, such as Stats SA. We believe that the positive findings from these initial experiments regarding the AZTool applications in the creation of census output areas in South Africa will encourage future possible collaboration between researchers and government stakeholders (such as Stats SA) as well as other South African census data users.

Regarding the limitations of our study, the poor accessibility of data at lower geographical levels (such as household and EA levels) and from the most recent 2011 census posed a problem. Hence, only the 2001 census EA estimates data were used as building blocks. Based on literature, there seems to be a challenge with regard to accessing census data at lower geographical levels for research purposes or for other purposes such as business and marketing, due to confidentiality.[20,25,34] Alternatively, dwelling-frame data could have been used, but this was challenging because the data had a lot of missing information or dwellings that were not captured in some areas across the country. The use of household-level data would have minimised the flaws carried by administrative data (EAs), which were created for a different purpose, when using EAs as building blocks for the created output areas. Therefore, caution should be taken when using pre-existing input areas to aggregate them into larger areas, as the flaws that are inherent in the building blocks will be carried over into the output areas – as well as possible bias and potential errors associated with the MAUP[20,30,34].

Conclusion

The success of this study is evident in the fact that the primary criterion of minimum population threshold of 500 people was kept and not breached throughout all newly created output areas, at different geographical levels, and for both rural and urban areas. In addition, the second most prioritised criterion, homogeneity of output areas, showed IACs of 0.45 for Gauteng, 0.52 for Free State, and 0.62 for both provinces combined. These IAC values are encouraging, as international studies show that an IAC of 0.5 indicates a highly acceptable degree of homogeneity within output areas. Based on these findings from different spatial settings and different geographical levels, we conclude that the AZTool software could be used to effectively and objectively create optimised output areas in South Africa. Further research should explore comparisons of the newly created output areas with existing output areas in South Africa. The availability or accessibility of data at lower geographical levels, such as the household level (or updated dwelling-frame data in South Africa), is highly recommended. The use of such data would improve the development of robust and optimised output areas using automated zone design techniques.

Acknowledgements

We sincerely thank Prof. David Martin and Dr Samantha Cockings for their permission to use the AZTool software (which is copyright to David Martin, Samantha Cockings and the University of Southampton). We also thank Ms Leana Meiring and Prof. Zitha Mokomane for editing and proofreading this paper.

Authors' contributions

T.M. prepared and analysed the data. F.A. and M.O. supervised the work and guided the main author. T.M. wrote the manuscript, and F.A. and M.O. provided valuable contributions to the final manuscript.

References

1. Martin D. 2001 Census output areas: From concept to prototype. Population Trends. 1998;94:19–24.
2. Hofstee P, Islam M. Disaggregation of census districts: Better population information for urban risk management. 25th Asian Conference on Remote Sensing, 22–26 November 2004, Thailand.
3. Martin D. Optimizing census geography: The separation of collection and output geographies. Int J Geogr Inf Sci. 1998;12(7):673–685. http://dx.doi.org/10.1080/136588198241590
4. Verhoef H, Grobbelaar N. The development of a small area layer for South Africa for census data dissemination. [Document on the Internet]. Pretoria: Statistics South Africa; 2005 [cited 2014 May 31]. Available from: http://www.cartesia.org/geodoc/icc2005/pdf/poster/TEMA26/HELENE%20VERHOEF.pdf
5. Martin D. Neighbourhoods and area statistics in the post 2001 census era. Area. 2004;36(2):136–145. http://dx.doi.org/10.1111/j.0004-0894.2004.00210.x
6. Duke-Williams O, Rees P. Can offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure. Int J Geogr Inf Sci. 1998;12(6):579–605. http://dx.doi.org/10.1080/136588198241680
7. Martin D, Nolan A, Tranmer M. The application of zone-design methodology in the 2001 UK Census. Environ Plann A. 2001;33(11):1949–1962. http://dx.doi.org/10.1068/a3497
8. Martin D. Geography for the 2001 census in England and Wales. Population Trends. 2002;108:7–15.
9. Martin D, Cockings S, Harfoot A. Development of a geographic framework for census workplace data. J Roy Stat Soc. 2013;176(2):585–602. http://dx.doi.org/10.1111/j.1467-985X.2012.01054.x
10. Cockings S, Harfoot A, Martin D, Hornby D. Maintaining existing zoning systems using automated zone-design techniques: Methods for creating the 2011 Census output geographies for England and Wales. Environ Plann A. 2011;43(10):2399–2418. http://dx.doi.org/10.1068/a43601
11. Openshaw S. A geographical solution to scale and aggregation problems in region building, partitioning and spatial modelling. T I Brit Geogr. 1977;2(4):459–472. http://dx.doi.org/10.2307/622300
12. Gehlke CE, Biehl K. Certain effects of grouping upon the size of the correlation coefficient in census tract material. J Am Stat Assoc. 1934;29:169–170.
13. Openshaw S. The modifiable areal unit problem. Norwich, UK: Geo Books; 1984. http://dx.doi.org/10.1068/a160017
14. Reynolds HD. The modifiable unit problem: Empirical analysis by statistical simulation [PhD Thesis]. Toronto, Canada: University of Toronto; 1998.
15. Ratcliffe JH, McCullagh MJ. Hotbeds of crime and the search for spatial accuracy. J Geograph Syst. 1999;1(4):385–398. http://dx.doi.org/10.1007/s101090050020
16. Kitchin R, Tate NJ. Conducting research into human geography, theory, methodology & practice. London: Prentice Hall, Essex; 2000.
17. Heywood I, Cornelius S, Carver S. An introduction to geographical information systems. 2nd ed. London: Pearson Prentice Hall; 2002.
18. Duque JC, Ramos R, Surinach J. Supervised regionalization methods: A survey. Int Regional Sci Rev. 2007;30(3):195–220. http://dx.doi.org/10.1177/0160017607301605
19. Dumedah G, Schuurman N, Yang W. Minimizing effects of scale distortion for spatially grouped census data using rough sets. J Geograph Syst. 2008;10(1):47–69. http://dx.doi.org/10.1007/s10109-007-0056-y
20. Ralphs M, Ang L. Optimized geographies for data reporting: Zone design tools for census output geographies. Statistics New Zealand Working Paper No 09–01. Wellington: Statistics New Zealand; 2009.
21. Martin D. Extending the automated zoning procedure to reconcile incompatible zoning systems. Int J Geogr Inf Sci. 2003;17(2):181–196. http://dx.doi.org/10.1080/713811750
22. Sabel CE, Kihal W, Bard D, Weber C. Creation of synthetic homogeneous neighbourhoods using zone design algorithms to explore relationships between asthma and deprivation in Strasbourg, France. Soc Sci Med. 2013;91:110–121. http://dx.doi.org/10.1016/j.socscimed.2012.11.018
23. Openshaw S, Rao L. Algorithms for reengineering 1991 Census geography. Environ Plann A. 1995;27(3):425–446. http://dx.doi.org/10.1068/a270425
24. Martin D, Office for National Statistics. Output area production: Overview of the system used to create the Output Areas. 2002 [cited 2016 February 1]. Available from: https://www.ons.gov.uk/census/2001censusandearlier/dataandproducts/outputgeography/outputareaproduction
25. Cockings S, Martin D. Zone design for environment and health studies using pre-aggregated data. Soc Sci Med. 2005;60(12):2729–2742. http://dx.doi.org/10.1016/j.socscimed.2004.11.005
26. Flowerdew R, Manley DJ, Sabel CE. Neighbourhood effects on health: Does it matter where you draw the boundaries? Soc Sci Med. 2008;66(6):1241–1255. http://dx.doi.org/10.1016/j.socscimed.2007.11.042
27. Haynes R, Daras K, Reading R, Jones A. Modifiable neighbourhood units, zone design and residents' perceptions. Health Place. 2007;13(4):812–825. http://dx.doi.org/10.1016/j.healthplace.2007.01.002
28. Daras K. An information statistics approach to zone design in the geography of health outcomes and provision [PhD Thesis]. Newcastle, England: University of Newcastle; 2006.
29. HSRC. 2001 census EA estimates. Human Sciences Research Council in collaboration with Prof DJ Stoker. Pretoria, South Africa; 2005.
30. Drackley A, Newbold KB, Taylor C. Defining socially-based spatial boundaries in the Region of Peel, Ontario, Canada. Int J Health Geogr. 2011;10(38):[12], 12 pages. http://dx.doi.org/10.1186/1476-072X-10-38
31. Flowerdew R. How serious is the modifiable areal unit problem for analysis of English census data? Population Trends. 2011;145(1):106–118. http://dx.doi.org/10.1057/pt.2011.20
32. Tranmer M, Steel D. Using local census data to investigate scale effects. In Tate NJ, Atkinson PM, editors. Modelling scale in geographical information science. Chichester: John Wiley & Sons; 2001. p. 105–122.
33. Tranmer M, Steel D. Using census data to investigate the causes of the ecological fallacy. Environ Plann A. 1998;30(5):817–831. http://dx.doi.org/10.1068/a300817
34. Cockings S, Harfoot A, Martin D, Hornby D. Getting the foundations right: Spatial building blocks for official population statistics. Environ Plann A. 2013;45(6):1403–1420. http://dx.doi.org/10.1068/a45276

Estimated abundance and diversity of heterotrophic protists in South African biocrusts

AUTHORS:
Kenneth Dumack[1]
Robert Koller[1,2]
Bettina Weber[3]
Michael Bonkowski[1]

AFFILIATIONS:
[1]Biocenter Cologne, Faculty of Zoology, Department of Terrestrial Ecology, University of Cologne, Cologne, Germany
[2]Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich GmbH, Jülich, 52425, Germany
[3]Multiphase Chemistry Department, Max Planck Institute for Chemistry, Max-Planck-Gesellschaft, Mainz, Germany

CORRESPONDENCE TO:
Kenneth Dumack

EMAIL:
kenneth.dumack@uni-koeln.de

POSTAL ADDRESS:
University of Cologne, Zülpicher Strasse 47B, Cologne NRW 50674, Germany

DATES:
Received: 13 Aug. 2015
Revised: 01 Feb. 2016
Accepted: 23 Feb. 2016

KEYWORDS:
biological soil crusts; biodiversity; protozoa; cryptogams; cyanobacteria

HOW TO CITE:
Dumack K, Koller R, Weber B, Bonkowski M. Estimated abundance and diversity of heterotrophic protists in South African biocrusts. S Afr J Sci. 2016;112(7/8), Art. #2015-0302, 5 pages. http://dx.doi.org/10.17159/sajs.2016/20150302

Biological soil crusts (biocrusts) occur widely in the uppermost millimeters of the soil in arid and semi-arid systems. Worldwide they cover large terrestrial areas and play a major role in the global terrestrial carbon and nitrogen cycles. However, knowledge of the microbial decomposer foodwebs within biocrusts is particularly scarce. Heterotrophic protists in soil are predominantly bacterivores, and because of their high biomass compared with other soil fauna and fast turnover rates, protists are considered an important factor for soil nutrient cycling and energy fluxes. Thus, knowledge of their biodiversity, abundance and functional roles is important to understand soil ecosystem functions. We investigated the diversity and abundance of heterotrophic soil protists in different types of biocrusts from the Succulent Karoo, South Africa. With an overall diversity of 23 distinct morphotypes, soil protist biodiversity was shown to be high. The most abundant groups were *Spumella*-like chrysomonads, gliding bodonids, glissomonads and heteroloboseans. Protist abundance was highly variable among samples. The abundance and diversity did not differ significantly among different types of biocrusts, indicating that microscale differences, but not macroscopic soil crust builders (e.g. cyanobacteria, lichens and bryophytes), have a major impact on the protist community.

Introduction

Biological soil crusts (biocrusts) are microscopic ecosystems. They comprise primary producers such as cyanobacteria, algae, lichens, and bryophytes together with decomposers such as fungi, bacteria and archaea. The microscopic consumers in biocrust foodwebs are adapted to arid soil conditions and taxonomically diverse, including protists, nematodes, rotifers, tardigrades, and microarthropods in variable proportions. Biocrusts grow within the uppermost millimeters of the soil in arid and semiarid regions throughout the world,[1] where they fulfil several highly relevant ecosystem services. They limit soil erosion by both wind and water[2-4]; influence water runoff, infiltration and retention within the uppermost soil layer[5-7]; and fertilise impoverished desert soils[8]. On a global scale, biocrusts cover over 10% of the terrestrial surface area, also influencing global nutrient cycling and climate processes.[9,10] Biocrusts form one subgroup within cryptogamic covers, with the latter also comprising cryptogamic communities on rock and epiphytically on plants. These cryptogamic covers have been estimated to account for 7% of the net primary production fulfilled by plants, and fix about 49 Tg N per year, corresponding to about half the maximum value estimated for the total terrestrial biological nitrogen fixation.[11]

Detailed information is available on the diversity and species composition of the primary producers in biocrusts. According to the dominant photoautotrophic organism group (i.e. the dominant soil crust builder), they have been coarsely defined as cyanobacteria-dominated, lichen-dominated and bryophyte-dominated soil crusts.[12] However, data on the microbial decomposer foodwebs, especially heterotrophic protists, are scarce. Nematodes, tardigrades, rotifers, mites, collembolans, heterotrophic protists, and even larger arthropods and molluscs have been observed to utilise biocrusts as a habitat.[13-16] However, their diversity, frequency, geographical distribution and feeding behaviour have been investigated in only a few local studies.[13-16]

Protists – with their high abundance, turnover and diversity – have been increasingly studied by soil ecologists in recent years.[17,18] Based on their general morphology and means of locomotion, soil-inhabiting protists comprise amoebae, flagellates and ciliates. Most heterotrophic protists in soil are bacterivores. As a result of their high numbers and turnover rates, protists are considered to play a major ecological role in soil foodwebs by the release of nutrients from consumed microbial biomass.[13,18]

Heterotrophic protists have been determined in arid soils in a number of habitats among different continents, namely southwestern USA, the Negev Desert of Israel, arid Australia, China and Antarctica.[13,19-23] However, data on biocrusts from Africa are lacking. Based on 73 soil samples from the Etosha region and the Namib Desert of Namibia, Foissner et al.[24] identified 365 ciliate species, of which 35% had been undescribed, including a new order and suborder, three new families, and 34 new genera and subgenera of soil ciliates. These findings suggest that biocrusts in South Africa can harbour a substantial undiscovered diversity of protists.

We investigated the heterotrophic protist community of biocrusts in the Succulent Karoo, South Africa. Besides determination of the diversity and abundance of amoeboid and flagellated protists, we evaluated whether they are affected by the identity of the major primary producers in primary producers in biocrusts that are dominated by cyanobacteria, chlorolichens and bryophytes.

Material and methods

The study site was situated in the Succulent Karoo biome, South Africa, in the vicinity of the village Soebatsfontein, about 60 km south of Springbok. The Succulent Karoo biome is a unique dryland system hosting a biodiversity hotspot with an extraordinarily high plant diversity and a unique flora of succulent plants.[25] Samples were collected next to the BIOTA observatory in Soebatsfontein (observatory number S22 at 30.19° S, 17.54° E, altitude 392 m). The hilly region comprises soils of sandy texture, some granite inselbergs and a dense pattern of fossil termite mounds.[26] The semi-arid climate of the region is characterised by mild winter and hot summer conditions, with

air temperatures below 2 °C in July and sometimes above 44 °C in February.[27] The study site is located in the winter rainfall area, with a mean of 129 mm rainfall which falls mainly during the cool winter months (July to August) and a second smaller peak in autumn (April to May). The study area is densely covered by diverse communities of biocrusts, reaching an overall surface coverage above 25% in regions without inselbergs and roads.[28]

Biocrust communities and sampling

Biocrust communities within southern Africa have been divided into seven main biocrust types,[12] of which we investigated three. First we analysed well-established cyanobacteria-dominated biocrusts, which are characterised by a more-or-less uniform dark surface coloration caused by cyanobacteria growing close to the surface. When being removed, the biocrust forms relatively large flakes of up to 3.9 mm thick. Dominating cyanobacterial genera are *Nostoc*, *Phormidium*, *Scytonema*, *Microcoleus* and *Leptolyngbya*, with the latter two already occurring in early, initially formed biocrusts. As a second type we investigated chlorolichen-dominated biocrusts, which can only form on stabilised surfaces, normally developing from well-established cyanobacteria-dominated biocrusts[12]. The dominating chlorolichen species was *Psora decipiens*, which has a particularly wide geographical distribution and occurs frequently within the study area[11]. Bryophyte-dominated biocrusts were the third type we analysed. This biocrust type is a late-successional stage that develops from previously well-established biocrusts, and because bryophytes require somewhat more water than the other biocrust components, they frequently occur in the shade or vicinity of small shrubs. The dominating moss species in these samples was *Ceratodon purpureus*, a small species with an almost global distribution[29].

The three different types of biocrusts were collected in 10-cm Petri dishes, with five replicates each. For each sample, the lower lid of the Petri dish was pressed approximately 1 cm deep into the soil, then a trowel was pushed below the lid and together with the sample was lifted from the surrounding soil. The Petri dish was carefully turned around, surplus soil was removed, and the dish was covered with the upper lid, which was subsequently sealed with parafilm and taping band. As the biocrusts were completely dry during sampling, no additional drying of the samples was necessary. The samples were transported to Germany and stored at 4 °C in the dark for 6 weeks.

Microbial determination of taxonomic units

Protist abundance and community composition were assessed by a liquid aliquot method according to Butler and Rogerson[30]. Briefly, 1 g of a homogenised surface soil (uppermost 2 mm) was suspended in 350 ml of sterile distilled water and shaken for 20 min. For incubation, the suspension was diluted by a factor of 4, and 20 μL of the suspension was added to 180 μL of wheat grass medium (WG). The WG was made by adding 0.15% dried wheat grass powder (Weizengras, Sanatur GmbH, D-78224 Singen) to PJ medium[31]. In total 144 wells per sample were stored at 15 °C in the dark. The plates were inspected for protists after 7 and 21 days using an inverted microscope (Nikon Eclipse TS100) at 100x and 400x magnification. Protist morphotypes were determined according to Jeuck and Arndt[32], Bass et al.[33], Smirnov[34] and Smirnov and Brown[35].

Protist diversity was determined by the Shannon-Weaver-Index[36]:

$$H = -\sum^{i} (pi * \ln pi) \qquad \text{Equation 1}$$

where *pi* is the proportion of the morphotype.

Furthermore, the evenness was calculated by the following formula:

$$J = H/\ln S \qquad \text{Equation 2}$$

where S is the total number of morphotypes.

Change in protist abundance between treatments was analysed by analysis of variance (ANOVA) followed by Tukey tests, using R software (Version 3.1.0; package: agricolae).

Results

The investigation of biocrust samples revealed abundances between 21×10^3 and 27×10^3 individuals per gram dry weight for the cultivable heterotrophic protist communities. However, numbers varied widely among the samples, from 7.0×10^3 to 56×10^3 individuals per gram dry weight. We observed similar abundances for flagellates and amoebae, followed by ciliates (Table 1).

Table 1: Mean (± *s.d.*) heterotrophic protist abundance per gram dry weight of substrate in the three ($n=5$) tested types of biocrusts.

Type of biocrust	Amoebae	Flagellates	Ciliates
bryophyte-dominated	9498 ±4741	13284 ±10429	679 ±1068
chlorolichen-dominated	9562 ±7510	11464 ±10875	665 ±700
cyanobacteria-dominated	11442 ±10154	15044 ±15144	1157 ±1383

The investigation of the biocrusts revealed high protist diversity in soil crusts, with 23 different morphotypes from various morphologies such as testate amoebae, naked amoebae, flagellates, amoeboflagellates and ciliates (Figure 1).

Independent of sampling sites, *Spumella*-like chrysomonads, gliding bodonids (*Neobodo*- and *Parabodo*-like), glissomonads and heterolobosean were the most abundant groups of protists. Among the less abundant protists, we found high diversities enabling a high taxonomic resolution of amoeboid organisms (Figure 1). Diversity and evenness values were similar among all types of biocrusts, with an average Shannon-Weaver index of $H=2.1 \pm 0.1$ ($F=0.36$, $p=0.71$) and evenness of 0.83 ± 0.5 ($F=0.74$, $p=0.50$). There was no significant difference in morphotype composition of protists among different biocrust types (data not shown).

Discussion

Protist abundance

Protist abundance ranged between 7.0×10^3 and 56×10^3 individuals per gram dry weight, with amoebae and flagellates sharing similar abundances (12.5×10^3 and 16×10^3 individuals per gram dry weight, respectively). These values were more than 10 times higher than protist abundances reported from biocrusts of the Colorado Plateau and the Chihuahua Desert in the Western USA[13], and up to more than 3 to 30 times higher than Bamforth[37] reported (ciliates and flagellates) for arid desert soils and litters of northern Arizona. However, total protist abundances were low compared with those in studies of soils in temperate regions. For example, Domonell et al.[38] reported protist abundances in the range of 17×10^3 to 127×10^3 individuals per gram dry weight for grassland and forest soils in Germany, whereas Finlay et al.[39] reported soil heterotrophic (flagellated) protist abundances of up to 539×10^3 individuals per gram dry weight for grassland soils, and up to 417×10^3 individuals per gram dry weight in forest soils, respectively.

Although we homogenised our sampled soil to minimise the effect of microscale differences, variation in protist abundances (by a factor of 8) was nevertheless very high in biocrusts. This indicates large differences in the small-scale distribution patterns of soil protists in these mostly dry soil systems. It is known that protist abundance tends to vary with a number of environmental parameters such as local humidity, distance to vascular plants, disturbance and distribution of bacterial food.[17,18,40,41]

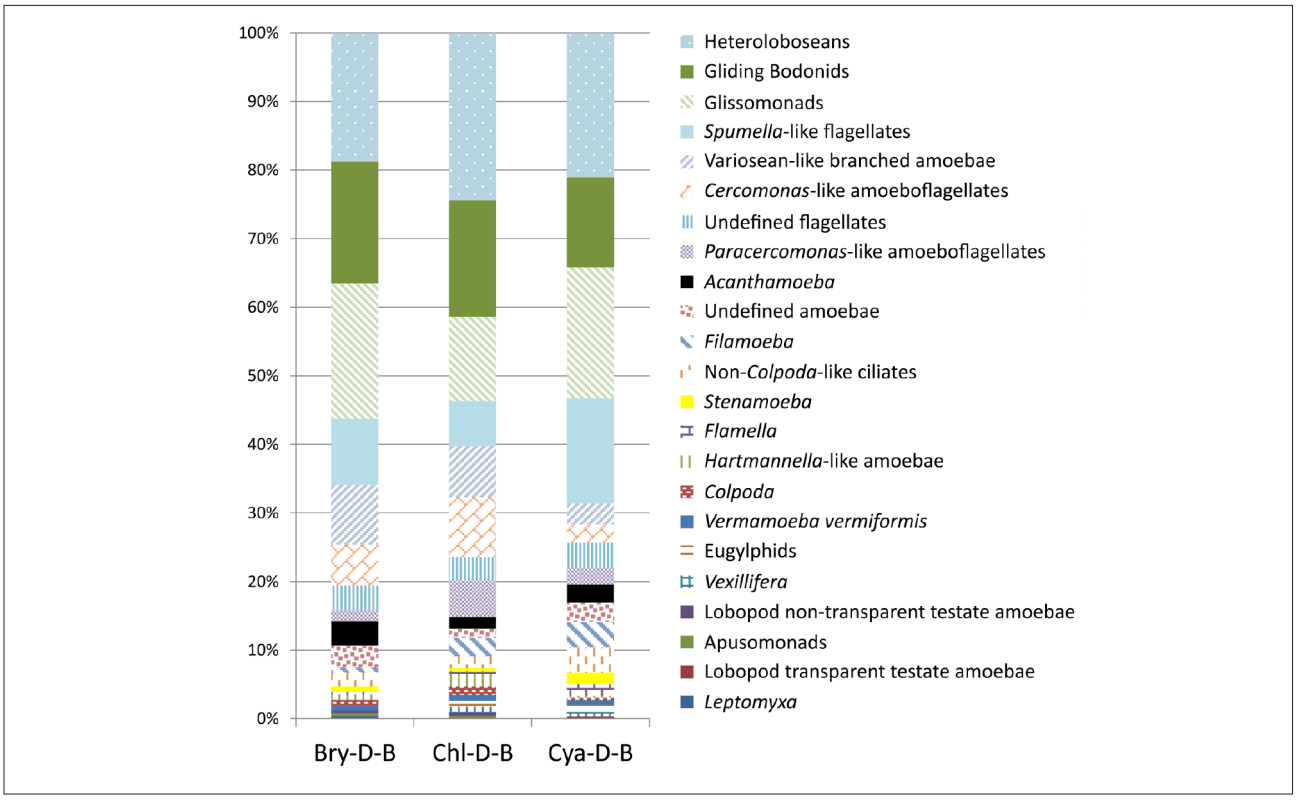


Figure 1: Heterotrophic protist community in bryophyte-(Bry-D-B), chlorolichen-(Chl-D-B) and cyanobacteria (Cya-D-B)-dominated biocrusts. Morphotypes are sorted from high to low mean abundances.

In the Succulent Karroo, short periods of moisture (soils are wet for roughly 45 days per year[42]) and large temperature extremes between day and night[27] mean that the protist community is likely to be influenced more by an accumulation of morning dew during the night than by periodic rainfall events[43]. Small protists with a rapid life cycle will have an adaptive advantage under these conditions.

All the protists we identified were cyst-forming bacterivores. Cysts enable protists to tolerate frequent wetting–drying cycles[17] and must be considered a major functional adaptation in desert protists[37,44]. Non-encysting protists that feed on cyanobacterial filaments identified in cyanobacterial crusts of Australia are likely an exception.[20] Cysts of some protist taxa are viable over decades[45,46] and will gradually accumulate under conditions where microbial production and environmental conditions are favourable. Cultivation-based methods therefore estimate the abundance of cysts together with active stages of protists, which can be regarded as an integrated measure of past microbial production in these soils.

Protist diversity

The diversity of heterotrophic protists in biocrusts comprised 'typical' soil protists such as acanthamoebae, vermamoebae and cercozoans (especially glissomonads and cercomonads), as well as less-frequently found and likely rarer taxa such as apusomonads.[39] A high abundance of glissomonads and heteroloboseans is commonly observed in soils, but the general lack of thaumatomonads and vannellids was quite surprising. It has recently been shown that protist diversity responds to changes in soil dryness, especially with regard to larger protists, which quickly disappear with decreasing soil moisture content.[47] Prolonged dryness therefore might have far-reaching negative effects on some protist groups, but more detailed studies are necessary to confirm this possibility.

Knowledge on the functional roles of protists is particularly scarce. Generally, chrysomonads, bicosoecids and some bodonids are considered as interception feeders in biofilms.[48,49] They create a water flow with their flagella to capture bacteria in water films.[48] Organisms that depend on a water current for feeding may be considered more strongly moisture-dependent than (for example) amoebae that graze within biofilms.[49] Therefore it was not surprising that bicosoecids and the mostly swimming or sessile bodonids, such as *Bodo saltans*, were missing from our samples. However, *Spumella*-like chrysomonads, which usually show high abundances in soil systems,[38,39] could be confirmed for soil crusts.

Most amoebae and amoeboflagellates (Cercomonads, i.e. *Cercomonas*-like and *Paracercomonas*-like amoeboflagellates) can attach to particles and feed on bacteria in biofilms and in tiny soil pores,[49] resulting in relatively high abundances in drier soil. In addition, Darby et al.[50] found amoebae (rather than nematodes and other protists such as ciliates and flagellates) to be highly tolerant to extreme environmental conditions, including increased temperatures and altered precipitation. In line with these results, our findings also showed that amoebae and amoeboflagellates were the most abundant functional groups in our study.

The comparison of protist diversity among the three different biocrust types revealed no statistically significant differences. This finding indicates that the different macroscopic soil crust builders might have no major effect on bacterivorous protist community composition. However, data on protists in soil are too scarce for us to propose a general conclusion. To our knowledge, ours is the first study on protist communities in biocrusts on the African continent, and is one of few studies that give detailed quantitative estimates on the cultivable amoeboid and flagellated protists.[13,38,39,51] Future studies need to include molecular approaches in order to estimate the non-cultivable taxa to the protist community.

Conclusion

This is the first detailed study on protist abundance and their morphotypes in biocrusts on the African continent. The diversity of protists was high, comprising solely cyst-forming protists. The abundance of heterotrophic

protists in biocrusts was found to be more than 10 times higher than in comparative studies on desert biocrusts in the south-western USA. Protist diversity and abundance varied substantially among samples, and there was no significant difference between types of biocrusts. This indicates that protist abundance and diversity are regulated at much smaller spatial scales (e.g. within single pores or on single leaves), and that the macroscopically visible biocrust components are likely of little relevance for protist occurrence. Further studies, especially on the combination of taxonomic and next-generation sequencing techniques, are necessary to characterise the protist communities of arid soil systems.

Acknowledgements

This study was supported by the German Research Foundation (project numbers WE 2393/2-1, WE 2393/2-2) and the Max Planck Society (Nobel Laureate Fellowship). Research in South Africa was conducted with South African research (No. 86/2013) and export permits (No. 87/2013).

Authors' contributions

K.D. performed the experimental analyses and wrote the article, R.K. and M.B. guided the experimental analyses and gave support in writing the article, and B.W. collected the samples and provided substantial assistance with writing the article. B.W. and M.B. contributed to an equal extent in expertise, time and resources for this article.

References

1. Belnap J, Büdel B, Lange OL. Biological soil crusts: Characteristics and distribution. In: Belnap J, Lange OL, editors. Ecological studies 150: Biological soil crusts: Structure, function, and management. Berlin, Heidelberg, New York: Springer; 2003. p. 3–30. http://dx.doi.org/10.1007/978-3-642-56475-8
2. Belnap J, Gillette DA. Disturbance of biological soil crusts: Impacts on potential wind erodibility of sandy desert soils in Southeastern Utah. Land Degrad Dev. 1997;8(4):355–362. http://dx.doi.org/10.1002/(SICI)1099-145X(199712)8:4<355::AID-LDR266>3.0.CO;2-H
3. Gaskin S, Gardner R. The role of cryptogams in runoff and erosion control on bariland in the Nepal Middle Hills of the Southern Himalaya. Earth Surf Proc Land. 2001;6(12):1303–1315. http://dx.doi.org/10.1002/esp.277
4. Zhao Y, Qin N, Weber B, Xu M. Response of biological soil crusts to raindrop erosivity and underlying influences in the hilly Loess Plateau region, China. Biodivers Conserv. 2014;23(7):1669–1686. http://dx.doi.org/10.1007/s10531-014-0680-z
5. Belnap J. The potential roles of biological soil crusts in dryland hydrologic cycles. Hydrol Process. 2006;20(15):3159–3178. http://dx.doi.org/10.1002/hyp.6325
6. Chamizo S, Cantón Y, Domingo F, Belnap J. Evaporative losses from soils covered by physical and different types of biological soil crusts. Hydrol Process. 2013;27(3):324–332. http://dx.doi.org/10.1002/hyp.8421
7. Eldridge DJ, Zaady E, Shachak M. Infiltration through three contrasting biological soil crusts in patterned landscapes in the Negev, Israel. Catena. 2000;40(3):323–336. http://dx.doi.org/10.1016/S0341-8162(00)00082-5
8. Evans RD, Lange OL. Biological soil crusts and ecosystem nitrogen and carbon dynamics. In: Belnap J, Lange OL editors. Ecological studies 150: Biological soil crusts: Structure, function, and management. Berlin, Heidelberg, New York: Springer; 2003. p. 263–279.
9. Porada P, Weber B, Elbert W, Pöschl U, Kleidon A. Estimating global carbon uptake by lichens and bryophytes with a process-based model. Biogeosc. 2013;10(11):6989–7033. http://dx.doi.org/10.5194/bg-10-6989-2013
10. Porada P, Weber B, Elbert W, Pöschl U, Kleidon A. Estimating impacts of lichens and bryophytes on global biogeochemical cycles. Global Biogeochem Cy. 2014;28(2):71–85. http://dx.doi.org/10.1002/2013GB004705
11. Elbert W, Weber B, Burrows S, Steinkamp J, Büdel B, Andreae MO, et al. Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. Nat Geosc. 2012;5:459–462. http://dx.doi.org/10.1038/ngeo1486
12. Büdel B, Darienko T, Deutschewitz K, Dojani S, Friedl T, Mohr K, et al. Southern African biological soil crusts are ubiquitous and highly diverse in drylands, being restricted by rainfall frequency. Microbial Ecol. 2009;57(2):229–247. http://dx.doi.org/10.1007/s00248-008-9449-9
13. Darby BJ, Housman DC, Zaki AM, Shamout Y, Adl,SM Belnap J, et al. Effects of altered temperature and precipitation on desert protozoa associated with biological soil crusts. J Eukaryot Microbiol. 2006;53(6):507–514. http://dx.doi.org/10.1111/j.1550-7408.2006.00134.x
14. Darby BJ, Neher DA. Stable isotope composition of microfauna supports the occurrence of biologically fixed nitrogen from cyanobacteria in desert soil food webs. J Arid Environ. 2012;85(1):76–78. http://dx.doi.org/10.1016/j.jaridenv.2012.06.006
15. Shachak M, Steinberger Y. An algae - desert snail food chain: Energy flow and soil turnover. Oecologia. 1980;46(3):402–411. http://dx.doi.org/10.1007/BF00346271
16. Shepherd UL, Brantley SL, Tarleton CA. Species richness and abundance patterns of microarthropods on cryptobiotic crusts in a piñon-juniper habitat: A call for greater knowledge. J Arid Environ. 2012;52(3):349–360. http://dx.doi.org/10.1006/jare.2002.1003
17. Adl MS, Gupta V. Protists in soil ecology and forest nutrient cycling. Can J Forest Res. 2016;36(7):1805–1817. http://dx.doi.org/10.1139/x06-056
18. Bonkowski M. Protozoa and plant growth: The microbial loop in soil revisited. New Phytol. 2014;162(3):617–631. http://dx.doi.org/10.1111/j.1469-8137.2004.01066.x
19. Rodriguez-Zaragoza S, Mayzlish E, Steinberger Y. Vertical distribution of the free-living amoeba population in soil under desert shrubs in the Negev Desert, Israel. Appl Environ Microb. 2005;71(4):2053–2060. http://dx.doi.org/10.1128/AEM.71.4.2053-2060.2005
20. Robinson BS, Bamforth SS, Dobson PJ. Density and diversity of protozoa in some arid australian soils. J Eukaryot Microbiol. 2002;49(6):449–453. http://dx.doi.org/10.1111/j.1550-7408.2002.tb00227.x
21. Colesie C, Green TGA, Turk, R, Hogg ID, Sancho LG, Budel B. Terrestrial biodiversity along the Ross Sea coastline, Antarctica: Lack of a latitudinal gradient and potential limits of bioclimatic modeling. Polar Biol. 2014;37(8):1197–1208. http://dx.doi.org/10.1007/s00300-014-1513-y
22. Bamforth SS, Wall DH, Virginia RA. Distribution and diversity of soil protozoa in the McMurdo Dry Valleys of Antarctica. Polar Biol. 2005;28(10):756–762. http://dx.doi.org/10.1007/s00300-005-0006-4
23. Liu Y, Li X, Jia R, Huang L, Zhou Y, Gao Y. Effects of biological soil crusts on soil nematode communities following dune stabilization in the Tengger Desert, Northern China. Appl Soil Eco. 2011;49(1):118–124. http://dx.doi.org/10.1016/j.apsoil.2011.06.007
24. Foissner W, Agatha S, Berger H. Soil ciliates (Protozoa, Ciliophora) from Namibia (Southwest Africa), with emphasis on two contrasting environments, the Etosha Region and the Namib Desert. Denisia. 2002;5:1–1459.
25. Brooks TM, Mittermeier RA, Mittermeier CG, Da Fonseca GAB, Rylands AB, Konstant WR, et al. Habitat loss and extinction in the hotspots of biodiversity. Conserv Biol. 2002;16(4):909–923. http://dx.doi.org/10.1046/j.1523-1739.2002.00530.x
26. Weber B, Graf T, Bass M. Ecophysiological analysis of moss-dominated biological soil crusts and their separate components from the Succulent Karoo, South Africa. Planta. 2012;236(1):129–139. http://dx.doi.org/10.1007/s00425-012-1595-0
27. BIOTA Southern Africa, weather data station: Soebatsfontein, as at 10 March 2011. Available at: http://www.biota-africa.org
28. Weber B, Olehowski C, Knerr T, Hill J, Deutschwitz K, Wessels DCJ, et al. A new approach for mapping of biological soil crusts in semidesert areas with hyperspectral imagery. Remote Sens Environ. 2008;112(5):2187–2201. http://dx.doi.org/10.1016/j.rse.2007.09.014
29. Ireland RR. Moss flora of the Maritime provinces. Publications in Botany, No. 13. National Museums of Canada. Ottawa 1982
30. Butler H, Rogerson A. Temporal and spatial abundance of naked amoebae (Gymnamoebae) in marine benthic sediments of the Clyde Sea Area, Scotland. J Eukaryot Microbiol. 1995;42(6):724–730. http://dx.doi.org/10.1111/j.1550-7408.1995.tb01624.x
31. Prescott DM, James TW. Culturing of *Amoeba proteus* on *Tetrahymena*. Exp Cell Res. 1955;8(1):256–258. http://dx.doi.org/10.1016/0014-4827(55)90067-7
32. Jeuck A, Arndt H. A short guide to common heterotrophic flagellates of freshwater habitats based on the morphology of living organisms. Protist. 2013;164(6):842–860. http://dx.doi.org/10.1016/j.protis.2013.08.003

33. Bass D, Howe AT, Mylnikov AP, Vickerman K, Chao EE, Smallbone JE, et al. Phylogeny and classification of Cercomonadida (Protozoa, Cercozoa): *Cercomonas, Eocercomonas, Paracercomonas*, and *Cavernomonas* gen. nov. Protist. 2009;160(4):483–521. http://dx.doi.org/10.1016/j.protis.2009.01.004

34. Smirnov A. A revised classification of naked lobose amoebae (Amoebozoa: Lobosa). Protist. 2011;162(4):545–570. http://dx.doi.org/10.1016/j.protis.2011.04.004

35. Smirnov A, Brown S. Guide to the methods of study and identification of soil gymnamoebae. Protistology. 2004;3(3):148–190.

36. Shannon CE, Weaver W. The mathematical theory of communication. Illinois: University of Illinois Press; 1949.

37. Bamforth SS. Microbial distributions in Arizona deserts and woodlands. Soil Biol & Biochem. 1984;16(2):133–137. http://dx.doi.org/10.1016/0038-0717(84)90103-2

38. Domonell A, Brabender M, Nitsche F, Bonkowski M, Arndt H. Community structure of cultivable protists in different grassland and forest soils of Thuringia. Pedobiologia. 2013;56(1):1–7. http://dx.doi.org/10.1016/j.pedobi.2012.07.001

39. Finlay BJ, Black HIJ, Brown S, Clarke KJ, Esteban GF, Hindle RM, et al. Estimating the growth potential of the soil protozoan community. Protist. 2000;151(1):69–80. http://dx.doi.org/10.1078/1434-4610-00008

40. Adl MS, Coleman DC. Dynamics of soil protozoa using a direct count method. Biol Fert Soils. 2005;42(2):168–171. http://dx.doi.org/10.1007/s00374-005-0009-x

41. Clarholm M, Popovic B, Rosswall T, Soderstrom B, Sohlenius B, Staaf H, et al. Biological aspects of nitrogen mineralization in humus from a pine forest podsol incubated under different moisture and temperature conditions. Oikos. 1981;37(2):137–145. http://dx.doi.org/10.2307/3544457

42. Weber, personal communication.

43. Mueller JA, Mueller WP. *Colpoda cucullus*: A terrestrial aquatic. Am Midl Nat. 1970;84(1):1– 12. http://dx.doi.org/10.2307/2423721

44. Bamforth SS. Water film fauna of microbiotic crusts of a warm desert. J Arid Environ. 2004;56(3):413-423. http://dx.doi.org/10.1016/S0140-1963(03)00065-X

45. Goodey T. Note on the remarkable retention of vitality by protozoa from old stored soils. Ann Appl Biol. 1915;1(3-4):395-399. http://dx.doi.org/10.1111/j.1744-7348.1915.tb08008.x

46. Moon-van der Staay SY, Tzeneva VA, Van Der Staay GWM, De Vos WM, Smidt H, Hackstein JHP. Eukaryotic diversity in historical soil samples. FEMS Microbiol Ecol. 2006;57(1):420–428. http://dx.doi.org/10.1111/j.1574-6941.2006.00130.x

47. Geisen S, Bandow C, Römbke J, Bonkowski M. Soil water availability strongly alters the community composition of soil protists. Pedobiologia. 2014;57(4-6):205–213. http://dx.doi.org/10.1016/j.pedobi.2014.10.001

48. Boenigk J, Arndt H. Bacterivory by heterotrophic flagellates: Community structure and feeding strategies. Anton Leeuw J Microb. 2002;81(1):465–480. http://dx.doi.org/10.1023/A:1020509305868

49. Parry JD. Protozoan grazing of freshwater biofilms. Adv Appl Microbiol. 2004;54(1):167–195. http://dx.doi.org/10.1016/S0065-2164(04)54007-8

50. Darby BJ, Neher DA, Housman DC, Belnap J. Few apparent short-term effects of elevated soil temperature and increased frequency of summer precipitation on the abundance and taxonomic diversity of desert soil micro- and meso-fauna. Soil Biol & Biochem. 2001;43(7):1474–1481. http://dx.doi.org/10.1016/j.soilbio.2011.03.020

51. Ekelund F, Rønn R, Griffiths BS. Quantitative estimation of flagellate community structure and diversity in soil samples. Protist. 2001;152(4):301–314. http://dx.doi.org/10.1078/1434-4610-00069

Diverse diets of the Mio-Pliocene carnivorans of Langebaanweg, South Africa

AUTHORS:
Adam Hartstone-Rose[1,2]
Katheryne N. Brown[3]
Carissa L. Leischner[1,2]
Ka'la D. Drayton[4]

AFFILIATIONS:
[1]Department of Cell Biology and Anatomy, University of South Carolina School of Medicine, University of South Carolina, Columbia, South Carolina, USA
[2]Department of Anthropology, University of South Carolina, Columbia, South Carolina, USA
[3]Medical Degree Program, University of South Carolina School of Medicine, Greenville, South Carolina, USA
[4]Health Information Technology, University of South Carolina, Columbia, South Carolina, USA

CORRESPONDENCE TO:
Adam Hartstone-Rose

EMAIL:
AdamHR@sc.edu

POSTAL ADDRESS:
Department of Cell Biology & Anatomy, University of South Carolina School of Medicine, University of South Carolina, 6439 Garners Ferry Road, Building 1, CBA rm C-36, Columbia, South Carolina 29209-1445, USA

DATES:
Received: 15 Dec. 2015
Revised: 01 Mar. 2016
Accepted: 05 Mar. 2016

KEYWORDS:
Carnivora; *Agriotherium*; Hayeanidae; Felidae; palaeodiet reconstruction

HOW TO CITE:
Hartstone-Rose A, Brown KN, Leischner CL, Drayton KD. Diverse diets of the Mio-Pliocene carnivorans of Langebaanweg, South Africa. S Afr J Sci. 2016;112(7/8), Art. #2015-0480, 14 pages. http://dx.doi.org/10.17159/sajs.2016/20150480

The Mio-Pliocene guild of carnivorans of Langebaanweg (LBW), South Africa, is phylogenetically and ecologically diverse. Unlike modern African fauna, this fossil sample contains a large ursid; although there are mustelids, herpestids and viverrids in Africa today, some of the LBW members of those families were much larger than their modern confamilials. There were also numerous felid species, including some that possess a more sabretoothed dental morphology, as well as several species of hyaenids that were very different from their modern confamilials. Questions remain about the dietary morphospace occupied by these fossils. Which taxa were predominately durophagous and which were the most hypercarnivorous? Did the level of durophagy and hypercarnivory in the LBW taxa reach the level of specialisation found in modern carnivores? In the current study, we evaluate the dietary specialisations of all the large terrestrial LBW carnivorans through analysis of the radii-of-curvature and intercuspid notches present in the mandibular dentition. We found that the LBW carnivorans had less sharp premolars than do their modern confamilials – an indication of greater durophagy. However, some families contain individuals with more extreme intercuspid notch patterns, indicating greater hypercarnivory. The ursid also possessed a suite of morphology unlike any modern carnivoran, exhibiting some morphology conducive to durophagy and some that places it functionally among the most hypercarnivorous of modern carnivorans. Thus it was likely capable of consuming high levels of both flesh and bone.

Introduction

Members of the order Carnivora exhibit a wide range of dietary adaptation and specialisation. Some members are 'hypercarnivorous', specialising in the consumption of vertebrate flesh[1-4] (e.g. felids), while others are 'durophagous', specialising in the consumption of bone or other obdurate foods[1-4] (e.g. most hyaenids), insectivorous (e.g. the aardwolf, *Proteles*), frugivorous (e.g. the binturong and kinkajou, *Arctictis* and *Potos*), and even strictly folivorous (e.g. the bamboo-eating red and giant pandas, *Ailurus* and *Ailuropoda*). Most taxa in the order represent a mixture of several of these specialisations. As carnivorans have diversified, their craniodental morphologies have adapted to diverse dietary niches. Important dietary information can be gained from analysing carnivoran teeth, especially in relation to tearing, shearing, cracking or grinding habits during food processing. These morphological changes and their significance can be evaluated by analysing the craniodental adaptations, and correlating these adaptions with dietary categories (e.g. durophagous or hypercarnivorous).

Analyses of premolar radii-of-curvature (ROC) and intercuspid notches (ICNs) show that tooth geometries correlate with today's observed diets of modern carnivorans; more hypercarnivorous taxa have sharper (higher ROC scores) teeth that are comparatively more intricately notched (higher ICN scores) than durophagous taxa.[1-3] The functional inferences are clear: sharper teeth are best for slicing ductile flesh, and yet this sharp edge is vulnerable to damage by particularly hard or tough foods. Likewise, numerous and sharp ICNs function to increase the linear sharp crests available for flesh slicing and also help focus slicing forces for ductile foods toward pinch points like the diamond-shaped slicing notches found in wire cutters. However, the well-developed, small and unsupported accessory cusps would be particularly vulnerable to breakage against hard foods like bone.

Goal and objective

The overarching goal of this study was to expand on the assessment of LBW hyaenid diets[3] to compare the morphology of the members of the terrestrial LBW carnivoran guild to modern carnivorans using analyses of dental ICN[2] and ROC[1] with the objective of evaluating the dietary niche space occupied by the large LBW carnivorans relative to that occupied by modern carnivoran lineages. Although both the geological context and taxonomic systematics of the LBW carnivorans are discussed more thoroughly in the previous literature (see below), we briefly discuss these in order to contextualise our fundamentally morphological study with particular emphasis on visually displaying some of the impressive specimens from this important location.

Background

Langebaanweg – a brief context

Originally discovered during phosphate mining operations in the early half of the 20th century, the fossiliferous deposits at LBW are known for their diverse terrestrial and marine taxa that make LBW one of the most significant Mio-Pliocene fossil-bearing sites in Africa. The broader Langebaanweg deposits include four distinct formations in geological succession: Elandsfontyn (underlies the formation in question and rests on local bedrock); Saldanha (early- to mid-Miocene in age); Varswater (most intensively studied Tertiary sequence in southern Africa); and Bredasdorp (post-early Pliocene deposits). The Varswater formation is divided into three distinct Members that comprise the LBW 'E' Quarry – the focal sample of the present study – from where most of the highly studied and best understood Mio-Pliocene fossils in southern Africa have been unearthed.[5]

The Langeberg Quartzose Sand Member (LQSM) was deposited during the late Miocene and early Pliocene and is abundant in fossils from both vertebrates and invertebrates and is thought to have resulted from repeated sedimentation events and the marine transgression of the Berg River.[6] Hundreds of thousands of fossils have thus far been recovered from 'E' Quarry belonging to over 200 species, more than 80 of which are mammals.[5] The remains of 13 out of 14 mammalian orders (excluding Sirenia) are represented, some of which no longer have relatives living south of the Sahara.[7]

The LBW 'E' Quarry is known for its fossils that have been found nowhere else in southwest Africa in such unprecedented quantity and quality.[5] Fossils from this deposit reflect what has been interpreted as a deteriorating environment during the early Pliocene as well as its subsequent change to fynbos and open grasslands that can currently be seen in LBW.[7] Within these two deposits, the following carnivoran families have been identified: Felidae (cats), Viverridae (civets), Hyaenidae (hyenas), Herpestidae (mongooses), Canidae (dogs), Ursidae (bears), Mustelidae (weasels) and Phocidae (seals).[5,7-25]

Carnivorans of Langebaanweg

'E' Quarry is especially well known for its fossil carnivorans, which comprise one third of the mammals from the locality and make Carnivora the most diverse order within the assemblage.[5] The fact that carnivorans are generally poorly represented at other late Miocene and early Pliocene sites in Africa makes the 'E' Quarry assemblage that much more valuable.[7] Unique in such high numbers to this fossil site are the remains of at least 14 individuals of the giant extinct African bear *Agriotherium africanum*, which are not generally found together in such large numbers.[20,26,27] This bear and many other large carnivores found at LBW (e.g. *Ikelohyaena, Hyaenictitherium, Plesiogulo* and *Dinofelis*) originally came from Eurasia when the drop in sea level at the end of the Miocene allowed passage into Africa via the Mediterranean land bridge.[7] The subsequent rise of global sea levels at the beginning of the Pliocene drastically changed the climate in LBW from tropical to temperate – thus, the 'E' Quarry assemblage dates from a period when fauna were either adapting to changing environmental conditions or going extinct.[7]

Ursidae

Agriotherium was the most geographically dispersed member of the family Ursidae, with fossils found at sites in North America, Europe, Asia and Africa.[28] It was the only ursid known to have inhabited sub-Saharan Africa.[28] *Agriotherium africanum*, the species found in LBW, was very large (~540 kg[29]) and exhibited several relatively unique craniodental characteristics: a short-snouted, robust skull; a 'chin'-like protrusion at the symphyseal region of the mandible; well-developed sectorial carnassials; and a premasseteric fossa.[5,20,27,29,30] The only extant ursid to also have premasseteric fossae is *Tremarctos*, which also has sectorial carnassials reminiscent of *Agriotherium.* This has led to the hypothesis that *Agriotherium* may have given rise to the *Tremarctos* lineage and possibly the other extant ursids.[29] *Agriotherium* may have been outcompeted and replaced by modern bears.[28] Ursidae is a fairly young family, and likely evolved from caniform ancestors only 23–24 million years ago, and *Agriotherium* exhibited many craniodental characteristics that are reminiscent of these ancestors (such as a robust P4) which has caused controversy over whether or not *Agriotherium* should be considered an ursid or an amphicyonid.[27,30]

Previous research on *Agriotherium* has focused largely on estimating its bite force in relation to body mass for the purposes of reconstructing evolutionary anatomical adaptations and hypothetical feeding ecology. For instance, one study[30] concluded that *Agriotherium* had an absolute bite force higher than that of any other mammalian carnivore and that, even relative to body mass, *Agriotherium*'s bite force was still formidable. However, the wide range of feeding niches and masticatory adaptations occupied by modern ursids make it difficult to predict *Agriotherium*'s diet. For example, the giant panda is able to generate and sustain a powerful bite force comparable to *Agriotherium* in order to exploit bamboo, its primary food source, although some studies have claimed that *Agriotherium* actively hunted and consumed large vertebrates.[20,30] Other studies[29] have suggested that the diets of these ursids included a lot of vegetation, while still others[29,31] have argued that *Agriotherium* is adapted largely for durophagy. Because of the presence of a true carnassial complex and robust molars fit for durophagy (Figure 1b), the growing consensus seems to be that *Agriotherium* was a predator-scavenger.[32]

Canidae

Canids are fairly poorly represented in the LBW assemblage, as they were not widespread in southern Africa during the Mio-Pliocene. There are at least two genera of canids: a larger one represented by a remarkably complete (although fragile) specimen (Figure 1a) that has been ascribed to the genus *Eucyon*, which is the earliest known canid from Africa; and a smaller, fox-like sample tentatively ascribed to the genus *Vulpes*.[5,7,15,33] Craniodental characteristics unique to the *Vulpes* specimen include short development of the P4 protocone and the presence of a second distal p4 cuspid.[27,34] A third specimen may also represent an additional species of canid, but the inability to distinguish the lower dental elements from that of a viverrid (a morphologically and taxonomically diverse group in the LBW sample) have prevented a confident diagnosis[7] and these specimens were not complete enough for our analyses anyway.

Felidae

The taxonomy and systematics of the LBW felids have yet to be fully resolved. The majority of the literature on the LBW felids focuses on the larger specimens ascribed to the machairodont taxon *Dinofelis* (Figure 1d).[3,35-38] Less complete remains of a second large sabretoothed cat, ascribed to *Homotherium,* have also been found at LBW.[7] There also are smaller felids in the sample – most of which have been questionably assigned to the modern genus *Felis* (Figure 1c), namely *F. obscura* and *F. issiodorensis* (both of which are lynx-like), '*Felis* sp.', which resembles a wildcat, and a felid of unknown genus and species.[7,16] As members of Felidae are highly specific to particular environments, a more complete systematic and functional understanding of the LBW species would help us understand the diversification and niche partitioning of this African lineage from the Quaternary.[6]

The wide distribution of *Dinofelis* and its especially broad presence (both in number of localities and specimens at those localities) in Africa suggest that it could be the most common fossil felid in the African Neogene.[27,38] The specimens from LBW represent the oldest known individuals of *Dinofelis*[15] and have been ascribed to the taxon *D.* cf. *diastemata*[27] which differs from its Eurasian congeners by a smaller body size and the presence of craniodental characteristics similar to those of the non-sabretooth genus *Panthera* in that it possessed a short upper canine, a P3 less reduced than in other machairodonts, and a less elongated m1 than those found in other machairodonts.[38-40] A notable evolutionary change that has been observed in LBW *Dinofelis* is a lengthening of the carnassials in relation to the rest of the tooth row – indicative of evolution from the more primitive state toward the more machairodont traits that are generally observed in other sabretoothed cats.[38] Like modern felids, *Dinofelis* was almost certainly hypercarnivorous, although it possessed a reduced post-canine dentition in comparison to other members of Felidae to accommodate the maxillary canines.[2] The rise of felids belonging to the genus *Panthera* likely led to the decline of *Dinofelis* and its eventual extinction.[7]

Hyaenidae

Members of Hyaenidae are the most commonly found terrestrial carnivores in the LBW assemblage. At least five species of hyena (*Chasmaporthetes australis, Hyaenictitherium namaquensis, Hyaenictis hendeyi*, *Ikelohyaena abronia,* 'Hyaenidae sp. E'[41]; Figure 1f–g) are represented in these assemblages that coexisted during the late Miocene and early Pliocene in southern Africa.[3,7,14,18,27] These species are believed to have evolved from civet-like ancestors on more than one occasion during the early Miocene.[3,18,41] Our recent study[3], which utilised the same methods as our current study, found that the LBW hyaenids were poorly adapted for bone cracking and were better suited for a more generalised carnivoran diet than the highly derived extant durophagous hyaenids. This finding corresponds with the conclusion that modern levels of durophagy are known to have appeared only recently in the fossil record.[42]

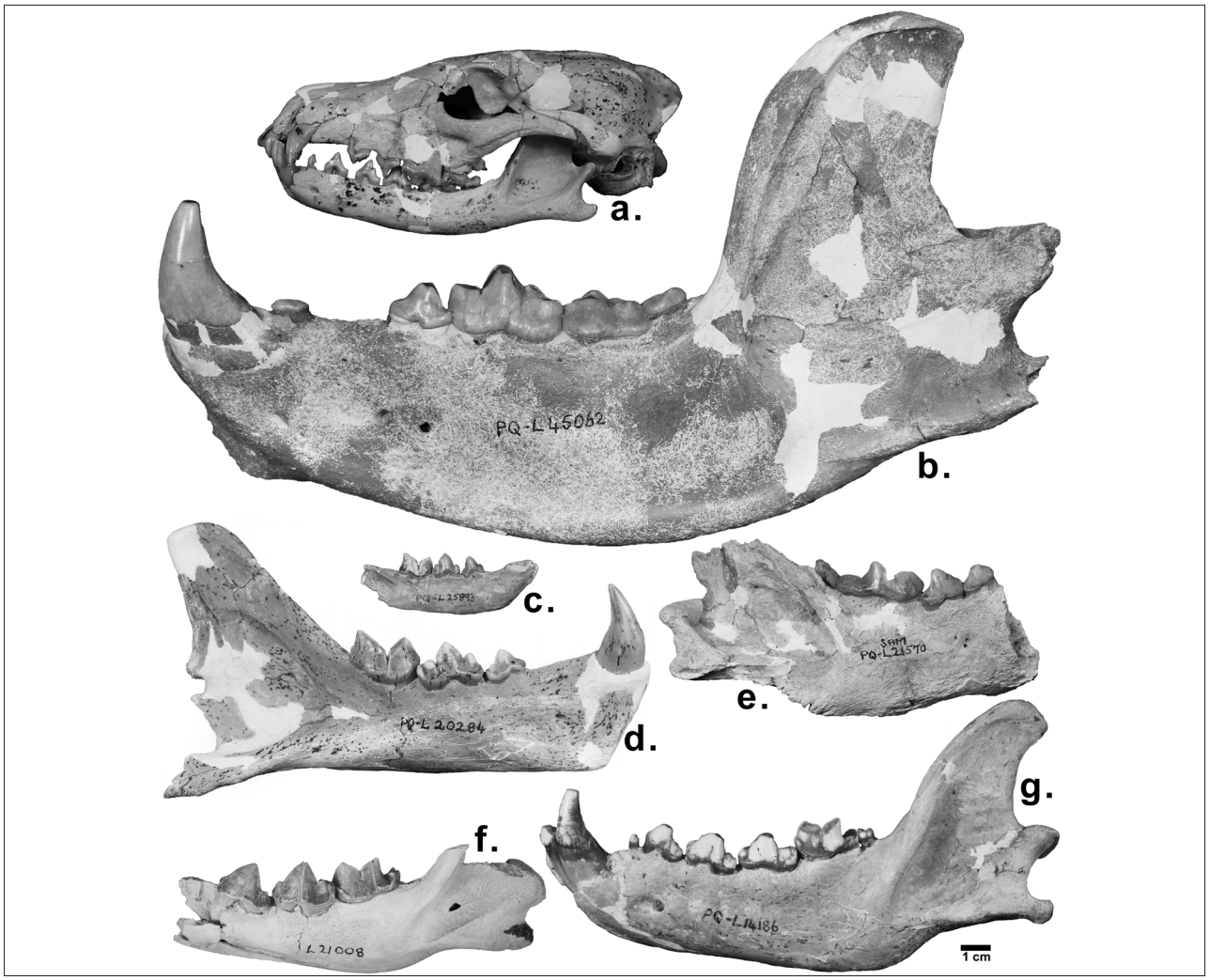


Figure 1: (a) The remarkably complete larger Langebaanweg (LBW) canid specimen (PQ-L 31272); (b) the most complete LBW mandibular specimen of the ursid *Agriotherium africanum* (PQ-L 45062); (c and d) representative specimens of the LBW felids including (c) the smaller (PQ-L 25893) and (d) the larger (PQ-L 20284) morphs; (e) representative specimen of a LBW mustelid, *Plesiogulo monpessulanus* (PQ-L 21570); and (f and g) representative specimens of LBW hyaenids, (f) *Hyaenictitherium namaquensis* (PQ-L 21008) and (g) *Ikelohyaena abronia* (PQ-L 14186).

The ability of durophagous hyaenids to efficiently process and consume obdurate foods like bone is enabled by the presence of broad, stout premolars that can transmit large forces without breaking.[1-3] However, gnaw marks and acid etching on the fossilised bones of prey species (consistent with being partially digested by hyenas) have been found at LBW, indicating that at least some of these extinct species engaged in the same behaviours and occupied part of the same scavenging niche as most of their extant confamilials.[7,43]

Informally known as the 'hunting hyena', *Chasmaporthetes australis* was the largest, yet most gracile of the LBW hyaenids and was likely adapted to a more cursorial lifestyle.[15] Craniodental characteristics of this species include: smaller cheek teeth and a markedly weaker masticatory apparatus when compared to the extant durophagous hyaenids (*Crocuta, Hyaena* and *Parahyaena*); strong posterior accessory cusps on p2 and p3; a large P4 protocone; a very large anterior p4 accessory cusp; and an m1 with a single, blade-like talonid cusp.[15,27,44] Analysis of the premolar ICN and occlusal ROC scores of this species suggests that *Chasmaporthetes australis* was more hypercarnivorous than modern hyenas and may have had a diet similar to those of modern felids.[1-3]

The holotype of *Ikelohyaena abronia* represents the most complete skeleton found to date within the LBW 'E' Quarry and is comprised of the skull and more than 80 postcranial bones from the same individual.[7] In a study by Werdelin and Solounias[41], *Ictitherium preforfex* was synonymised with *I. abronia* – these were originally thought to be two separate species because of their large variation in size.[15,18] Craniodental characteristics unique to *I. abronia* include an enlarged p3 and P3, a lack of anterior accessory cusps on p2 and P2, a large and roughly triangular-shaped m1, and smaller, lower crowned anterior premolars than *H. hyaena*.[15,27] *Ikelohyaena abronia* was the most durophagous of all LBW hyaenids[3] although it lacked the strong bite force of the modern durophagous hyenas.[3,45] Along with the not yet formally named Hyaenid sp. E from LBW, *I. abronia* belonged to a clade of early or transitional bone-cracking hyaenids.[46]

Hyaenictitherium namaquensis was the most widespread African member of the ictitherines, a clade of hyenas with canid characteristics that was most prominent in Eurasia from the middle of the Miocene until its end.[27] Like other ictitherines, the genus *Hyaenictitherium* possessed a large m1 talonid as well as both an m2 and M2.[27,41,47] Unlike its more doglike Eurasian counterparts, however, African species within *Hyaenictitherium* possessed crushing dentition and shorter, more massive canines along with a stronger mandibular ramus – all adaptations conducive to consuming obdurate foods.[27] Excluding many dental characteristics, the skull of *H. namaquensis* shares many similarities with the sympatric *I. abronia* and the extant *Hyaena hyaena*.[15] Craniodental characteristics specific to this species include: a marked shortening of the snout; a significantly larger M1 than that of both *I. abronia* and *H. hyaena*; and higher crowned premolars and molars than *I. abronia*.[15]

Hyaenictis hendeyi, formerly described as *Euryboas* sp. by Hendey[15], was a gracile, long-limbed LBW hyaenid that was believed to be both cursorial and actively predacious[15,44], much like *C. australis*. This species retained an m2 (a primitive character in hyenas), small or non-existent anterior accessory cuspids on p2 and p3, and a distinct anterior accessory cuspid on p4 that is notably smaller than its posterior accessory cusp.[44] Because of the absence of extant lineages of big cats at the time when *H. hendeyi* lived, it has been proposed that this species could have occupied the ecological niche that is now occupied by the extant cheetah *Acinonyx.*[7]

Only a single mandibular fragment of Hyaenid sp. E with an intact p2 and p3 from an immature individual has been discovered and described.[15,18] Although characteristics of the dental elements of this specimen indicate that it is distinct from all known hyaenids, there is not enough material present to completely name and describe it as a species.[41] Hyaenid sp. E displays the following dental characteristics: an absent p1; no noticeable anterior accessory cusp on p2; and a p3 with no anterior accessory cusp and a greatly reduced posterior accessory cusp.[15] These characteristics, at the least, suggest that Hyaenid sp. E was highly durophagous and belonged to the same clade of early/transitional bone-cracking hyaenids as *I. abronia.*[3,46] Because this specimen is rather fragile and subadult, it was excluded from our sample.

Although we have previously published these analyses for the LBW hyaenid sample[3], we are including them in the present investigation to be able to more thoroughly discuss the morphospace occupied by all of the LBW large carnivore guild. As described in our previous study, the only LBW hyaenid specimens complete enough for our analyses were from *Chasmaporthetes* and *Ikelohyaena*.

Mustelidae

Although mustelids were widespread throughout Africa and the northern continents during the late Tertiary, they are relatively poorly represented at 'E' Quarry (in terms of numbers of species and overall specimens) in comparison to the species diversity and abundance of other families.[19] Three species of large mustelid have thus far been identified from LBW 'E' Quarry: *Plesiogulo monpessulanus*, *Mellivora benfieldi* and *Sivaonyx hendeyi.*[15,19,48] The discovery of the remains of at least three individual *P. monpessulanus* (Figure 1e) at LBW marks the southernmost known record of a wolverine.[15,19] As members of its genus were widely spread throughout Eurasia and North America during the late Miocene and Pliocene, it has been suggested that *Plesiogulo* could also be found throughout the African continent.[19,23] It is generally accepted that *P. monpessulanus* is ancestral to the extant *Gulo gulo*, as a result of both morphological similarities and assumed behavioural similarities.[19,23] What so obviously sets this species apart from its modern descendant is its large size, apparently being only slightly smaller than the largest recorded fossil mustelid, the North American *Megalictis ferox.*[18] Dental characteristics of *P. monpessulanus* include a relatively long m1 talonid, shorter carnassials relative to those of *G. gulo*, and larger maxillary and mandibular cheek teeth relative to those of the extant *G. gulo.*[27] While modern *Gulo* is extremely powerful – capable of cracking bones[49] – the dental characteristics of *P. monpessulanus* likely indicate even greater durophagy abilities.

Mellivora benfieldi, likely ancestral to the extant *Mellivora capensis*, was somewhat smaller in size than its descendant and is tentatively represented by as many as six individuals from the LBW 'E' Quarry. Dentition from these fossil specimens shows a loss of p1 (an advanced character in mellivorines), narrow premolars, and sharp-pointed cusps and sharp-edged keels on the cheek teeth compared to those of *M. capensis*. *Mellivora benfieldi* is the less specialised of the two species in terms of the lack of defined p4 accessory cusps and the presence of a comparatively unspecialised m1 talonid.[19]

The Mio-Pliocene otter *Sivaonyx hendeyi* of 'E' Quarry was removed from its original classification because its unique m1 presents characteristics too different to warrant inclusion within the pre-existing species *Enhydriodon africanus.*[48] It is believed to be an intermediate between *E. iluecai* (late Miocene) and *E. sivalensis* (late Pliocene or early Pleistocene) in terms of its craniodental characteristics as well as its temporal occupation.[19] *Sivaonyx hendeyi* was slightly larger than the extant *Aonyx cinerea* in terms of body size, and its cheek teeth exhibit a primitive inflation of cusps and lowering of crown height that is much more developed in extant members of *Enhydra.*[19]

Herpestidae

Herpestids are relatively common in LBW 'E' Quarry (Figure 2a); however, the fragmentary nature and similarity among specimens makes classifying them to a species level difficult.[15] Two unidentified or undescribed species of Pliocene *Herpestes* (*H.* sp. A and *H.* sp. B), have been described from LBW. *Herpestes* Species A lacks an anterior accessory cusp on P3 and displays a P1 that is triangular in outline. *Herpestes* Species B was smaller in size than Species A, and lacks the first mandibular premolar. *Herpestes ichneumon*, *H. pulverulentus,* and *Suricata major* have been identified from 'E' Quarry. *Herpestes ichneumon* has been hypothesised to be ancestral to the modern Egyptian mongoose and displays many similar characteristics in its mandibular dentition, but also has a larger m2 than its modern conspecifics. *Herpestes pulverulentus* has been hypothesised to be ancestral to the extant Cape grey mongoose (*Galerella pulverulenta*) and is represented by a skull with incomplete maxillary dentition and a mandibular fragment. As in *H.* Species B and the modern *G. pulverulus*, *H. pulverulentus* lacks p1; it was also likely somewhat larger than its modern conspecific. *Suricata major* was of large size and had a p4 without an anterior accessory cusp and also displayed a p1.[15]

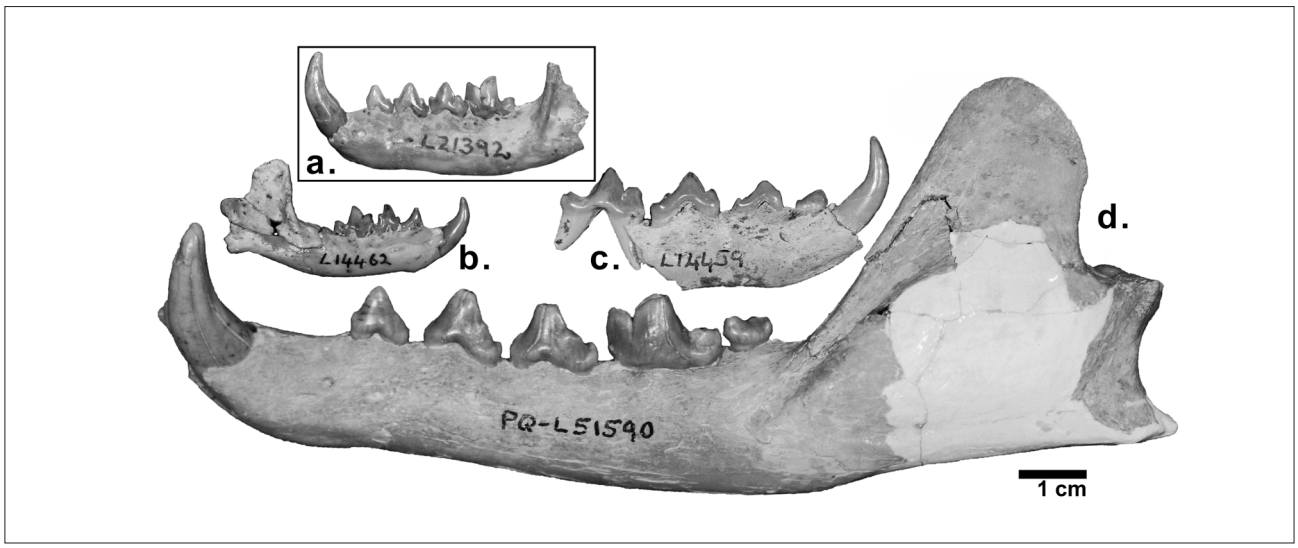


Figure 2: Representative of Langebaanweg (a) Herpestidae (PQ-L 21392) and (b–d) Viverridae (b: PQ-L 14462, c: PQ-L 14459, d: PQ-L 51590).

Viverridae

One of the more diverse families of carnivorans represented in the LBW sample is the Viverridae[15] (Figure 2b–d). However, most of the smaller specimens have not been classified beyond the level of family.[27] Of the identified LBW specimens, one was the giant viverrid *Viverra leakeyi*[50] (Figure 2d). Judging from one nearly complete LBW cranium, *V. leakeyi* had both a high and narrow skull as well as a relatively long mandible.[15] The postcranial remains of this species suggest that it was nearly the same size as the extant *Civettictis civetta*, while its craniodental characteristics share a strong similarity with those of the modern *Canis mesomelas.*[15]

Genetta sp. is represented by a single mandibular fragment from 'E' Quarry and appears to have been about one third of the size of extant genets. Despite the disparity in size, the dentition of this specimen is very similar to that of *Genetta tigrina* in that its p2 has one posterior accessory cusp, while both the p3 and p4 have two. The mandibular fragment also displays a convex curve along the entire length of its inferior margin. It is the only known fossil genet to have been described in South Africa, but the lack of additional material has prevented it from being named.[15]

Materials and methods

Methods

Dietary niche is assessed using two previously developed methods: premolar ROC and ICN score analyses.[1-3] Both methods examine dietary correlates of mandibular dentition. Although many studies of dietary correlates have focused on lower dentition (most notably the primate work of Kay and followers thereof[51-53]), this focus on only maxillary dentition limits dietary reconstruction of fossils – a constraint to be remedied in further studies. Furthermore, we studied only adult specimens with relatively unworn and taphonomically unmodified crowns that are strong enough for physical moulding. (See the Discussion for notes about the inclusion of maxillary dentition and fragile specimens.)

As in our previous studies[1-3], modern carnivorans are ascribed to general dietary categories based on the primary literature and previous categorisation schemes[54]. The first category consists of 'hypercarnivores', or species whose diet consists of more than 70% vertebrate flesh. The second category consists of 'durophages', or species that consume bone or other hard dietary products. The last two categories are 'meat/non-vertebrate' and 'non-vertebrate/meat', which contain species that consume fruit and/or insects as either secondary or predominant components of their diets, respectively.[54]

Premolar radii-of-curvature

Premolar ROC is a dietary evaluation method used to predict diet and methods of consumption of flesh or bone. Post-canine specimens containing p3, p4 and the carnassial (m1) cusps are moulded using regular-body President Jet (Coltene Whaledent, Cuyahoga Falls, OH, USA), a quick-setting product used primarily in the dental industry, and then cast using Smooth-Cast (Smooth-On, Easton, PA, USA), a fast setting, low-viscosity, two-part resin. The casts are sectioned at the central cusp of each premolar and the two mesial cusps of the carnassial using a thin kerf saw in the coronal plane according to their own orientation.

These sections are then scanned onto a flatbed scanner and measured in ImageJ (NIH, Bethesda, MD, USA) by fitting the cross section of the cusp with the smallest circle that fills the tip and then taking the inverse of the radius of the fitted circle (1/radius) as the ROC. Unfortunately, teeth are not simple geometric shapes and fitting a circle to a cusp point is clearly an oversimplification. Other techniques have been employed to fit more complex shapes (e.g. a parabola) to tooth cross sections, but these too are only slightly more accurate in describing the shape. Indeed, each additional term that could be added to the equation would make the line fit the curve of the tooth more accurately. However, the fitting of circles to premolar cross sections has been shown to be reproducible and indicative of diet, and, because it is easier and faster than digitising the whole outline of the tooth and fitting lines with algorithms, more data can be incorporated into the sample more efficiently.[1,3]

Intercuspid notch

Intercuspid notch scoring is a dietary evaluation method in which the notches between cusps of premolars are scored on a scale from 1 to 5 with sharpness of crests leading into the notch being considered as perpendicular to the slope of the crest itself.[2,3] A score of 1 represents a barely visible notch; 2 represents a notch between two rounded, dull crests; 3 represents a notch between one dull crest and one sharp crest at $<45°$; 4 represents a notch between two sharp crests; and 5 represents a 'carnassiform notch' [2,3] or a notch with a keyhole pattern formed at the bottom of a sharp vertical notch at the convergence of crests.

Interpretation of metrics

For both ROC and ICN, modern species that have higher values tend toward hypercarnivory and modern species with lower values tend toward durophagy.[1-3] The p3 is most indicative of diet for ROC analysis[1] because the carnassial is generally utilised for meat slicing in large carnivores and therefore the greatest level of durophagy is restricted away from this fragile tooth – especially in feloids. In some canoids, durophagy is done in the posterior-most portions of the tooth row – the carnassial talonid and posterior molars. However, because these regions are not universal in carnivores, dietary correlation has been restricted to premolars and carnassials. The p4 correlates most closely with diet along the hypercarnivory–durophagy scale for ICN analysis[2] because this tooth contains the most variability in ICN and therefore is most discriminant along the continuum.

All statistical significance was evaluated using a non-parametric Wilcoxon test. When specific p-values are not given, $p<0.05$ is taken to be significant and $p<0.01$ is taken to be highly significant.

Sample

A total of 264 specimens of 35 modern taxa were included from the collections of the American Museum of Natural History (New York) and the National Museum of Natural History (Washington, DC) – a sample that we fully described in our previous publications.[1,2] These were compared to a total of 65 LBW specimens that were complete enough to evaluate our variables of interest at least in part (Table 1).

Unfortunately, our methods were developed to evaluate dietary specialisation in relatively large carnivorans the size of black-backed jackals, *Lupullela mesomelas* (~8kg[55]) or larger – essentially focusing on carnivoran specialisations in hypercarnivory, durophagy and moderate to extreme herbivory – to the exclusion of the dependence on insectivory found extensively in small members of the order. Thus, the smallest LBW viverrids and mustelids, the smaller of the two canids, and all of the herpestids are essentially outside of the body size range of applicability of these methods. However, in an effort to address the guild as completely as possible, some individuals below this size range (one LBW herpestid and one small LBW viverrid) were included. Because of this scaling issue, their dietary assessments, as discussed at the end of this paper, should be taken as tentative. Much more work needs to be done on the small LBW carnivorans including extending methods to account for the effects of insectivory on dental sharpness and cusp notch morphology (see future directions).

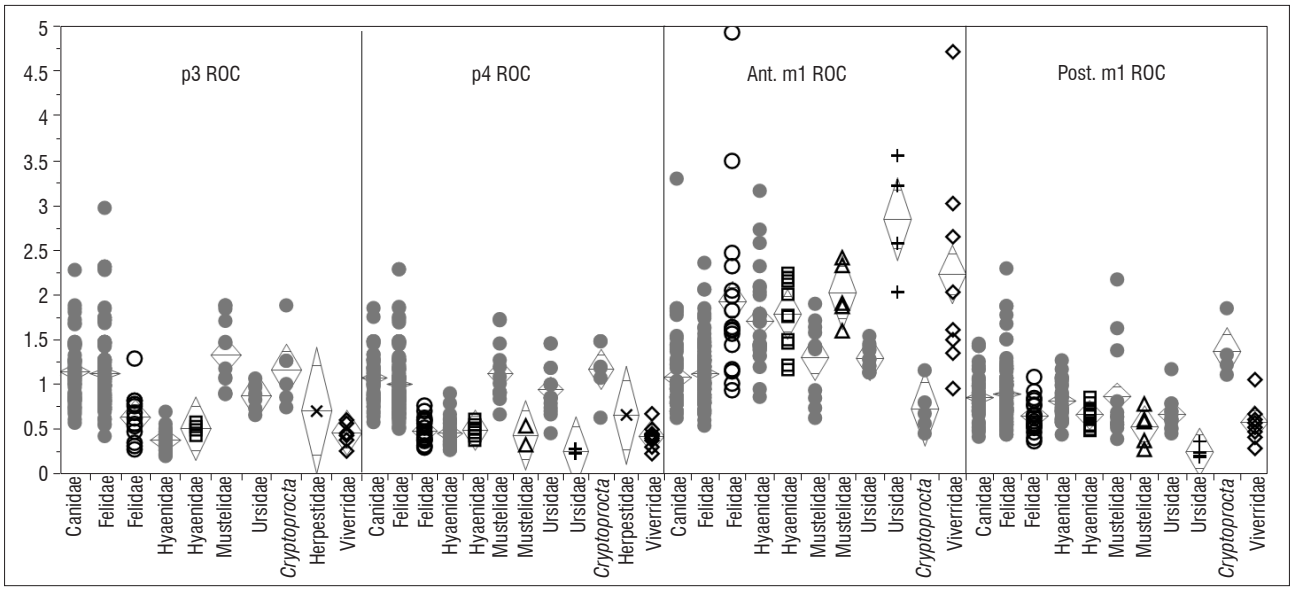
Results

Family-level radii-of-curvature results

The ROC results for p3, p4 and the anterior and posterior cusps of the carnassial (m1) occupy similar morphospace as those of modern carnivorans (Figure 3). As we found previously[3], the modern and fossil hyenas overlap in the sharpness of their teeth. However, comparisons of specific ROC data of the other families reveal significant differences between fossil and modern confamilials. Namely, at the family-level, the LBW felids have lower p3 and p4 ROC (i.e. are duller) than the modern felid sample ($p=0.0002$ and $p<0.0001$, respectively). Likewise, the LBW viverrids also have significantly ($p=0.0131$ for p3 ROC and $p<0.0001$ for p4 ROC) less-sharp premolars than those of *Cryptoprocta*, the most comparable extant viverroid in our comparative sample.

Table 1: Specimens studied by family: Extant species and Langebaanweg (LBW) confamilials

Family	Species	Common name	Number of specimens	
			Female	Male
Extant Canidae	*Canis lupus*	Grey wolf	3	4
	Canis mesomelas	Black-backed jackal	2	3
	Canis rufus	Red wolf	9	9
	Canis simensis	Ethiopian wolf	2	2
	Chrysocyon brachyurus	Maned wolf	4	6
	Cuon alpinus	Dhole	3	3
	Lycaon pictus	African hunting dog, African painted dog	3	8
LBW Canidae			2	
Extant Felidae	*Acinonyx jubatus*	Cheetah	9	7
	Caracal caracal	Caracal, African lynx	2	2
	Catopuma temmincki	Asiatic golden cat	6	2
	Neofelis nebulosa	Clouded leopard	10	7
	Panthera leo	Lion	5	6
	Panthera onca	Jaguar	6	8
	Panthera pardus	Leopard	6	14
	Panthera tigris	Tiger	4	6
	Panthera uncia	Snow leopard	6	6
	Prionailurus viverrinus	Fishing cat	1	2
	Profelis aurata	African golden cat	0	3
	Puma concolor	Mountain lion, cougar, puma, catamount	10	4
LBW Felidae			26	
Extant Hyaenidae	*Crocuta crocuta*	Spotted hyena	4	8
	Hyaena hyaena	Striped hyena	9	3
	Parahyaena brunnea	Brown hyena	6	3
LBW Hyaenidae			15	
Extant Mustelidae	*Enhydra lutris*	Sea otter	1	1
	Gulo gulo	Wolverine	4	5
	Mellivora capensis	Ratel or honey badger	1	1
	Pteronura brasiliensis	Giant otter	0	2
LBW Mustelidae			5	
Extant Ursidae	*Ailuropoda melanoleuca*	Giant panda or panda bear	1	2
	Helarctos malayanus	Malaysian sun bear	1	1
	Melursus ursinus	Sloth bear	1	1
	Tremarctos ornatus	Spectacled bear	1	1
	Ursus americanus	Black bear	1	1
	Ursus arctos	Brown or grizzly bear	1	2
	Ursus maritimus	Polar bear	1	1
	Ursus thibetanus	Asiatic black bear	1	1
LBW Ursidae			6	
Extant Eupleridae	*Cryptoprocta ferox*	Fossa	3	2
LBW Herpestidae			1	
LBW Viverridae			11	



*Extant taxa are represented by solid grey circles, Langebaanweg (LBW) felids by open circles, LBW hyaenids by open squares, LBW mustelids by open triangles, LBW ursids (*Agriotherium*) by '+' signs, the single analysed LBW herpestid by an 'x' sign, and LBW viverrids by open diamonds.*

Because the only euplurid in the extant sample is represented by Cryptoprocta*, that column is labelled with the generic and not familial name.*

No LBW canid fossil preserved cusps that are strong enough for moulding for the ROC measurement method.

Each dotted diamond represents the mean and 95% confidence interval vertical span; the horizontal lines are the mean and 'overlap marks' that indicate statistical separation.

Figure 3: Radii-of-curvature (ROC) of the third (p3) and fourth (p4) lower premolars and the anterior (paraconid) and posterior (protoconid) carnassial (m1) cusps by family.

Although the LBW ursids (*Agriotherium*) and mustelids do not preserve third premolars that are complete enough for ROC analysis, the p4 ROC results are interesting; in the ursid comparison, the *Agriotherium* specimens have significantly less-sharp premolars than their modern confamilials ($p=0.0048$), and while there are only two LBW mustelids with well-preserved fourth premolars, these two teeth are less-sharp than those of their modern confamilials, and despite the small sample, this difference does approach significance ($p=0.0567$). Similar patterns exist when considering the sharpness of the posterior cusp (protoconid) of the carnassial; the LBW felids and viverrids have significantly less sharp posterior carnassial cusps than do modern felids and *Cryptoprocta* ($p=0.0239$ and $p<0.0001$, respectively). The fossil mustelids and ursids have duller posterior carnassial cusps than do their modern confamilials, although these differences are not significant at these small sample sizes.

The patterns notably shift when considering the anterior carnassial (paraconid) cusps (Figure 3); while the LBW taxa tend to have duller premolars than their modern confamilials, their posterior carnassial cusps are *sharper* for all families other than the hyaenids, significantly so for the felids, ursids and non-hyaenid viverroids ($p<0.0001$ for all three comparisons).

Species-level radii-of-curvature results

With some ambiguity about the taxonomic diversity within the LBW families, comparing the fossil families to individual modern species yields interesting results. For instance, the LBW felids have significantly less-sharp premolars (Figures 4 and 5) than those of the sharpest felids (e.g. *Acinonyx, Caracal, Neofelis, Prionailurus* and *Profelis*) and indeed overlap with all of the modern and fossil hyaenids – a group that overlaps in premolar ROC only with the most robust of the felids (e.g. *Panthera leo*, *P. onca* and *P. tigris*). Likewise, the LBW viverrids have significantly less sharp premolars than those of the modern *Cryptoprocta* ($p=0.0035$ and $p<0.0001$ for p3 and p4 ROC, respectively). When examining the ROC of mustelid and ursid fourth premolars (again, the only tooth well enough preserved in the LBW sample for these families), they are indeed statistically distinct from some of their confamilials despite the small comparative sample size.

When examining the ROC of the carnassial cusps by species (Figures 6 and 7), as was evident in the family level analysis, the LBW felids have significantly sharper anterior carnassial cusps than all but a few felid species (namely *Panthera leo*, *P. tigris* and *Profelis aurata*). The LBW mustelids have relatively sharper anterior carnassial cusps than their modern congeners, but this find is not statistically significant (probably as a consequence of the small sample sizes). However, despite the extremely small sample sizes of modern ursids, the *Agriotherium* specimens are significantly sharper in this cusp than those of most of the ursids.

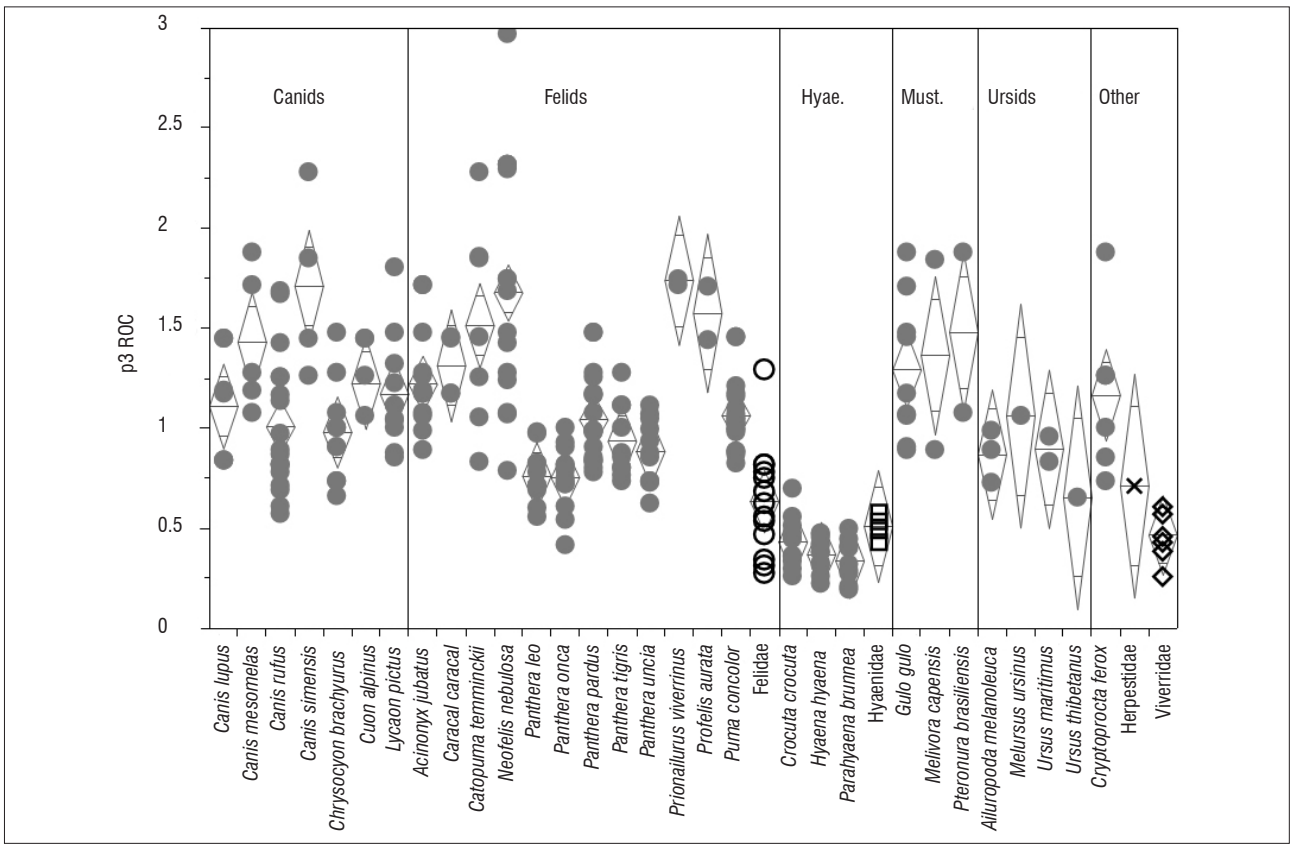
In terms of the posterior carnassial cusp ROC scores (Figure 7), the LBW specimens are statistically indistinguishable from almost all of their modern analogues. There are two notable exceptions: the LBW mustelids have significantly sharper posterior carnassial cusps than do the modern *Mellivora,* as do the LBW viverrids relative to *Cryptoprocta* ($p>0.0001$ for both comparisons).

Premolar intercuspid notch results

At the family level, none of the LBW samples can be statistically distinguished from their confamilials (Figure 8). However, comparing the fossil samples to individual modern species does yield significant findings (Figures 9 and 10). Namely, the LBW felids have significantly less notched third premolars than those of *Acinonyx* ($p<0.0001$) – an unsurprising finding given that *Acinonyx* has a significantly higher p3 ICN score than every other carnivoran represented by an adequate sample size.

However, the LBW felids also have significantly lower p3 ICN scores than those of *Lycaon* ($p=0.0434$) – the most hypercarnivorous of the canids and a taxon with a p3 ICN score significantly higher than that of about half of the modern felids.

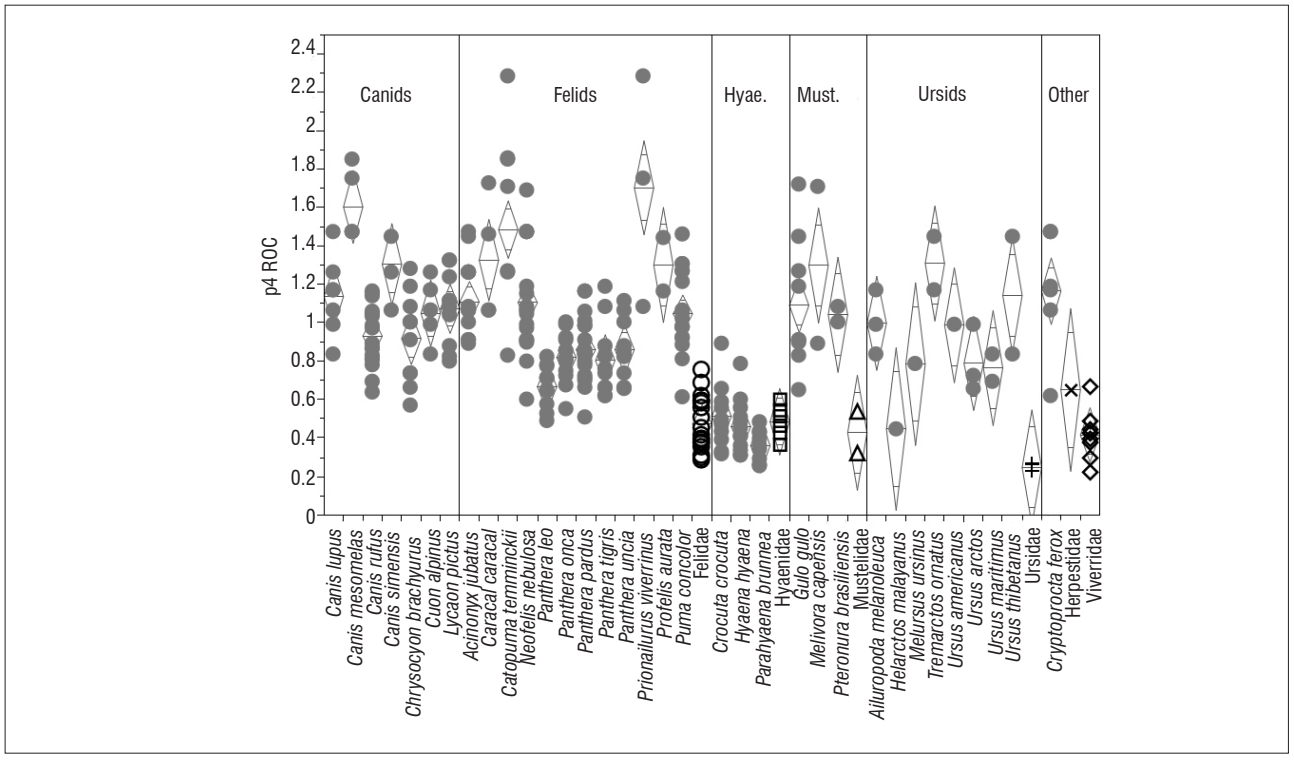
The only two well-represented LBW canids show vastly different morphospace with the small specimen (previously ascribed to *Vulpes*[5,7]) with premolar notch morphology similar to the most durophagous modern carnivorans, and the larger specimen (previously ascribed to *Eucyon*[5,7]) exhibiting notch morphology that would place it among the most hypercarnivorous modern carnivorans. Although the LBW hyaenids have slightly more notched third premolars than their confamilials, they are not significantly so.



Extant taxa are represented by solid grey circles, Langebaanweg (LBW) felids by open circles, LBW hyaenids by open squares, LBW mustelids by open triangles, the single analysed LBW herpestid by an 'x' sign, and LBW viverrids by open diamonds.

Each dotted diamond represents the mean and 95% confidence interval vertical span; the horizontal lines are the mean and 'overlap marks' that indicate statistical separation.

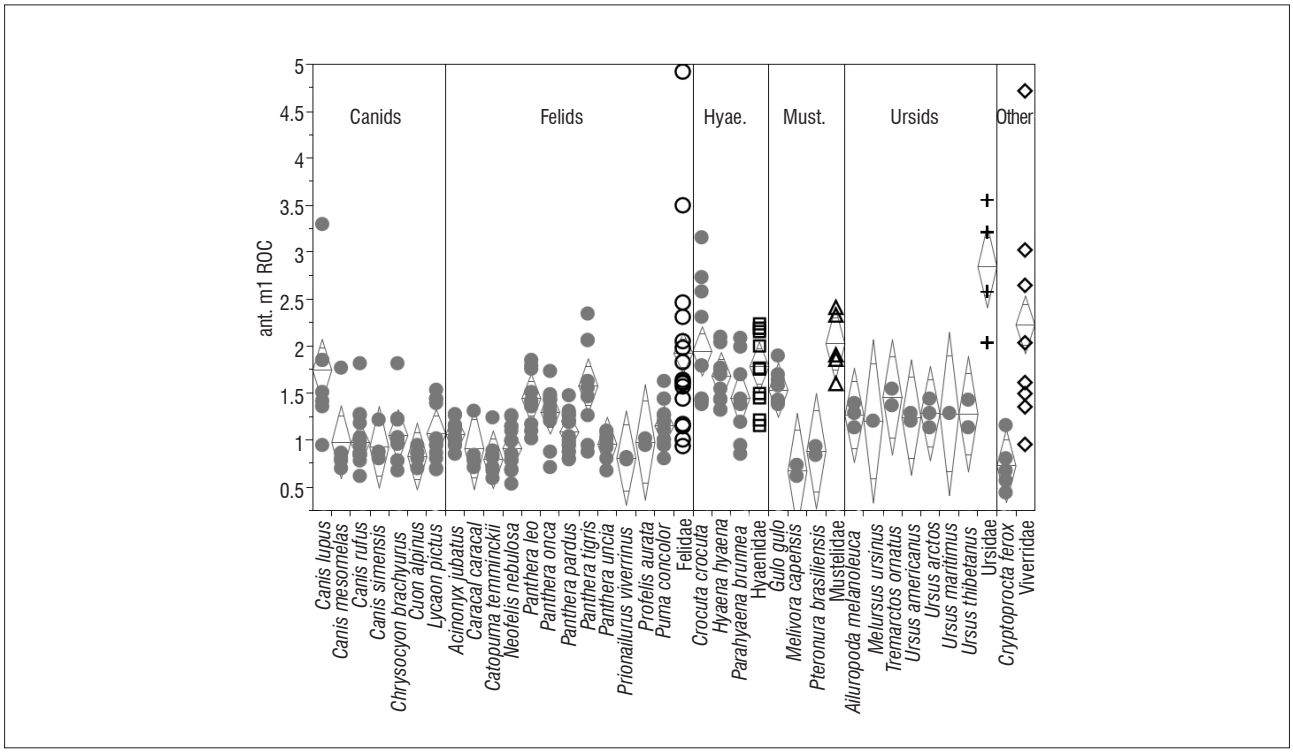
Figure 4: Radii-of-curvature (ROC) score of the third (p3) lower premolars by species and/or family.



*Extant taxa are represented by solid grey circles, Langebaanweg (LBW) felids by open circles, LBW hyaenids by open squares, LBW mustelids by open triangles, LBW ursids (*Agriotherium*) by '+' signs, the single analysed LBW herpestid by an 'x' sign, and LBW viverrids by open diamonds.*

Each dotted diamond represents the mean and 95% confidence interval vertical span; the horizontal lines are the mean and 'overlap marks' that indicate statistical separation.

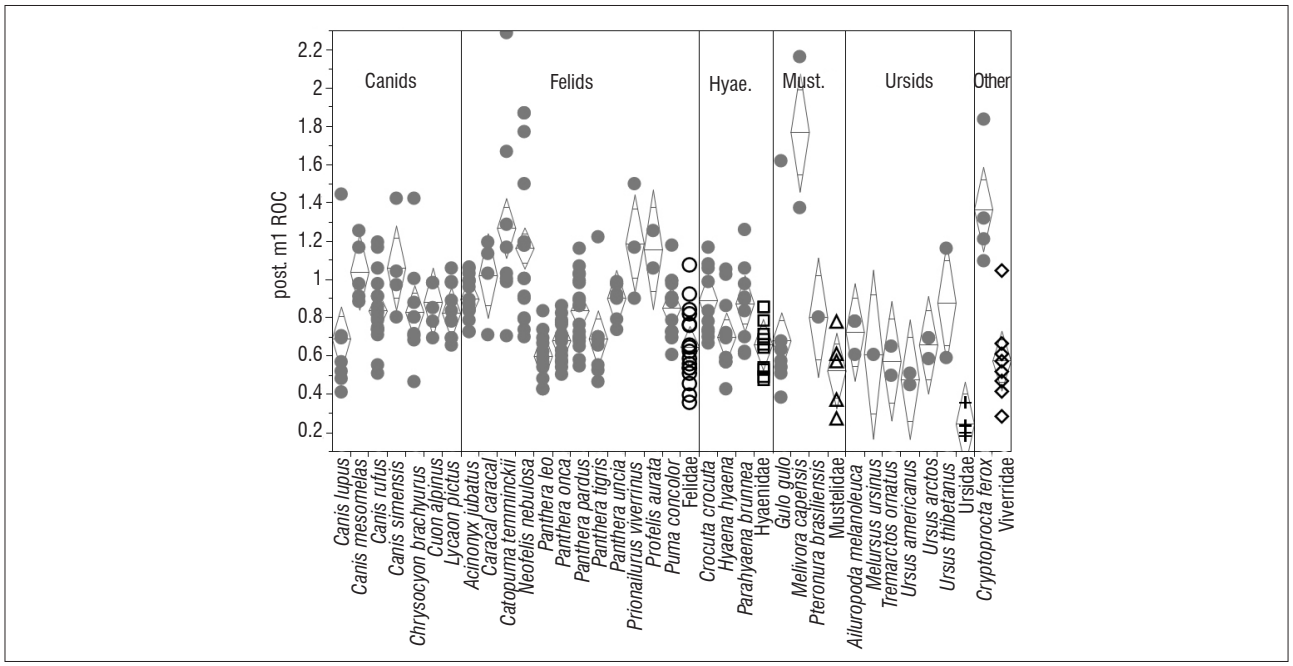
Figure 5: Radii-of-curvature (ROC) score of the fourth (p4) lower premolars by species.



*Extant taxa are represented by solid grey circles, Langebaanweg (LBW) felids by open circles, LBW hyaenids by open squares, LBW mustelids by open triangles, LBW ursids (*Agriotherium*) by '+' signs, and LBW viverrids by open diamonds.*

Each dotted diamond represents the mean and 95% confidence interval vertical span; the horizontal lines are the mean and 'overlap marks' that indicate statistical separation.

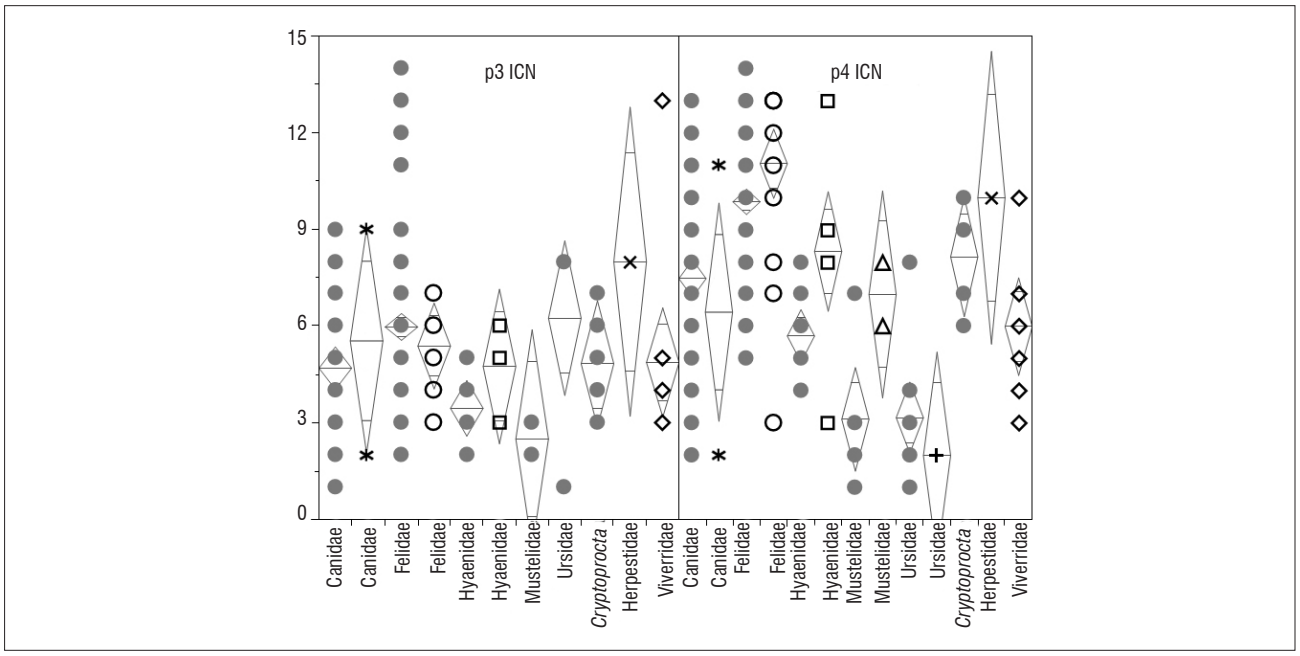
Figure 6: Radii-of-curvature (ROC) score of the anterior (paraconid) carnassial (m1) cusp by species.



*Extant taxa are represented by solid grey circles, Langebaanweg (LBW) felids by open circles, LBW hyaenids by open squares, LBW mustelids by open triangles, LBW ursids (*Agriotherium*) by '+' signs, and LBW viverrids by open diamonds.*

Each dotted diamond represents the mean and 95% confidence interval vertical span; the horizontal lines are the mean and 'overlap marks' that indicate statistical separation.

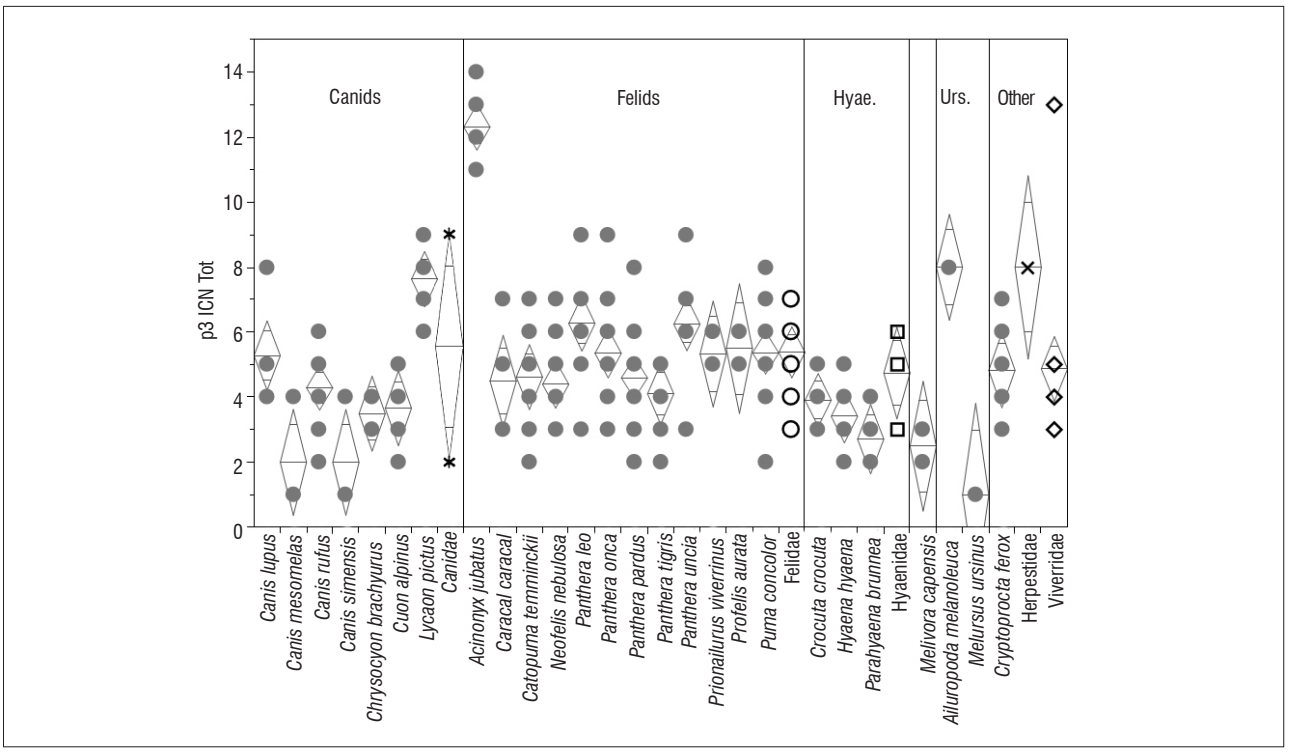
Figure 7: Radii-of-curvature (ROC) score of the posterior (protoconid) carnassial (m1) cusp by species.



*Extant taxa are represented by solid grey circles, Langebaanweg (LBW) canids are represented by asterisks, LBW felids by open circles, LBW hyaenids by open squares, LBW mustelids by open triangles, LBW ursids (*Agriotherium*) by '+' signs, the single analysed LBW herpestid by an 'x' sign, and LBW viverrids by open diamonds.*

Each dotted diamond represents the mean and 95% confidence interval vertical span; the horizontal lines are the mean and 'overlap marks' that indicate statistical separation.

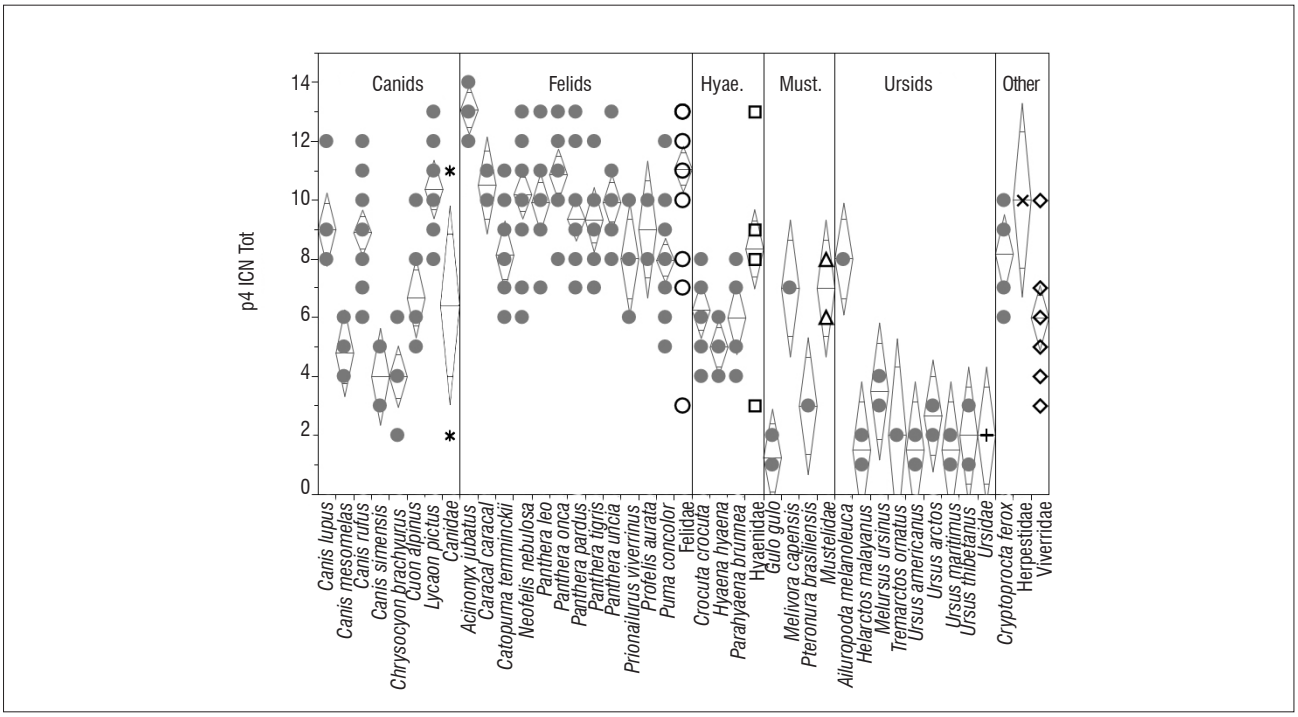
Figure 8: Intercuspid notch (ICN) scores of the third (p3) and fourth (p4) lower premolars by family.



*Extant taxa are represented by solid grey circles, Langebaanweg (LBW) canids are represented by asterisks, LBW felids by open circles, LBW hyaenids by open squares, LBW mustelids by open triangles, LBW ursids (*Agriotherium*) by '+' signs, the single analysed LBW herpestid by an 'x' sign, and LBW viverrids by open diamonds.*

Each dotted diamond represents the mean and 95% confidence interval vertical span; the horizontal lines are the mean and 'overlap marks' that indicate statistical separation.

Figure 9: Intercuspid notch (ICN) scores of the third (p3) lower premolar by species.



*Extant taxa are represented by solid grey circles, Langebaanweg (LBW) canids are represented by asterisks, LBW felids by open circles, LBW hyaenids by open squares, LBW mustelids by open triangles, LBW ursids (*Agriotherium*) by '+' signs, the single analysed LBW herpestid by an 'x' sign, and LBW viverrids by open diamonds.*

Each dotted diamond represents the mean and 95% confidence interval vertical span; the horizontal lines are the mean and 'overlap marks' that indicate statistical separation.

Figure 10: Intercuspid notch (ICN) scores of the fourth (p4) lower premolar by species.

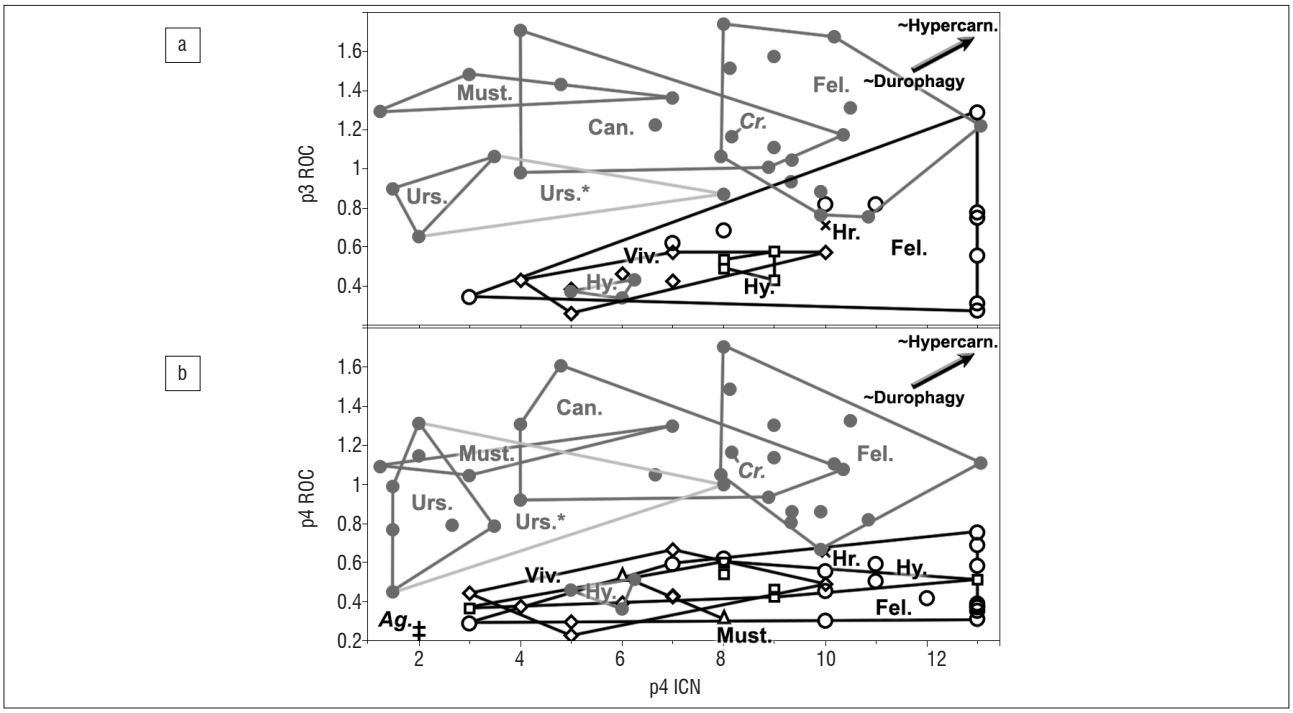
As has been found previously[1-3] (and discussed above), the p4 ICN scores are actually more telling (Figure 10). By this metric, the LBW felids have significantly higher scores than all but the most hypercarnivorous carnivorans (most of the modern felids, *Canis lupus* and *C. rufus*, *Lycaon* and *Cryptoprocta*). However, surprisingly (although as discussed previously[3]), the LBW hyaenids also have significantly higher p4 ICN scores than many modern carnivorans including their confamilial *Hyaena,* placing them statistically within the range of the more hypercarnivorous lineages (e.g. most of the felids) – and significantly more hypercarnivorous even than some of the canids, e.g. *Canis simensis* ($p=0.0378$) and *Chrysocyon* ($p=0.0007$) and all of the ursids other than the unique dentition of *Ailuropoda*. Among the mustelids, the LBW sample has statistically higher p4 ICN scores than those of *Gulo* ($p=0.0439$). As seen with the family-level analysis, the two LBW canids have very different p4 notch scores with the larger specimen plotting among the modern *Lycaon* and most hypercarnivorous felids, and the smaller specimen plotting below even the most durophagous hyaenids and sorting among *Gulo* and most of the bears (Figure 10). According to its p4 ICN score, *Agriotherium* falls well within the modern ursid with the exception of the highly derived *Ailuropoda* – a taxon with highly notched premolars perhaps as an adaptation for severing tough bamboo fibres.[5,6] Despite the small ursid sample size, the difference between the *Agriotherium* and *Ailuropoda* p4 ICN scores is nearly significant ($p=0.0502$) as it is between *Ailuropoda* and all of the other modern ursids.

Overall dental geometry morphospace

Comparing the LBW carnivorans using individual measures of ROC and ICN relative to whole samples of modern analogues yields mixed statistical results: by some measures some of the LBW lineages are statistically distinct from their modern analogues, and by some measures they are not. However, comparing the LBW specimens to the morphospace represented by the mean of individual modern species within the families, it is readily apparent that the Mio-Pliocene carnivorans from Langebaanweg had drastically different dental morphology (Figure 11).

Although Figure 11 compares the morphospace of species means of modern carnivorans to all analysable individuals of the fossil sample (i.e. an admittedly unequal comparison), for graphical purposes, it is informative. In this respect, with the notable exceptions of the hyaenids and felids, the LBW carnivores do not overlap with the mean clouds of the modern carnivorans. That is, the *Agriotherium* specimens, to the extent that this small sample could be analysed, fall outside of the morphospace of the averages of the modern ursid taxa. Likewise, the modern and fossil mustelids are also separate. In both cases, the separation is predominantly in terms of their premolar ROC – namely the fossils have less sharp fourth premolars than their modern confamilials. The same is true of the LBW herpestids and viverrids. Because the ROC scores of the LBW canids could not be evaluated, they are not on these plots. However, their respective notch scores of 2 and 9 for their third premolars and 2 and 11 for their fourth premolars are more divergent than the species averages within any of the modern families. In other words, had their premolars been preserved enough for evaluation, these two specimens represent an amount of variation more extreme than that found between average individuals of any two modern confamilials in our sample.

The notable difference between the LBW and modern families, the hyaenids, falls along the other scale: the fossil specimens have similar premolar sharpness (ROC) as their modern relatives, but their teeth generally have higher notch scores (ICN; Figure 11a). Although there is overlap in some measures of morphospace (as exemplified by Figure 11b), the ICN difference is notable. Namely, this measure documents the shift that occurred in the lineage: a transformation from the highly cusped feliform ancestral hyaenids to the most durophagous modern carnivorans – the latter representative of an ecological niche greatly facilitated by the notable reduction in premolar accessory cusps. The many caniforms that also have low ICN scores achieve that status – not by reducing the accessory cusps independently, but rather by having overall reduced premolar morphology. This reflects their greater reliance on post-carnassial dental elaboration – a condition not found in feliforms and essentially the opposite emphasis compared to that found in hyaenids.



*Extant taxa are represented by solid grey circles, Langebaanweg (LBW) canids are represented by asterisks, LBW felids by open circles, LBW hyaenids by open squares, LBW mustelids by open triangles, LBW ursids (*Agriotherium*) by '+' signs, the single analysed LBW herpestid by an 'x' sign, and LBW viverrids by open diamonds.*

Points for the modern families (represented in grey) are for species averages, while points for the LBW sample represent individual specimens as the fossil samples clearly include multiple taxa within each family.

Lines delineate minimum convex units for each family: Can., Canidae; Fel., Felidae; Hr., Herpestidae; Hy., Hyaenidae; Must., Mustelidae; Urs., Ursidae; Viv., Viverridae. The individual taxa Agriotherium *and* Cryptoprocta *are denoted 'Ag' and 'Cr.', respectively. 'Urs.*' denotes the Ursid morphospace if the unique* Ailuropoda *(connected by lighter lines) is included with its confamilials.*

Figure 11: Dental geometry morphospace of fourth premolar intercuspid notch (p4 ICN) relative to (a) third (p3 ROC) and (b) fourth premolar radii-of-curvature (p4 ROC). As discussed by Hartstone-Rose[1] and Hartstone-Rose and Stynder[3], modern taxa falling toward the top right of the plot tend toward hypercarnivory and taxa falling toward the bottom left of the plot tend toward durophagy with osteophages generally falling low on the ROC scale and omnivores (particularly those with greater dental heterogeneity) falling further to the left on the ICN scale.

Discussion

Our previous research has shown that recent carnivoran (i.e. Plio-Pleistocene and Pleisto-Holocene) fossil taxa generally sort within the morphospace of the modern congeners and confamilials.[1,2,57] However, the pattern seems to be slightly different in older carnivoran clades; for instance, although in one recent analysis two large Mio-Pliocene felids from Langebaanweg fell at the most hypercarnivorous end of the modern felid range, a sample of sympatric hyaenids exhibited significantly higher ICN scores than their modern confamilials.[3] As would be expected, detailed analysis of these dental geometries documented the state of the teeth of the ancestors in this lineage before the highly derived modern taxa reached their most durophagous state among modern carnivores. However, questions remained about the dietary patterns of the other enigmatic and diverse members of the LBW carnivore guild – the subject of our current study.

As seen in the previous study of the LBW hyaenids,[3] the other LBW carnivorans are very different from their modern analogues. In fact, in some respects, the LBW hyaenids are more like modern hyaenids in their tooth morphology than are the other LWB carnivorans with their modern relatives; that is, in terms of ROC, the LBW hyaenids are indistinguishable from their modern confamilials, and they occupy overlapping morphospace while all of the other LBW carnivorans occupy fairly distinct morphospace (Figure 11). In this respect, even the LBW felids appear quite different from their modern confamilials and the larger *Dinofelis*-like LBW felids included in the previous analysis[3] are more similar to the modern felids than are the other LBW felids that tend to have significantly lower premolar ROC scores and occupy a much wider range of variation in ICN scores than their modern counterparts.

The other LWB carnivorans are harder to compare to a modern sample because they are even more morphologically and phylogenetically distinct than either the LBW felids or hyaenids. Namely, the diverse sample of LBW viverrids, which includes some exceptionally large specimens, occupies morphospace (and likely ecology) that is unoccupied by any modern viverrid. Although the LBW mustelids are not as richly represented in the fossil record, their morphology is also impressive in its extreme divergence from the modern confamilials with giant forms significantly exceeding the range of modern mustelids. Likewise, the two LBW canids seem to represent extreme dietary divergence. Whether *Agriotherium* is an ursid or hemicyonid it is unlike any modern bear or dog or any other carnivore; like other bears, it has relatively low ICN scores but exceptionally low premolar ROC, which would seem to indicate an extreme adaptation for durophagy, even more so than the modern hyaenids with the exception of its anterior carnassial cusp (the paraconid) ROC score which makes it significantly sharper than not only the modern ursids, but all of the modern carnivoran families including the hypercarnivorous felids. Thus, in some regards its dental morphology appears optimally adapted for extreme durophagy while in other ways the taxon exhibits just the opposite. In the final consideration, this large, impressive carnivoran was likely able to dominate the entire carnivore guild, utilising the whole carcass of its prey possibly more efficiently than any modern carnivoran.

There is still much to be analysed in terms of the dietary niche partitioning of the Mio-Pliocene carnivores from Langebaanweg. For instance, the LBW carnivoran assemblage includes many small specimens, particularly those specimens ascribed to Herpestidae as well as small Viverridae, which are unfortunately beyond the comparative scope of

our modern sample. However, the small carnivoran sample from LBW is apparently both taxonomically and morphologically diverse and is worthy of deeper evaluation both in terms of its systematics and ecodiversity. It is our hope that an extension of the ROC and ICN methodology, perhaps with a refinement of its approach to geometric analyses, and with an expansion of the modern comparative sample, will be conducted to evaluate the dietary specialisations of small fossil carnivorans.

Another constraint of this study is that the current analyses are limited in focus by the ICN and ROC metrics. While these have been proven to be highly informative indicators of dietary specialisation, they can only be assessed for specimens that have well-preserved lower premolar and carnassial morphology. But the LBW sample contains elements beyond these teeth that could aid in the dietary reconstruction of its carnivorans. Namely, there are numerous well-preserved maxillary teeth and remarkable crania from most of the represented families. These elements open up possibilities for analyses based on cranial biomechanics and masticatory muscle reconstruction – potentially in three dimensions.

In conclusion, the Mio-Pliocene carnivoran guild from Langebaanweg, South Africa, is clearly impressive both in its systematic and ecological diversity. Although it consists of only a few hundred, mostly fragmentary and isolated specimens, some of the fossils are truly astonishing in terms of their preserved completeness, including postcranial elements beyond the scope of this paper, as well as magnificent whole crania preserved or fully reconstructed for several of the enigmatic species. This unique sample, for its place and time, gives us an unparalleled window into a remarkably complete carnivore guild allowing us to examine subtle niche partitioning between sympatric taxa at a level that exists for few other older palaeontological locations in the world.

Acknowledgements

We thank the staff of Iziko Museums of South Africa and the West Coast Fossil Park, particularly Romala Govender, Graham Avery and Pippa Haarhoff for granting us access to the collections, providing us with valuable feedback on some of our ideas and being wonderful and welcoming hosts. We also thank Deano Stynder for providing valuable discussion and conversation about this research, Francis Thackeray for providing editorial advice, an anonymous reviewer for providing valuable recommendations, and the University of South Carolina for funding this research, particularly the students involved.

Authors' contributions

A.H-R. is the principal investigator and mentor to the other authors (who are his students); he conceptualised the project, taught the students the methods and oversaw all aspects of the data collection; he conducted the analyses and wrote the results and discussion sections and rewrote all other parts of the paper; and he created all of the figures. K.N.B. and K.D.D. performed the majority of data collection. K.N.B. wrote most of the first draft of the introduction and methods. C.L.L. wrote the majority of the introduction and methods as they stand in the current version and undertook the most substantial revision to the manuscript other than the contribution of A.H-R. K.D.D. also contributed to the background as well as editing of the manuscript.

References

1. Hartstone-Rose A, Wahl S. Using radii-of-curvature for the reconstruction of extinct South African carnivoran masticatory behavior. C R Palevol. 2008;7(8):629–643. http://dx.doi.org/10.1016/j.crpv.2008.09.015
2. Hartstone-Rose A. Reconstructing the diets of extinct South African carnivorans from premolar 'intercuspid notch' morphology. J Zool (Lond). 2011;285:119–27. http://dx.doi.org/10.1111/j.1469-7998.2011.00821.x
3. Hartstone-Rose A, Stynder D. Hypercarnivory, durophagy or generalised carnivory in the Mio-Pliocene hyaenids of South Africa? S Afr J Sci. 2013;109(5/6):77–87. http://dx.doi.org/10.1590/sajs.2013/20120040
4. Van Valkenburgh B. Iterative evolution of hypercarnivory in canids (Mammalia: Carnivora): Evolutionary interactions among sympatric predators. Paleobiology. 1991;17:340–362.
5. Hendey QB. Langebaanweg: A record of past life. Cape Town: South African Museum; 1982.
6. Roberts DL, Matthews T, Herries AIR, Boulter C, Scott L, Dondo C, et al. Regional and Global Palaeoenvironmental and Sea Level Context of the Late Cenozoic Langebaanweg (LBW) Palaeontological Site: West Coast of South Africa. Earth-Sci Rev. 2011;106:191–214. http://dx.doi.org/10.1016/j.earscirev.2011.02.002
7. Hendey QB. Palaeoecology of the late tertiary fossil occurrences in 'E' Quarry, Langebaanweg, South Africa and a reinterpretation of their geological context. Ann S Afr Mus. 1981;84:1–104.
8. Hendey QB. Quaternary vertebrate fossil sites in the south-western Cape Province. S Afr Archaeol Bull. 1969:96–105. http://dx.doi.org/10.2307/3888285
9. Hendey QB. A review of the geology and palaeontology of the Plio/Pleistocene deposits at Langebaanweg, Cape Province. Ann S Afr Mus. 1970;56:75–117.
10. Hendey QB. Further observations on the age of the mammalian fauna from Langebaanweg, Cape Province. Paleoecology. 1972;6:172–175.
11. Hendey QB. A Pliocene ursid from South Africa. Ann S Afr Mus. 1972;59(6):115–132.
12. Hendey QB, Repenning CA. A Pliocene phocid from South Africa. Ann S Afr Mus. 1972;59:71–98.
13. Hendey QB. Fossil occurrences at Langebaanweg, Cape Province. Nature. 1973;244:13–14. http://dx.doi.org/10.1038/244013a0
14. Hendey QB. New fossil carnivores from the Swartkrans australopithecine site (Mammalia, Carnivora). Ann Transv Mus. 1974;29(3):27–48.
15. Hendey QB. The late Cenozoic Carnivora of the south-western Cape Province. Ann S Afr Mus. 1974;63:1–369.
16. Hendey QB. Faunal dating of the late Cenozoic of southern Africa, with special reference to the Carnivora. Quaternary Res. 1974;4(2):149–161. http://dx.doi.org/10.1016/0033-5894(74)90004-0
17. Hendey QB. The Pliocene fossil occurrences in' E' Quarry, Langebaanweg, South Africa. Cape Town: South African Museum; 1976.
18. Hendey QB. Late Tertiary Hyaenidae from Langebaanweg, South Africa, and their relevance to the phylogeny of the family. Ann S Afr Mus. 1978;76:265–297.
19. Hendey QB. Late Tertiary Mustelidae (Mammalia, Carnivora) from Langebaanweg, South Africa. Ann S Afr Mus. 1978;76:329–357.
20. Hendey QB. *Agriotherium* (Mammalia, Ursidae) from Langebaanweg, South Africa, and relationships of the genus. Ann S Afr Mus. 1980;81:1–109.
21. Hendey QB. Southern African Late Tertiary vertebrates. In: Klein RG, editor. Southern African prehistory and palaeoenvironments. Rotterdam: Balkema; 1984. p. 81–106.
22. Hunt R Jr. Basicranial anatomy of the giant viverrid from 'E' Quarry, Langebaanweg, South Africa. In: Stewart KM, Seymour KL, Churcher CS, editors. Palaeoecology and palaeoenvironments of Late Cenozoic mammals: Tributes to the career of CS Churcher. Toronto: University of Toronto Press; 1996. p. 588–597.
23. Kurtén B. The Neogene wolverine *Plesiogulo* and the origin of *Gulo* (Carnivora, Mammalia). Helsinki: Societas pro Fauna et Flora Fennica; 1970.
24. Peigné S, De Bonis L, Likius A, Mackaye HT, Vignaud P, Brunet M. The earliest modern mongoose (Carnivora, Herpestidae) from Africa (late Miocene of Chad). Naturwissenschaften. 2005;92(6):287–292. http://dx.doi.org/10.1007/s00114-005-0626-0
25. Peigné S, De Bonis L, Mackaye HT, Likius A, Vignaud P, Brunet M. Late Miocene Carnivora from Chad: Herpestidae, Viverridae and small-sized Felidae. C R Palevol. 2008;7(8):499–527. http://dx.doi.org/10.1016/j.crpv.2008.10.002
26. Howell FC, García N. Carnivora (Mammalia) from Lemudong'o (Late Miocene: Narok District, Kenya). Kirtlandia. 2007;56:121–139.
27. Werdelin L, Sanders WJ. Carnivora. In: Werdelin L, editor. Cenozoic mammals of Africa. Berkeley, CA: University of California Press; 2010. p. 609–663. http://dx.doi.org/10.1525/california/9780520257214.003.0032

28. Samuels JX, Meachen-Samuels JA, Gensler PA. The first mid-Blancan occurrence of *Agriotherium* (Ursidae) in North America: A record from Hagerman fossil beds National Monument, Idaho. J Paleontol. 2009;83(4):597–603. http://dx.doi.org/10.1666/08-112R.1

29. Sorkin B. Ecomorphology of the giant short-faced bears *Agriotherium* and *Arctodus*. Hist Biol. 2006;18(1):1–20. http://dx.doi.org/10.1080/08912960500476366

30. Oldfield C, McHenry C, Clausen P, Chamoli U, Parr W, Stynder D, et al. Finite element analysis of ursid cranial mechanics and the prediction of feeding behaviour in the extinct giant *Agriotherium africanum*. J Zool (Lond). 2012;286:93–171. http://dx.doi.org/10.1111/j.1469-7998.2011.00862.x

31. Stynder DD, Kupczik K. Tooth root morphology in the early Pliocene African bear *Agriotherium africanum* (Mammalia, Carnívora, Ursidae) and its implications for feeding ecology. J Mammal Evol. 2013;20(3):227–237. http://dx.doi.org/10.1007/s10914-012-9218-x

32. Miller WE, Carranza-Castaóeda O. *Agriotherium schneideri* from the Hemphillian of Central Mexico. J Mammal. 1996;77(2):568–577. http://dx.doi.org/10.2307/1382830

33. Pickford M, Soria D. Carnivores from the late Miocene and basal Pliocene of the Tugen Hills, Kenya. Revista de la Sociedad Geológica de España. 2005;18(1):39–61.

34. Montoya P, Morales J, Abella J. *Eucyon debonisi* n. sp., a new Canidae (Mammalia, Carnivora) from the latest Miocene of Venta del Moro (Valencia, Spain). Geodiversitas. 2009;31(4):709–722. http://dx.doi.org/10.5252/g2009n4a709

35. Lacruz R, Turner A, Berger LR. New *Dinofelis* (Carnivora: Machairodontinae) remains from Sterkfontein Valley sites and a taxonomic revision of the genus in southern Africa. Ann Transv Mus. 2006;43:89–106.

36. Turner A. *Megantereon cultridens* (Cuvier) (Mammalia, Felidae, Machairodontinae) from Plio-Pleistocene deposits in Africa and Eurasia, with comments on dispersal and the possibility of a New World origin. J Paleon. 1987;61:1256–1268.

37. Turner A, Antón M. The big cats and their fossil relatives. New York: Columbia University Press; 1997.

38. Werdelin L, Lewis ME. A revision of the genus *Dinofelis* (Mammalia, Felidae). Zool J Linn Soc. 2001;132(2):147–258. http://dx.doi.org/10.1111/j.1096-3642.2001.tb02465.x

39. Hemmer H. Zur nomenklatur und verbreitung des genus *Dinofelis* Zdansky, 1924 (Therailurus Priviteau, 1948) [For nomenclature and spread of the genus *Dinofelis* Zdansky, 1924 (Therailurus Priviteau, 1948)]. Palaeontol Afr. 1965;9:75–89. German.

40. Hemmer H. Neue Befunde zur Verbreitung und Evolution der pliozan-pleistozanen Gattung *Dinofelis* (Mammalia, Carnivora, Felidae) [New findings on the distribution and evolution of the Pliocene-Pleistocene genus *Dinofelis* (Mammalia, Carnivora, Felidae)]. Neues Jb Geol Palaeont Mh. 1973;1973(3):157–169. German.

41. Werdelin L, Solounias N. The Hyaenidae: Taxonomy, systematics, and evolution. Fossils Strata. 1991;30:1–104.

42. Stynder DD, Ungar PS, Scott JR, Schubert BW. A dental microwear texture analysis of the Langebaanweg (South African Mio-Pliocene) fossil Hyaenidae. Acta Palaeontol Pol. 2012;57(3):485–496. http://dx.doi.org/10.4202/app.2011.0053

43. Cruz-Uribe K. Distinguishing hyena from hominid bone accumulations. J Field Archaeol. 1991;18(4):467–486. http://dx.doi.org/10.2307/530408

44. Werdelin L, Turner A, Solounias N. Studies of fossil hyaenids: The genera *Hyaenictis* Gaudry and *Chasmaporthetes* Hay, with a reconsideration of the Hyaenidae of Langebaanweg, South Africa. Zool J Linn Soc. 1994;111(3):197–217. http://dx.doi.org/10.1006/zjls.1994.1022

45. Tseng ZJ, Stynder D. Mosaic functionality in a transitional ecomorphology: Skull biomechanics in stem Hyaeninae compared to modern South African carnivorans. Biol J Linn Soc. 2011;102:540–559. http://dx.doi.org/10.1111/j.1095-8312.2010.01602.x

46. Stynder DD. Tooth crown form as an indicator of niche partitioning among Late Miocene/Early Pliocene hyenas from 'E' Quarry, Langebaanweg, South Africa. Paleogeogr Paleoclimatol Paleoecol. 2009;283(3–4):148–159. http://dx.doi.org/10.1016/j.palaeo.2009.09.017

47. Kretzoi M. Die Raubtiere von Gombaszög nebst einer Übersicht der Gesamtfauna [The predators of Gombaszög along with an overview of the total fauna]. Hist Nat Mus NationHungarici, Pars Miner Geol Paleont. 1938;31:88–157. German.

48. Morales J, Pickford M. Giant bunodont Lutrinae from the Mio-Pliocene of Kenya and Uganda. Estudios Geologicos (Madrid). 2005;61(3–6):233–246.

49. Pasitschniak-Arts M, Larivière S. *Gulo gulo*. Mamm Species. 1995;499:1–10.

50. Werdelin L. The position of Langebaanweg in the evolution of Carnivora in Africa. Afr Nat History. 2006;2:201–202.

51. Kay RF. Molar structure and diet in extant Cercopithecidae. In: Butler PM, Joysey KA, editors. Development, function, and evolution of teeth. New York: Academic Press; 1978. p. 309–340.

52. Kay RF, Simons EL. The ecology of Oligocene African Anthropoidea. Int J Primatol. 1980;1:21–37. http://dx.doi.org/10.1007/BF02692256

53. Kirk EC, Simons EL. Diets of fossil primates from the Fayum Depression of Egypt: A quantitative analysis of molar shearing. J Hum Evol. 2001;40(3):203–229. http://dx.doi.org/10.1006/jhev.2000.0450

54. Van Valkenburgh B. Trophic diversity in past and present guilds of large predatory mammals. Paleobiology. 1988;14:155–173.

55. Walton LR, Joly DO. *Canis mesomelas*. Mamm Species. 2003;715:1–9. http://dx.doi.org/10.1644/715

56. Hartstone-Rose A. Evaluating the hominin scavenging niche through analysis of the carcass-processing abilities of the carnivore guild [MSc dissertation]. Durham, NC: Duke University; 2008.

57. Simpler EM, Perry JMG, Hartstone-Rose A. Determining diet from the crossectional shape and intercuspid notches of the teeth of the carnivores of Rancho La Brea. J Vertebr Paleontol. 2011;31(Suppl. 3):195A.

Reptiles sold as traditional medicine in Xipamanine and Xiquelene Markets (Maputo, Mozambique)

AUTHORS:
Vivienne L. Williams[1]
Thibedi J. Moshoeu[1]
Graham J. Alexander[1]

AFFILIATION:
[1]School of Animal, Plant & Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Vivienne Williams

EMAIL:
vivwill@netdial.co.za

POSTAL ADDRESS:
School of Animal, Plant & Environmental Sciences, University of the Witwatersrand, Private Bag 3, Wits 2050, South Africa

DATES:
Received: 03 Nov. 2015
Revised: 01 Mar. 2016
Accepted: 08 Mar. 2016

KEYWORDS:
Crocodylus; ethnoherpetology; python; wildlife trade; zootherapy

HOW TO CITE:
Williams VL, Moshoeu TJ, Alexander GJ. Reptiles sold as traditional medicine in Xipamanine and Xiquelene Markets (Maputo, Mozambique). S Afr J Sci. 2016;112(7/8), Art. #2015-0416, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20150416

Zootherapy plays a role in healing practices in Mozambican society. Although several studies have focused on ethnobotany and traditional medicine in the country, little research has been conducted on the use of reptiles in zootherapy. The aim of this study was therefore to fill this gap by assessing the reptile species traded for traditional medicine in the Xipamanine and Xiquelene Markets in Maputo, Mozambique. We found that few reptile species are traded domestically for traditional medicine and that their use appears to be in decline in Mozambique. Our findings also suggest that the domestic trade of reptiles for traditional medicines in Maputo markets is unlikely to have a significant impact on the conservation of reptiles in Mozambique. However, we suggest that international trade with South Africa is likely having a larger impact, given observations of Mozambican nationals selling a diverse range of fauna in urban traditional medicine markets in Johannesburg and Durban.

Introduction

The use of animal products for healing purposes is an ancient practice complementary to the body of knowledge on plant-based medicines[1], and trade in wildlife products for these purposes, especially in parts of Asia and Africa, is increasing[2]. Despite the rise of zootherapeutic studies to address a paucity of information on the subject of traditional animal therapies, studies have rarely addressed the use of reptiles by African communities.[2,3] These interrelationships between humans and herpetofauna are referred to as 'ethnoherpetology'.[4]

Reptiles, especially crocodiles and pythons, are typically present and frequently sought after in African traditional medicine markets.[2,3,5-11] Focused ethnoherpetological research and quantitative studies have, however, been largely overshadowed by the generalised ethnozoological studies that document and inventory multiple vertebrate classes.[2] This broad approach may signify a bias towards investigating aesthetically pleasing or charismatic species such as birds and mammals, or fauna that are of conservation concern. What is more, the scarcity of ethnoherpetological records contributes to the traditional importance of herpetofauna (and the degree to which they are exploited) being overlooked and underestimated.[12,13]

Decades of civil war in Mozambique – and the resulting impoverishment – has resulted in reduced public access to conventional Western medicines and, as a consequence, the healthcare system is dominated by a reliance on traditional medicines.[13,14] Consequently, most Mozambicans' first encounter with healthcare is allegedly through a network of traditional medical practitioners.[15] While plants are the primary source of traditional remedies in the country[16] and dominate the products sold by traders in the markets (e.g. Krog et al.[7] calculated that traders on average sell 2 ± 3.5 different animal products compared to 27 ± 12.5 different plant products), animals used in zootherapeutic preparations also play a role[5]. However, there is a shortage of accessible and/or published material on the subject of Mozambican zootherapeutics, particularly on the nature and dynamics of the reptile trade. Studies that have included some information on the sale of herpetofauna in Mozambican markets include reports by Chauqúe[5], Krog et al.[7], Marshall[13] and Cunningham and Zondi[17]. Thus the aim of our study was to document the reptile taxa in urban markets selling traditional medicine in Maputo, and to consider the implications that this trade may have for reptile conservation in Mozambique.

Methods

Market surveys were carried out in the traditional medicine section of informal markets in the capital city of Maputo, Mozambique, in March 2015 (Xipamanine Market) and January 2016 (Xipamanine, Xiquelene and Adelino Markets). Ethics clearance to conduct the surveys was granted by the University of the Witwatersrand Human Non-medical Ethics Screening Committee (protocol number H14/06/02). Permission to conduct the market interviews was granted by the local heads of the traditional healers association (Associação dos Médicos Tradicionais de Moçambique, or AMETRAMO), and the association for retailers of traditional medicines of Mozambique (Associação des Vendedores de Medicamentos Tradicionais de Moçambique, or AVEMETRAMO). These organisations represent the collective interests of Mozambique's traditional healers and vendors in the Xipamanine and Xiquelene Markets, respectively.

Market identification

There are three informal markets selling traditional medicine in Maputo (Table 1). The biggest, Xipamanine, is probably also the largest market for medicinal plants in Mozambique[7]; traders here sell a variety of products in distinct sections of the market, including traditional medicine, meat, livestock, clothing, crafts and raw materials such as charcoal. Xiquelene (or Xikalene) is the second largest market in Maputo, but it has less than half the number of traditional medicine traders as Xipamanine (Table 1). This market is also arranged according to the type of product sold, with traditional medicine vendors clustered together. Adelino is the third informal market in Maputo, but it is minor in size compared with the other two and sells mainly textiles and charcoal. Only Xipamanine and Xiquelene Markets sell animal products for traditional medicine (Table 1).

Table 1: Number of stalls selling traditional medicines in three Maputo markets in January 2016

Market	Stalls registered to sell traditional medicine	Stalls openly selling animals as traditional medicine	Stalls openly selling reptiles as traditional medicine (% of animal traders with reptiles)[a]	Male traders selling reptiles
Xipamanine	150[b]	>36	9 (<25%)	6
Xiquelene	69[c]	36	5 (14%)	3
Adelino	2	0	0	0
Total	**221**	**>72**	**14 (<19%)**	**9**

[a]*These stalls were sampled in January 2016.*

[b]*Information supplied by AMETRAMO. Traders allegedly all have permits to sell plants and animals in the market.*

[c]*Information supplied by AVEMETRAMO.*

There are no significant markets for traditional medicines in Maputo Province outside the city of Maputo, including rural markets (Falcâo MP 2015, written communication, December 22). There are anecdotal reports of a rural traditional medicine market operating once a week near Ponta do Ouro (a coastal town about 130 km south of Maputo and approximately 15 km from the South African border), which also attracts South African traditional healers. However, there was no evidence for its existence when the region was visited in January 2016 (Falcâo MP 2016, oral communication, January 20). Given the notable cross-border trade of ethnomedicinal resources between Mozambique and South Africa (Williams VL, personal observation), it is plausible that such a place where these resources can be occasionally traded exists; however, further investigations are warranted to assess whether these anecdotes are exaggerated.

Based on the available market information, we undertook our first market survey in Xipamanine in March 2015 with seven traders. We had intended to conduct a comparative study in South Africa in 2015 (specifically in Faraday Market in Johannesburg and Warwick Market in Durban), but the Chairpersons of these markets denied permission for the study. Hence we limited the trade study to Maputo and followed up with a second survey conducted there in January 2016, this time in two of the three markets and including all 14 traders that sold reptiles as traditional medicine in the study (Table 1).

First market survey

A semi-quantitative questionnaire-based market survey was carried out in Xipamanine over 2 days in March 2015. Permission to speak to the traders was granted beforehand, and thereafter a discreet pre-survey assessment was conducted (initially without the AMETRAMO head) in order to become familiar with the market and to identify which traders were openly selling animal products. Four reptile traders were counted on the first pre-survey visit. When accompanied by AMETRAMO on the second pre-survey visit, a further three traders were identified (all by AMETRAMO); these traders stored animals out of sight and only reluctantly showed them to us on request of AMETRAMO. Whether there were more reptile traders in the market not known to the AMETRAMO head could not be established. After the pre-survey visits, and before the interviews commenced, the research objectives of the study were explained to the seven known reptile traders to obtain their consent to participate.

An interpreter from Eduardo Mondlane University (Maputo), trained in interviewing techniques, assisted with the interviews. A semi-structured questionnaire (see Appendix 1 in the supplementary material) was used that was verbally translated into Portuguese and Xitsonga – the most commonly spoken languages in Maputo – by the interpreter during the interviews. The questionnaire consisted of three parts: (1) basic trader information (sex; whether traders were traditional healers or not; a participant code for anonymity); (2) market information (species preferences by traders and customers; how reptiles reach the market); and (3) species information (taxon sold; common names; parts sold; geographical harvesting origin; availability; uses; etc.) (see Appendix 1 in the supplementary material). The information was supplemented by informal discussions. All reptiles, except those that were easily identified, were photographed where consent was given and later identified to species level using the taxonomic arrangement of Bates et al.[18] and Uetz and Hošek[19].

Second market survey

A second survey was carried out in Xipamanine and Xiquelene Markets over 3 days in January 2016. Adelino Market was also considered, but none of the traders sold animal parts (Table 1). The intention had been to repeat the 2015 survey using a slightly extended questionnaire (see Appendix 2 of the supplementary material) that included questions added to clarify matters that had arisen during the first survey. As with the first survey, permission to interview traders was first sought from AMETRAMO in Xipamanine. Permission was also required from AVEMETRAMO to conduct the survey in Xiquelene. Although permission to conduct the study was granted, AMETRAMO and AVEMETRAMO requested that certain questions be shortened and/or omitted from the questionnaire; they also discouraged photographs and lengthy interviews. Hence, we were unable to replicate all of the first survey as certain questions had to be omitted, but we were able to include new informal questions on the pricing of body parts and cross-border trade and to record whether reptile fat was sold. Professor MP Falcâo from Eduardo Mondlane University, who has conducted ethnomedicinal surveys in Maputo markets before, assisted us by liaising with the associations and traders, and translating the interviews and discussions.

While we were able to conduct the questionnaire-based survey in Xiquelene with some success, the traders in Xipamanine were hostile and uncooperative on this occasion (despite mediation by AMETRAMO). Consequently, the questionnaire could not be used here and discreet mental and shorthand notes were made instead. The head of AMETRAMO also became a surrogate informant by supplying some answers to key questions. The resurvey of Xipamanine thus focused on recording species, the prices of body parts, and from where the animals had originated. An added challenge, however, was that many live animals were kept concealed – something AMETRAMO had warned us of beforehand. But whereas AMETRAMO intervened in the 2015 survey to coax traders into revealing hidden animals (albeit reluctantly), no such intervention was attempted by them in 2016. Accordingly, we do not consider the species inventory to be complete for the 2016 survey.

As traders in the markets were not traditional healers (and accordingly had limited knowledge about reptile zootherapeutics), we also attempted to interview suburban Mozambican traditional healers on the importance of reptiles to traditional healing. AMETRAMO recommended that we should first contact the Mozambican Ministry of Health for permission to conduct interviews, which we duly did. The representative of the Ministry agreed to assist, but imposed several last-minute administrative hurdles. Unfortunately, timeframes prohibited the completion of this part of the study.

Results and discussion

Reptile traders

There were at least 30 traders selling animal parts in Xipamanine Market in March 2015, but only 7 (all men) were identified as selling reptiles. In January 2016, 36 traders in Xiquelene and >36 traders in Xipamanine were recorded with animal parts; of these, a total of 14 vendors (64.3% men) had reptiles visible at their stalls and all of them were interviewed (Table 1). The sampled reptile traders thus accounted for 100% of all known reptile traders at the market, but <19% of all animal traders. The proportion of traders selling reptiles was unexpectedly low, but the extent of the covert trade (characterised by vendors concealing live animals) was difficult to gauge.

All market traders interviewed worked as permanent traders and lived in Maputo. While none of the respondents was a traditional healer, it appears that most of their customers for reptile parts were healers. None of the traders specialised in selling reptiles (mammals were the preferred vertebrates on sale), and medicinal plants dominated every stall. Xipamanine and Xiquelene are diversified markets with most men selling medicinal wildlife products and most women selling butchered domestic meat and clothes. The dominance of men selling traditional medicines corresponds with the findings for Xipamanine Market by Marshall[13] and Chauque[5] (100% and 95% male traders, respectively), and for traders in three Maputo markets by Krog et al.[7] (86% male traders).

Reptiles sold

In total, 10 reptile species belonging to eight families were observed for sale in the markets in 2015/2016 (Table 2; Figure 1). This figure is higher than the number of reptile taxa documented in previous studies for markets in Maputo.[4,13] The three most frequently recorded taxa were *Python natalensis* (southern African python), *Varanus* spp. (monitor lizards, *V. niloticus* and *V. albigularis*), and *Kinixys* sp. (hinge-back tortoise, most likely *K. zombensis* and/or *K. spekii*) (Table 2). Other species recorded in the markets more than once were *Crocodylus niloticus* (Nile crocodile), *Chamaeleo dilepis dilepis* (flap-necked chameleon) and *Broadleysaurus major* (rough-scaled plated lizard). One trader allegedly sold *Dendroaspis* sp. (mamba) fat.

Table 2: Reptile taxa, common names and observed frequencies in two surveys conducted in two markets

ORDER Family *Species*	English name	Common names (P=Portuguese; T=Xitsonga)	Respondent frequency in market surveys			
			Xipamanine 1 (*n*=7, Mar. 2015)	**Xipamanine 2** (*n*=9, Jan. 2016)	**Xiquelene** (*n*=5, Jan. 2016)	**Total** (*n*=21)
CROCODILIA						
Crocodylidae						
Crocodylus niloticus	Nile crocodile	Crocodilo (P); Nguenha (T)	2	1	0	3
SQUAMATA						
Pythonidae						
Python natalensis	Southern African python	Gibóia (P); Nhlaru (T)	1	4	4	9
Elapidae						
Dendroaspis sp.[a]	Mamba	Not known	0	0	1	1
Viperidae						
Bitis gabonica	Gaboon adder	Bululu (T)	1	0	0	1
Gerrhosauridae						
Broadleysaurus major	Rough-scaled plated lizard	Makokorombane (T)	1	1	1	3
Varanidae						
Varanus niloticus	Nile monitor	Nkwahle (T)	1	3	0	4
Varanus albigularis	Rock monitor	Nkwahle (T)	0	1	0	1
Chamaeleonidae						
Chamaeleo dilepis dilepis	Flap-necked chameleon	Lompfanhe (T)	2	1	0	3
TESTUDINES						
Testudinidae						
Kinixys sp.[b]	Hinge-back tortoise	Cágado (P); Futsu (T); Nfutso (T); Chibodze (T)	2	2	1	5
Stigmochelys pardalis	Leopard tortoise	Cágado (P); Futsu (T)	0	1	0	1

[a]*Allegedly mamba, but the fat was sold and thus could not be authenticated*

[b]*Includes* K. zombensis *and/or* K. spekii

Figure 1: Reptiles recorded in the 2015/2016 Xipamanine and Xiquelene Market surveys. (a) Live *Kinixys* sp.; (b) *Stigmochelys pardalis* carapaces; (c) concealed *Crocodylus niloticus* skin; (d) *Python natalensis* skin and vertebrae (rolled up); (e) handful of *Python natalensis* vertebrae; (f) allegedly *Python natalensis* fat; (g) whole *Broadleysaurus major*; (h) live *Chamaeleo dilepis dilepis* in a plastic bottle; (i) *Varanus niloticus* in a container; (j) whole *Varanus niloticus* skin; (k) whole *Varanus albigularis* skin.

Our species count and the number of animals or body parts (Table 3) is likely to be an underestimation of actual numbers because live animals were generally concealed from view and were sometimes only shown on request of the accompanying AMETRAMO representative. The willingness of traders to disclose their stock in the 2016 survey was markedly reduced compared to 2015. Chauqúe[5] similarly noted that some animals are stored out of sight in Xipamanine and are only shown to potential customers. Our role as researchers would thereby have obviated any incentive for vendors to show us the species for sale.

Previous studies in Maputo recorded *P. natalensis* (identified as *Boa constrictor*[5], but likely to be python because boas are not indigenous to Africa) and *C. niloticus* as the reptiles most commonly sold in the markets.[5,13] These studies also indicated the sporadic presence of *Chelonia mydas* (green turtle) and *Eretmochelys imbricata* (hawksbill turtle)[5] and the spiny lizard *Smaug warreni warreni* (Warren's dragon lizard)[13]. Krog et al.[7] also noted the presence of live chameleons and unidentified reptile skins. We recorded two taxa that were either always sold live, namely chameleons (Figure 1h), or generally sold live, namely tortoises (live tortoises were recorded in 2015, but only carapaces were observed in 2016) (Table 3). For all other species, usually the skins were sold (especially those of python, crocodile and monitor lizards) (Table 3). Python vertebrae were on display in 2016, but only in Xiquelene Market.

Bottled animal fats were sold at many of the stalls (ranging from 5 to 15 bottles per trader). We did not inventory reptile fat during the 2015 survey, but one respondent indicated that he sold 'quite a bit' of crocodile and python fat. During the 2016 survey, we recorded four traders willing to acknowledge that they had python fat (Figure 1f), one with crocodile fat, and another allegedly with mamba fat (Table 3). There is a strong likelihood that these fats are mostly fake, and that traders are unable to verify the species (see Chauqúe[5]). For example, one respondent pointed out many bottles of fat belonging to different vertebrates, including three reptiles – but he failed to match the same species to the same bottles when re-interviewed later.

Ethnoherpetological nomenclature

African traditional nomenclature generally limits cultures from recognising the full range of reptile species and classification is not hierarchical[3]; instead, species are generally named in terms of their appearance, habitat and/or behaviour, and some names represent multiple species that are superficially similar.[3,20,21] For example, the common name for sea turtles (*N'futso*) documented in Maputo by Chauqúe[5], is similar to that of the Mozambican Xitsonga name recorded for the tortoises *Kinixys* sp. and *Stigmochelys pardalis* in our study (i.e. *futsu*) (Table 2). However, tortoises are also called *chibodze* (Table 2), which is the same as the group name for *Kinixys* and *Testudo* spp. (now species of *Chersina*, *Homopus*, *Psammobates* and *Stigmochelys*) recorded by Taylor[21] for 'Shangaan' speakers in southeastern Zimbabwe (note: while linguistically similar, the name 'Shangaan' has derogatory meaning to speakers of Xitsonga/Tsonga dialects in southern Africa). All these traditional common names are consistent with the Tsonga names for tortoises in southern and South Africa.[22,23] Hence, the common names for the testudines identify them as a 'morphospecies' across the region and similar taxa are likely to be used interchangeably for the same purposes.

Table 3: Quantities of reptile body part(s) per survey

English name	Quantities of body parts per market survey [number of traders]		
	Xipamanine 1[a]	Xipamanine 2	Xiquelene
Nile crocodile	2 skin pieces [2]	1 skin piece [1]	Fat [1]
Southern African python	1 skin piece [1]	4 full skins [4]; fat [1]	13 vertebrae (1 handful) [1]; 1 full skin and spine, 1 full skin and fat [1]; 1 full skin [1]
Mamba	–	–	Fat [1]
Gaboon adder	1 full skin [1]	–	–
Rough-scaled plated lizard	1 whole body [1]	1 whole body [1]	1 whole body [1]
Nile monitor	1 whole body [1]	3 full skins [3]	–
Rock monitor	–	1 full skin [1]	–
Flap-necked chameleon	4 live [2]	1 live [1]	–
Hinge-back tortoise	2 live [2]	2 carapaces [2]	1 carapace [1]
Leopard tortoise	–	3 carapaces [1]	–

[a]*Did not inventory animal fat in this survey*

The naming of *Varanus* spp. (*nkwahle*), *C. d. dilepis* (*lompfanhe*) and *C. crocodilus* (*nguenha*) in this study was also consistent with the names recorded in Taylor[21], namely *gwahli, limvani* and *ngwenya*, respectively. The naming of the *Varanus* spp. is supposedly based on their habitat; in South African Tsonga, the Nile monitor is called *ngwehle* and the Rock monitor is called *ngwahle*.[23] The Nile monitor is also called *gwahli mati*,[21] where the term *mati* means water. Although the spelling of the names is slightly different, they are phonetically similar.

The Xitsonga name *bululu* for the Gaboon adder (*Bitis gabonica*) is allied with the isiZulu name *imbululu* or *ibululu* for species of adders (*B. arietans, B. atropos, B. cornuta*) in South Africa.[3] Because elements of the isiZulu language are linguistically related to Xitsonga, all adder-like species in the region are thus likely to be similarly named. The similar names for individual morphospecies reflect the distribution of taxa that resemble each other, as well as the distribution of Tsonga speakers, across southern Mozambique, southeastern Zimbabwe and eastern South Africa. However, *makokoromba* (*B. major*) is the only Mozambican Xitsonga common name for which we found no linguistically similar ethnospecies name in South African Tsonga or isiZulu.

Species use

Documenting species uses is a controversial matter that risks breaching rights concerning indigenous knowledge.[24] We were informally made aware of the therapeutic use for only one species in 2015, namely chameleons for asthma (Table 1), and we were discouraged from discussing uses in 2016. Asthma is allegedly widespread in Mozambique according to one trader, and the tails are used to treat the condition. Non-therapeutic uses for two other species were also cited, namely tortoises for food, and python skins as part of the traditional attire of healers. Uses for the body parts of other species were either unknown (which is plausible as none of the traders was a traditional healer) or deliberately not divulged. The use of python skin in ceremonial regalia is a common practice among traditional healers and has been recorded in South Africa among Zulu-speaking people.[16]

Although the uses for the other species and their parts were not documented, their therapeutic values are less likely to be species-specific and more likely to be similar to previously documented uses for allied morphospecies with similar-sounding common names in the region. For example, uses for *B. gabonica* skin were not mentioned in our survey, but they are likely to be similar to other *Bitis* spp. and taxa collectively known as *bululu* that are used as snakebite antidotes and protective charms.[3,17] However, primary uses for certain taxa can be highly variable and characterised by a low degree of consensus among consumers.[24] What is more, uses are generally associated with the 'Doctrine of Signatures' and the complete or partial resemblance of a species to a bodily attribute or function, hence some uses can in some cases be deduced – such as the use of pythons to imbue strength.[24]

Sale prices of reptile parts

The prices of reptile parts seem variable and negotiable, and are dependent on the customer and on the freshness of the material. Fresher, more recently harvested, animals sell for higher prices because the medicines are seen as more effective. Because the respondents in Xipamanine were generally non-responsive to questions, we could only document prices for three species sold in Xiquelene in 2016. Prices of tortoise carapaces were quoted as MZN300 (ZAR95; USD6) for the whole carapace and MZN100–MZN150 per scute (where 1 Mozambican metical (MZN) = 0.3156 South African rand (ZAR) = 0.02 US dollars (USD)). The whole body of a relatively fresh-plated lizard (*B. major*) was quoted as MZN500. Python bones were sold for MZN10 per vertebra, and MZN150 per handful of vertebrae (see Figure 1e).

The unpredictability of the prices is evident when comparing information reported in trade studies conducted previously. In January 2016, prices quoted to us for python skins were: (1) whole: MZN300; (2) 200-mm piece: MZN100; (3) unknown size: MZN100–MZN150. In 2010, however, Chauqúe[5] listed the average price of a smaller 100 x 50 mm piece of python skin to be MZN208.33 (ZAR48; USD6) (exchange rate for June 2010; MZN1 = ZAR0.2304 = USD0.00288). Chauqúe[5] also listed the price of crocodile skin (no size specified) to be MZN587.5, ranging from MZN500 to MZN750 per unit. This range in pricing and the practice of negotiating is not uncommon throughout informal African markets.

Trade dynamics and procurement

Vendors report that the trade in animals, especially reptiles, is largely sporadic and not as profitable as that in medicinal plants. This irregular demand translates into some specimens being kept in the market for more than a year (Chauqúe[5] reported 3 years) before being sold and/or discarded if the stock becomes too decomposed. Given the losses associated with disposal of stock, these factors partially explain the preferences for live animals.

It is clear that traditional healers are the primary, if not sole, purchasers of reptile-based medicines in the markets. Crocodile and python were cited in the 2015 survey as the species that healers request most often, and also the species the traders would prefer to sell if they could acquire the stock (Figure 2) (note: these questions were deleted on request in the 2016 survey). Chameleons and tortoises were also mentioned as species that are in demand (although less often than crocodiles and pythons, and perhaps the reason for keeping them alive).

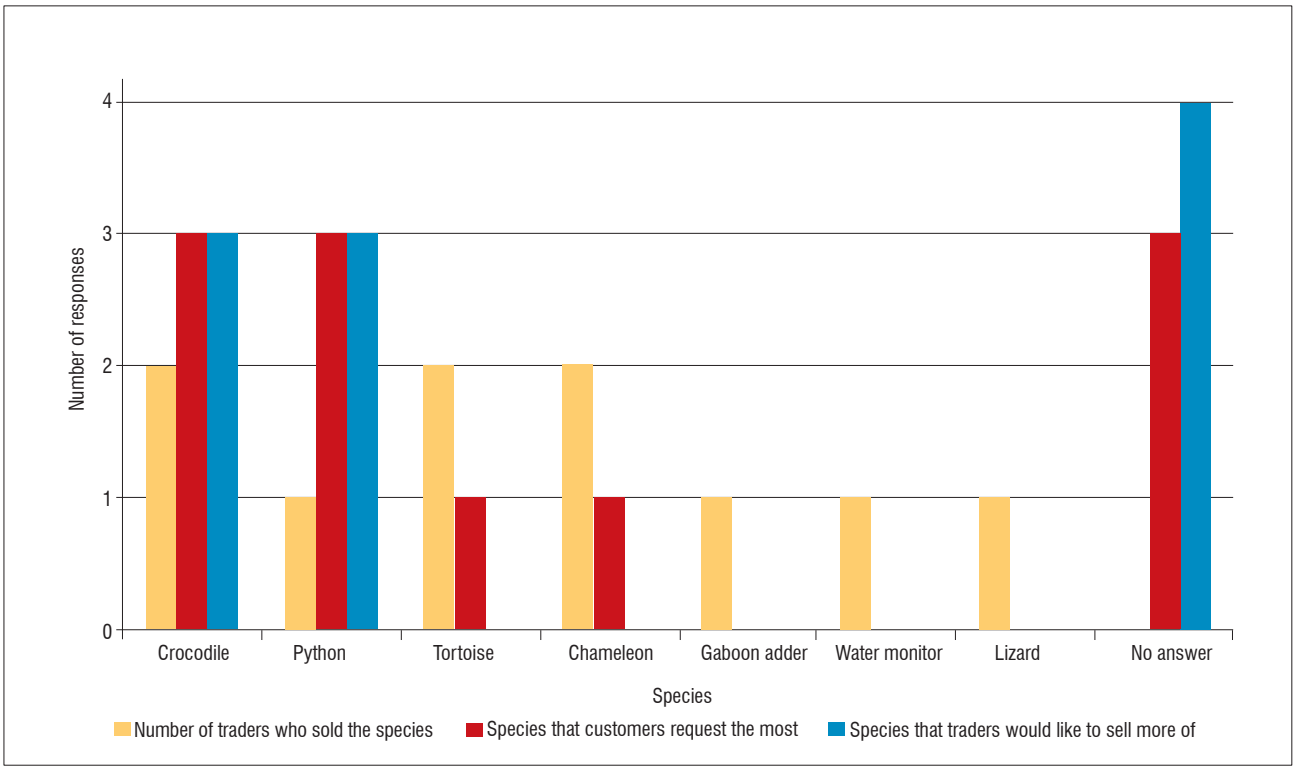


Figure 2: Variation in the number of traders selling reptile species, and the customers' and traders' preferences for medicinal reptiles. (Xipamanine Survey 1, March 2015. A comparative study in January 2016 was not permitted.)

Rural harvesters and suppliers are central to the provision of animal products to the market. None of the respondents harvested animals themselves – they bought stock directly from sellers or harvesters at the market. While some traders know rural harvesters they can contact to place an order, no special deliveries of reptiles are made and they almost always accompany a much larger consignment of medicinal plants. Smaller reptiles such as chameleons and tortoises tend to be acquired opportunistically year-round when harvesters find them whilst collecting medicinal plants. Other species are mostly brought to the market in summer – usually because frequent veld fires and brumation of reptiles during winter decreases their seasonal detectability. The hunting of crocodiles, however, is not opportunistic nor incidental because of the effort required to kill them. Large reptiles tend to be killed in situ and the products (e.g. skin, fat, bones) are transported to the markets hidden in bags of charcoal or medicinal plants. Confiscation of material at roadblocks is a constant risk to the suppliers, which adds to their costs along with the cost of transport from the harvesting catchment to the markets. Poor and/or rural-based harvesters are thus especially vulnerable to the risks of transporting large and/or protected species without the necessary permits.

Origin of acquired reptiles

The procurement of reptiles from three provinces (Gaza, Inhambane, Maputo) and 10 districts (including all 7 districts within Maputo Province) (Figure 3; Table 4) highlights the exchange of resources between urban and rural areas and the cultural ties that exist between them. The use of the same species for traditional medicine in rural and urban areas also suggests that zootherapeutic practices may function as a social conduit between rural people in remote rural areas and people (such as migrants) living in urban areas that helps maintain traditional culture and values, as well as information on illnesses and potential treatments.[12]

Most specimens were collected from districts within Maputo Province (Table 4), but python, crocodile and the Nile monitor were also acquired from the adjacent provinces of Gaza and Inhambane (particularly along the Save River in northern Inhambane). The Chicualacuala District in Gaza is approximately 430 km from Xipamanine and was noted by Chauqúe[5] to be a major supply area for all animals sold in Maputo markets – probably because of the high faunal diversity in the region and its proximity to the Greater Limpopo Transfrontier Park. However, Krog et al.[7] also noted that most traders (of medicinal plants) interviewed in Maputo markets originally came from the neighbouring provinces of Gaza and Inhambane before becoming permanent traders at the market, and that the majority of plant products was acquired from these two provinces and Maputo. Although we did not record the respondents' ethnic group, Chauqúe[5] further reported that *Mashope* vendors (Chopi-speaking and originating from northern Gaza and southern Inhambane) were the second-most common ethnic group of the animal traders. Therefore, links to 'home' and familial ethnic ties to certain rural areas may also function as important harvesting conduits for acquiring species.

Cross-border trade

There is a cross-border trade in reptiles between neighbouring countries and/or customers from Tanzania, Zambia, Malawi, Zimbabwe and Swaziland, and especially with traditional healers from South Africa. However, respondents indicated that medicinal plants are the main focus of this cross-border exchange of resources and that animals are a minor accompaniment. When considering the number of ethnomedicinal animal traders in the Johannesburg and Durban markets (≈60 in Durban; Moshoeu TJ, personal observation), and the presence of fauna procured from Mozambique (Williams VL, personal observation), then rumours of a weekly ethnomedicinal market operating at Ponta do Ouro, only 15 km from the South African border, are plausible. It is suspected, however, that these Mozambican resources are mostly smuggled through the nearby Kosi Bay border area given its proximity to South Africa and the relative remoteness of the region. Transport of fauna and flora across the Lebombo/Ressano Garcia border is less likely given the higher levels of security there.

Table 4: Harvesting localities/sources for reptiles sold in the markets

English name	Localities per survey Province (district, number of traders)		
	Xipamanine 1	**Xipamanine 2**	**Xiquelene**
Nile crocodile	Gaza (Chicualacuala, 1) Not known (1)	Maputo (Moamba, 1)	–
Southern African python	Not known (1)	Inhambane (Govuro[c], 1; along the Save River, 1) Maputo (Boane, 1; Manhiça, 1)	Gaza (Chibuto, 1; Chicualacuala, 1; not known, 1) Maputo (Namaacha, 1)
Mamba	–	–	Not known (1)
Gaboon adder	Maputo (Manhiça [a], 1)	–	–
Rough-scaled plated lizard	Not known (1)	Not known (1)	Maputo (Magude, 1)
Nile monitor	Maputo (along rivers, 1)	Inhambane (along the Save River, 1) Maputo (Manhiça, 1) Not known (1)	–
Rock monitor	–	Maputo (Marracuene, 1)	–
Flap-necked chameleon	Maputo (Marracuene, 2)	Not known (1)	–
Hinge-back tortoise	Maputo (Maputo[b], 1; Namaacha, 1)	Maputo (Marracuene, 1; Namaacha[d], 1)	Maputo (Boane, 1)
Leopard tortoise	–	Maputo (Matatuine[e], 1)	–

Specific localities mentioned: a, Nhongonhane; b, outskirts of the city of Maputo; c, Mambone; d, Changalane; e, Ponta do Ouro

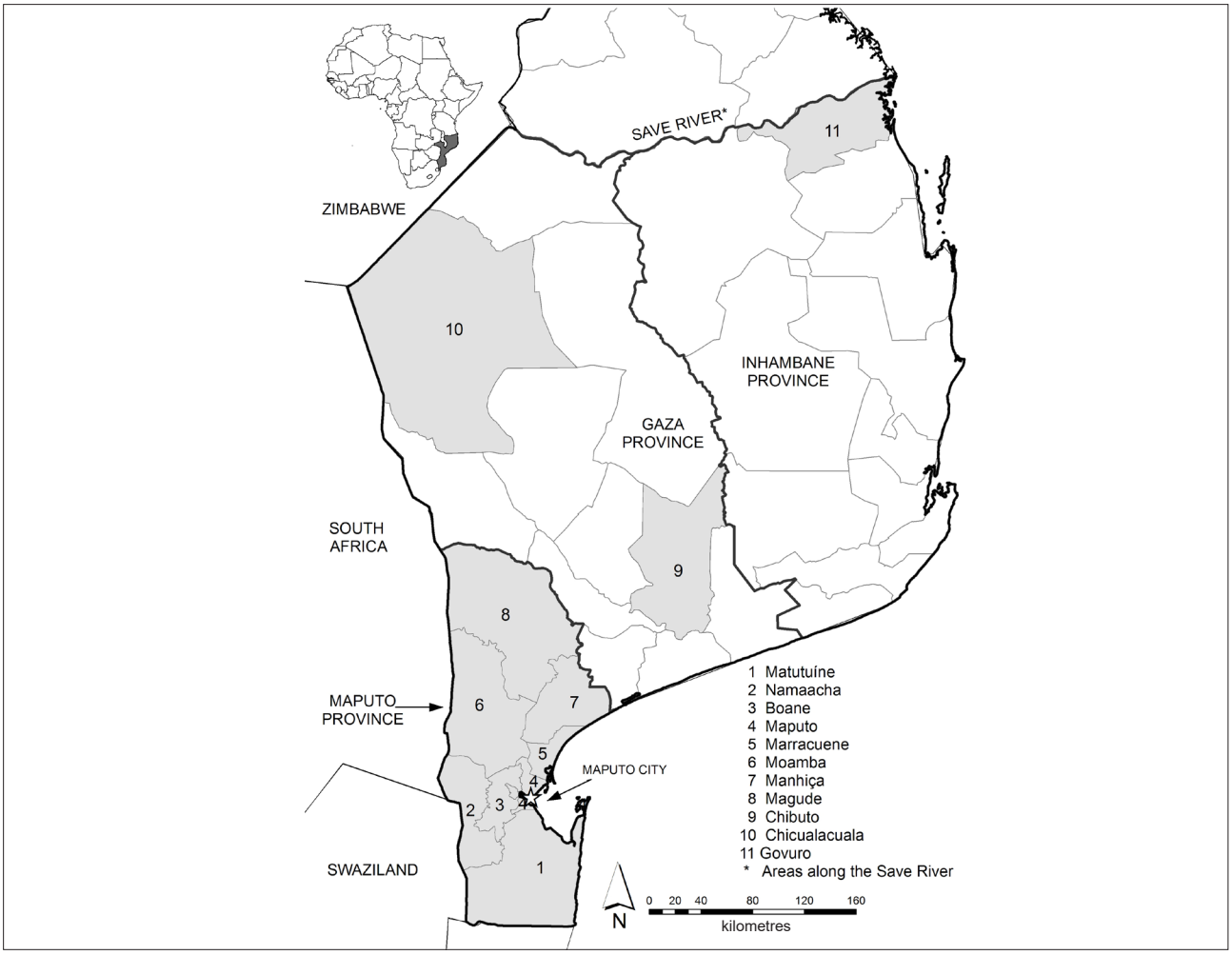


Figure 3: Map of Mozambique indicating the provinces and districts cited as source areas for the reptiles documented during the 2015/2016 market surveys.

Perceptions of resource availability

During the 2015 Xipamanine survey, all the traded reptile species were perceived by the sellers to be declining in the wild, and frequent veld fires were cited as the primary cause for the decline, especially in the case of chameleons. As we were prevented from asking this question during the second survey, no comparative results are available for 2016. Difficulties in the acquisition of medicinal animals sold in Mozambique markets owing to their scarcity were also noted by Marshall[13]; of special concern were *P. natalensis* and *S. w. warreni*. Chauqúe[5] further noted that the turtles *C. mydas* and *E. imbricata* were in 'danger of extinction' and urgent action was required to reverse the situation.

Despite the reported 'shortages' and declines, most respondents reported that reptiles were no longer in great demand – suggesting that there has been a change or adaptation in consumer requirements, and accordingly a drop in stock numbers compared to previous years. Is diminishing indigenous knowledge of reptiles' therapeutic and symbolic properties one reason for the reduced demand? Or, are consumers increasingly less likely to utilise zootherapeutic remedies relative to plant-based remedies? We do know that there is a large demand for certain reptiles (especially crocodiles, pythons and tortoises) in the urban traditional markets of South Africa.[11,17,24,25] Even though elements of the trade within Mozambique appear to be small, we believe that trade with South African consumers (in both urban and rural areas) is negatively impacting some species. For example, Ngwenya[25] reported that there is a high demand for turtle products and eggs in Maputaland (a region in the northern part of the South African province of KwaZulu-Natal adjacent to Mozambique and between Swaziland and the coast). Turtle eggs are alleged to make poultry more fertile and consumption of turtle meat improves longevity[25]; hence, turtles are killed and their nests are robbed in neighbouring Mozambique and along the northern KwaZulu-Natal coastline.[25]

Implications for conservation

Of the 10 reptile taxa recorded, only *C. dilepis*, *C. niloticus* and *S. pardalis* have been subjected to global conservation assessments by the International Union for Conservation of Nature and Natural Resources (IUCN), and are all listed as Least Concern (as of 24 January 2016). Regional assessments have been conducted on nine species for South Africa, Lesotho and Swaziland.[18] *Crocodylus niloticus* is listed regionally as Vulnerable in this assessment[26], *B. gabonica* as Near Threatened[27] and the remainder as Least Concerned[28-32]. No corresponding national IUCN Red List assessment has been done for Mozambique. Given the wide distribution of *B. gabonica*, and lack of genetic differences between populations,[27] it is likely that it too would be listed as Least Concern in a global assessment. Thus, none of the species that we recorded being traded in Maputo markets in 2015/2016 are of particular global conservation concern. Given the low numbers of individual reptiles traded at the market, the IUCN status of each, and the apparent long-term decrease in the importance of the reptiles for zootherapeutic needs, we believe that the reptile trade in these markets is unlikely to pose any grave conservation concern.

Conclusion

We recorded few reptile species for sale in the Xipamanine and Xiquelene Markets and traders appear to have limited knowledge on the traditional uses of the species that they sell. Traders also claim that there is a declining demand for reptiles, which is why some of them prefer to keep the animals alive and have fresh material available for when customers (usually traditional healers) do request them. However, trade in animals is mostly illegal and traders routinely store species out of sight. It is thus possible, in addition to there being a limited urban demand in the markets, that our survey did not document the full extent of the domestic trade.

With these caveats in mind, our findings suggest that trade of reptiles via the markets in Maputo is unlikely to have any significant conservation consequences for the species concerned. However, we believe that the ethnoherpetological trade between Mozambique and South Africa is active, especially to the large urban traditional medicine markets in Johannesburg and Durban where there are more than 60 traders (some of them Mozambican) in each market selling animal products. Consequently, it is likely that, while the domestic market in Maputo is quite small, consumption in the rural areas and the international trade are having a larger impact and that warrants further investigation, possibly from the South African end of the supply chain.

Acknowledgements

We thank the traders of the Xipamanine and Xikelene Markets for their time, and Professor Mário Paulo Falcão from the Faculty of Agronomy and Forestry Engineering (Eduardo Mondlane University, Mozambique) and Mr Gediao Penicela (student at Eduardo Mondlane University) for their assistance during the fieldwork. We thank Dr Kaera Coetzer from Wits University for preparing the map.

Authors' contributions

Design and planning was carried out by T.J.M. and V.L.W. The market survey was carried out by T.J.M. Species were identified by G.J.A. V.L.W., T.J.M. and G.J.A. wrote the manuscript. All authors read and approved the manuscript.

References

1. Lev E. Traditional healing with animals (zootherapy): Medieval to present day Levantine practice. J Ethnopharmacol. 2003;85:107–118. http://dx.doi.org/10.1016/S0378-8741(02)00377-X
2. Segniagbeto HA, Petrozzi F, Aïdam A, Luicelli L. Reptiles traded in the Fetish Market of Lomé, Togo (West Africa). Herpetol Conserv Biol. 2013;8(2):400–408.
3. Simelane TS, Kerley GI. Recognition of reptiles by Xhosa and Zulu communities in South Africa, with notes on traditional beliefs and uses. Afr J Herpetol. 1997;46(1):49–53. http://dx.doi.org/10.1080/21564574.1997.9649975
4. Mendonça LET, Vieura WLS, Alves RRA. Caatinga ethnoherpetology: Relationships between herpetofauna and people in semiarid region of northeastern Brazil. Amphib Reptile Conserv. 2014;8(1):24–32.
5. Chauqúe M. Análise da comercialização de animais ou partes dos mesmos para fins medicinais na cidade de Maputo [Analysis of marketing animals or parts thereof for medicinal purposes in Maputo]. Maputo: Eduardo Mondlane University; 2010. Portuguese.
6. Dedeke GA, Soewu DA, Lawal OA, Ola M. Pilot survey of the ethnozoological utilization of vertebrates in South Western Nigeria. Indilinga – Afr J Indig Knowl Syst. 2006;5(1):87–96.
7. Krog MP, Falcão MP, Olsen CS. Medicinal plant markets and trade in Maputo, Mozambique. Hørsholm: Christian Ejlers; 2006.
8. Nevadomsky J. Kemwin-Kemwin: The Apothecary Shop in Benin City. Afr Arts. 1988;22(1):72–100. http://dx.doi.org/10.2307/3336693
9. Sodeinde OA, Soewu DA. Pilot study of the traditional medicine trade in Nigeria. TRAFFIC Bull. 1999;18(1):35–40.
10. Vats R, Thomas S. A study on use of animals as traditional medicine by Sekuma tribe of Busega district in North-western Tanzania. J Ethnobiol Ethnomed. 2014;11, Art. #38, 11 pages. http://dx.doi.org/10.1186/s13002-015-0001-y
11. Whiting MJ, Williams VL, Hibbitts TJ. Animals traded for traditional medicine at the Faraday market in South Africa: Species diversity and conservation implications. J Zool. 2011;284:84–96. http://dx.doi.org/10.1111/j.1469-7998.2010.00784.x
12. Alves RRN, Rosa IL. Trade of animals used in Brazilian traditional medicine: Trends and implications for conservation. Hum Ecol. 2010;38:691–704. http://dx.doi.org/10.1007/s10745-010-9352-0
13. Marshall N. Searching for a cure: Conservation of medicinal wildlife resources in East Southern Africa. Cambridge: Traffic International; 1998.
14. Bruschi P, Morganti M, Mancini M, Signorini MA. Traditional healers and laypeople: A qualitative and quantitative approach to local knowledge on medicinal plants in Muda (Mozambique). J Ethnopharmacol. 2011;138:543–563. http://dx.doi.org/10.1016/j.jep.2011.09.055

15. World Health Organization Country Office – Mozambique. Health systems service delivery [homepage on the Internet]. No date [cited 2015 Oct 02]. Available from http://www.afro.who.int/en/mozambique/country-programmes/health-systems/service-delivery.html

16. Luo X, Pires D, Aínsa JA, Gracia B, Mulhovo S, Duarte A, et al. Antimycobacterial evaluation and preliminary phytochemical investigation of selected medicinal plants traditionally used in Mozambique. J Ethnopharmacol. 2011;137:114–120. http://dx.doi.org/10.1016/j.jep.2011.04.062

17. Cunningham AB, Zondi AS. Use of animal parts for commercial trade in traditional medicine. Working Paper 76. Pietermaritzburg: Institute of Natural Resources, University of Natal; 1991.

18. Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alexander GH, et al. Atlas and Red List of the reptiles of South Africa, Lesotho and Swaziland. Pretoria: South African National Biodiversity Institute; 2014.

19. Uetz P, Hošek J. The reptile database [homepage on the Internet]. No date [cited 2015 March 23]. Available from: http://www.reptile-database.org.

20. Smart R, Whiting MJ, Twine W. Lizards and landscapes: Integrating field surveys and interviews to assess the impact of human disturbance on lizard assemblages and selected reptiles in a savanna in South Africa. Biol Conserv. 2005;122:23–31. http://dx.doi.org/10.1016/j.biocon.2004.06.016

21. Taylor P. Shangaan herpetology. J Herpetol Assoc Rhodesia. 1962;19(1):11–12. http://dx.doi.org/10.1080/0440730x.1962.9650493

22. Anthony BP, Bellinger EG. Importance value of landscapes, flora and fauna to Tsonga communities in the rural areas of Limpopo Province, South Africa. S Afr J Sci. 2007;103:148–154.

23. Wild Earth. Shangaan [homepage on the Internet]. No date [cited 2015 May 25]. Available from: http://blog.wildearth.tv/2008/06/shangaan.html

24. Williams VL, Whiting MJ. A picture of health? Animal use and the Faraday traditional medicine market, South Africa. J Ethnopharmacol. 2016;179:265–273. http://dx.doi.org/10.1016/j.jep.2015.12.024

25. Ngwenya MP. Implications for the medicinal animal trade for nature conservation in KwaZulu-Natal. Unpublished report no. NA/124/04. Pietermaritzburg: Ezemvelo KZN Wildlife; 2001.

26. Marais J. *Crocodylus niloticus* (Laurenti, 1768). In: Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alexander GJ, et al., editors. Atlas and Red List of the reptiles of South Africa, Lesotho and Swaziland. Suricata 1. Pretoria: South African National Biodiversity Institute; 2014. p. 87–88.

27. Turner AA. *Bitis gabonica* (Linnaeus, 1762). In: Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alexander GJ, et al., editors. Atlas and Red List of the reptiles of South Africa, Lesotho and Swaziland. Suricata 1. Pretoria: South African National Biodiversity Institute; 2014. p. 335–336.

28. Alexander GJ. *Varanus niloticus* (Linnaeus, 1762). In: Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alexander GJ, et al., editors. Atlas and Red List of the reptiles of South Africa, Lesotho and Swaziland. Suricata 1. Pretoria: South African National Biodiversity Institute; 2014. p. 283–284.

29. Alexander GJ. *Python natalensis* (A. Smith, 1840). In: Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alexander GJ, et al., editors. Atlas and Red List of the reptiles of South Africa, Lesotho and Swaziland. Suricata 1. Pretoria: South African National Biodiversity Institute; 2014. p. 238.

30. Bates MF. *Broadleysaurus major* (Duméril, 1851). In: Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alexander GJ, et al., editors. Atlas and Red List of the reptiles of South Africa, Lesotho and Swaziland. Suricata 1. Pretoria: South African National Biodiversity Institute; 2014. p. 255.

31. Boycott RC. *Kinixys zombensis* (Hewitt, 1931). In: Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alexander GJ, et al., editors. Atlas and Red List of the reptiles of South Africa, Lesotho and Swaziland. Suricata 1. Pretoria: South African National Biodiversity Institute; 2014. p. 79.

32. Tolley KA. *Chamaeleo dilepis* (Leach, 1819). In: Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alexander GJ, et al., editors. Atlas and Red List of the reptiles of South Africa, Lesotho and Swaziland. Suricata 1. Pretoria: South African National Biodiversity Institute; 2014. p. 300.

Note: This article is accompanied by supplementary material.

Developmental simulation of the adult cranial morphology of *Australopithecus sediba*

AUTHORS:
Keely B. Carlson[1]
Darryl J. de Ruiter[1,2]
Thomas J. DeWitt[3]
Kieran P. McNulty[4]
Kristian J. Carlson[1,5]
Paul Tafforeau[6]
Lee R. Berger[2]

AFFILIATIONS:
[1]Department of Anthropology, Texas A&M University, College Station, Texas, USA
[2]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa
[3]Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, Texas, USA
[4]Department of Anthropology, University of Minnesota, Minneapolis, Minnesota, USA
[5]Department of Anthropology, Indiana University, Bloomington, Indiana, USA
[6]European Synchrotron Radiation Facility, Grenoble, France

CORRESPONDENCE TO:
Keely Carlson

EMAIL:
keelyc@tamu.edu

POSTAL ADDRESS:
Department of Anthropology, Texas A&M University, MS 4352 TAMU College Station, Texas 77845, USA

DATES:
Received: 12 Jan. 2016
Revised: 16 Mar. 2016
Accepted: 17 Mar. 2016

KEYWORDS:
Malapa; geometric morphometrics; craniofacial morphology; hominin evolution; ontogenetic projection

HOW TO CITE:
Carlson KB, De Ruiter DJ, DeWitt TJ, McNulty KP, Carlson KJ, Tafforeau P, et al. Developmental simulation of the adult cranial morphology of *Australopithecus sediba*. S Afr J Sci. 2016;112(7/8), Art. #2016-0012, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20160012

The type specimen of *Australopithecus sediba* (MH1) is a late juvenile, prompting some commentators to suggest that had it lived to adulthood its morphology would have changed sufficiently so as to render hypotheses regarding its phylogenetic relations suspect. Considering the potentially critical position of this species with regard to the origins of the genus *Homo*, a deeper understanding of this change is especially vital. As an empirical response to this critique, a developmental simulation of the MH1 cranium was carried out using geometric morphometric techniques to extrapolate adult morphology using extant male and female chimpanzees, gorillas and humans by modelling remaining development. Multivariate comparisons of the simulated adult *A. sediba* crania with other early hominin taxa indicate that subsequent cranial development primarily reflects development of secondary sexual characteristics and would not likely be substantial enough to alter suggested morphological affinities of *A. sediba*. This study also illustrates the importance of separating developmental vectors by sex when estimating ontogenetic change. Results of the ontogenetic projections concur with those from mandible morphology, and jointly affirm the taxonomic validity of *A. sediba*.

Introduction

To date, only a single, relatively complete cranium of *Australopithecus sediba* has been recovered from the Malapa fossil site, belonging to the type specimen MH1.[1] Dating to 1.977±0.002 Ma,[2] the Malapa hominins exhibit a unique, mosaic morphology, possessing features that align them with both the genus *Homo* as well as other species of australopith[1]. Based on this intermediate morphology, Berger et al.[1] suggested a possible ancestor–descendant relationship between *A. sediba* and the genus *Homo*, with the possibility of *A. sediba* representing the direct ancestor to *H. erectus*, or otherwise a close sister group to that ancestor. The cranium itself shows remarkable preservation, possessing a complete facial skeleton and detailed surface anatomy with clearly visible suture lines. The second molars are in occlusion in this specimen, while the third molars remain in the crypt[1], indicating its sub-adult status. With an estimated cranial capacity of 420 cm^3, it is estimated that MH1 had achieved approximately 95% of its expected brain growth at age of death.[1,3]

In response to the announcement, several outside commentators immediately questioned the distinctiveness of *A. sediba* as a unique species separate from *A. africanus*.[4,5] In a diametrically opposed argument, others suggest that this species should have been assigned to the genus *Homo*.[4] Because of the young age and hence incomplete growth of the type specimen, several commentators have further questioned the reliability of phylogenetic interpretations based on the MH1 fossil.[4-8] Critics of the Berger et al.[1] interpretations argue that the degree of development expected to occur between second and third molar eruption would have been substantial enough to alter our current interpretations of the morphological affinities of *A. sediba*, especially those features thought to resemble later *Homo*. For example, Spoor[6] argued the possibility for increased constriction of the MH1 brain case had it reached full adulthood. More recently, Kimbel[8] criticised the use of the juvenile mandible in multivariate statistical comparisons with other species, based on the potential of continued growth and development for impacting linear measurements.[9]

To address this possibility, we used established geometric morphometric techniques[10-16] to produce 3D renderings of the inferred adult cranial morphology of *A. sediba* based on regression of developmental samples in accordance with dental eruption sequence, or dental stage. The goal of this procedure was to generate a series of developmental trajectories for both male and female extant apes and humans (Table 1), and then apply these trajectories to MH1 to create a series of virtual adult crania. This procedure, in turn, allowed us to empirically establish the estimated adult form of *A. sediba*, and test whether or not current interpretations regarding the cranial morphology of this species should be modified as a result of future developmental changes.

Table 1: Summary of hominoid sample used to create developmental vectors

Species	Juveniles†	Male adults‡	Female adults‡
Pan troglodytes	13	6	7
Gorilla gorilla	15	6	4
Homo sapiens	8	16	11

†M2 indicates second molars are erupted and in occlusion; ‡M3 indicates that the third molars are erupted and in occlusion.

Materials and methods

Reconstruction

A reconstruction of the MH1 cranium was carried out with the goal of correcting for distortion in the cranium and producing a more complete rendered model of the *A. sediba* skull (Figure 1a,b). Rapidform® software (now 3D Systems Geomagic; see http://www.rapidform.com/home) was employed in this reconstruction to refine and process 3D models. The original 3D model of the MH1 cranium employed in this reconstruction was created using synchrotron image data generated on the beamline ID17 at the European Synchrotron Radiation Facility located in Grenoble, France.[17] Synchrotron image data were segmented using Avizo 6.3® software, resulting in a 3D virtual rendering of the MH1 cranium, removed from the encasing breccia.[17] This imaging allowed for the collection of landmark data in areas that were previously obscured by matrix.

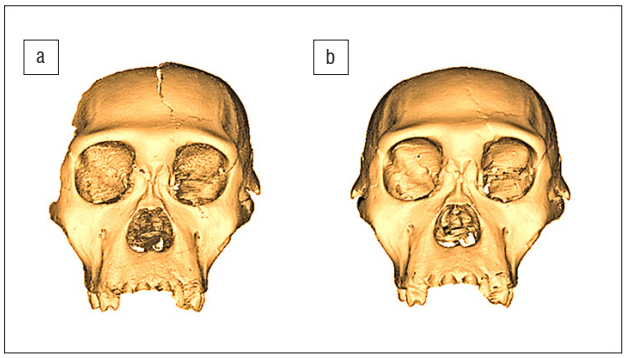


Figure 1: MH1 cranium (a) before and (b) after reconstruction.

When examining the juvenile cranium, several preservation issues impacting the integrity of the fossil were apparent. The most prominent among these is a large crack, originating at the left supraorbital torus, which runs posteromedially across the frontal, widening as it continues to bregma to reach a maximum breadth of approximately 7 mm. An additional crack affecting landmark placement extends from the medial margin of the right orbit, inferomedially across the frontal process of the maxilla, breaking across the nasal bridge. Best viewed from frontal perspective, the crack obscures the right frontomaxillary suture and laterally displaces the inferior portion of both nasals.

Entire cranial bones were also displaced. As discussed above, the large crack extending across the frontal has resulted in the displacement of the left portion of the frontal bone. The displacement extended laterally from above the left orbit to the articulation with the zygomatic bone. The zygomatic bone was displaced posteroinferiorly, disarticulating the bone from the zygomatic arch of the temporal and the zygomatic process of the frontal. This distortion can also be noted from the frontal aspect when examining the inferior margin of the left orbit. Reconstruction correcting for the cracks and displacements as just described was a major focus of the present project. An additional goal was to produce a more complete calvaria by reflecting the preserved portions of the left parietal and temporal bones. Fortunately, any plastic deformation was deemed extremely minor, if present at all.

The 3D model of the MH1 cranium was first imported into Rapidform®, and all adjustments were conducted in 'mesh mode'. It was decided that the most efficient and effective way to account for cracks and displacements was by selecting the affected areas that required adjustment, copying and pasting these regions into a separate window, and then re-aligning the selected area with the original model. The selected areas were reoriented using the 'scan tools' property and 'align between scan data'. Using these tools, one selects a reference scan and a moving scan, which are then aligned in accordance with selected analogous points that serve to stitch the scans together.

To correct for the large crack across the frontal bone, we first selected the preserved portions of the frontal, parietal, sphenoid and temporal bones extending to the left of the crack. These portions were then copied into a separate window and selected as the moving scan to align with the remaining portion of the cranium using the 'align between scan data' tool. Once the crack had been corrected, we were then able to move and realign the zygomatic using the same process, in which the zygomatic bone was removed and then rearticulated with the zygomaticotemporal and zygomaticofrontal sutures. The portion of the nasal bones laterally displaced inferior to the break was additionally removed and realigned using the same process used for correcting the zygomatic and frontal bones.

The mirror tool was used to correct for the distortion along the nasal bridge, by reflecting the left side of the superior portion of the nasal bridge to remove the crack across the frontomaxillary suture. After correcting for displacement and distortion in the cranium, the mirror tool was used to reflect the left half of the calvaria posterior to the coronal suture to produce a more complete calvaria. As a final step in the reconstruction, the scans from the corrected model were then merged into a single scan and the resulting model was run through 'global remesh'. This command regenerates the mesh structure with removed defects in accordance with the model's curvature flow. A final product is shown in Figure 1b.

Sample

A summary of the hominoid comparative samples – comprising male and female chimpanzees (*Pan troglodytes*), gorillas (*Gorilla gorilla gorilla*) and modern humans (*Homo sapiens*) – is listed in Table 1. Ape data were collected from wild-shot specimens housed at the Cleveland Museum of Natural History (Cleveland, OH, USA). Records for chimpanzees indicate that specimens were collected in Ebolwa, Cameroon, as well as Abong Mbong, French Cameroons, and Abong Mbong, Djaposten, Cameroons, West A, with the exception of three specimens for which no geographical data are available. The records for the gorilla crania used in the comparative sample indicate that specimens were collected in French West Africa, French Congo, Ebolwa, Cameroon, Abong Mbong, French Cameroons, and Abong Mbong, Djaposten, Cameroons, West A. Both chimpanzee and gorilla crania were sexed using cranial remains. The human sample primarily included cadaver crania obtained from the Raymond A. Dart Collection of Human Skeletons at the School of Anatomical Sciences at the University of the Witwatersrand (Johannesburg, South Africa), although several well-preserved archaeological crania were included from both the Raymond A. Dart Collection of Human Skeletons and the Cleveland Museum of Natural History. Specimens with obvious pathologies or abnormalities were excluded from the study sample. Archaeological crania were only used if they could be confidently sexed using available records or standard cranial sexing criteria.

Specimens were assigned to developmental categories based on dental eruption sequence. Sub-adults were judged to be of the same developmental age as MH1, and therefore suitable for inclusion within the study, if the second molars were erupted and in occlusion, while the third molars had not yet erupted. Specimens were designated as adults if the third molars were erupted and in occlusion. Both male and female specimens were sampled for parity, although the juvenile sample was pooled for the purpose of developmental simulation. Considerable research has been conducted on both the timing and development of sexual dimorphism among hominoids.[18-21] Results of this research indicate that differences in craniofacial morphology of great apes and humans is established early in ontogeny.[20] Research further identifies the effects of allometric scaling, as well as disparities in growth rates and growth spurts of male and female hominoids as considerable influences in the onset and resulting degree of sexual dimorphism among species.[18,19,21] However, as a result of both the limited number of juveniles, and the fact that sex was often unknown for these specimens, we maintain that pooling of the juvenile sample was appropriate in the present study. We further note that previous studies employing developmental simulation have used pooled juvenile samples for similar reasons.[13]

The hominin fossil sample used in geometric morphometric comparative analyses included a sample of non-'robust' australopith and early

Homo crania from five species: *A. sediba* (MH1), *A. africanus* (Sts 5, Sts 71, Stw 53), *H. habilis* (OH 24, KNM-ER 1813), *H. rudolfensis* (KNM-ER 1470) and *H. erectus* (D2700, KNM-WT 15000, KNM-ER 3733). Laser surface scans of *A. africanus* fossils were collected from original fossil material at the Ditsong Museum of Natural History in Pretoria, South Africa. The original OH 24 fossil, referred to as *H. habilis*, was scanned at the National Museum of Tanzania in Dar es Salaam. Scans from the Dmanisi and Kenyan fossil material were obtained from casts, as the original fossil material was not available.

Methods

The ontogenetic samples for extant chimpanzees, gorillas and humans and fossil hominin crania were collected using a NextEngine 3D laser scanner®. Surface scans were stitched together and fused using Scanstudio HD Pro software®. These scans were later imported into Geomagic® where they were smoothed and polished using the 'mesh doctor' tool. 'Mesh doctor' provides an automatic polygon mesh improvement tool, which both detects errors in the mesh and corrects them. All holes in the mesh were filled-in using Geomagic® to prevent any landmarks or semi-landmarks from 'falling through' the mesh during landmark placement.

After fusing and polishing, each scan was imported into Landmark Editor 3.6®[22] software where 76 traditional landmarks and semi-landmarks were then placed (Table 2). Three semi-landmark curves were placed along the mid-sagittal plane between rhinion and bregma. These landmarks were not slid, but rather served as traditional landmarks. Landmark files for each hominoid species were exported to MorphoJ®[23] software where they were aligned through generalised Procrustes analysis. Generalised Procrustes analysis minimises the sum of squared distances between homologous points on each specimen and the imputed mean configuration through translation, rotation and scaling, thus controlling for the effects of location, orientation and size within the sample.[24-27] Estimated adult MH1 crania produced from each ontogenetic trajectory (male and female chimpanzee, gorilla and human) were visualised in Landmark Editor 3.6® software[22] using thin-plate spline interpolations of the corrected MH1 cranial surface warped into the estimated adult configuration (Figure 2a,b).

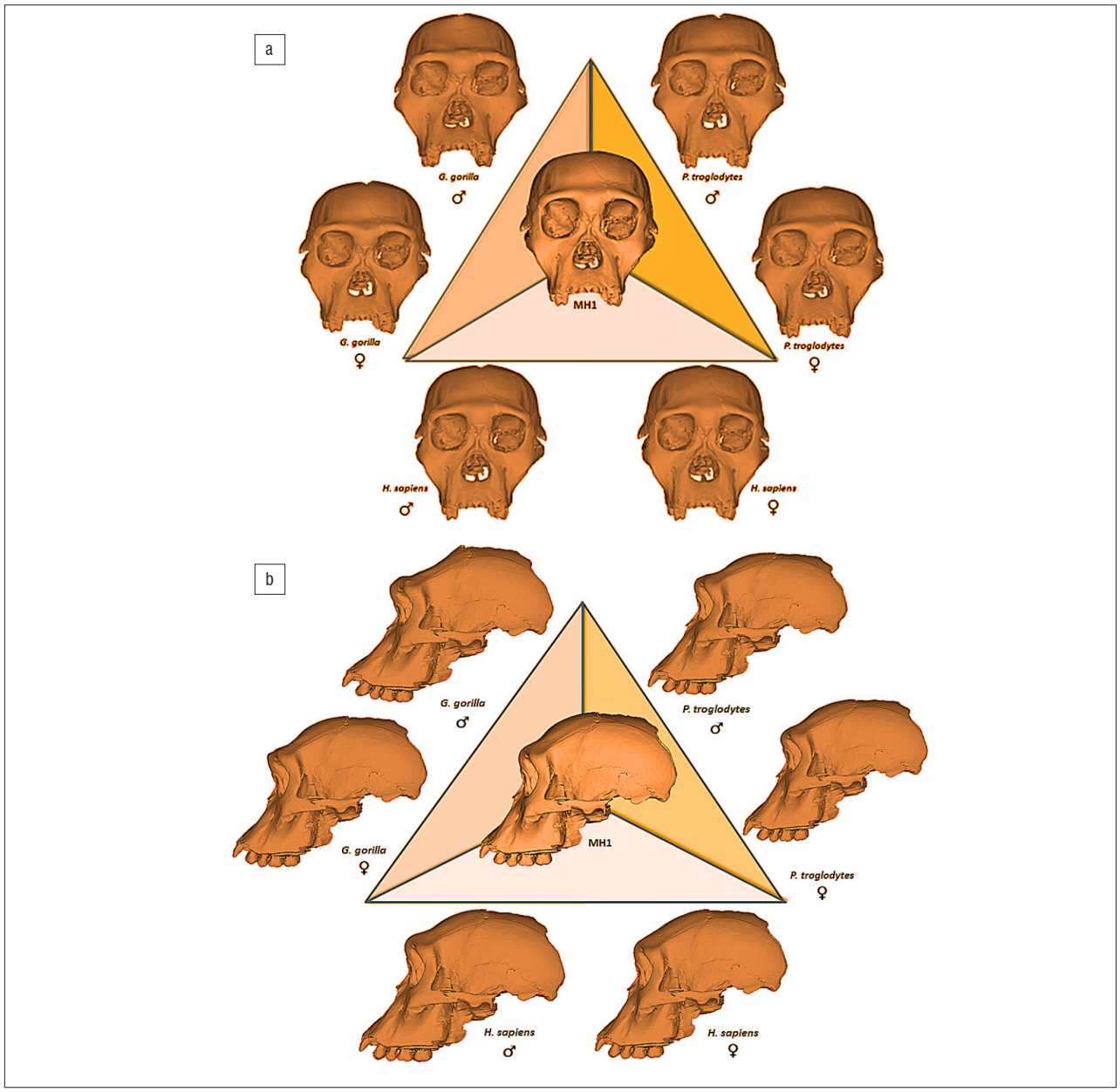


Figure 2: Visualisations of MH1 and its simulated adult forms from hominoid developmental trajectories in (a) frontal and (b) lateral aspects.

Table 2: Landmark definitions for landmarks used in developmental simulation

Landmark	Definition
1,2	Mid-torus inferior (right and left)
3,4	Mid-torus superior (right and left)
5,6	Dacryon (right and left)
7,8	Zygoorbitale (right and left)
9,10	Frontomalare orbitale (left and right)
11,12	Infraorbital foramen (right and left)
13,14	Zygomaxillare (right and left)
15,16	Alare (right and left)
17	Anterior attachment of nasal septum
18	Prosthion
19,20	I1-I2 contact (left and right)
21,22	I2-canine contact (left and right)
23,28	Canine-P3 contact
24,29	P3-P4 contact
25,30	P4-M1 contact
26,31	M1-M2 contact
27,32	M2-M3 contact
33,34	Jugale (left and right)
35	Zygomatico-temporal suture superior
36	Zygomatico-temporal suture inferior
37,38	Pterion
39,40	Inferior-most point of post-glenoid process
41	Incisivion
42	Alveolon
43,44	Inferolateral junction of nasal with maxilla (right and left)
45,46	Frontomalare temporale (left and right)

Note: In addition to the landmarks listed here, three curves of semi-landmarks were added along the mid-sagittal curve between rhinion and bregma, with a density of 10 equidistant spaced semi-landmarks each, making for a total of 76 landmarks. Curve one begins at rhinion and ends at the posterior junction of the glabella and the slope of the frontal bone. Curve two begins at this point and ends at the midpoint along the mid-sagittal curve of the frontal bone. Curve three begins at this midpoint on the frontal bone and ends at bregma.

A subset of 32 landmarks across the calvaria, face and palate of specimens was then used for morphometric comparisons between hominin crania and simulated versions of *A. sediba* individuals (Table 3). Producing this subset was necessary for comparing extant taxa landmarks to landmarks obtainable on the occasionally incomplete fossil specimens selected for analysis. In order to include specimens Sts 71 and KNM-WT 15000, missing landmarks were estimated through reflection of antimeres.

Table 3: Landmark definitions for the subset of landmarks used in morphometric comparison

Landmark	Definition
1	Rhinion
2	Nasion
3	Glabella
4	Bregma
5	Anterior attachment of nasal septum
6	Prosthion
7,8	Mid-torus inferior (right and left)
9,10	Mid-torus superior (right and left)
11,12	Frontomalare orbitale (left and right)
13,14	Frontomalare temporale (left and right)
15,16	Dacryon (left and right)
17,18	Zygoorbitale (right and left)
19,20	Zygomaxillare (right and left)
21,22	Alare (right and left)
23,24	Malar root origin (right and left)
25,26	I1-I2 contact
27,28	C-P3 contact
29	Incisivion
30	Alveolon
31	Left distal palate
32	Right distal palate

Principal component analysis (PCA) of covariances was conducted on the Procrustes shape coordinates using the subset of 32 landmarks collected from the extant hominoid and fossil sample. The adult gorilla sample was excluded from the PCA because of concerns that, if included, these specimens would largely distort the results as a result of their more divergent morphology compared to other species included in the sample.[18,21,28] Analyses were conducted both with and without the extant hominoids to control for the likely dominance of these samples in driving definition of the major principal component axes. In other words, one could ask, to what extent was the position of the fossil specimens affected by the ordination of modern chimpanzees and humans in the same data space? To control for this effect, we also conducted a PCA using only fossil specimens to assess the distribution of crania in the context of fossil hominin variation.

Average Procrustes chord distances both within and between taxa were also calculated using Excel software on the subset of 32 landmarks for the same chimpanzee, human and fossil sample that was used in the PCA. The purpose of this test was to determine if the overall distance within the *A. sediba* sample (i.e. intragroup variation), including all simulated adults and the reconstructed juvenile cranium, exceeds that observed in extant hominoid species. One can further compare the average distance between *A. sediba* and other individual hominin species to assess group similarities.

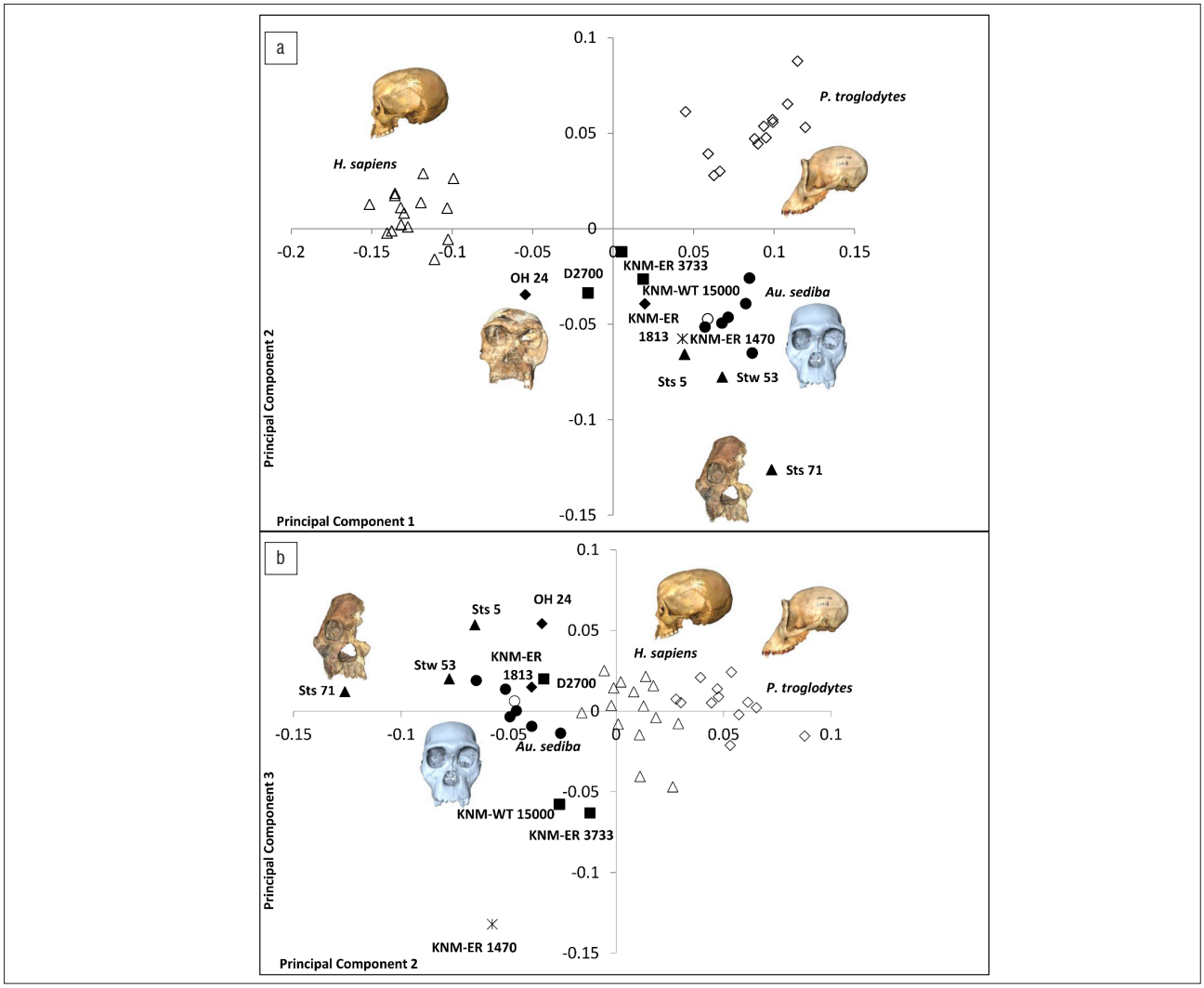
Results and discussion

Developmental simulations

Qualitative assessment of the simulated adult crania indicated that the majority of morphological changes expected to occur between second and third molar eruption are related to the development of secondary sexual characteristics (Figure 2a,b). The most significant changes were observed along the male gorilla developmental vector, for which we see the glabella becoming more pronounced, the supraorbital torus thickening, and the zygomatic increasing in its superior-inferior dimension. Increased horizontal angulation, or bending across the mid-face, is also apparent, as is increased lower facial prognathism. Similar transformations were also observed, although to a lesser extent, along the male chimpanzee developmental vector, where one again observes development of moderate lower facial prognathism and a general enhancement of facial robusticity, such as a slight thickening of the supraorbital torus and glabella, as well as a more rugged appearance in the zygomatics. The magnitude of morphological change observed for the female developmental vectors of the gorilla and chimpanzee simulations was comparatively much less. Changes associated with both the male and female human developmental vectors were minimal, producing no notable deviations from the original juvenile form. Thus, while the choice of sex and species did result in observable differences in the estimated adult form of MH1 when applying the chimpanzee and gorilla developmental vectors, applying the human developmental vectors to MH1 resulted in relatively more modest variation in adult simulated form. Given that *A. sediba* evinces a pattern of canine size dimorphism and facial robusticity similar to other australopiths and to early *Homo*,[1,9] we consider the gorilla developmental pattern, in particular that of the male gorilla, to be the least suitable model, while chimpanzee or human developmental patterns would likely provide better estimates of the onset of secondary sexual characteristics in MH1.

Morphometric comparisons

Results of the PCA are illustrated in Figure 3 using adult chimpanzee, human and non-robust hominin crania, including MH1 and its simulated adult conformations. Visualisations of principal component shape change for the first three principal component axes are provided in Figure 4. The first axis of variation is dominated by differences in the length and orientation of the frontal bone, in accordance with the high loading for the bregma landmark, whereas the second axis also summarises changes in cheek morphology, with specimens separated based on the overall gracility or robusticity of the zygomatic. This latter interpretation is based on the high loadings for zygomaxillary landmarks as well as the relative distribution of specimens. The third axis is dominated by differences in morphology of the anterior and posterior palate, with the highest loadings being observed for landmarks in this region. These interpretations correspond to the visualisations of shape change for each axis provided in Figure 4.



A. africanus *(filled triangles);* A. sediba *simulated adults (filled circles); MH1 juvenile (open circle);* Homo habilis *(filled diamonds);* H. rudolfensis *(star);* H. erectus *(filled squares);* H. sapiens *(open triangles);* Pan troglodytes *(open diamonds).*

Figure 3: Major (principal component) axes of cranial shape for hominoids, including simulated adult crania of *Australopithecus sediba*. (a) Principal components 1 and 2 (58.7% and 11.0% of variance, respectively). (b) Principal components 2 and 3 (11.0% and 6.5% of total variance, respectively).

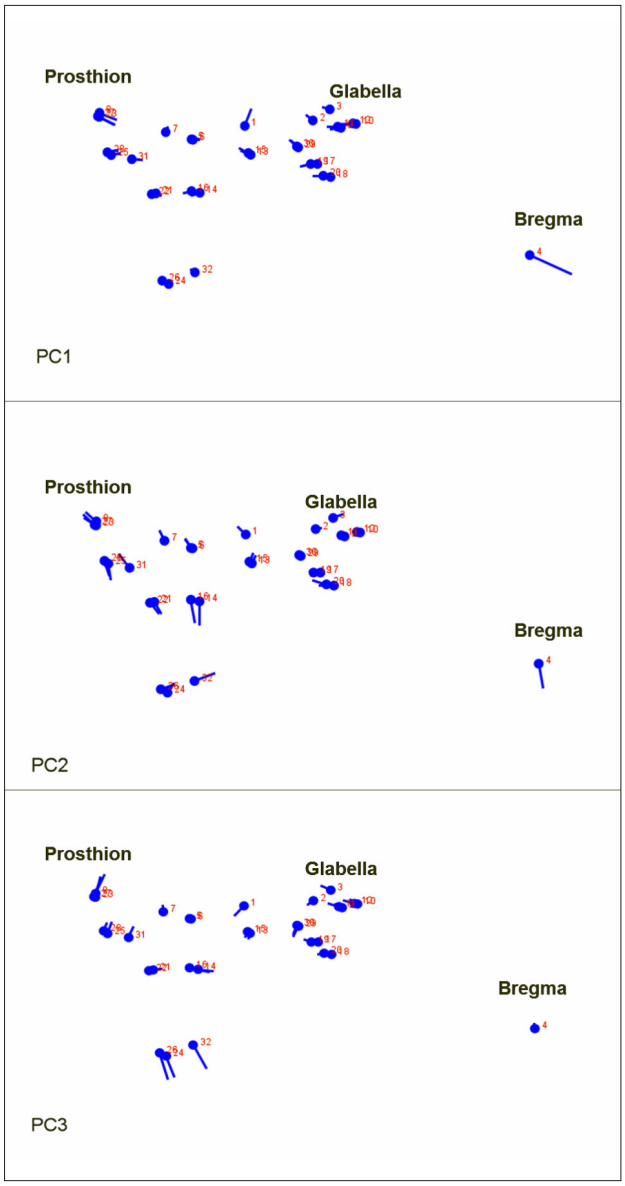


Figure 4: Visualisations of shape change for principal component axes 1–3. Axes correspond to Figure 3.

All six simulated adult *A. sediba* crania – regardless of the species or sex used to estimate them – fall out along both axes with the original, juvenile MH1 cranium in a discrete group relative to other apes or fossil taxa in the analysis (Figure 3a). Early *Homo* crania tend to cluster together with the exception of KNM-ER 1470, which appears to cluster with Sts 5 on the first two major axes (Figure 3a). However, this proximity vanishes when we examine the third principal component, wherein KNM-ER 1470 plots as an outlier (Figure 3b). In addition, considering component 2 versus component 3 (Figure 3b), KNM-ER 1813, OH 24, and the Dmanisi specimen D2700 are separated from *H. erectus sensu lato* crania KNM-WT 15000 and KNM-ER 3733, instead plotting near the cluster of simulated *A. sediba* crania. One outlier to this latter group is the *A. sediba* cranium simulated using a male gorilla developmental vector, which plots in this space nearest the *A. africanus* cranium Stw 53. The exceptional location for the MH1 adult version simulated from male gorillas is not entirely surprising, given that one observes the greatest degree of morphological change when applying this vector, although again we consider a male gorilla-like pattern of development to be the least likely for *A. sediba*. Of perhaps more interest is the fact that a male chimpanzee developmental trajectory makes MH1 appear more similar to early *Homo* and modern humans.

A second PCA on the fossil sample, excluding chimpanzees and humans, is illustrated in Figure 5, with visualisations of shape change for the first three principal component axes provided in Figure 6. Similar to the PCA including chimpanzees and humans, the first axis of variation is dominated by changes in morphology of the frontal bone, but is also influenced by morphological variation of the posterior palate, likely related to the relative degree of facial prognathism among specimens. The second axis is also largely dominated by the morphology of the posterior palate, with high loadings for landmarks in this region. While the *A. sediba* crania generated using chimpanzee and modern human developmental vectors continue to cluster closely to one another, the simulated adult produced from the male gorilla developmental vector shows dissimilarity relative to other *A. sediba* specimens, plotting nearest *A. africanus* specimen Sts 71 (Figure 5a).

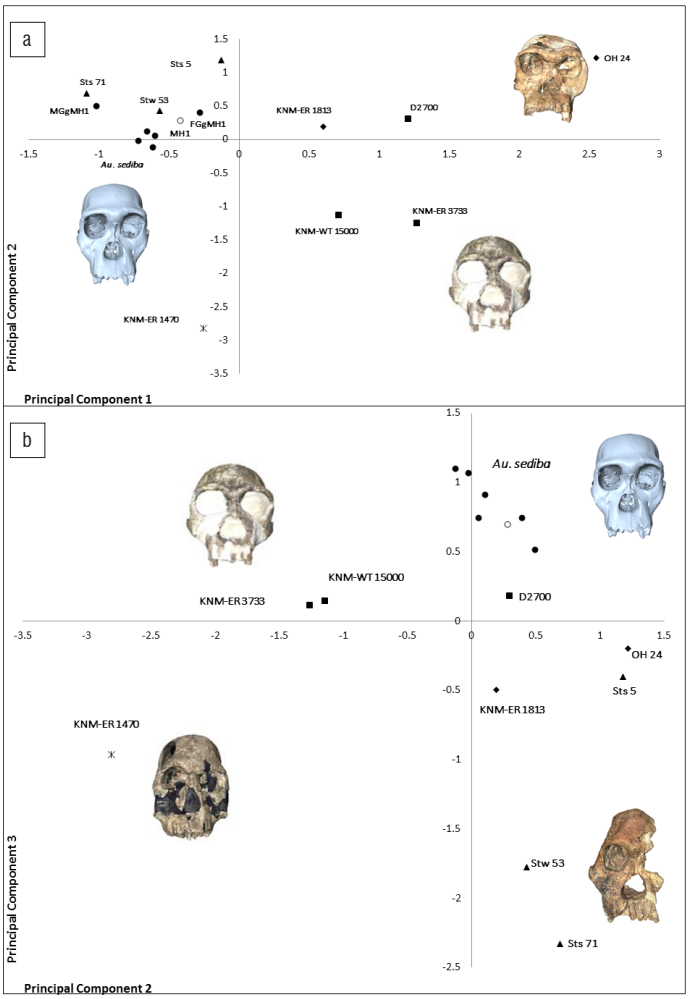
As the third principal component also accounts for a substantial amount of the total variation (15.2%) in the PCA using only fossils, we examined the distribution of fossil taxa along this axis. This third axis of variation was also influenced primarily by changes in the frontal bone and palate. When comparing principal component 2 versus component 3, all *A. sediba* estimated adults, along with the original juvenile cranium, clustered together, separated from *A. africanus* and other hominin taxa along the third axis of variation (Figure 5b).

The results of the above PCA are supported across the full shape space by an analysis of shape differences within and among taxa using average Procrustes distances (Table 4). Examination of the Procrustes chord distances indicates that the within-species variation for *A. sediba* simulated adults and the juvenile cranium is actually less than that of the modern human and chimpanzee samples used in the study.

Conclusions

The results of this study support the hypothesis that the expected degree and nature of development that would have occurred between second and third molar eruption in MH1 would not have been substantial enough relative to other hominin taxa to alter our current craniofacial understanding of the phylogenetic relationships of *A. sediba*. In other words, had MH1 lived to adulthood, its craniofacial morphology would look similar enough to its present, juvenile cranial morphology that we can reliably differentiate the taxon using the cranium of this particular specimen. Any future changes would likely be related to the onset of puberty, and the concomitant development of secondary sexual characteristics. As a result, this study further reinforces the importance of separating developmental vectors by sex to avoid obscuring this type of variation in ontogenetic projections.

The fact that the juvenile MH1 cranium continues to cluster with simulated *A. sediba* adults in PCAs conducted in this study reinforces the above position. While we cannot ultimately rule out the scenario that qualitative apomorphies shared with other fossil taxa (e.g. *A. africanus*) might have developed and become expressed only during this later stage of ontogeny, the expression of such features in developmentally much younger fossil hominin specimens[29,30] argues against the likelihood of this possibility. For example, anterior pillars are already present and easily observable in the Taung child, for which only the first molars are in occlusion.[30] Therefore, the late juvenile status of the type specimen, MH1, would not likely influence our diagnosis of this species. These traits on the reconstructed cranium of MH1 and its ontogenetic simulated extrapolations concur with those for the mandibles of MH1 and MH2.[9] Both studies indicate the verity of *A. sediba* as a unique taxon. Based upon our current understanding of the postcranial anatomy of the Malapa hominins, which indicates an *Australopithecus*-level adaptive grade, the *Homo*-like features observed in the simulated adults generated when applying the *H. sapiens* and *P. troglodytes* trajectories reinforce the transitional nature of this species.



Australopithecus africanus *(filled triangles);* A. sediba *simulated adults (filled circles); MH1 juvenile (open circle);* Homo habilis *(filled diamonds);* H. rudolfensis *(star);* H. erectus *(filled squares).*

MGgMH1 indicates the adult A. sediba *specimen generated from the male gorilla developmental vector; FGgMH1 indicates the adult* A. sediba *specimen generated from the female gorilla developmental vector.*

Figure 5: Major (principal component) axes of cranial shape for non-robust fossil crania. (a) Principal components 1 (28.3% of variance) and 2 (21.7% of variance). (b) Principal components 2 and 3 (15.2% of variance).

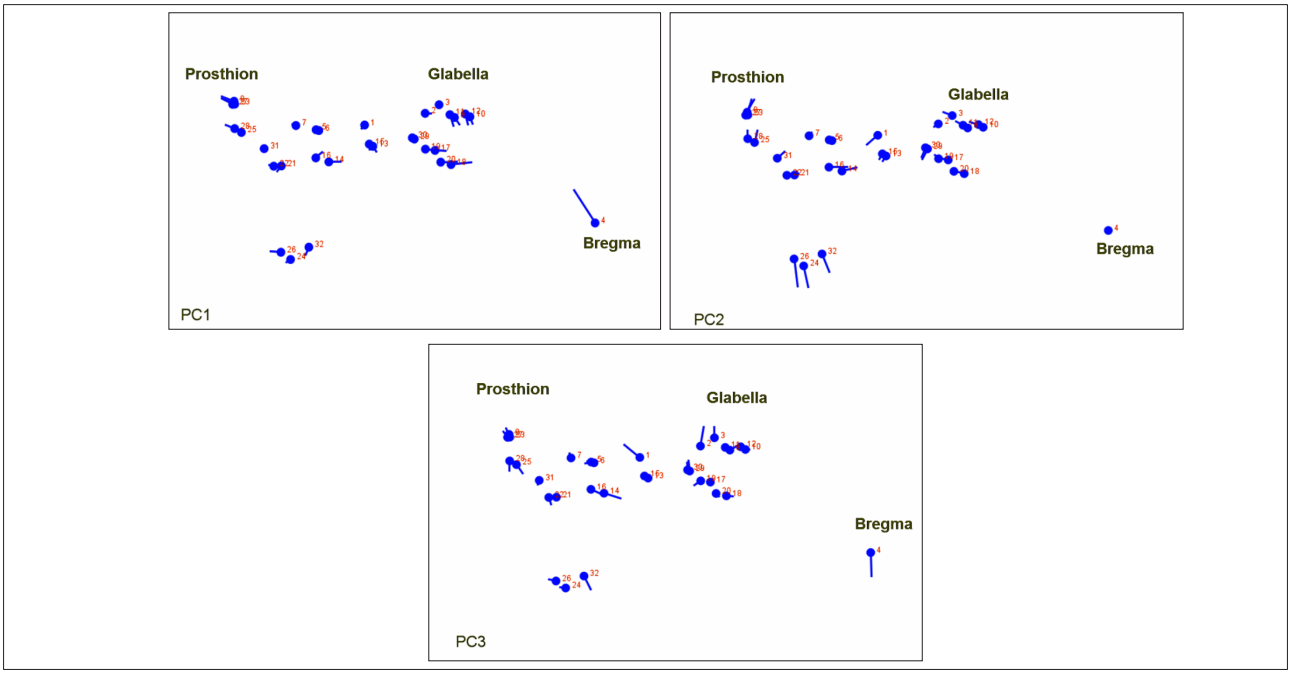


Figure 6: Visualisations of shape change for principal component axes 1–3. Axes correspond to Figure 5.

Table 4: Average Procrustes chord distances within taxa (on diagonal) and among taxa (off diagonal)

	Australopithecus africanus	*Australopithecus sediba*	*Homo erectus*	*Homo habilis*	*Homo sapiens*	*Pan troglodytes*	*Homo rudolfensis*
A. africanus	**0.145**						
A. sediba	0.101	**0.059**					
H. erectus	0.138	0.112	**0.144**				
H. habilis	0.137	0.138	0.098	**0.163**			
H. sapiens	0.226	0.21	0.149	0.146	**0.089**		
P. troglodytes	0.152	0.11	0.133	0.159	0.218	**0.092**	
H. rudolfensis	0.188	0.173	0.158	0.208	0.236	0.2	–

Bolded numbers indicate average within species variation. The range of variation within P. troglodytes *is 0.0597–0.14093. The range of variation within* H. sapiens *is 0.06654–0.12589. Distances were calculated using Excel software for the subset of 32 landmarks collected from the adult* P. troglodytes, H. sapiens *and fossil sample.*

Acknowledgements

We thank the South African Heritage Resource agency for permits to work on the Malapa site, and the Nash family for granting access to the Malapa site and the continued support of research on their reserve. The South African Department of Science and Technology, and the African Origins Platform, the South African National Research Foundation, the Evolutionary Studies Institute, the Palaeontological Scientific Trust, the Andrew W. Mellon Foundation, the United States Diplomatic Mission to South Africa, the National Geographic Society, the A. H. Schultz Foundation, the Oppenheimer and Ackerman families, Sir Richard Branson, Sigma Xi, Texas Academy of Science, Vision 2020 Dissertation Enhancement Award of Texas A&M University, and the Program to Enhance Scholarly and Creative Activities and the International Research Travel Award Grant of Texas A&M University all provided funding for this research. D.J.d.R. holds a Liberal Arts Cornerstone Faculty Fellowship at Texas A&M University, which provided additional funding. The University of the Witwatersrand's Schools of Geosciences and Anatomical Sciences and the Evolutionary Studies Institute provided support and facilities. We thank the European Synchrotron Radiation Facility for granting beam time on ID17 for experiment #EC521. We thank E. Mbua, P. Kiura, V. Iminjili, and the National Museums of Kenya, Dr. P. Msemwa and the National Museum and House of Culture of Tanzania, S. Potze of the Ditsong Museum, B. Billings of the School of Anatomical Sciences of the University of the Witwatersrand, W. Seconna of the Iziko South African Museum, and L.M. Jellema of the Cleveland Museum of Natural History for access to comparative fossil and extant primate and human materials in their care. Numerous individuals have been involved in the ongoing preparation and excavation of these fossils including C. Dube, B. Eloff, C. Kemp, M. Kgasi, M. Languza, J. Malaza, G. Mokoma, P. Mukanela, T. Nemvhundi, M. Ngcamphalala, S. Jirah, S. Tshabalala and C. Yates. Other individuals who have given significant support to this project include B. de Klerk, C. Steininger, B. Kuhn, L. Pollarolo, B. Zipfel, J. Kretzen, D. Conforti, J. McCaffery, C. Dlamini, H. Visser, R. McCrae-Samuel, B. Nkosi, B. Louw, L. Backwell, F. Thackeray, M. Peltier, J. Soderberg and D. Roach. The *A. sediba* specimens are archived at the Evolutionary Studies Institute at the University of the Witwatersrand. All data used in this study are available upon request, including access to the original specimens, by bona fide scientists.

Authors' contributions

All authors contributed to the work presented in this paper and the writing of the document. K.B.C., D.J.d.R., T.J.D. and K.P.M. conceived the project. K.B.C., T.J.D. and K.P.M. carried out all statistical analyses and comparisons of the MH1 fossil and simulated adults. P.T. and L.R.B. carried out the synchrotron scan of the MH1 fossil cranium. K.J.C. analysed and segmented the synchrotron scan data providing the 3D model that was used for reconstruction and comparison. K.B.C. and K.P.M. reconstructed the synchrotron scan of the MH1 cranium, correcting for cracks and displacements in the fossil.

References

1. Berger LR, De Ruiter DJ, Churchill SE, Schmid P, Carlson KJ, Dirks PHGM, et al. *Australopithecus sediba*: A new species of *Homo*-like australopith from South Africa. Science. 2010;328:195–204. http://dx.doi.org/10.1126/science.1184944
2. Pickering R, Dirks PHGM, Jinnah Z, De Ruiter DJ, Churchill SE, Herries AIR, et al. *Australopithecus sediba* at 1.977 M and implications for the origins of the genus *Homo*. Science. 2011;333:1421–1423. http://dx.doi.org/10.1126/science.1203697
3. Tobias PV. The brain in hominid evolution. New York: Columbia University Press; 1971.
4. Balter M. Candidate human ancestor from South Africa sparks praise and debate. Science. 2010;328:154–155. http://dx.doi.org/10.1126/science.328.5975.154
5. Cherry M. Claim over 'human ancestor' sparks furore. Nature. 2010 April 08; News. http://dx.doi.org/10.1038/news.2010.171
6. Spoor F. Malapa and the genus *Homo.* Nature. 2011;478:44–45. http://dx.doi.org/10.1038/478044a
7. Wood B, Harrison T. The evolutionary context of the first hominins. Nature. 2011;470:347–352. http://dx.doi.org/10.1038/nature09709
8. Kimbel WH. Hesitation on hominin history. Nature. 2013;497:573–574. http://dx.doi.org/10.1038/497573a
9. De Ruiter DJ, DeWitt TJ, Carlson KB, Brophy JK, Schroeder L, Ackermann RR, et al. Mandibular remains support taxonomic validity of *Australopithecus sediba*. Science. 2013;340(6129). http://dx.doi.org/10.1126/science.1232997
10. McNulty KP, Frost SR, Strait DS. Examining affinities of the Taung child by developmental simulation. J Hum Evol. 2006;51:274–296. http://dx.doi.org/10.1016/j.jhevol.2006.04.005
11. Neubauer S, Gunz P, Hublin J. Endocranial shape changes during growth in chimpanzees: A morphometric analysis of unique and shared aspects. J Hum Evol. 2010;59:555–566. http://dx.doi.org/10.1016/j.jhevol.2010.06.011
12. Gunz P, Neubauer S, Maureille B, Hublin JJ. Brain development after birth differs between Neanderthals and modern humans. Curr Biol. 2010;20:921–922. http://dx.doi.org/10.1016/j.cub.2010.10.018
13. Singleton M, McNulty KP, Frost SR, Soderberg J, Guthrie EM. Bringing up baby: Developmental simulation of the adult cranial morphology of *Rungwecebus kipunji*. Anat Rec. 2010;293:388–401. http://dx.doi.org/10.1002/ar.21076
14. Gunz P, Bulygina E. The Mousterian child from TeshikTash is a Neanderthal: A geometric morphometric study of the frontal bone. Am J Phys Anthropol. 2012;149:365–379. http://dx.doi.org/10.1002/ajpa.22133
15. Gunz P, Neubauer S, Golovanova L, Doronichev V, Maureille B, Hublin JJ. A uniquely modern human pattern of endocranial development. Insights from a new cranial reconstruction of the Neandertal newborn from Mezmaiskaya. J Hum Evol. 2012;62:300–313. http://dx.doi.org/10.1016/j.jhevol.2011.11.013

16. Scott N, Neubauer S, Hublin JJ, Gunz P. A shared pattern of postnatal endocranial development in extant hominoids. Evol Biol. 2014;41:572–594. http://dx.doi.org/10.1007/s11692-014-9290-7

17. Carlson KJ, Stout D, Jashashvili T, De Ruiter DJ, Tafforeau P, Carlson K, et al. The endocast of MH1, *Australopithecus sediba*. Science. 2011;333:1402–1407. http://dx.doi.org/10.1126/science.1203922

18. Shea BT. The ontogeny of sexual dimorphism in the African apes. Am J Primatol. 1985;8:183–188. http://dx.doi.org/10.1002/ajp.1350080208

19. Shea BT. Ontogenetic approaches to sexual dimorphism in anthropoids. Hum Evol. 1986;1(2):97–110. http://dx.doi.org/10.1007/BF02437489

20. Mitteroecker P, Gunz P, Bernhard M, Schaefer K, Bookstein FL. Comparison of cranial ontogenetic trajectories among great apes and humans. J Hum Evol. 2004;46:679–698. http://dx.doi.org/10.1016/j.jhevol.2004.03.006

21. Schaefer K, Mitteroecker P, Gunz P, Bernhard M, Bookstein FL. Craniofacial sexual dimorphism patterns and allometry among extant hominids. Ann Anat. 2004;186:471–478. http://dx.doi.org/10.1016/S0940-9602(04)80086-4

22. Wiley DF, Amenta N, Alcantara DA, Ghosh D, Kil YJ, Delson E, et al. Evolutionary morphing. Proceedings of IEEE Visualization 2005; 2005 October 23–28; Minneapolis, MN, USA. New York: IEEE; 2005. p. 431–438. http://dx.doi.org/10.1109/VISUAL.2005.1532826

23. Klingenberg CP. MorphoJ: An integrated software package for geometric morphometrics. Mol Ecol Resour. 2011;11(2):353–357. http://dx.doi.org/10.1111/j.1755-0998.2010.02924.x

24. Gower JC. Generalized Procrustes analysis. Psychometrika. 1975;40:33–51. http://dx.doi.org/10.1007/BF02291478

25. Bookstein FL. Size and shape spaces for landmark data in two dimensions. Stat Sci. 1986;1:181–222. http://dx.doi.org/10.1214/ss/1177013696

26. Rohlf FJ. Fitting curves to outlines. In: Rohlf FJ, Bookstein FL, editors. Proceedings of the Michigan Morphometrics Workshop. Special Publication no. 2. Ann Arbor, MI: University of Michigan Museum of Zoology; 1990. p. 167–177.

27. Adams DC, Rohlf FJ, Slice DE. Geometric morphometrics: Ten years of progress following the 'revolution'. Ital J Zool. 2004;71:5–16. http://dx.doi.org/10.1080/11250000409356545

28. O'Higgins P, Dryden IL. Sexual dimorphism in hominoids: Further studies of craniofacial shape differences in *Pan, Gorilla* and Pongo. J Hum Evol. 1993;24:183–205. http://dx.doi.org/10.1006/jhev.1993.1014

29. Clarke RJ. Latest information on Sterkfontein's *Australopithecus* skeleton and a new look at *Australopithecus.* S Afr J Sci. 2008;104:443–449.

30. Rak Y. The australopithecine face. New York: Academic Press; 1984.

AUTHORS:
Patrick S. Randolph-Quinney[1,2*]
Scott A. Williams[2,3]
Maryna Steyn[1]
Marc R. Meyer[4]
Jacqueline S. Smilg[2,5,6]
Steven E. Churchill[2,7]
Edward J. Odes[1,2]
Tanya Augustine[1]
Paul Tafforeau[8]
Lee R. Berger[2,9]

AFFILIATIONS:
[1]School of Anatomical Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa
[3]Center for the Study of Human Origins, Department of Anthropology, New York University, New York, New York, USA
[4]Department of Anthropology, School of Social & Behavioral Sciences, Chaffey College, Rancho Cucamonga, California, USA
[5]School of Radiation Sciences, University of the Witwatersrand, Johannesburg, South Africa
[6]Department of Radiology, Charlotte Maxeke Academic Hospital, Johannesburg, South Africa
[7]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, USA
[8]European Synchrotron Radiation Facility, Grenoble, France
[9]DST/NRF South African Centre of Excellence in Palaeosciences, University of the Witwatersrand, Johannesburg, South Africa
*Current address: School of Forensic and Applied Sciences, University of Central Lancashire, Preston, Lancashire, United Kingdom

CORRESPONDENCE TO:
Patrick Randolph-Quinney

EMAIL:
prandolph-quinney@uclan.ac.uk

POSTAL ADDRESS:
School of Forensic and Applied Sciences, University of Central Lancashire, Preston, Lancashire, PR1 2HE, UK

DATES:
Received: 11 Dec. 2015
Revised: 16 Mar. 2016
Accepted: 17 Mar. 2016

KEYWORDS:
Malapa; palaeopathology; neoplasia; taphonomy; osteoma; malignant

HOW TO CITE:
Randolph-Quinney PS, Williams SA, Steyn M, Meyer MR, Smilg JS, Churchill SE, et al. Osteogenic tumour in *Australopithecus sediba*: Earliest hominin evidence for neoplastic disease. S Afr J Sci. 2016;112(7/8), Art. #2015-0470, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20150470

Osteogenic tumour in *Australopithecus sediba*: Earliest hominin evidence for neoplastic disease

We describe the earliest evidence for neoplastic disease in the hominin lineage. This is reported from the type specimen of the extinct hominin *Australopithecus sediba* from Malapa, South Africa, dated to 1.98 million years ago. The affected individual was male and developmentally equivalent to a human child of 12 to 13 years of age. A penetrating lytic lesion affected the sixth thoracic vertebra. The lesion was macroscopically evaluated and internally imaged through phase-contrast X-ray synchrotron microtomography. A comprehensive differential diagnosis was undertaken based on gross- and micro-morphology of the lesion, leading to a probable diagnosis of osteoid osteoma. These neoplasms are solitary, benign, osteoid and bone-forming tumours, formed from well-vascularised connective tissue within which there is active production of osteoid and woven bone. Tumours of any kind are rare in archaeological populations, and are all but unknown in the hominin record, highlighting the importance of this discovery. The presence of this disease at Malapa predates the earliest evidence of malignant neoplasia in the hominin fossil record by perhaps 200 000 years.

Introduction

A neoplasm ('new-growth' or tumour) is defined as a mass of localised tissue growth, the cellular proliferation of which is no longer subject to the effects of normal growth-regulating mechanisms.[1-3] A neoplasm may be benign or malignant. Malignant tumours are often referred to colloquially as cancer, although the term 'malignant neoplasia' is more clinically appropriate.[1] In the developed world, death from malignancy is second only to cardiovascular disease and is often perceived as a disease of modernity.[4] Neoplastic disease would have been prevalent in the past (e.g. Odes et al.[5]), but most likely occurred at much lower levels of incidence than today, given the shorter life expectancy for victims[1,6,7] and the differing environmental context. Both these factors strongly influence the incidence and prognosis of any cancer.[3,8] The preserved signatures of neoplasms of any kind are rare in archaeological populations, and are all but unknown in the hominin record. Here we present the earliest fossil evidence for neoplastic disease in the human lineage, with a detailed description and diagnosis of a tumorous lesion affecting the spine of a juvenile male *Australopithecus sediba,* Malapa Hominin 1 (MH1).[9,10] This species has been postulated as a possible ancestor of the genus *Homo.*[9] The clinical and evolutionary implications of the diagnosed condition are discussed.

The Malapa hominin site

The Malapa site is one of several hominin-bearing Plio-Pleistocene cave deposits located within the Cradle of Humankind World Heritage Site to the northwest of Johannesburg, South Africa. The region includes sites such as Sterkfontein[11], Swartkrans[12], Kromdraai[13], Gladysvale[14] and Rising Star[15]. The fossil deposits in these caves were formed in roughly similar fashion as debris cone accumulations deposited beneath vertical cave openings, which formed phreatically within the dolomites of the Malmani Subgroup.[15,16] At Malapa, the main hominin-bearing deposits have been dated using uranium-lead dating of flowstones, combined with palaeomagnetic and stratigraphic analyses of flowstones and underlying sediments, to 1.977 ± 0.002 million years ago (Ma).[17] The cave deposits comprise five sedimentary facies, termed A to E, from stratigraphically lowest to highest.

Facies A and B occur below a central flowstone sheet, and are overlain by an erosion remnant (facies C), which in turn is overlain by the main hominin-bearing breccia, facies D. This has yielded well-preserved macro- and micro-mammal fossils (such as carnivores, equids and bovids[18]), including the fossilised remains of at least six hominins. Two of these, MH1 and MH2, have been reported in the literature as representatives of a new hominin species, *Australopithecus sediba*[9]. Taphonomically the site has been interpreted as a complex cave system with open deep vertical shafts that operated as death traps for animals on the surface of the landscape. This death-trap scenario might have been the process by which the Malapa hominins entered the cave system[17,18], as evidenced by peri-mortem damage on the skeletons of MH1 and MH2, consistent with a fatal fall[19]. Furthermore, both skeletons present partial anatomical articulation consistent with rapid incorporation into the cave sediments early in the decomposition process.[18]

Case study: Vertebra U.W. 88-37

A pathological lesion affects the spine of Malapa Hominin 1 (MH1), the type specimen of *Australopithecus sediba*. This individual (Figure 1) was male, and at death he was at a developmental stage equivalent to that of a human child aged 12 to 13 years[9]. The pathological specimen (U.W. 88-37) is a complete vertebra originally assigned to T5-T7[10], now considered to represent the sixth thoracic vertebra[10]. The dorsal surface of the right-side lamina exhibits a rounded penetrating defect (Figure 2), measuring approximately 6.7 mm supero-inferiorly and 5.9 mm medio-laterally.

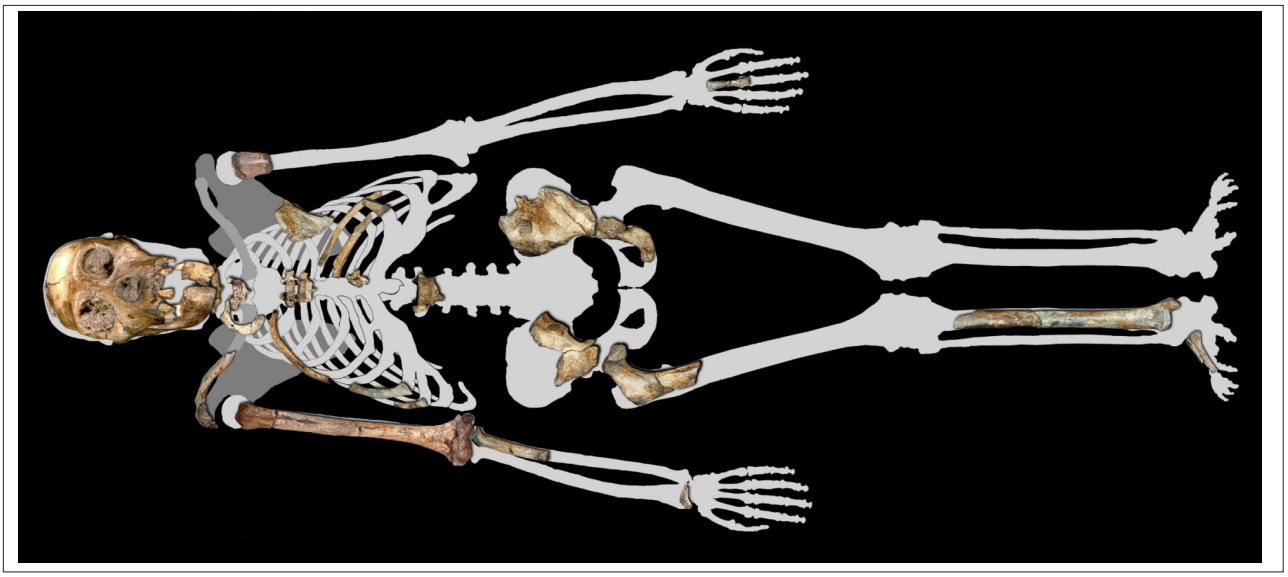


Figure 1: Surviving skeletal elements attributed to Malapa Hominin 1 (at time of writing).

Figure 2: Vertebra U.W. 88-37. Photographs of surface morphology of U.W. 88-3 showing position of lesion on right side of vertebral lamina: (a) right lateral aspect, (b) left lateral aspect, (c) inferior aspect, (d) superior aspect, (e) posterior aspect, (f) anterior aspect. Note that apertures seen on lateral aspects of the vertebral body in images (a), (b) and (f) represent normal vascular foramina infilled with residual breccia matrix. Images produced by Peter Schmidt.

The defect presents as a lytic lesion that extends ventrally into the lamina for much of its length, the most anterior portion of which remains infilled with breccia matrix (Figure 3). On the surface, the lesion has well-rounded edges with a somewhat sclerotic appearance. There is no evidence of periosteal or reactive bone formation on the cortex of the specimen. Viewing the right lamina from above, it appears thicker than the left lamina and bulges laterally over the lesion, indicating a reactive remodelling response to the presence of the defect.

Figure 3: Vertebra U.W. 88-37. Multi-focus (composite image stack) micrograph of surface morphology of U.W. 88-37 showing sub-angular penetrating defect on the right vertebral lamina. The lesion has well-rounded edges with lateral bulging of the cortex over the lesion, indicating a reactive remodelling response to the presence of the defect. Note that anterior portion of defect remains infilled with breccia matrix. Micrograph taken with Olympus SZX Multi-focus microscope, magnification 7x. Scale bar = 10 mm. Image courtesy of Alexander Parkinson.

The lesion initially widens directly under the oval opening, but then narrows as it progresses anteriorly. The base of the lesion appears smooth and sclerotic under microscopic evaluation insofar as the presence of residual breccia allows. The spinous process deviates slightly to the right, but appears in keeping with slight asymmetry noted elsewhere in the surviving thoracic vertebrae. This deviation falls within normal variation; we do not consider it significant enough to cause scoliosis or other vertebral misalignment, and it is unlikely that this asymmetry was related to the pathology.

Because of the presence of breccia within the lesion, the internal morphology of the specimen was assessed using phase-contrast X-ray synchrotron microtomography (performed at the European Synchrotron Radiation Facility, ESRF) and a specific acquisition protocol applied for high-quality imaging of large fossils (see Supplementary Appendix materials and methods). From the microtomographic volume, the maximum long axis of the lesion in the transverse plane measures 11.8 mm x 4.9 mm along the minor axis, with a cross-sectional area of 45.6 mm, and in the sagittal plane the lesion measures 14.7 mm x 7.9 mm, with a cross-sectional area of 68.6 mm. The internal linear dimensions are consistently less than 20 mm in diameter, which has important implications for final diagnosis.

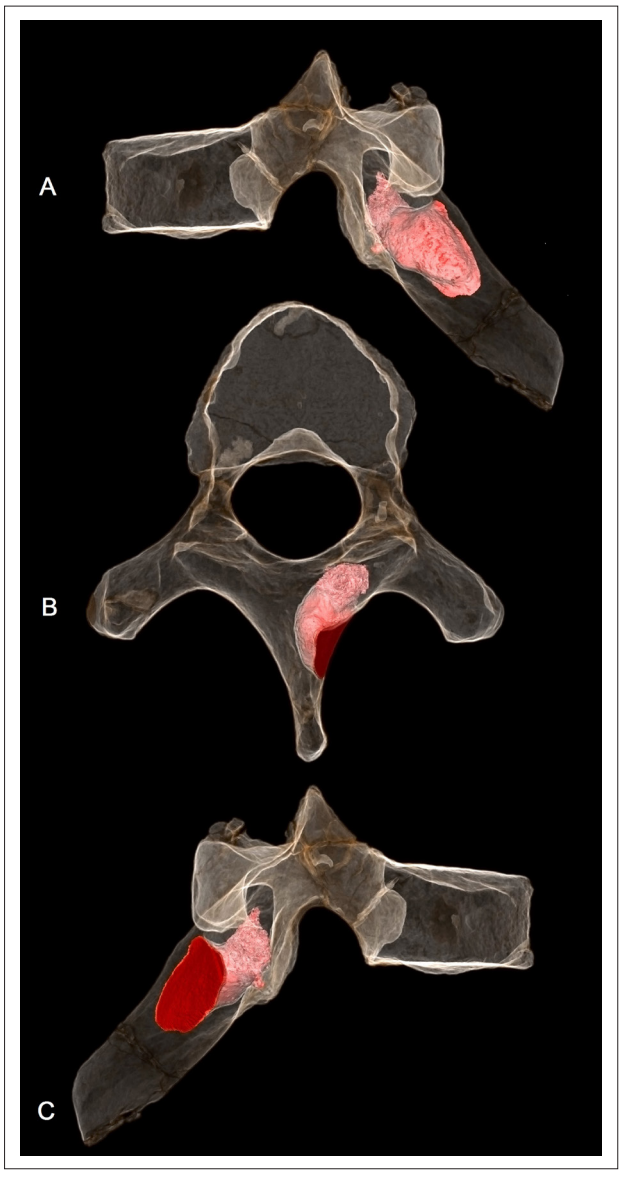


Figure 4: Vertebra U.W. 88-37. Sixth thoracic vertebra of juvenile *Australopithecus sediba* (Malapa Hominin 1). Partially transparent image volume with the segmented boundaries of the lesion rendered solid pink. Volume data derived from phase-contrast X-ray synchrotron microtomography. (a) left lateral view, (b) superior view, (c) right lateral view. Images produced by P.T.

Figure 4 shows the microtomographic imaging, with a semi-transparent volume-rendered image row. The imaging indicates that the lesion is highly penetrative and extends ventrally within the right-side of the spinous process, penetrating the lamina before terminating at the approximate level of the superior articular facet. The internal morphology shows no involvement of the transverse process or pedicle, and the lesion does not penetrate the vertebral canal. No mineralised focal point or nidus was discerned. The edges of the first two-thirds of the lesion (moving dorsal to ventral) display sclerotic characteristics, with circumscribed margins of well-integrated cortical bone, abutted and intersected by trabecular striae (Figure 5 and Supplementary Appendix). This pattern is indicative of a slow-forming bony process, with remodelling and reorganisation of posterior aspects of the lesion. The shape of a lesion is indicative of its growth rate, with lesions that are long and oriented with the long axis of a bone indicating a nonaggressive benign process. The ventral third of the lesion, however, displays a geographic pattern of bone destruction, showing a sharp non-sclerotic margin and evidence of active osteolytic processes, with sharply-defined transection of individual trabeculae, and active osteolytic penetration into the anterior portion of the lamina. A volume-rendered negative surface model of the lesion (Figure 6) demonstrates the clear distinction between the dorsal sclerotic zone and the ventral lytic zone within the body of the active lesion.

Key: S – quiescent sclerotic zone, O – active osteolytic zone, B – remaining breccia matrix infill.

Figure 5: Transverse slices through vertebra U.W. 88-37 derived from phase-contrast X-ray synchrotron microtomography. Relative position and anatomical orientation of orthoslices (a), (b) and (c) shown on the volume-rendered model. The posterior portion of the lesion is sclerotic with circumscribed margins of well-integrated cortical bone, abutted and intersected by trabecular striae, with remodelling and reorganisation of the cortex. The anterior portion of the lesion displays a geographic pattern of bone destruction, showing a sharp non-sclerotic margin and evidence of active osteolytic processes, with sharply defined transection of individual trabeculae and active osteolytic penetration into the anterior portion of the lamina. Image produced by P.S.R.Q.

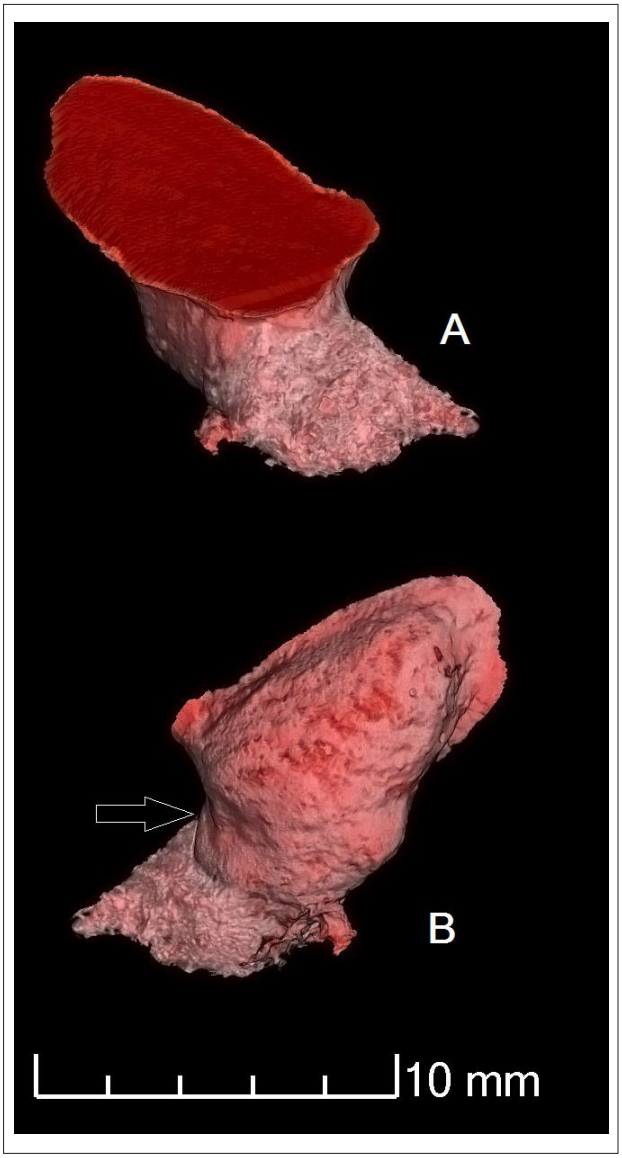


Figure 6: Surface rendered image volume of the U.W. 88-37 lesion derived from phase-contrast X-ray synchrotron microtomography. Images show isosurface derived from segmented boundaries of the lesion (remaining breccia infill removed). The arrow denotes the interface between the smoother dorsal sclerotic zone and the disorganised ventral lytic zone within the body of the lesion. (a) right lateral view, (b) medial view. Images produced by P.T.

Differential diagnosis

Diagnosis was undertaken using palaeopathological and clinical diagnostic criteria[1,2,20-34]. The accumulated evidence for osteolytic and osteosclerotic processes indicates that the disease process was both chronic and active at the time of death of MH1 (as mentioned, at a developmentally equivalent stage to a modern human child of 12 to 13 years of age). The lesion was less than 15 mm at the largest diameter, extending deep into the right side of the spinous process and involving only the vertebral lamina. The presence of reorganised sclerotic bone indicates a reactive ante-mortem process, and the lesion can therefore not be attributed to taphonomic, diagenetic or pathology-mimicking effects or processes[1].

The morphology of the lesion externally and internally is inconsistent with vertebral osteomyelitis. The absence of a proliferative cortical inflammatory response (such as periosteal and/or endosteal bone hypertrophy) or secondary lytic lesions across both the U.W. 88-37 vertebra and the surviving cranial and post-cranial elements of MH1 excludes a diagnosis of specific or non-specific systemic infection, such as brucellosis, non-specific osteitis, haematogenous osteomyelitis or treponemal osteitis. There is no evidence of deformation or callus formation associated with skeletal trauma such as a healed fracture, and the lesion does not present morphology consistent with post-traumatic processes such as cortical hypertrophy or the development of a cloaca. It is therefore most likely that this condition represents a primary osteogenic or osseous tumour of the spine. These are rare lesions with a much lower incidence than metastases, multiple myeloma or lymphoma.[1,2,20,21,23,27,32] Based on age at death, sex, anatomical location of the lesion, and specific patterns of expression and skeletal involvement, conditions such as osteosarcoma, chondrosarcoma or Ewing's sarcoma can be excluded; these neoplasms are often more aggressive, with destruction of the cortex[1,21,23].

Included in the differential diagnosis as the most likely cause of the observed lesion are osteoid osteoma, osteoblastoma, giant cell tumour and aneurysmal bone cyst. A number of secondary diagnoses are possible, specifically enostosis (compact bone island), fibrous cortical defect (fibroxanthoma), plasmacytoma, eosinophilic granuloma, and hydatid cyst infection. The range of possible differential diagnoses and primary diagnostic criteria are detailed in Table S1 (Supplementary Appendix).

Based on the observed pathological, morphological, and life-history criteria, the two most likely diagnoses are osteoid osteoma and osteoblastoma. Taking the demographic data for these two tumour types into account, both options seem possible: both are primary bone-forming tumours, osteoblastic in nature; benign; have a predilection for males; and show the highest prevalence in juveniles and adolescents. Osteoid osteoma resembles the observed lesion in terms of size, as these tumours are usually less than 20 mm in diameter, with well-circumscribed margins and being round or oval in form[23].

McCall[22] notes that computed tomography is the most valuable method to investigate this type of lesion. Under CT imaging of osteoid osteoma a small lucency is often recorded, which may have a central high attenuation as a result of mineralisation, and surrounding sclerotic bone is noted with some thickening of the lamina or pedicle. These are features seen in MH1 (Figure 4). On plain radiographs, most osteoid osteomas are osteosclerotic, with or without a visible nidus. By contrast, Kan and Schmidt[35] suggest that osteoblastomas are predominantly lucent or lytic in roughly 50% of cases, sclerotic in 30% of cases, and mixed in the remaining 20% of cases. On plain radiographs, osteoblastomas are typically expansile with a scalloped or lobulated appearance, and their margins are well-defined, with a sclerotic rim evident in approximately 30% of patients. A sclerotic rim is therefore much more common in osteoid osteomas than in osteoblastomas. The smooth, sclerotic, well-defined posterior margins of the lesion we studied are fully consistent with a resolving osteoid osteoma. However, the skeletal distribution of osteoid osteoma might argue against this being the most likely diagnosis, as osteoid osteomas are most commonly found in the lower extremities; occurrence in the spine is less likely than that exhibited in osteoblastoma[22].

To quantifiably assess the differential diagnosis, we applied Bayes Theorem of conditional probability to the diagnosis of osteoid osteoma and osteoblastoma. Using absolute clinical incidence data of osteoid osteoma[36-38] and osteoblastoma[25,37-42] to calculate prior and conditional probabilities of the disease expression in the vertebral column (as opposed to elsewhere in the skeleton), a conditional probability of 0.214 was derived for the likelihood of osteoid osteoma, and 0.068 for osteoblastoma. These results indicate a 3.75-fold higher likelihood that osteoid osteoma was represented in this case than osteoblastoma (see supplementary online material Table S2 for discussion of Bayes parameters and probability functions used). Given the morphological and pathological similarities between the two tumour types, and the age and nature of the specimen under analysis, the results suggest osteoid osteoma firstly and osteoblastoma secondly as the most likely diagnoses of what was clearly a benign entity of abnormal nature.

Discussion

MH1 suffered from a primary osteogenic tumour, which affected the right lamina of the sixth thoracic vertebra. The neoplastic lesion was chronic and was still active at the time of his death. From modern clinical studies[36-38] it is likely that osteoid osteoma may have taken months, rather than years, to develop. This neoplastic condition may involve neurological deficits, although this is unlikely as the lesion did not penetrate the neural canal, and no scoliosis was noted. However, the position of the lesion may have affected normal musculoskeletal function and movement of both the shoulder-blade and the upper right quadrant of the back. The tumour may have invoked a number of physiological responses including acute or chronic pain, muscular disturbance and pain-provoked muscular spasm, as discussed in clinical case studies.[21,36-38,40] A close association exists between the affected region and overlaying or closely inserting muscles such as trapezius, erector spinae, and rhomboid major, and this might have led to limitations on normal movement, given the likely arboreal component in the locomotor repertoire of *A. sediba.*[9,43]

The presence of a primary bone-forming tumour of the spine presents a number of considerations with regard to both the life-history of *Australopithecus sediba,* and evidence for neoplasia elsewhere in the deep past. Evidence for neoplastic disease is not unknown in the fossil, archaeological and historical records[1,8,44]. However, preservational factors limit the study of neoplasms to the skeleton (with the rare exception of naturally and artificially mummified bodies that may preserve pathological soft tissues) from which the confident diagnosis of tumours has been problematical[45]. The earliest skeletal evidence for neoplastic disease comes from pre-Cenozoic contexts, with purported cases of neoplasm found in fossil fish from the Upper Devonian. The earliest unequivocal case dates from 300 Ma, with evidence of benign osteoma with focal hyperostosis affecting the skeleton of *Phanerosteon mirabile* from the North American Lower Carboniferous[3]. Later terrestrial cases include diagnoses of benign haemangioma and eosinophilic granuloma in Jurassic dinosaurs; benign osteoma in mosasaurs; and haemangioma, metastatic cancer, desmoplastic fibroma and osteoblastoma in Cretaceous hadrosaurs.[46,47] In the more recent past, benign osteoid osteoma and osteoblastoma have been identified in European mammoths dating from 24 000 to 23 000 years ago (ka).[48]

The presence of neoplastic disease in the hominin fossil record is highly contentious. Until recently, the earliest purported evidence was suggested to be from a mandible of archaic *Homo* from Kanam, Kenya. This fossil is generally thought to derive from the Lower or Middle Pleistocene, and expresses pathological growth in the symphysial region. The lesion has been attributed to osteosarcoma, bone keloid, or Burkitt's lymphoma, although some researchers have diagnosed it as osteomyelitis resulting from a facial fracture[49-52]. The first substantive evidence for malignant neoplasia in hominins is derived from the SK7923 metatarsal fragment, dated to 1.8 to 1.6 Ma, from the site of Swartkrans, South Africa; a bony cortical exostosis together with osseous infilling of the medullary cavity of the shaft of the bone has been attributed by Odes and colleagues to osteosarcoma[5].

The next significant evidence for near-human neoplastic disease is suggested by Monge and colleagues, who present a case of fibrous dysplasia in a rib of *Homo neanderthalensis* dated to 120 ka from the European site of Krapina.[53] The Middle Pleistocene site of Atapuerca (Sima de los Huesos) evidenced small benign osteoid osteomata affecting the orbital roof of crania AT-777 and the endocranial surface of Cranium 4.[54] Other evidence comes from the Vogelherd (Stetten) II parietal bone, initially thought to represent a 35-ka-old Neanderthal, but now known to be Neolithic in origin[55]; in this specimen new bone formation has been linked to a possible meningioma although the final diagnosis remains equivocal[56]. The most significant evidence for neoplastic disease in antiquity derives from the bio-archaeological record of the recent Holocene (and the last four millennia in particular) and is detailed in a number of historical reviews and texts[1-3,46] to which the reader is directed.

As noted above, neoplastic disease in various forms, including osteoid osteoma and osteoblastoma, is an ancient phenomenon. It first appeared during the Palaeozoic and Mesozoic in extinct fish and members of the Dinosauria respectively.[3,46,47] However, the fact that reports of cancers or neoplasms remain exceedingly rare in the fossil record of almost any geological epoch[1,3,8,46,47,53] may be due to a number of factors, exacerbated by the relative disjunction between osseous tumours and all other forms of neoplasms. Primary bone tumours are rare compared with other neoplasms and account for around 7% of all soft and hard tissue cancers.[22] Neoplasms are historically reported to be rare in wild living mammals, with only 1.8% of deaths in chimpanzee communities reportedly resulting from cancer.[3] A mere handful of neoplastic cases have been recorded based on observational studies of camels, deer, gibbons, tigers, kangaroos, pacaranas, fur seals, ferrets, killer whales, harbour seals, sea lions and harp seals.[3] However, recent reviews of neoplasms in wild non-human primates[57] have shown that neoplastic disease might be far more widespread than previous studies suggest, in both monkeys and great apes; however, the vast majority of such cases involve benign soft tissue rather than malignant tumours. When bone tumours have been noted, they have tended to present as small benign growths such as button osteomata, which have been observed in both gorilla subspecies but have not been seen in either chimpanzees or orangutans[57]. An isolated case of benign osteochondroma was observed in the Gombe chimpanzee 'Old Female'.[58] Whilst these rare cases of neoplasia in non-human primates share morphological homology with human disease expression, it is unclear whether they share a common genetic basis or evolutionary history.

With regard to osteoid osteoma in humans, cytogenetic chromosomal studies indicate some degree of a genetic basis. This includes the involvement of chromosome 22, 22q monosomy and trisomy aberrations[59]; aberrant expression of transcription factors *Runx2* and *Osterix*, both of which are master regulators of osteoblastic lineage differentiation[60]; and duplications and deletions at 22q13.1[59], the locus of which reflects genes that play a role directly in osteogenesis (PDGF-B and ATF-4). The involvement of the latter suite of genes may suggest a degree of evolutionary conservatism, which warrants further investigation across primate taxa. As noted by Odes et al.[5], whilst the expression of neoplasia is rare in prehistory, the *capacity* for neoplastic disease (as evidenced by both fossil evidence and oncogenes) was present in deep-time.

It is no surprise that metastatic bone tumours are rare or absent in the archaeological and fossil records, because of the limited life expectancies of our ancestors[6,7] and the low incidence generally of skeletally forming or affecting neoplasms[1,3,20,23,46]. It is well known that primary bone tumours mostly occur in younger individuals[1,20,21,27,37,40], and it can therefore be expected that such tumours would have been present and have a similar prevalence to what is observed among modern individuals. It seems likely that neoplastic disease was as prevalent in ancient hominin populations as that expressed today in wild primate groups, but for various reasons it left little fossil trace. One reason might be the sheer paucity of individuals recovered from the hominin record, which would represent an issue of epidemiological sampling[6].

With regard to the earliest evidence for neoplastic disease in the hominin fossil record reported here, the fact that primary bone neoplasms are so rare makes this an important discovery. Whilst we consider it unlikely that neoplastic disease would have played a major role in the evolutionary forces operating on the Homininae, this case provides a unique glimpse into the individual life experience of a single extinct hominin. MH1 provides a window onto the expression and evolution of neoplastic disease in the human lineage, and highlights the utility of multidisciplinary clinical studies applied to the understanding of the evolution and development of disease in the human lineage.

Acknowledgements

We would like to acknowledge the assistance and help of the following people in the production of this research: Bonita de Klerk, Wilma Lawrence and Jennifer Randolph-Quinney. S.A.W. would like to acknowledge the help of Morgan Hill and Eric Mazelis from the

Microscopy and Imaging Facility, and Neil Duncan and Eileen Westwig of the Department of Mammalogy at the American Museum of Natural History. Further support in scanning time was given by the European Synchrotron Radiation Facility and by the Department of Radiology at Charlotte Maxeke Academic Hospital in Johannesburg. The research was funded by grants to L.R.B. from the National Geographic Society, the National Research Foundation of South Africa and the DST/NRF South African Centre of Excellence in Palaeosciences. Additional direct support for E.J.O. was received from the National Research Foundation of South Africa and the DST/NRF South African Centre of Excellence in Palaeosciences. M.S. was supported by the National Research Foundation of South Africa.

Authors' contributions

P.S.R.Q. coordinated the research and wrote the original draft of the manuscript, incorporating additional case notes and observational data on U.W. 88-37 as provided by S.A.W., M.S., M.R.M., J.S.S., S.E.C. and L.R.B. Detailed discussion of oncogenetics was provided by T.A, and E.J.O. provided detailed discussion of the historical data on early hominin palaeopathology. P.T. undertook the synchrotron scanning of the specimen, and primary reconstruction, segmentation and imaging. All authors contributed equally to data acquisition and analysis, and to editing.

References

1. Ortner DJ. Tumors and tumor-like lesions of bone. In: Ortner DJ, editor. Identification of Pathological Conditions in Human Skeletal Remains. Amsterdam: Academic Press; 2003. p. 503–544. http://dx.doi.org/10.1016/B978-012528628-2/50057-0
2. Aufderheide AC, Rodríguez-Martin C. The Cambridge encylopedia of human paleopathology. Cambridge: Cambridge University Press; 1998.
3. Capasso LL. Antiquity of cancer. Int J Cancer. 2005;113(1):2–13. http://dx.doi.org/10.1002/ijc.20610
4. Jemal A, Bray F, Center M, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69–90. http://dx.doi.org/10.3322/caac.20107
5. Odes EJ, Randolph-Quinney PS, Steyn M, Throckmorton Z, Smilg JS, Zipfel B, et al. Earliest hominin cancer: 1.7-milllion-year-old osteosarcoma from Swartkrans Cave, South Africa. S Afr J Sci. 2016;112(7/8), Art. #2015-0471, 5 pages. http://dx.doi.org/10.17159/sajs.2016/20150471
6. Chamberlain AT. Demography in archaeology. Cambridge: Cambridge University Press; 2006. http://dx.doi.org/10.1017/CBO9780511607165
7. Mann A. Paleodemographic aspects of the South African australopithecines. Philadelphia: University of Pennsylvania; 1975.
8. Retief FP, Cilliers L. Breast cancer in antiquity. S Afr Med J. 2011;101(8):513–515.
9. Berger LR, de Ruiter DJ, Churchill SE, Schmid P, Carlson KJ, Dirks PHGM, et al. *Australopithecus* sediba: A new species of *Homo*-like australopith from South Africa. Science. 2010;328(5975):195–204. http://dx.doi.org/10.1126/science.1184944
10. Williams SA, Ostrofsky KR, Frater N, Churchill SE, Schmid P, Berger LR. The vertebral column of *Australopithecus sediba*. Science. 2013;340(6129). http://dx.doi.org/10.1126/science.1232996
11. Berger L, Lacruz R, De Ruiter D. Revised age estimates of *Australopithecus*-bearing deposits at Sterkfontein, South Africa. Am J Phys Anthropol. 2002;119:192–197. http://dx.doi.org/10.1002/ajpa.10156
12. Brain CK. Structure and stratigraphy of the Swartkrans Cave in the light of the new excavations. In: Brain CK, editor. Swartkrans. Pretoria: Transvaal Museum Monograph 8; 1993. p. 23–34.
13. Berger LR, Menter CG, Thackeray JF. The renewal of excavation activities at Kromdraai, South Africa. S Afr J Sci. 1994;90:209–210.
14. Lacruz RS, Brink JS, Hancox J, Skinner AS, Herries A, Schmidt P, et al. Paleontology, geological context and paleoenvironmental implications of a Middle Pleistocene faunal assemblage from the Gladysvale Cave, South Africa. Paleontologia Africana. 2002;38:99–114.
15. Dirks PHGM, Berger LR, Roberts EM, Kramers JD, Hawks J, Randolph-Quinney PS, et al. Geological and taphonomic context for the new hominin species *Homo naledi* from the Dinaledi Chamber, South Africa. eLife. 2015;4, Art. #e09561. http://dx.doi.org/10.7554/eLife.09561
16. Dirks P, Berger L. Hominin-bearing caves and landscape dynamics in the Cradle of Humankind, South Africa. J Afr Earth Sci. 2012;78:109–131. http://dx.doi.org/10.1016/j.jafrearsci.2012.09.012
17. Dirks PHGM, Kibii JM, Kuhn BF, Steininger C, Churchill SE, Kramers JD, et al. Geological setting and age of *Australopithecus sediba* from Southern Africa. Science. 2010;328(5975):205–208. http://dx.doi.org/10.1126/science.1184950
18. Val A, Dirks PHGM, Backwell LR, d'Errico F, Berger LR. Taphonomic analysis of the faunal assemblage associated with the hominins (*Australopithecus sediba*) from the Early Pleistocene cave deposits of Malapa, South Africa. PLoS ONE. 2015;10(6), Art. #e0126904. http://dx.doi.org/10.1371/journal.pone.0126904
19. L'Abbé EN, Symes SA, Pokines JT, Cabo LL, Stull KE, Kuo S, et al. Evidence of fatal skeletal injuries on Malapa Hominins 1 and 2. Sci Rep. 2015;5, Art. #15120. http://dx.doi.org/10.1038/srep15120
20. Bahk Y-W. Combined scintigraphic and radiographic diagnosis of bone and joint diseases. New York: Springer; 2007. http://dx.doi.org/10.1007/978-3-540-68732-0_4
21. Khurana JS. The surgical pathology of bone tumors and tumor-like lesions. In: Khurana JS, editor. Bone pathology. London: Humana Press; 2009. p. 285–345. http://dx.doi.org/10.1007/978-1-59745-347-9_19
22. McCall IW. Radiology and pathological correlations of bone tumours of the spine. In: Gourtsoyiannis NC, Ros PR, editors. Radiologic-pathologic correlations from head to toe: Understanding the manifestations of disease. Berlin: Springer; 2005. p. 721–742. http://dx.doi.org/10.1007/3-540-26664-X_32
23. Vigorita VJ. Orthopaedic pathology. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
24. Baba H, Maezawa Y, Furusawa N, Wada M, Kokubo Y, Imura S, et al. Solitary plasmacytomas of the spine associated with neurological complications. Spinal Cord. 1998;36(7):470–475. http://dx.doi.org/10.1038/sj.sc.3100572
25. Berry M, Mankin H, Gebhardt M, Rosenberg A, Hornicek F. Osteoblastoma: a 30-year study of 99 cases. J Surg Oncol. 2008;98:179–183. http://dx.doi.org/10.1002/jso.21105
26. Betsy M, Kupersmith LM, Springfield DS. Metaphyseal fibrous defects. J Am Acad Orthop Surg. 2004;12(2):89–95. http://dx.doi.org/10.5435/00124635-200403000-00004
27. Binning M, Klimo Jr P, Gluf W, Goumnerova L. Spinal tumors in children. Neurosurg Clin N Am. 2007;18(4):631–658. http://dx.doi.org/10.1016/j.nec.2007.07.001
28. McLain RF, Weinstein JN. Solitary plasmacytomas of the spine: A review of 84 cases. J Spinal Disord. 1989;2(2):69–74. http://dx.doi.org/10.1097/00002517-198906000-00001
29. Mendenhall WM, Zlotecki RA, Scarborough MT, Gibbs CP, Mendenhall NP. Giant cell tumor of bone. Am J Clin Oncol. 2006;29(1):96–99. http://dx.doi.org/10.1097/01.coc.0000195089.11620.b7
30. Celik C, Münevver FS, Ucan H. Spinal hydatid cyst: Review. Türkiye Klinikleri J Med Sci. 2010;30(3):1073–1077. http://dx.doi.org/10.5336/medsci.2009-13722
31. Kransdorf MJ, Sweet DE. Aneurysmal bone cyst: Concept, controversy, clinical presentation, and imaging. Am J Roentgenol. 1995;164(3):573–580. http://dx.doi.org/10.2214/ajr.164.3.7863874
32. Osborn AG. Tumors, cysts, and tumorlike lesions of the spine and spinal cord. In: Osborn AG, editor. Diagnostic neuroradiology. St. Louis: Mosby; 1994. p. 876–918.
33. Kuruvath S, O'Donovan DG, Aspoas AR, David KM. Benign fibrous histiocytoma of the thoracic spine: Case report and review of the literature. J Neurosurg Spine. 2006;4(3):260–264. http://dx.doi.org/10.3171/spi.2006.4.3.260
34. Kaye JJ, Freiberger RH. Eosinophilic granuloma of the spine without vertebra plana: A report of two unusual cases. Radiology. 1969;92(6):1188–1191. http://dx.doi.org/10.1148/92.6.1188

35. Kan P, Schmidt MH. Osteoid osteoma and osteoblastoma of the spine. Neurosurg Clin North Am. 2008;19(1):65–70. http://dx.doi.org/10.1016/j.nec.2007.09.003

36. de Souza Dias L, Frost HM. Osteoid Osteoma–Osteoblastoma. Cancer. 1974;33(4):1075–1081. http://dx.doi.org/10.1002/1097-0142(197404)33:4<1075::AID-CNCR2820330427>3.0.CO;2-4

37. Azouz EM, Kozlowski K, Marton D, Sprague P, Zerhouni A, Asselah F. Osteoid osteoma and osteoblastoma of the spine in children: Report of 22 cases with brief literature review. Pediatr Radiol. 1986;16(1):25–31. http://dx.doi.org/10.1007/BF02387501

38. Saccomanni B. Osteoid osteoma and osteoblastoma of the spine: A review of the literature. Curr Rev Musculoskelet Med. 2009;2(1):65–67. http://dx.doi.org/10.1007/s12178-009-9047-6

39. Lucas DR. Osteoblastoma. Arch Pathol Lab Med. 2010;134:1460–1466.

40. Amacher AL, Eltomey A. Spinal osteoblastoma in children and adolescents. Child Nerv Syst. 1985;1(1):29–32. http://dx.doi.org/10.1007/BF00706727

41. Aszódi K. Benign osteoblastoma: Quantitative histological distinction from osteoid osteoma. Arch Orthop Unfallchir. 1977;88(3):359–368. http://dx.doi.org/10.1007/BF00416625

42. Boriani S, Capanna R, Donati D, Levine A, Picci P, Savini R. Osteoblastoma of the spine. Clin Orthop Relat Res. 1992;278:37–45. http://dx.doi.org/10.1097/00003086-199205000-00006

43. Churchill SE, Holliday TW, Carlson KJ, Jashashvili T, Macias ME, Mathews S, et al. The upper limb of *Australopithecus sediba*. Science. 2013;340(6129), Art. #1233477. http://dx.doi.org/10.1126/science.1233477

44. Eshed V, Latimer B, Greenwald CM, Jellema LM, Rothschild BM, Wish-Baratz S, et al. Button osteoma: Its etiology and pathophysiology. Am J Phys Anthropol. 2002;118(3):217–230. http://dx.doi.org/10.1002/ajpa.10087

45. Rothschild BM, Rothschild C. Comparison of radiologic and gross examination for detection of cancer in defleshed skeletons. Am J Phys Anthropol. 1995;96(4):357–363. http://dx.doi.org/10.1002/ajpa.1330960404

46. Rothschild BM, Martin LD. Paleopathology: Disease in the fossil record. Boca Raton, FL: CRC Press; 1993.

47. Rothschild BM, Tanke DH, Hebling M, Martin LD. Epidemiologic study of tumors in dinosaurs. Naturwissenschaften. 2003;90(11):495–500. http://dx.doi.org/10.1007/s00114-003-0473-9

48. Leshchinsky SV. Paleoecological investigation of mammoth remains from the Kraków Spadzista Street (B) site. Quartern Int. 2012;276–277:155–169. http://dx.doi.org/10.1016/j.quaint.2012.05.025

49. Phelan J, Weiner MJ, Ricci JL, Plummer T, Gauld S, Potts R, et al. Diagnosis of the pathology of the Kanam Mandible. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2007;103(4), Art. #e20. http://dx.doi.org/10.1016/j.tripleo.2006.12.041

50. Tobias PV. The Kanam jaw. Nature. 1960;185:946–947. http://dx.doi.org/10.1038/185946a0

51. Tobias PV. A re-examination of the Kanam mandible. In: Mortelmans G, Nenquin J, editors. Actes du IVe Congrès Panafricain de Préhistoire et de l'Étude du Quaternaire, Sections I and II. Tervuren, Belgique: Annales Ser. Qu-8o, Sciences Humaines 40; 1962. p. 341–350.

52. Stathopoulos G. Kanam mandible's tumour. Lancet. 1975;305(7899). p. 165. http://dx.doi.org/10.1016/S0140-6736(75)91462-2

53. Monge J, Kricun M, Radovčić J, Radovčić D, Mann A, Frayer DW. Fibrous dysplasia in a 120,000+ year old Neandertal from Krapina, Croatia. PLoS ONE. 2013;8(6), Art. #e64539. http://dx.doi.org/10.1371/journal.pone.0064539

54. Pérez P, Gracia A, Martínez I, Arsuaga J. Paleopathological evidence of the cranial remains from the Sima de los Huesos Middle Pleistocene site (Sierra de Atapuerca, Spain). Description and preliminary inferences. J Hum Evol. 1997;33(2):409–421. http://dx.doi.org/10.1006/jhev.1997.0139

55. Conard NJ, Grootes PM, Smith FH. Unexpectedly recent dates for human remains from Vogelherd. Nature. 2004;430:198–201. http://dx.doi.org/10.1038/nature02690

56. Czarnetzki A. Pathological changes in the morphology of the Young Paleolithic skeletal remains from Stetten (South-west Germany). J Hum Evol. 1980;9(1):15–17. http://dx.doi.org/10.1016/0047-2484(80)90037-8

57. Lovell NC. Patterns of injury and illness in Great Apes. Washington, DC: Smithsonian Press; 1990.

58. Jurmain R. Trauma, degenerative disease and other pathologies among the Gombe chimpanzees. Am J Phys Anthropol. 1989;80(2):229–237. http://dx.doi.org/10.1002/ajpa.1330800211

59. López-Terrada D, Hicks JM. Genetics and molecular biology of bone and soft tissue tumors. In: Khurana JS, editor. Bone pathology. Dordrecht: Humana Press; 2009. p. 91–124. http://dx.doi.org/10.1007/978-1-59745-347-9_6

60. Dancer J, Henry S, Bondaruk J, Lee S, Ayala A, de Crombrugghe B, et al. Expression of master regulatory genes controlling skeletal development in benign cartilage and bone forming tumors. Hum Pathol. 2010;41(12):1788–1793. http://dx.doi.org/10.1016/j.humpath.2010.06.008

Note: This article is accompanied by supplementary material.

AUTHORS:
Edward J. Odes[1,2]
Patrick S. Randolph-Quinney[1,2*]
Maryna Steyn[1]
Zach Throckmorton[2,3]
Jacqueline S. Smilg[2,4,5]
Bernhard Zipfel[2,6]
Tanya N. Augustine[1]
Frikkie de Beer[7]
Jakobus W. Hoffman[7]
Ryan D. Franklin[8]
Lee R. Berger[2,6]

AFFILIATIONS:
[1]School of Anatomical Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa
[3]De Busk College of Osteopathic Medicine, Lincoln Memorial University, Harrogate, Tennessee, USA
[4]School of Radiation Sciences, University of the Witwatersrand, Johannesburg, South Africa
[5]Department of Radiology, Charlotte Maxeke Academic Hospital, Johannesburg, South Africa
[6]DST/NRF South African Centre of Excellence in Palaeosciences, University of the Witwatersrand, Johannesburg, South Africa
[7]Radiography/Tomography Section, South African Nuclear Energy Corporation (NECSA), Pelindaba, South Africa
[8]Archaeological and Historical Conservancy, Davie, Florida, USA
*Current address: School of Forensic and Applied Sciences, University of Central Lancashire, Preston, Lancashire, United Kingdom

CORRESPONDENCE TO:
Patrick Randolph-Quinney

EMAIL:
prandolph-Quinney@uclan.ac.uk

POSTAL ADDRESS:
School of Forensic and Applied Sciences, University of Central Lancashire, Preston, Lancashire, PR1 2HE, UK

DATES:
Received: 11 Dec. 2015
Revised: 16 Mar. 2016
Accepted: 17 Mar. 2016

KEYWORDS:
palaeopathology; oncology; malignant neoplasia; metatarsal; micro-computed tomography

HOW TO CITE:
Odes EJ, Randolph-Quinney PS, Steyn M, Throckmorton Z, Smilg JS, Zipfel B, et al. Earliest hominin cancer: 1.7-million-year-old osteosarcoma from Swartkrans Cave, South Africa. S Afr J Sci. 2016;112(7/8), Art. #2015-0471, 5 pages. http://dx.doi.org/10.17159/sajs.2016/20150471

Earliest hominin cancer: 1.7-million-year-old osteosarcoma from Swartkrans Cave, South Africa

The reported incidence of neoplasia in the extinct human lineage is rare, with only a few confirmed cases of Middle or Later Pleistocene dates reported. It has generally been assumed that pre-modern incidence of neoplastic disease of any kind is rare and limited to benign conditions, but new fossil evidence suggests otherwise. We here present the earliest identifiable case of malignant neoplastic disease from an early human ancestor dated to 1.8–1.6 million years old. The diagnosis has been made possible only by advances in 3D imaging methods as diagnostic aids. We present a case report based on re-analysis of a hominin metatarsal specimen (SK 7923) from the cave site of Swartkrans in the Cradle of Humankind, South Africa. The expression of malignant osteosarcoma in the Swartkrans specimen indicates that whilst the upsurge in malignancy incidence is correlated with modern lifestyles, there is no reason to suspect that primary bone tumours would have been any less frequent in ancient specimens. Such tumours are not related to lifestyle and often occur in younger individuals. As such, malignancy has a considerable antiquity in the fossil record, as evidenced by this specimen.

Introduction

The reported incidence of neoplastic disease in the extinct human lineage is rare. Only a few confirmed cases of Middle or Later Pleistocene dates (780 000 to 120 000 years old) have been reported.[1,2] It is generally assumed that pre-modern incidence of neoplastic disease of any kind is rare and limited to benign conditions, but new fossil evidence suggests this is not so. We here present the earliest identifiable case of malignant neoplastic disease from an early human ancestor dated to 1.8–1.6 million years old (Ma). The diagnosis is possible only because of advances in 3D imaging methods as an aid in diagnosis.

A neoplasm ('new-growth' or tumour) is defined as a mass of localised tissue growth in which cellular proliferation is no longer subject to the effects of normal growth-regulating mechanisms.[1,2] A tumour may be benign or malignant in nature; malignant tumours are often colloquially referred to as a cancer.[3] Malignancy is the primary cause of death in industrialised countries and the second foremost cause of death in developing countries.[4,5] Since 1999, malignancy has surpassed cardiac disease as the leading cause of death for humans younger than 85 years in the USA[6], and is often perceived to be a disease of modernity[7-9]. At present, true neoplastic diseases seem to be restricted to complex vertebrate animals. Only one observation of true malignancy has been described in one of the simpler living vertebrates, specifically hepatomas in the cartilaginous skeleton of the jawless hagfish.[2] This is a very important case for the comparative pathology of malignancy, because lampreys are among the simplest living vertebrates.

The fact that malignancy has great antiquity is demonstrated from the fossil record. The earliest definitive evidence for neoplastic disease comes from pre-Cenozoic contexts, with purported cases of neoplasm found in fossil fish from the Upper Devonian. The earliest unequivocal case dates from 300 Ma, with evidence of benign osteoma with focal hyperostosis affecting a partial skeleton of the fish *Phanerosteon mirabile* from the North American Lower Carboniferous.[2] Later cases include diagnoses of benign haemangioma and eosinophilic granuloma in Jurassic dinosaurs; benign osteoma in mosasaurs; and haemangioma, metastatic disease, desmoplastic fibroma, and osteoblastoma in Cretaceous hadrosaurs.[10,11] Benign osteoid osteoma and osteoblastoma have been identified in European mammoths dating from 24 to 23 thousand years ago (ka).[12] Evidence for neoplastic disease is not unknown in the human fossil, archaeological and historical records,[1,3,13] but is generally considered rare. Historically, the earliest fossil evidence for neoplastic disease in the human lineage was suggested to be from a mandible of archaic *Homo* from Kanam in Kenya. This lesion has variously been attributed to osteosarcoma, bone keloid, Burkitt's lymphoma, or traumatic osteomyelitis.[2,14-17] The first substantive evidence for hominin neoplastic disease is derived from a juvenile skeleton of *Australopithecus sediba,* dated to 1.98 Ma, from the site of Malapa in South Africa. An invasive spinal lesion has been attributed by Randolph-Quinney and colleagues to benign osteoid osteoma, a non-malignant tumour.[18] Later significant evidence for near-human neoplastic disease is suggested by Monge and colleagues'[19], who present a case of fibrous dysplasia from a Neanderthal rib dated to 120 000 ka from the site of Krapina.

Here we present the earliest fossil evidence for malignant neoplastic disease in the hominin record, with a detailed description and differential diagnosis of malignant osteosarcoma. Our report is based on re-analysis of a hominin metatarsal specimen (SK 7923) (*gen. et spec. indet.*) from the cave site of Swartkrans in the Cradle of Humankind, South Africa. SK 7923 is a metatarsal recovered from the Member 1 Hanging Remnant, which has yielded fossils of both *Homo ergaster* and *Paranthropus robustus*.[20] Several faunal estimates have indicated the age of the Hanging Remnant at between 1.5 Ma and 1.8 Ma.[21-23] Recent electron spin resonance dating has estimated the age of the Hanging Remnant of Swartkrans at 1.6 Ma.[24] The oldest southern African specimens of early *Homo* and *Paranthropus* present around 2.1–1.9 Ma in Member 1, and are recorded until around 1.0–0.6 Ma in Member 3 of the Swartkrans cave.[25]

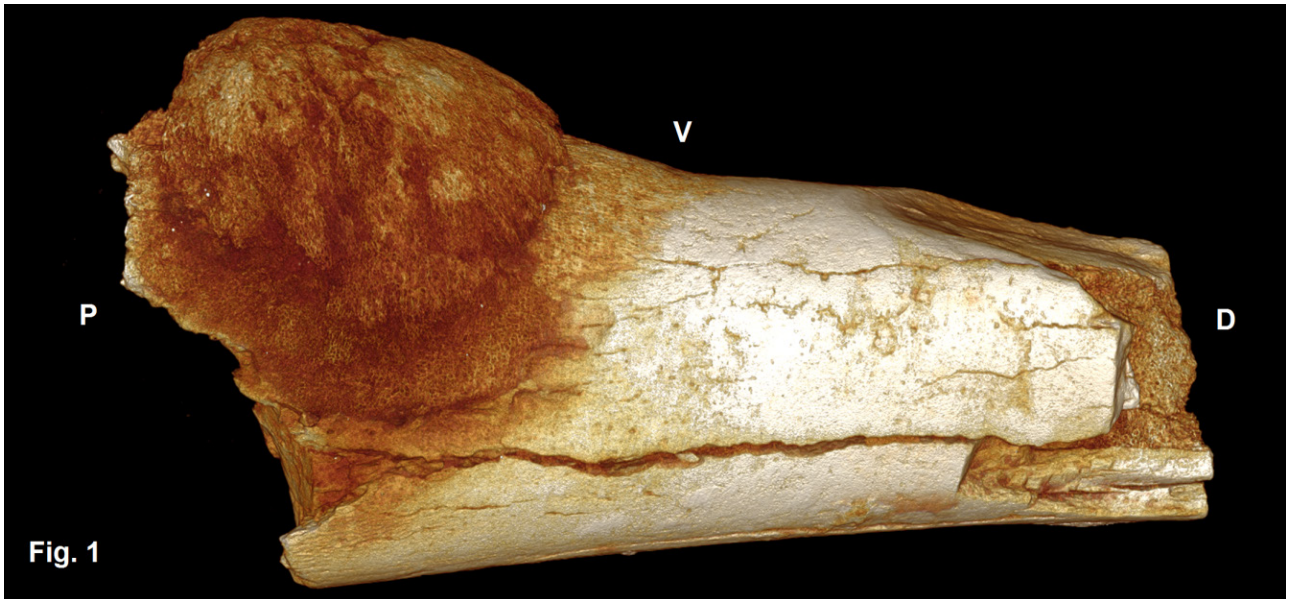


Figure 1: SK 7923 , a hominin 5th metatarsal, exhibits a hemi-spherical bony mass located on the proximo-ventral aspect of the shaft, abutting the cortical bone surface. P – proximal, D – distal, V – ventral.

Swartkrans hominin site

Swartkrans was discovered in 1948. The site is situated approximately 40 km northwest of Johannesburg, on the bank of the Blaauwbank River in the province of Gauteng, South Africa. It is arguably one of the most important palaeocave sites in southern Africa, if not globally, and is best known for its rich heritage of *Paranthropus robustus* fossils, and purported evidence for early hominin use of controlled pyrotechnology. During much of their early research, Broom and his assistant Robinson excavated a considerable sample of hominin remains attributable to *Paranthropus robustus* and *Homo* from the site.[26,27] Swartkrans was the first site where these two genera of hominins were considered to be contemporary.

After an approximately 12-year hiatus of non-activity, Brain's subsequent excavations at Swartkrans from the mid-1960s to the mid-1980s produced a significant addition to the faunal[20], fossil and archaeological collection. From the total number of specimens obtained at the Swartkrans site, 415 accessioned specimens are identified as hominin.[20] Brain's work demonstrated that the stratigraphy of the site was much more complex than originally posited. Brain subsequently demonstrated five members at Swartkrans.[22] Member 1 consists of two distinct masses, the Hanging Remnant and Lower Bank[22], each of them yielding *Homo ergaster* and *Paranthropus robustus*[28]. *Homo* and *Paranthropus* have been discovered from Member 2.[28]

SK 7923 case study

SK 7923 is a left 5th metatarsal, preserving the proximal diaphysis and much of the distal portion, but lacking the articular end. The specimen is hominin, but cannot be allocated to a specific taxon. Pathologically, SK 7923 presents a growth on the proximo-ventral aspect of the shaft. Here, an irregular hemi-spherical mass abuts the cortex (Figure 1), measuring 5.2 mm x 4.7 mm. The specimen was originally studied as part of an unpublished doctoral thesis, where the morphology led one of us (R.F.) to diagnose osteoid osteoma. However, recent internal imaging has led to a re-evaluation of the pathology.

The bone was examined using micro-focus X-ray computed tomography (μXCT) at the South African Nuclear Centre for Radiography and Tomography (located at the South African Nuclear Energy Corporation, NECSA). The bone was scanned by F.d.B. and J.W.H. using a Nikon XTH225ST μXCT system, at an energy potential of 100 kV and resolution of 17 microns. Reconstruction was performed by E.J.O. and P.R.Q. using Avizo Amira 5.4 to generate both 2D orthoslice and 3D surface rendered views. The cross-section shows that the hemispherical mass is not fully fused to or integrated with the cortex, but adheres to the bone surface, displaying an irregular spongy woven bone texture with a cauliflower-like external appearance (Figure 2a). The cortical bone directly underlying the mass is covered with a thin layer of new woven bone, with a Codman triangle displayed at the margins. The texture is granular and exhibits ellipsoid lytic lesions in transverse view. There is localised sub-periosteal invasion by the mass into the cortex (Figure 2a). Surprisingly, μXCT showed much of the medullary cavity to be infilled with bone, with clear internal remodelling and *de-novo* bone formation (Figure 2b). Inside the cortex are several irregular and large circular voids, caused by bony encapsulation of normal vascular channelling found within the endosteal surface of the medullary cavity. The remaining medullary space is obliterated by new bone growth.

A number of bone-forming conditions should be considered in the differential diagnosis: chondrosarcoma, Ewing's sarcoma, metastatic carcinoma, osteochondroma, osteoblastoma and osteosarcoma (see Supplementary Appendix for a more detailed breakdown).[3,29-46] Osteosarcoma usually starts in the medulla and characteristically arises within the metaphysis of long bones, growing circumferentially through the cortex into soft tissue and raising the periosteum. This seems to have occurred in SK 7923, with a Codman triangle visible. Osteosarcoma (osteogenic sarcoma) prefers fast-growing regions and usually occurs around the knee. It presents in metatarsals in less than 1% of clinical cases. According to Vigorita[47], the periosteal reaction may have a 'sunburst' appearance, which is to some extent visible in Figure 2a. Given the internal and external morphology of this specimen, it seems most likely that this pathology is attributable to osteosarcoma, with a strong possibility of the parosteal variant of this condition. Diagnosis was supported by μXCT imaging of a modern clinically diagnosed case of osteosarcoma of the distal femur. Comparison of Figures 2b and 2c shows clear similarities between the appearance of the fossil and modern medullary infills.

Osteosarcoma is a primary malignant tumour that typically exhibits cortical and medullary disruption and some degree of mineralisation. It also typically shows aggressive periosteal new bone reaction, which can include either lamination, Codman triangle, or spiculated sunburst reaction. The most common form is central osteogenic sarcoma, which could be osteoblastic or osteolytic in nature, or both. This typically occurs in the metaphyseal and diaphyseal area of the major long bones. Parosteal osteosarcoma is the most common type of juxtacortical or surface osteosarcoma. This often presents as a lobulated 'cauliflower' mass with central ossification adjacent to the bone, and may infiltrate the bone marrow. Bone destruction is rare, but when present is regional.

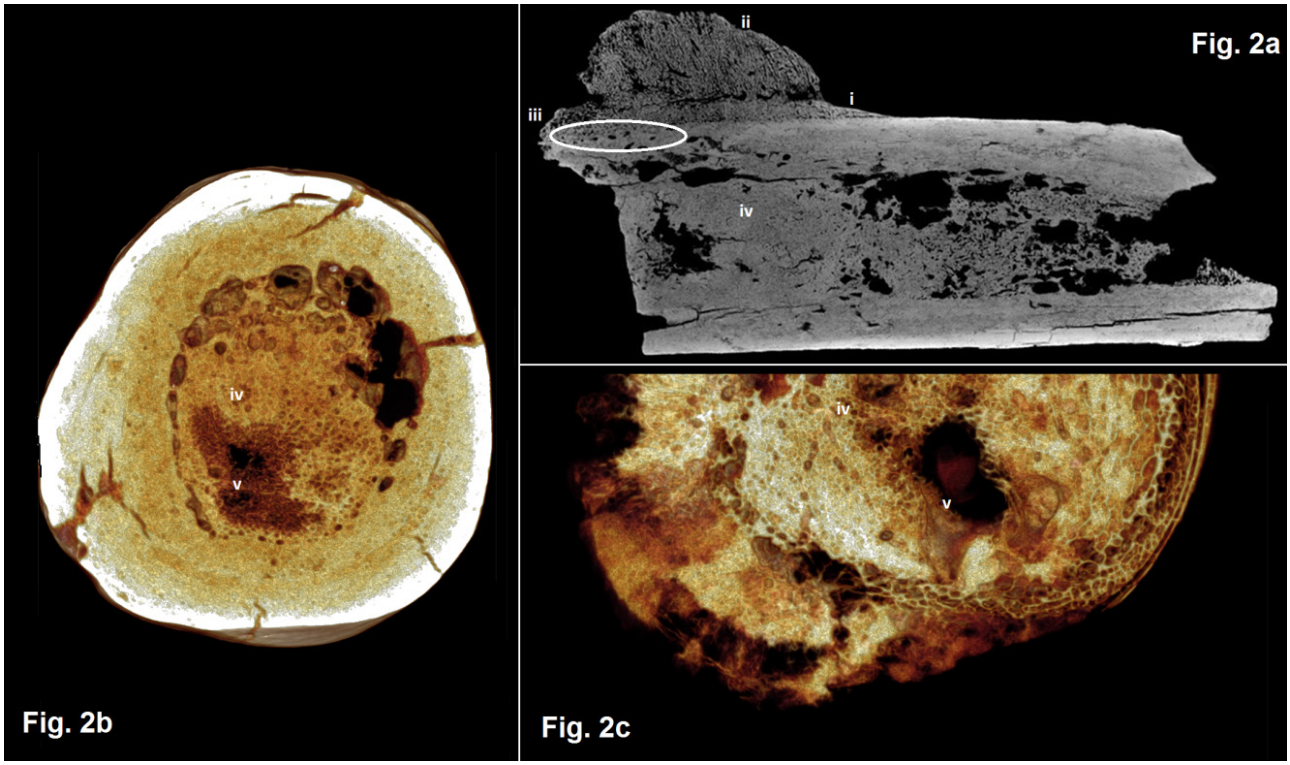


Figure 2: (a) Axial micro-CT orthoslice indicating (i) reactive new bone formation subperiosteally forming a Codman triangle, (ii) ossified exophytic (cauliflower-like) and/or spiculated mass adjacent to the bone, (iii) localised sub-periosteal invasion by the mass into the cortex, (iv) remodelled bone infill; (b) transverse rendered view of SK 7923; (c) transverse rendered view of modern clinical case of osteoblastic osteosarcoma with aggressive local medullary infilling (Courtesy: Department of Anatomy, University of Pretoria). Note (v) the homologous combination of spongy and solid bone between the fossil and clinical specimens.

Most parosteal osteosarcomas are found in the metaphyseal region of long tubular bones. The most common site is the distal femur, followed by the upper shaft of the tibia, and then the proximal humerus.

Functionally, this malignancy may have presented secondary consequences in our case study. The belly of *abductor digitii minimi* may have been displaced laterally, and with *fibularis brevis* and *f. tertius* inserting just proximal and distal to the growth, some influence on gait was likely. The presence of sub-periosteal bone formation in the form of a Codman triangle, cortical invasion, spiculated mass abutting the exosteum, and aggressive medullary infilling (including the combination of trabeculated and avascular bone) indicates that what was originally diagnosed as a benign exosteal growth is now shown to represent a malignant bone malignancy. This change in diagnosis is entirely the result of advances in high-resolution 3D imaging, together with the judicious use of comparative clinical pathology. It is thus possible that cases of malignancy might remain unknown in fossil assemblages awaiting imaging and discovery.

Discussion

This case highlights a significant issue with regard to modern clinical incidence and expression of neoplastic disease, and malignancy in particular. That is, how can we understand ancient disease evolution when sample sizes are extremely small? As noted above[7-9], malignancy is perceived as a disease of modernity. However, it is worth noting that primary skeletal malignant tumours are not commonly encountered in the modern clinical environment, and although rare they do feature in the archaeological and fossil record[3,18]. Historically, factors of preservation have limited the study of human neoplasia to the skeleton, from which the confident diagnosis of tumours has been problematical.[48] Recent work on artificially mummified Egyptian remains has suggested to some scholars[9,49] that malignancy was almost absent in pre-modern human populations. For example, Gray[49] reported no radiological confirmation of malignant neoplasia among 193 examined Egyptian mummies. However, we view this assertion as tautological, because the samples on which the claim is based are not representative of the bulk of the human species living in antiquity; they represent only a small fraction of all humans living at that time[50]. From the global historical sample available to us for study, malignancy does exist, albeit rarely – and a subset of those cases include osteosarcoma. Probable pre-modern cases are reported from Hawaii[51], the Czech Republic[52], and the Peruvian Andes[1].

The precise range of causes underlying malignancies is still largely unknown. Where causes have been established, these are generally understood to fall into three categories: physical, chemical, and viral.[1] Physical causes include being exposed to ultraviolet light (which increases the risk of basal cell carcinoma and malignant melanoma) and levels of background radiation.[1] Causative examples in the historical modern environment include the effects of radiation from the dropping of atomic bombs on Japan during World War II, which has been linked to an escalation in myelogenous leukaemia and thyroidal malignancy.[1] Chemical carcinogens in humans have historically been known to target skin and lungs, urinary bladder, and nasal sinuses and pleura. Potts demonstrated a relationship between scrotal malignancy in chimney sweepers and exposure to soot as early as 1775.[53] Many chemicals are widely accepted to be carcinogenic. Radon from granite, for example, is a radioactive gas causing lung malignancy.[1] Several viruses can cause malignancy in animals, and the association of some viruses with human malignancy is considerable. Examples include human papilloma virus (cervix malignancy), hepatitis B or C virus (liver malignancy), Epstein-Barr virus (non-Hodgkin lymphomas and nasopharyngeal malignancy), and human immunodeficiency virus (linked to non-Hodgkin lymphoma and Kaposi sarcoma).[1]

Whilst most modern human malignancies are thought to be caused by environmental agents of a chemical nature, the evidence for this is not entirely conclusive.[1,p.373] The internal environment, namely

diet and lifestyle, is thought to play a significant role in malignancy disposition. Some malignancies are certainly triggered by modern lifestyle factors, such as smoking, drinking (which can lead to liver and oesophageal malignancy), sunbathing, and obesity (which can lead to gastrointestinal gut malignancy).[1] Pesticides and industrial chemicals can cause malignancy, but their contribution is thought to be relatively minor. Of perhaps greatest impact is colorectal malignancy, which kills approximately 700 000 people every year worldwide[8], and is most prevalent in developed countries such as USA and Europe. The lowest incidence occurs in underdeveloped countries in Africa, which have vastly different dietary regimes available to the population. Brody[8] has named colorectal cancer a disease of modernity and development. In other words, economic growth is associated with a rise in the incidence of colorectal malignancy, with corresponding lifestyle changes possibly playing a strong role in the prevalence of this disease. An example is China, where a marked increase in the prevalence of colorectal malignancy is occurring.[8]

In addition, whilst the modern lifestyles of humans may enhance the frequency of cancer, longer life expectancies mean malignancy would logically occur at a higher rate among modern people than in our prehistory. We are unable to assess the age at death of the Swartkrans SK 7923 hominin (other than skeletally adult), but it was likely to have been substantially less than modern life expectancy, based on demographic studies of early hominin taxa[54].

The expression of malignant osteosarcoma in the Swartkrans SK 7923 specimen indicates that whilst the explosion of malignancy incidence is clearly correlated with the hazards of the modern world and increased life expectancy, primary bone tumours evidently occurred throughout history. Then, as now, such tumours would have occurred predominantly in younger individuals. Neoplastic disease has considerable antiquity, as evidenced by this specimen and further supported by numerous published case studies of benign neoplasms with deep antiquity in the fossil record, as noted above.[1-3,10-12,18,55,56] The theory that the almost total lack of malignancies in Egyptian mummies indicates that the disease occurs only in industrialised societies[9] is thus questionable.

The lack of substantial evidence of malignancy in the fossil and bio-archaeological records might arguably be an artefact of preservation, or of sampling bias and tiny sample sizes, or of analytical techniques or the application of inefficient imaging modalities[57]. For example, the results would differ if plain radiography or clinical magnetic resonance imaging are used rather than micro-computed tomography. The absolutely small sample sizes arise from short life expectancy among pre-modern societies. Hence, the rare incidence of cancer cases found in the record can be seen as most likely non-representative, and should not be construed as indicating the true prevalence of disease. It is important to note that because of the worldwide demographic transition and average increased age at death for humans, non-primary bone malignancy can be expected to occur at a higher rate today than in pre-transition populations.[50]

As highlighted in the introduction, malignancy occurs in almost all metazoans, suggesting that the mechanisms of malignancy have an extremely old evolutionary history. For example, neoplasms have been recorded in dinosaurs and other fossil forebears.[2,11,56] A number of oncogenes are particularly archaic, with their antecedents exhibited in some form in primitive common ancestor metazoans of chordates and arthropods.[55,58] Thus, the *capacity* for malignancy is ancient, and the higher incidence of malignancy in today's developed and developing world may be related to the unique interaction between environmental factors – which have no parallel in prehistory[59]. It is also important to note that modern non-invasive imaging techniques (such as micro-computed tomography or phase-contrast synchrotron tomography), together with the appropriate use of clinical homologues, play a considerable role in enabling accurate diagnoses. These new techniques hold considerable potential for re-investigating previously reported palaeopathological lesions.

Acknowledgements

We would like to thank Stephany Potze and Lazarus Kgasi of the Ditsong National Museum of Natural History for providing access to the Swartkrans fossil specimen. E.J.O. would like to acknowledge the National Research Foundation of South Africa, the DST/NRF South African Centre of Excellence in Palaeosciences, and the University of the Witwatersrand for funding this research. We would like to thank the Department of Anatomy, University of Pretoria for permission to image and analyse the comparative clinical specimen used in this study. M.S. is supported by the National Research Foundation of South Africa.

Authors' contributions

E.J.O. and P.S.R.Q. wrote the original draft of the manuscript, incorporating additional information and data on SK 7923 obtained from M.S., J.S., Z.T., B.Z., T.N.A. and L.R.B. P.S.R.Q. and L.R.B. supervised the research. Critical input on the pathology of malignant neoplasms was provided by M.S. and J.S., and Z.T. and B.Z. gave input on functional aspects of pedal pathology. E.J.O. provided the differential diagnoses and historical pathological and palaeoanthropological data. J.W.H. and F.d.B. undertook the micro-computed tomographic scanning of the hominin specimen and the primary reconstruction. E.J.O. and P.S.R.Q. undertook secondary reconstruction from orthoslice and 3D volume data. All authors contributed equally to analysis and editing.

References

1. Aufderheide AC, Rodríguez-Martin C. The Cambridge Encylopedia of human paleopathology. Cambridge: Cambridge University Press; 1998.
2. Capasso LL. Antiquity of cancer. Int J Cancer. 2005;113:2–13. http://dx.doi.org/10.1002/ijc.20610
3. Ortner DJ. Tumors and tumor-like lesions of bone. In: Ortner DJ, editor. Identification of pathological conditions in human skeletal remains. Amsterdam: Academic press; 2003. p. 503–544. http://dx.doi.org/10.1016/b978-012528628-2/50057-0
4. Jemal A, Bray F, Center M, Ferlay J, Ward E, Forman D. Global cancer statistics. CA-Cancer J Clin. 2011;61:69–90. http://dx.doi.org/10.3322/caac.20107
5. World Health Organization. The global burden of disease: 2004 update. Geneva: WHO Press; 2008. 153 pages.
6. Jemal A, Siegal R, Xu J, Ward E. Cancer statistics, 2010. CA-Cancer J Clin. 2010;60:277–300. http://dx.doi.org/10.3322/caac.20073
7. Binder M, Roberts C, Antoine D, Cartwright C. On the antiquity of cancer: Evidence for metastatic carcinoma in a young man from Ancient Nubia (c. 1200BC). PloS ONE [Internet]. 2014;9, Art. #e90924 11 pages. http://dx.doi.org/10.1371/journal.pone.0090924
8. Brody H. Colorectal cancer. Nature. 2015;521, Art. #a7551. http://www.nature.com/nature/journal/v521/n7551_supp/full/521S1a.html
9. David RA, Zimmerman MR. Cancer: An old disease, a new disease or something in between? Nat Rev Cancer. 2010;10:728–32. http://dx.doi.org/10.1038/nrc2914
10. Rothschild BM, Tanke DH, Hebling M, Martin LD. Epidemiologic study of tumors in dinosaurs. Naturwissenschaften. 2003;90:495–500. http://dx.doi.org/10.1007/s00114-003-0473-9
11. Rothschild BM, Martin LD. Paleopathology: Disease in the fossil record. Boca Raton, FL: CRC Press; 1993.
12. Leshchinsky SV. Paleoecological investigation of mammoth remains from the Kraków Spadzista Street (B) site. Quatern Int. 2012;(276–277):155–169. http://dx.doi.org/10.1016/j.quaint.2012.05.025
13. Retief FP, Cilliers L. Breast cancer in antiquity. S Afr Med J. 2011;101(8):513–515.
14. Phelan J, Weiner MJ, Ricci JL, Plummer T, Gauld S, Potts R, et al. Diagnosis of the pathology of the Kanam mandible. Oral Surg Oral Med O. 2007;103(4), Art. #e20. http://dx.doi.org/10.1016/j.tripleo.2006.12.041
15. Stathopoulos G. Kanam mandible's tumour. Lancet. 1975;305(7899). p. 165. http://dx.doi.org/10.1016/S0140-6736(75)91462-2

16. Tobias PV. The Kanam jaw. Nature. 1960;185:946-947. http://dx.doi.org/10.1038/185946a0

17. Tobias PV. A re-examination of the Kanam mandible. In: Mortelmans G, Nenquin J, editors. Actes du IVe Congrès Panafricain de Préhistoire et de l'Étude du Quaternaire, Sections I and II. Tervuren, Belgique: Annales Ser. Qu-8o, Sciences Humaines 40; 1962. p. 341–350. [French].

18. Randolph-Quinney PS, Williams SA, Steyn M, Meyer MR, Smilg JS, Churchill SE, et al. Osteogenic tumour in *Australopithecus sediba*: Earliest hominin evidence for neoplastic disease. S Afr J Sci. 2016;112(7/8), Art. #2015-0470, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20150470

19. Monge J, Kricun M, Radovčić J, Radovčić D, Mann A, Frayer DW. Fibrous dysplasia in a 120,000+ year old Neandertal from Krapina, Croatia. PLoS ONE. 2013;8(6), Art. #e64539. http://dx.doi.org/10.1371/journal.pone.0064539

20. de Ruiter DJ. Revised faunal lists for Members 1–3 of Swartkrans, South Africa. Ann Trans Mus. 2003;40:29–41.

21. Berger L, Lacruz R, De Ruiter D. Revised age estimates of *Australopithecus*-bearing deposits at Sterkfontein, South Africa. Am J Phys Anthropol. 2002;119:192–197. http://dx.doi.org/10.1002/ajpa.10156

22. Brain C. Structure and stratigraphy of the Swartkrans cave in light of the new excavations. In: Brain C, editor. Swartkrans: A cave's chronicle of early man. Pretoria: Transvaal Museum; 1993. p. 23–34.

23. Vrba E. Early hominids in southern Africa: Updated observations on chronological and ecological background. In: Tobias P, editor. Hominid evolution: Past, present and future. New York: Alan R. Liss; 1985. p. 195–200.

24. Curnoe D, Grun R, Taylor L, Thackeray J. Direct ESR dating of a Pliocene hominin from Swartkrans. J Hum Evol. 2001;40:379–391. http://dx.doi.org/10.1006/jhev.2000.0459

25. Herries A, Curnoe D, Adams J. A multi-disciplinary seriation of early Homo and Paranthropus bearing palaeocaves in southern Africa. Quatern Int. 2009;202(1):14–28. http://dx.doi.org/10.1016/j.quaint.2008.05.017

26. Brain C. New finds at the Swartkrans australopithecine site. Nature. 1970;225:1112–1119. http://dx.doi.org/10.1038/2251112a0

27. Broom R, Robinson J. A new type of fossil man. Nature. 1949;164:322. http://www.nature.com/nature/journal/v164/n4164/pdf/164322a0.pdf

28. Brain C. A taphonomic overview of the Swartkrans fossil assemblages. In: Brain C, editor. Swartkrans: A cave's chronicle of early man. Pretoria: Transvaal Museum; 1993. p. 257–264.

29. Burchill S. Ewing's sarcoma: Diagnostic, prognostic, and therapeutic implications of molecular abnormalities. J Clin Pathol. 2003;56(2):96–102. http://dx.doi.org/10.1136/jcp.56.2.96

30. Eisenberg R. Bubbly lesions of bone. Am J Roentgenol. 2009;193(2):W79–94. http://dx.doi.org/10.2214/AJR.09.2964

31. Geirnaerdt M, Hogendoorn P, Bloem J, Taminiau A, van der Woude H. Cartilaginous tumors: Fast contrast-enhanced MR Imaging 1. Radiology. 2000;214(2):539–546. http://dx.doi.org/10.1148/radiology.214.2.r00fe12539

32. Huvos A. Ewing's sarcoma. In: Huvos A, editor. Bone tumors: Diagnosis, treatment and prognosis. 2nd ed. Philadelphia PA: Sanders; 1991. p. 523–552.

33. Llauger J, Palmer J, Amores S, Bague S, Camins A. Primary tumors of the sacrum: Diagnostic imaging. Am J Roentgenol. 2000;174(2):417–424. http://dx.doi.org/10.2214/ajr.174.2.1740417

34. Lucas D, Unni K, McLeod R, O'Connor M, Sim F. Osteoblastoma: Clinicopathologic study of 306 cases. Hum Pathol. 1994;25(2):117–134. http://dx.doi.org/10.1016/0046-8177(94)90267-4

35. McLeod R, Dahlin D, Beabout J. The spectrum of osteoblastoma. Am J Roentgenol. 1976;126:321–325. http://dx.doi.org/10.2214/ajr.126.2.321

36. Murphey M, Walker E, Wilson A, Kransdorf M, Temple H, Gannon F. From the archives of the AFIP: Imaging of primary chondrosarcoma: Radiologic–pathologic correlation. Radiographics. 2003;23(5):1245–1278. http://dx.doi.org/10.1148/rg.235035134

37. Osborne R. The differential radiologic diagnosis of bone tumors. CA-Cancer J Clin 1974;24(4):194–211. http://dx.doi.org/10.3322/canjclin.24.4.194

38. Trübenbach J, Nägele T, Bauer T, Ernemann U. Preoperative embolization of cervical spine osteoblastomas: Report of three cases. Am J Neuroradiol. 2006;27(9):1910–1912.

39. Unni K, Dahlin D. Osteosarcoma: Pathology and classification. Semin Roentgenol. 1989;24(3):143–152. http://dx.doi.org/10.1016/0037-198X(89)90010-2

40. Berry M, Mankin H, Gebhardt M, Rosenberg A, Hornicek F. Osteoblastoma: A 30-year study of 99 cases. J Surg Oncol. 2008;98:179–183. http://dx.doi.org/10.1002/jso.21105

41. Binning M, Klimo Jr P, Gluf W, Goumnerova L. Spinal tumors in children. Neurosurg Clin N Am. 2007;18(4):631–658. http://dx.doi.org/10.1016/j.nec.2007.07.001

42. de Souza Dias L, Frost HM. Osteoid Osteoma – Osteoblastoma. Cancer. 1974;33:1075–1081. http://dx.doi.org/10.1002/1097-0142(197404)33:4<1075::AID-CNCR2820330427>3.0.CO;2-4

43. Giannico G, Holt G, Holmar K, Johnsson J, Pinnt J, Bridge J. Osteoblastoma characterized by a three-way translocation: Report of a case and review of the literature. Cancer Genet Cytogen. 2009;195:168–171. http://dx.doi.org/10.1016/j.cancergencyto.2009.06.024

44. Lucas DR. Osteoblastoma. Arch Pathol Lab Med. 2010;134:1460–1466.

45. Osborn AG. Tumors, cysts, and tumorlike lesions of the spine and spinal cord. In: Osborn AG, editor. Diagnostic neuroradiology. St. Louis: Mosby; 1994. p. 876–918.

46. Saccomanni B. Osteoid osteoma and osteoblastoma of the spine: A review of the literature. Curr Rev Musculoskel Med. 2009;2:65–67. http://dx.doi.org/10.1007/s12178-009-9047-6

47. Vigorita VJ. Orthopaedic pathology. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.

48. Rothschild BM, Rothschild C. Comparison of radiologic and gross examination for detection of cancer in defleshed skeletons. Am J Phys Anth. 1995;96(4):357–363. http://dx.doi.org/10.1002/ajpa.1330960404

49. Gray PHK. The radiography of mummies of ancient Egyptians. J Hum Evol. 1973;2:51–53. http://dx.doi.org/10.1016/0047-2484(73)90098-5

50. Chamberlain AT. Demography in archaeology. Cambridge: Cambridge University Press; 2006. http://dx.doi.org/10.1017/CBO9780511607165

51. Suzuki T. Paleopathological study on a case of osteosarcoma. Am J Phys Anth. 1987;74:309–318. http://dx.doi.org/10.1002/ajpa.1330740305

52. Strouhal E, Vyhnanek L, Horackova L, Benesova L, Nemeckova A. A case of osteosarcoma in a late medieval–early modern skull from Kyjov (Czech Republic). Int J Osteoarchaeol. 1997;7:82–90. http://dx.doi.org/10.1002/(SICI)1099-1212(199701)7:1<82::AID-OA320>3.0.CO;2-J

53. Clayson D. Chemical carcinogenesis. Boston: Little, Brown & Co; 1962.

54. Mann A. Paleodemographic aspects of the South African australopithecines. Philadelphia: University of Pennsylvania; 1975.

55. Davies PCW, Lineweaver CH. Cancer tumors as Metazoa 1.0: Tapping genes of ancestors. Phys Biol. 2011;8, Art. #a015001. http://dx.doi.org/10.1088/1478-3975/8/1/015001

56. Rothschild BM, Witzke BJ, Hershkovitz I. Metastatic cancer in the Jurassic. Lancet. 1999;354. p. 398. http://dx.doi.org/10.1016/S0140-6736(99)01019-3

57. Beckett RG, Conlogue GJ. Paleoimaging: Field applications for cultural remains and artefacts. Boca Raton, FL: CRC Press; 2010.

58. Weinberg R. Oncogenes and the molecular biology of cancer. J Cell Biol. 1983;97:1661–1662. http://dx.doi.org/10.1083/jcb.97.6.1661

59. Klepinger L. The evolution of human disease: New findings and problems. J Biosoc Sci. 1980;12(4):481–486. http://dx.doi.org/10.1017/s0021932000013067

Note: This article is accompanied by supplementary material.

Effects of animal class and genotype on beef muscle nanostructure, pH_u, colour and tenderness

AUTHORS:
Amanda Y. Chulayo[1*]
Voster Muchenje[1]

AFFILIATION:
[1]Livestock and Pasture Science, Faculty of Science and Agriculture, University of Fort Hare, Alice, South Africa
[*]Current address: Animal, Wildlife and Grassland Science, Faculty of Natural and Agricultural Science, University of the Free State, South Africa

CORRESPONDENCE TO:
Voster Muchenje

EMAIL:
vmuchenje@ufh.ac.za

POSTAL ADDRESS:
Livestock and Pasture Science, Faculty of Science and Agriculture, University of Fort Hare, Private Bag X1314, Alice 5700, South Africa

DATES:
Received: 02 Feb. 2015
Revised: 01 Nov. 2015
Accepted: 29 Mar. 2016

KEYWORDS:
meat quality; muscle fibre; nanostructure; principal component analysis; scanning electron microscopy

HOW TO CITE:
Chulayo AY, Muchenje V. Effects of animal class and genotype on beef muscle nanostructure, pH_u, colour and tenderness. S Afr J Sci. 2016;112(7/8), Art. #2015-0056, 9 pages. http://dx.doi.org/10.17159/sajs.2016/20150056

The objective of the study was to determine the effects of animal class and genotype of cattle on *Muscularis longissimus thoracis et lumborum* (LTL) nanostructure, ultimate pH (pH_u), colour and tenderness of beef. We found significant positive relationships between distance travelled (DT) and meat temperature (T_m) ($p<0.01$); lairage duration (LD_{hr}) and lightness of colour (L*) ($p<0.01$); ambient temperature (T_a) and L* ($p<0.05$) and LD_{hr} and yellowness (b*) ($p<0.05$) of beef from Bonsmara cattle. Positive linear relationships were observed between DT and T_m ($p<0.05$) and DT and L* ($p<0.01$) of the non-descript cattle. There were no significant relationships between pre-slaughter stress and other beef quality parameters (pH_u, Warner–Bratzler shear force [WBSF], redness [a*] and b*) of Bonsmara, Nguni and non-descript cattle. Muscle fibres differed among class and genotype and had an effect on meat quality. The Bonsmara, non-descript and Nguni cows and heifers had visible skeletal muscle fibres which were thin and long, promising improved tenderness of beef. Genotype and class had significant effects on meat quality parameters (T_m, pH_u, L*, a*, b* and WBSF). The first important principal components as they appeared from the analysis were pH_u, T_m, L*, a*, b* and WBSF. Therefore, animal class did not affect the nanostructure of beef; instead, meat tenderness was enhanced by the longer and visible muscle fibres. Nguni cattle produced meat of superior quality to that of the Bonsmara and the non-descript cattle.

Introduction

Meat is an important source of nutrients for people and most consume it also for its flavour, aroma and tenderness. Meat is composed of about 300 distinct muscles of the carcass including the fat deposited in the muscles and connective tissues. Consumers appreciate meat with a bright red colour and a medium amount of fat, referred to as marbling, because it enhances the flavour, juiciness and tenderness of meat.[1-3] However, factors that interplay 'from farm to fork' – such as sex, age, nutrition, rearing conditions, weight at slaughter and genetics of the cattle as well as environmental conditions – usually have an impact on the quality of the meat.[2,4,5] Other variables affected by these factors include the amount of external and intramuscular fat, the appearance of the meat and sensory properties (such as the aroma, flavour, texture, first impression of juiciness and off-flavours).[6] Moreover, the structure of the muscle and the size of the bundles affect the connective tissue and therefore tenderness. Muscles are differentiated into smooth, cardiac and skeletal depending on their structure, contractile properties and mechanism of control. They are responsible for the support and movement of the animal skeleton; movement is initiated voluntarily through impulses of the neurons in the nervous system.[6-8]

The chemical composition and biological properties of these skeletal muscles are important for a better understanding of the major causes of variations in meat quality, particularly colour, intramuscular fat and tenderness. Meat tenderness is influenced by the amount of myofibrillar and connective tissue of the muscle tissue.[9] Furthermore, muscle development in cattle has an effect on muscle structure, connective tissue and tenderness and the quantity of meat produced for human consumption.[10] Mapiye et al.[11] highlighted that the appropriate utilisation of cattle genotypes with the necessary dietary regimes could be the way forward in order to meet the quality that is required by consumers. According to Strydom et al.[12] and Muchenje et al.[13], among the local genotypes, Nguni cattle have favourable genes that contribute to better outcomes of this genotype in terms of meat quality. However, besides genotype influences, there are other factors such as pre-slaughter conditions, stunning, animal class, age and nutritional factors which influence the acceptability of beef for consumers.[14-16] In addition, it is not possible to improve the tenderness of meat through genetic manipulation of cattle if the nanostructure components – such as sarcomere length, muscle fibre orientation and fibre texture – are not known.[17]

Some muscles are less favoured by consumers as meat because of their reduced tenderness and discolouration. The meat industry therefore requires reliable information on meat quality throughout the production process that will ensure high-quality meat for consumers.[16,18] Variability in meat products prevents the industry from marketing its produce according to quality (tenderness, colour, water-holding capacity and juiciness). Previous studies have been conducted on the improvement of meat quality using genetic variation[16,19], optical scattering and absorption coefficients[9], sensorial consumer evaluation[13,20], and use of video image analysis[13,11,21] to measure sarcomere length – with a longer sarcomere resulting in more tender beef. However, there is little information with regard to the influence of genotype and distance travelled by cattle on the tenderness of beef. Therefore, the objective of this study was to determine the quality of beef as affected by genotype, animal class and distance travelled by cattle.

Materials and methods

Study site

The study was conducted in Buffalo City Municipality at a commercial East London abattoir situated in the Eastern Cape Province of South Africa. The abattoir is governed by the *Meat Safety Act*[22] and the South African Meat Industry Company[23]. The abattoir is located 120 km from the University of Fort Hare in Alice (Nkonkobe Municipality in the Eastern Cape).

Animal management

Cattle (n=170) consisting of three breeds (56 Bonsmara, 65 non-descript and 49 Nguni) were brought to the abattoir in trucks by road from different environments at different times. The animal classes per breed (or genotype) are highlighted in Table 1. Upon arrival at the abattoir, distance travelled by the animals was obtained from the drivers. The minimum and maximum distance travelled (DT) and lairage duration (LD_{hr}) recorded ranged from 183 km to 300 km and from 12 h to 16 h, respectively.

Table 1: Number of animals per class and genotype

Animal class	Genotype of cattle			Total
	Bonsmara	Non-descript	Nguni	
Cows	19	22	18	59
Bulls	17	25	16	58
Heifers	20	18	15	53
Total	56	65	49	170

The ambient temperatures at the abattoir ranged from 14.5 °C to 19.5 °C in the months of May to July 2014 (autumn–winter season). On arrival, the animals were allowed to rest at the lairages and had ad-libitum access to water overnight. The captive bolt method of stunning was used to stun the animals before they were slaughtered, following the standard slaughter procedures at the abattoir.

Meat quality parameters

At 48 h after slaughter, the hanging carcasses were split in half following the abattoir's procedures. A total of 170 representative samples of the *Muscularis longissimus thoracis et lumborum* (LTL) muscle (100-mm thick) were cut between the tenth rib and the third lumbar vertebra of the carcass. The samples were then used to determine ultimate pH (pH_u) and colour coordinates (lightness, L*; redness, a*; and yellowness, b*). From each breed, 36 samples (from 4 bulls, 4 cows and 4 heifers each) were randomly selected to determine the nanostructure of beef.

Determination of pH_u, colour and Warner–Bratzler shear force

A portable fibre-optic pH and T_m meter probe with a sharp metal sheath to prevent damage from raw meat (pH 25, CRISON Instruments S.A., Alella, Spain) was used to measure the ultimate pH and temperature of the carcasses 48 h post-mortem. The pH meter was calibrated before taking measurements using pH 4, pH 7 and pH 9 standard solutions (CRISON Instruments S.A.). The L*, a* and b* colour coordinates were determined at 48 h after slaughter using a Minolta colour guide 45/0 BYK-Gardener GmbH machine with a 20-mm diameter measurement area and illuminant D65-day light, 100 standard observer. The machine was calibrated each day before taking measurements using the green, black and white colour standard samples provided for this purpose. The readings were taken by rotating the colour guide 90° between measurements so as to obtain the average value for the colour. The samples were frozen at -20 °C (for 7 days) until tenderness was measured.

The LTL beef samples were weighed after 1 week of freezing and then thawed for 10 h and weighed again. The samples were placed in a plastic bag and cooked in a water bath at 85 °C for 45 min. After cooking, three sub-samples of meat were cored, parallel to the grain of the meat using a specified core diameter (100 mm). The samples were mounted on an Instron 3344 Universal Testing System (with a crosshead speed of 400 mm/min) and sheared perpendicular to the fibre direction (one shear in the centre of each core) using a Warner–Bratzler shear device. The mean maximum load recorded for the three cores represented the average of the peak force in Newtons (N) for each sample.

Determination of beef nanostructure

LTL muscle samples from four bulls, four cows and four heifers randomly selected from each breed (n=36) were used to determine the nanostructure of beef. The LTL samples were immediately put in a small bottle containing 10% formalin for fixation. The samples were then dehydrated to remove the formalin and kept in ethanol of increasing concentration from 10% to 100% for 20 min at each concentration. In order to improve the electrical conductivity of the sample surface in the scanning electron microscope, a thin film of gold–palladium was used for sputter coating to enhance the analysis. Critical point drying was performed using the Hitachi critical point dryer HCP-2 (Hitachi Koki Co Ltd, Tokyo, Japan) to prevent the samples from alteration and to boost good structural preservation. This was done by mounting the samples on aluminium stubs with double-sided carbon tape then sputter coating with gold–palladium (Au–Pb) using the Eiko IB.3 Ion Coater (EIKO Engineering Co TD, Japan). The samples were then observed under the JEOL JSM-6390LV scanning electron microscope for determination of the skeletal surface area of beef muscles. The nanostructure of the skeletal surface area of beef samples was then viewed using a JEOL JM-5600 scanning electron microscope at a magnification of x5000.

Statistical analysis

PROC GLM[24] (general linear model) was used to determine the influence of class and breed on meat quality parameters. PROC REG[24] (regression) was performed to determine the potential relationships between pre-slaughter conditions (distance travelled, lairage duration and ambient temperature) and meat quality characteristics (pH_u, T_m, L*, a*, b* and Warner–Bratzler shear force [WBSF]). The relationships between pH_u, L*, a*, b* and WBSF, and the width and length between intercalated discs of muscle fibres, among animal classes where genotype was used as a random variable, were determined using principal component analysis in JMP 9.0.[25]

Results and discussion

Animal-related factors, pre-slaughter conditions and beef quality

Table 2 shows the relationships between pre-slaughter conditions – distance travelled (DT), lairage duration (LD_{hr}) and ambient temperature (T_a) – and beef quality characteristics (pH_u, T_m, L*, a*, b* and WBSF). Significant positive relationships were observed between DT and T_m ($p<0.01$), LD_{hr} and L* ($p<0.01$), T_a and L* ($p<0.05$) and between LD_{hr} and b* ($p<0.05$) for Bonsmara cattle. For Nguni cattle, positive relationships were observed between DT and pH_u ($p<0.001$), LD_{hr} and pH_u ($p<0.01$), DT and T_m ($p<0.01$) and between DT and L* ($p<0.01$). A negative linear relationship was observed between DT and a* ($p<0.05$) for Nguni cattle. For the non-descript cattle, positive linear relationships were observed between DT and T_m ($p<0.05$) and DT and L* ($p<0.01$), while significant negative relationships were observed between DT and a* ($p<0.01$), LD_{hr} and T_m ($p<0.001$) and DT and WBSF ($p<0.01$). There were no significant relationships between pre-slaughter stress and other beef quality parameters (pH_u, WBSF, a* and b*) for Bonsmara, Nguni and non-descript cattle. According to Muchenje et al.[13] who observed negative relationships between pre-slaughter conditions and beef quality parameters, poor meat quality can result when animals are exposed to high temperature and humidity during transportation, which causes heat production and reduced energy levels in the muscles.[26]

The biochemical and physiological changes have a detrimental effect on the amount of muscle glycogen which increases pH_u and leads to 'dark, firm and dry' beef. During long transportation distances, animals experience undue stress which reduces meat quality. Stressed animals produce darker and tougher meat, with reduced water-holding capacity.[27] Short lairage duration was highlighted to have a negative impact on meat quality. Similar results were observed in the current study in which LD_{hr} had a negative relationship with the L* and a* values of Bonsmara beef. This is an indication that animals did not experience an adequate lairage period to regain the energy levels lost during transportation. However, the beef parameters of other breeds were not significantly affected by lairage period; therefore the observed relationships could also be because of differences among breeds.

Table 2: Linear relationships between pre-slaughter conditions (distance travelled, lairage duration and ambient temperature) and beef quality characteristics by genotype

Parameter	Relationship	*p*-value
	Bonsmara (n = 56)	
pH_u	$Y = 6.06^{ns} - 0.01^{ns}_{X1} - 0.05^{ns}_{X2} - 0.10^{ns}_{X3}$	0.9735
T_m	$Y = -15.58^{***} + 1.30^{**}_{X1} + 6.19^{**}_{X2} - 6.60^{**}_{X3}$	<0.0001
L*	$Y = 45.37^{***} + 0.12^{ns}_{X1} - 8.03^{**}_{X2} - 5.63^{*}_{X3}$	0.0031
a*	$Y = 9.31^{ns} - 0.36^{ns}_{X1} + 6.24^{ns}_{X2} 5.09^{ns}_{X3}$	0.3471
b*	$Y = 21.96^{*} - 0.42^{ns}_{X1} - 3.70^{*}_{X2} - 0.35^{ns}_{X3}$	0.0054
WBSF	$Y = -28.26^{ns} + 2.63^{ns}_{X1} + 7.38^{ns}_{X2} + 25.26^{ns}_{X3}$	0.0645
	Nguni (n=49)	
pH_u	$Y = 5.61^{***} + 0.05^{***}_{X1} + 0.60^{**}_{X2} - 0.03^{ns}_{X3}$	<0.0001
T_m	$Y = 11.36^{***} + 0.83^{**}_{X1} + 0.40^{ns}_{X2} - 0.25^{ns}_{X3}$	0.0036
L*	$Y = 27.33^{***} + 0.74^{**}_{X1} + 0.70^{ns}_{X2} + 0.59^{ns}_{X3}$	0.0436
a*	$Y = 20.70^{***} - 0.45^{*}_{X1} + 0.56^{ns}_{X2} - 0.65^{ns}_{X3}$	0.0389
b*	$Y = 13.46^{ns} + 0.05^{ns}_{X1} - 0.60^{ns}_{X2} - 0.16^{ns}_{X3}$	0.9106
WBSF	$Y = 46.08^{ns} - 1.08^{ns}_{X1} + 0.60^{ns}_{X2} - 2.00^{ns}_{X3}$	0.2548
	Non-descript (n=65)	
pH_u	$Y = 5.81^{***} - 0.07^{ns}_{X1} + 0.03^{ns}_{X2} + 0.08^{ns}_{X3}$	<0.0001
T_m	$Y = 17.24^{***} + 1.85^{*}_{X1} - 1.69^{ns}_{X2} - 3.55^{***}_{X3}$	<0.0001
L*	$Y = 29.21^{***} + 2.30^{**}_{X1} - 0.91^{ns}_{X2} - 1.07^{ns}_{X3}$	<0.0001
a*	$Y = 9.81^{***} - 2.04^{**}_{X1} + 0.89^{ns}_{X2} + 0070^{ns}_{X3}$	<0.0001
b*	$Y = 13.25^{ns} - 0.07^{ns}_{X1} + 0.34^{ns}_{X2} + 1.16^{ns}_{X3}$	0.0737
WBSF	$Y = 56.67^{***} - 12.58^{**}_{X1} + 3.84^{ns}_{X2} - 0.39^{ns}_{X3}$	<0.0001

pH_u, ultimate pH; T_m, meat temperature; L, lightness; a*, redness; b*, yellowness; WBSF, Warner–Bratzler shear force*

$_{X1}$, distance travelled; $_{X2}$, lairage duration; $_{X3}$, ambient temperature

*Significance level: *p<0.05; **p<0.01; ***p<0.001 and ns, not significant (p>0.05)*

The relationship observed between genotype and meat quality parameters was linked to the response of cattle to different environmental conditions prior to slaughter. There was a decline in muscle performance during long hours of transportation, which could be recovered through enough hours of rest at the lairages.[28] However, if care is not taken, some animals die before they arrive at the abattoir, which also is a loss to the industry.[29,30]

The effects of genotype on beef quality characteristics are presented in Table 3. There were significant effects of genotype on T_m ($p<0.001$), pH_u ($p=0.0452$), L* ($p<0.001$), a* ($p<0.001$), b* ($p<0.001$) and WBSF ($p<0.05$). T_m (16.5±0.18 °C) and L* (38.6±0.76) were higher in beef from Nguni cattle than that from the non-descript and Bonsmara cattle (Table 3). The difference in the L* values was a result of the amount of myoglobin in the muscle. The lightness of the colour is used by consumers when assessing the visual appearance of meat and thereby influences the purchasing decisions of consumers as reported by Vimiso et al.[3] The non-descript cattle had higher values for yellowness (18.4±0.39) and WBSF (54.4±1.69 N) than did Bonsmara (15.7±1.46 and 51.5±2.20 N, respectively) and Nguni (13.3±0.37 and 48.6±1.80 N, respectively) cattle. Genotype therefore had an effect on the WBSF values. In a previous study, genotype was the major determinant of meat quality, especially tenderness, which is considered to be one of the most important factors for consumer satisfaction.[31]

Genotype, age and weight of the animal at slaughter determines the type of meat grade that is produced.[32-34] Nguni cattle reportedly produce leaner meat grades than Bonsmara and Angus cattle because of variations in intramuscular fat, fibre types and moisture content.[1] However, different fibre types of the skeletal muscles could also have been the cause of variation in tenderness between these three breeds of cattle. A similar report[35] indicated that breed had an influence on fibre composition, protein availability, and molecular structure and thereby contractile texture of the muscle, all of which influences tenderness. It was also reported that as the animal ages, the meat becomes redder, tougher and darker as a result of the amount of collagen, haem and myoglobin in the muscles.[32,36,37]

Table 3: Least square means and standard errors of beef quality parameters from cattle by genotype

Parameter	Genotype			*p*-value
	Bonsmara (n=56)	Nguni (n=49)	Non-descript (n=65)	
Ultimate pH	5.7^{a} ± 0.04	5.7^{a} ± 0.03	5.8^{b} ± 0.03	0.0452
Meat temperature	14.5^{a} ± 0.22	16.5^{c} ± 0.18	15.5^{b} ± 0.17	0.0001
Lightness	32.1^{a} ± 1.40	38.6^{b} ± 0.76	33.5^{a} ± 0.54	<0.0001
Redness	18.8^{b} ± 0.52	16.1^{a} ± 0.42	18.4^{b} ± 0.48	<0.0001
Yellowness	15.7^{b} ± 1.46	13.3^{a} ± 0.37	18.4^{c} ± 0.39	0.0002
Warner–Bratzler shear force	51.5^{b} ± 2.20	48.6^{a} ± 1.80	54.4^{c} ± 1.69	0.043

^{a,b,c}Means with different superscripts within a row are significantly different (p<0.05)

Table 4: Least square means and standard errors of beef quality parameters from cattle by class

Parameter	Animal class			*p*-value
	Heifers (n=53)	Bulls (n=58)	Cows (n=59)	
Ultimate pH	5.7 ± 0.04	5.7 ± 0.07	5.7 ± 0.03	0.0689
Meat temperature	16.2^{b} ± 0.20	16.0^{b} ± 0.36	15.2^{a} ± 0.16	0.0007
Lightness	32.3 ± 0.56	31.7 ± 0.03	31.7 ± 0.44	0.6457
Redness	17.0 ± 0.45	18.1 ± 0.82	18.1 ± 0.36	0.0629
Yellowness	13.9 ± 0.40	15.5 ± 0.73	14.8 ± 0.37	0.0712
Warner-Bratzler shear force	44.2^{a} ± 1.80	57.7^{c} ± 0.37	48.8^{b} ± 1.42	0.0089

^{a,b,c}Means with different superscripts within a row are significantly different (p<0.05)

Monsón et al.[38] reported that double-muscled animals produced meat with improved tenderness compared to others. Sañudo et al.[33] also indicated that the high WBSF values from larger animals led to tougher beef as a result of a higher intake of grass all year round. The amount of collagen in the muscle may be reduced and transformed into soluble collagen during cooking.[39] Enhanced tenderness improves the value of the final meat product.

Table 4 shows the effect of animal class on the quality parameters of beef. Significant differences in beef quality parameters were observed between classes, with bulls having higher values of T_m (16.0±0.36 °C) and WBSF (57.7±3.30 N) than heifers (T_m=16.2±0.26 °C and WBSF=44.2±1.80 N) and cows (T_m=15.2±0.16 °C and WBSF=48.8±1.42 N). The results indicated that beef from bulls was relatively tough with less intramuscular

fat and reduced flavour. Animal classes of cattle did not exhibit the same body weight, conformation and fatness. Bulls were characterised by higher lean meat content, lower fat and higher bone content than heifers and cows. It has been reported that meat from bulls had undesirable traits such as high pH_u, dark colour and toughness, and was hence less desirable for direct market sales.[39,40] The observed greater tenderness of meat from heifers was a result of a higher intramuscular fat content and the smaller diameter of muscle fibres. In addition, tenderness increased with age as a result of the structural changes that occurred in the muscles, thus increasing the purchasing and acceptability of the meat because of its enhanced juiciness and flavour.[39,41]

Nanostructure of beef muscles

Figure 1 shows the nanostructure of LTL muscles from the non-descript, Bonsmara and Nguni cattle of different classes (bull, cow and heifer). Among the non-descript, Bonsmara and Nguni bulls, different fibre orientation and length between discs were revealed, with non-descript cattle having less visible intercalated discs than the other breeds. However, it was observed that Bonsmara, non-descript and Nguni cows had visible fibres which were thin and long for improved meat tenderness. The observed differences in the muscle fibres has been anecdotally reported to be because of the finer muscle grain that some genotypes possess, making the muscle fibres visible microscopically.[18,42] This characteristic is appreciated by consumers because it enhances tenderness.

The heifers of Nguni and non-descript cattle had thin and visible intercalated discs, with few traces of the intramuscular fat that enhances flavour and tenderness. Although conformation and fatness were not determined, it has been reported that these parameters have an effect on fibre orientation.[32] Intramuscular fat increases with age because it develops late in the maturation of cattle.[43-45] Intramuscular fat consists of fat cells which are situated in the perimysium and endomysium that surround myofibrils and muscle fibre bundles. The Bonsmara heifers had visible and thick muscle fibres which indicate tougher meat. Muscle fibre type is linked to variations in the glycolytic rate among animal classes of different genotypes, changes in growth, and meat quality traits. As an animal ages, the collagen content and stability of cross-bridges increases, leading to reduced meat tenderness.[18,46] Muscle structure and composition is influenced by breed, and is a factor which must be considered during the selection of meat animals to ensure beef quality as well as quantity.[1]

Figure 2 highlights the differences in length between intercalated discs of muscle from different breeds. The length between the z-line of cows of the non-descript breed ranged from 967.47 nm to 1.33 μm whereas the width ranged from 441.81 nm to 684.69 nm. The length between the z-lines of the Nguni cow meat ranged from 720 nm to 1.12 μm while that of the Bonsmara ranged from 1.39 μm to 1.72 μm. The fibre orientations were significantly different between cattle genotypes. Fibre orientation is linked to the tenderness of meat and is greatly influenced by muscle structure of different genotypes and age groups.[47] From our results, Bonsmara cows had elongated z-lines compared with the non-descript and Nguni cows, indicating that the sarcomere length of Bonsmara cows was longer, which results in more tender meat and higher acceptability by consumers. Fibres are essential units of all muscles; some are long and narrow with multinucleated cells.[48] Previous research has indicated that stress decreases the length of the fibres. The sliding motion of many cross-bridges forces the thin filaments (actin) towards the centre of a sarcomere, so shorter fibres affect the sarcomere length and produce tougher meat.[6] Swatland[17] has indicated that it is important to understand the function of sarcomere length which forms the length of fibres in order to improve the tenderness of meat. Toughness is associated with dark, firm and dry meat, found especially in animals that are starved for too long.[49,50] The collagen content and the stability of cross-bridges increases as the animal ages, causing reduced tenderness.[46] This finding was also reported by Koohmaraie and Geesink[43] who observed that proteins such as nebulin and desmin play a role in the post-mortem tenderisation of meat as they lie next to the z-line.

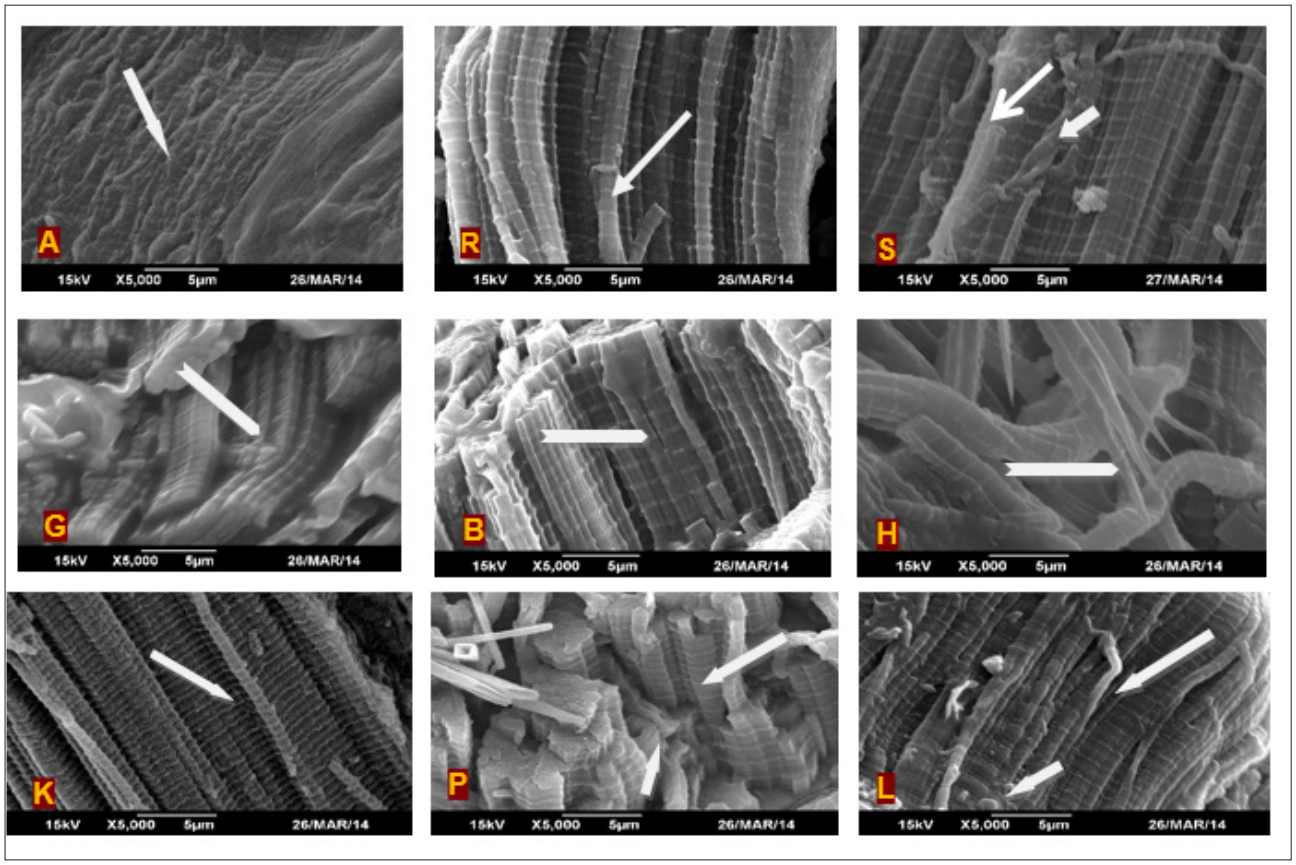


Figure 1: Nanostructure of the *Muscularis longissimus thoracis et lumborum* muscle of different cattle genotypes and classes: (A) bulls, (R) cows and (S) heifers of non-descript cattle; (G) bulls, (B) cows and (H) heifers of Bonsmara cattle; and (K) bulls, (P) cows and (L) heifers of Nguni cattle.

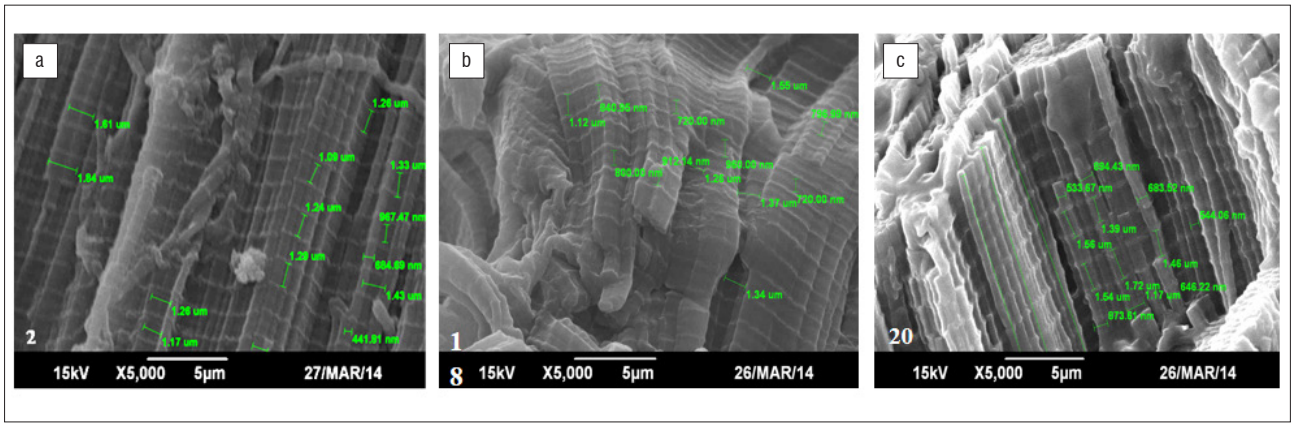


Figure 2: Length and width between intercalated discs of the *Muscularis longissimus thoracis et lumborum* muscle of cows from (a) non-descript, (b) Nguni and (c) Bonsmara breeds of cattle.

A study conducted in chicken by Łukasiewicz et al.[49] indicated that genotype had an influence on the histological structure of muscles, such that tenderness, moisture content and muscle pH were affected. Tougher meat was associated with lean and older animals, and especially those that were stressed in the pre-slaughter environment.[49] Less tender meat also resulted from muscle fatigue pre-slaughter. During this ante-mortem period, there was a rapid increase in the size of muscle fibres and an increase in the diameter through expansion of new myofibrils and sarcoplasm, which led to tougher meat.[10,48] Because meat is a complex structure that is affected by several factors during its production chain, it is of paramount importance to examine the sarcoplasmic changes of animals destined for slaughter.

The length between the z-line of the Nguni bulls was between 720 nm and 1.12 µm, whereas the Brahman cow had a length of 1.39–1.72 µm. The shorter distance between the z-lines for the Nguni bulls indicates tougher meat, because the length is below the normal values of 1.5–3.5 µm.[6] White et al.[20] indicated that in hot-boned and non-electrically stimulated beef, muscles had shorter sarcomeres which led to higher values for WBSF. In addition, it should be noted that bulls usually produce tougher meat than cows because of differences in the amount of fat between the sexes with cows having more intramuscular fat.[50,51] Studies have indicated that the microstructure alterations in beef intramuscular connective tissue are caused by hydrodynamic pressure processing.[52] Xia et al.[9] characterised cooked beef muscles using optical scattering and absorption coefficients to measure tenderness. However, more research is required to determine the effects of pre-slaughter stress on the sarcomere length of cooked muscles.

Beef quality parameters

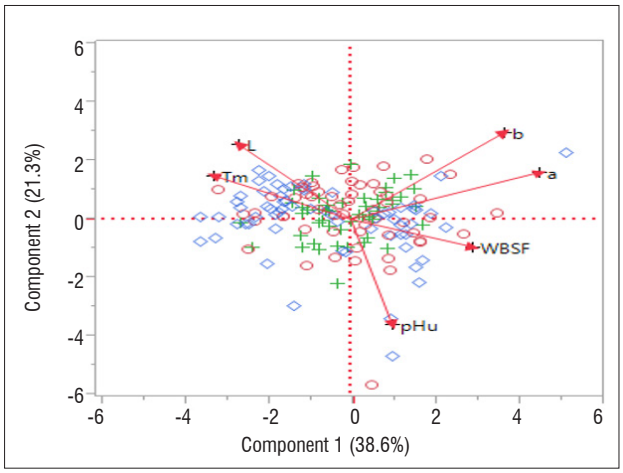
Table 5 presents the eigenvalues of beef from heifers, bulls and cows according to the percentage contribution to the total variance. The first important principal components as they appear from the principal component analysis were pH_u, T_m, L*, a*, b* and WBSF, with eigenvalues greater than 0.5. The literature indicates that principal components with an eigenvalue greater than 0.5 are considered to be important to contribute to the quality of beef.[53] The first two principal components – pH_u and T_m – of beef from heifers contributed about 94% of the total variance while pH_u and T_m of meat from bulls and cows contributed about 65% and 62%, respectively. Figure 3 indicates the relationship between tenderness and visual aspects of beef quality. The first principal component is explained by the pH_u, T_m and WBSF while the second principal component is explained by L*, a* and b*. pH_u was positively correlated with WBSF but negatively correlated with a* and b* values. Significant differences were observed between a* and b* and between L* and T_m.

Table 5: Contribution of the principal components to the total variance of beef quality by animal class

Principal component	Eigenvalue	Portion of variance (%)	Cumulative variance (%)
Heifers			
Ultimate pH	3.353	55.894	55.894
Meat temperature	2.332	38.871	94.765
Lightness	0.314	5.235	100.000
Bulls			
Ultimate pH	2.410	40.166	40.166
Meat temperature	1.546	25.768	65.934
Lightness	1.477	24.620	90.554
Redness	0.395	6.595	97.149
Yellowness	0.171	2.851	100.000
Cows			
Ultimate pH	2.487	41.460	41.460
Meat temperature	1.261	21.024	62.485
Lightness	0.942	15.704	78.189
Redness	0.714	11.914	90.103
Yellowness	0.547	9.124	99.226
Warner–Bratzler shear force	0.046	0.774	100.000

The observed negative relationships between pH_u and L* have been previously reported in beef.[54,55] This finding implies that the amount of myoglobin was affected by a pH decline in the muscles after slaughter.[56] Muscle pH decreased during post-mortem storage of meat as a result of accumulation of lactic acid during glycolysis. However, contrastingly, Purchas[57] indicated that ultimate pH and colour coordinates were not related. The technological or physico-chemical properties of meat included meat pH, water-holding capacity, colour, tenderness, thawing and cooking loss, evaporation loss and water distribution.[13,53,58]

According to Hoffman et al.[59], meat quality is the sum of all quality factors of meat in terms of the sensory, nutritive, hygienic, toxicological and technological properties. Muchenje et al.[13] first indicated that pH_u affected the colour and tenderness of beef. Contradictory reports have, however, shown that pH_u is a weak predictor of final meat quality.[60]



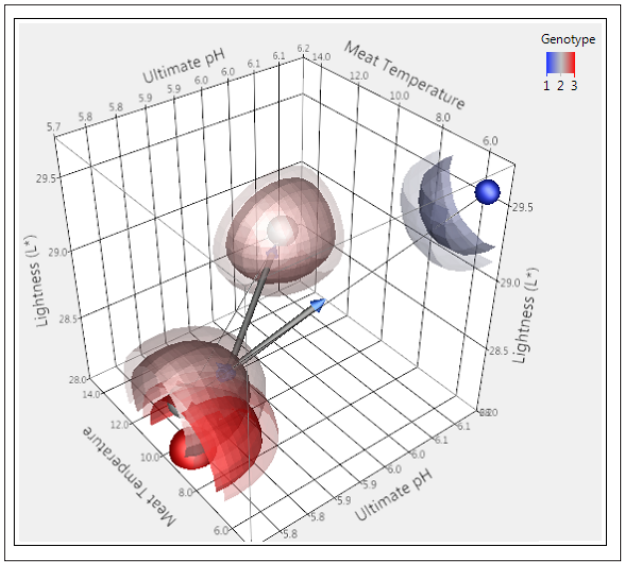
pH_u, ultimate pH; T_m, meat temperature; L, lightness; a*, redness; b*, yellowness; WBSF, Warner–Bratzler shear force*

Figure 3: Relationship between tenderness and visual qualities of beef.

Heifers, bulls and cows had the highest percentage variance for pH_u (55%, 40% and 41%, respectively) in comparison to the other meat quality parameters. It has been previously reported that pH_u made the highest contribution to beef quality in all animal classes (heifers, bulls and cows).[13] This finding agrees with that of Andrés-Bello et al.[61] who indicated that pH was the major determinant of functional foods because of the influence it had on other meat quality characteristics. The concentration of hydrogen ions in meat were the determinant of the proceeding chemical reaction, especially when the animal was in a stressful environment which apparently reduced the quality of meat.[62] Ultimate pH had an effect on the content of myoglobin which was responsible for the bright colour of meat. In addition, pH was responsible for the increased growth, while at the same time, death of microorganisms, to enhance the enzymatic activities which determine the shelf life of meat. Meat temperature also has an effect, because high meat temperatures favour the growth of microorganisms which eventually deteriorates the quality and shelf life of meat. Moreover, a higher meat temperature also increases the mobility of hydrogen ions as a result of the dissociation that occurs in the molecules, which leads to a change in pH.[63,64]

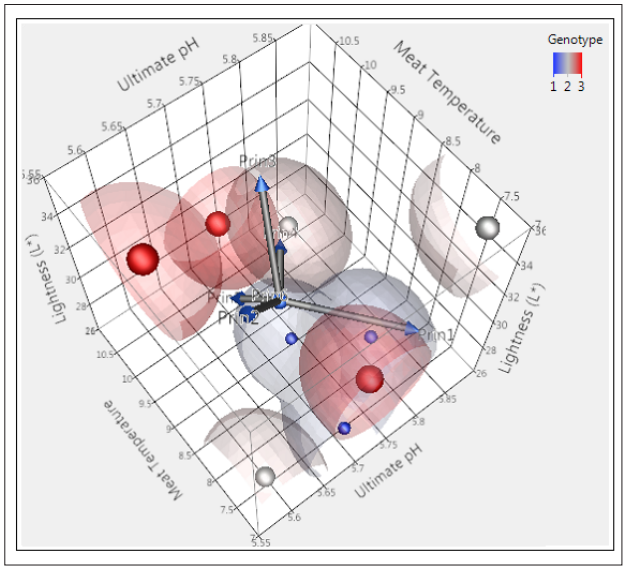
Lightness of colour contributed about 5% to meat quality in heifers, although in bulls and cows it contributed about 24% and 15% of the total variance, respectively. So although L* was less important in the quality of meat from heifers, it is an important parameter in meat quality in beef from bulls and cows. The L* values were dependent on the animal class, based on the accumulation of fat, amount of myoglobin and the type of feed consumed. Myoglobin is the principal protein (water soluble) found in muscles and is responsible for the bright colour of meat. Once the colour of meat is reduced, the acceptability of meat is also affected because colour is one of the sensory characteristics used by consumers to judge the freshness of beef. Andrés-Bello et al.[61] and Dewi et al.[62] agree that colour is important as it affects consumer acceptance of meat. The relationship observed between L* and T_m is associated with the handling of meat after slaughter. An increase in meat temperature reduces the colour of meat and provides a medium for meat spoilage by bacteria.[65] Mancini and Hunt[66] further indicated that the formation of metmyoglobin depended on temperature and pH, and that meat lightness was the third principal component after pH_u and T_m. However, further research is needed to determine the amount of myoglobin that is required to give a bright and acceptable colour to beef from different genotypes and classes of cattle.

Figures 4–6 present the first three eigenvalues of beef from heifers, bulls and cows, respectively. Heifers, bulls and cows of the Nguni cattle had normal pH_u, T_m and L*. Sañudo et al.[37] reported that meat becomes darker owing to different factors related to animal age and muscle energy which lead to poor colour stability. The results from the study were also related to age differences between classes as it was observed that older animals tend to have a higher myoglobin content which lowers the L* value leading to darker meat.[66] The observed darker meat in Bonsmara cattle was also linked to reduced intramuscular fat, as fat also plays a role in the brightness of meat.[11] In addition, darker meat was caused by an increased ultimate pH as a result of depleted glycogen levels and reduced production of lactic acid. Li et al.[53] highlighted that pH_u influenced the extent of protein denaturation, colour and water-holding capacity of fresh meat. In addition, pH variation is caused by mitochondria consumption[67], and the size, thickness and location of the muscle.



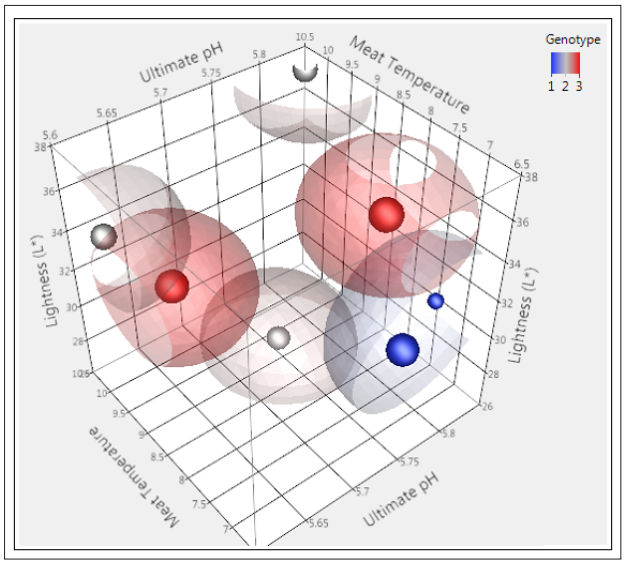
Genotype: 1, Bonsmara; 2, non-descript; 3, Nguni

Figure 4: The first three principal components of beef *Muscularis longissimus thoracis et lumborum* muscle from heifers for meat attributes related to meat quality.



Genotype: 1, Bonsmara; 2, non-descript; 3, Nguni

Figure 5: The first three principal components of beef *Muscularis longissimus thoracis et lumborum* muscle from bulls for meat attributes related to meat quality.



Genotype: 1, Bonsmara; 2, non-descript; 3, Nguni

Figure 6: The first three principal components of beef *Muscularis longissimus thoracis et lumborum* muscle from cows for meat attributes related to meat quality.

As observed in Table 5, pH_u is the primary attribute that controls the outcome of the other meat quality attributes. This implies that when considering the improvement of beef, one has to make sure that the factors that influence pH are minimised as far as possible. In agreement with this study, reports by Li et al.[53], Vimiso and Muchenje[55] and Scholtz[68] indicated that pH_u was the major factor in meat quality and the extent of protein denaturation. In addition, the amount of protein in the muscles was negatively affected by reduced metmyoglobin which affected the water-holding capacity and colour stability of the muscle.[61,69] This finding is in agreement with reports by Xia et al.[9] in which the variation of the pigment and oxidation state led to differences in the L* and a* values.

Conclusion

It could be inferred from this study that animal class and genotype do not affect the nanostructure of beef. Among the Bonsmara, non-descript and Nguni cattle, heifers had a better meat quality than bulls and cows. The first two principal components of beef from heifers had the highest contribution of the total variance followed by bulls and cows. Therefore, the nanostructure of beef was not affected by animal class, with heifers having the best meat followed by bulls and cows. Considering the most important beef quality traits, heifers produced better meat, with pH_u, T_m and L* contributing the highest percentages to the total variance.

Acknowledgements

We are grateful to the National Research Foundation of South Africa and the Red Meat Research and Development, Technology and Human Resources for Industry Programme (Projects T079 and T113) and Govan Mbeki Research and Development Centre, University of Fort Hare for funding this research. We thank Ms Elizabeth Famewo, Ms Papadi Nakalebe and Ms Maurine Mangoale for their support during the analysis of the nanostructure of beef at the Scanning Electron Microscopy Unit, University of Fort Hare.

Authors' contributions

A.Y.C. developed this project together with V.M. and designed the data collection process. A.Y.C. further analysed, interpreted and discussed the results while V.M. proofread the manuscript and submitted it to the journal. V.M. also used funds from his research grants to purchase the equipment needed for this research.

References

1. Sisson S, Grossman JD. The anatomy of domestic animals. 4th ed. Philadelphia, PA: WB Saunders; 1953.
2. Maltin CA, Sinclair KD, Warriss PD, Grant CM, Porter AD, Delday MI, et al. The effects of age at slaughter, genotype and finishing systems on the biochemical properties, muscle fibre type characteristics and eating quality of bull beef from suckled calves. Anim Sci. 1998;66:341–348. http://dx.doi.org/10.1017/S1357729800009462
3. Vimiso P, Muchenje V, Marume U, Chiruka R. Preliminary study on consumers' and meat traders' perceptions of beef quality and how the beef quality is affected by animal welfare practices. Sci Res Essays. 2012;7(22):2037–2048. http://dx.doi.org/10.5897/SRE12.071
4. Muchakilla MB, Asimwe L, Kimambo AE, Mtenga LA, Laswai GH. Effect of diet and muscle type on meat quality characteristics of Tanzania Shorthorn Zebu. Livestock Res Rural Dev. 2014;26(10), Art. #186. Available from: http://www.lrrd.org/lrrd26/10/much26186.html.
5. Iwanowska A, Pospiech E. Comparison of slaughter value and muscle properties of selected cattle breeds in Poland – Review. Acta Sci Pol Technol Aliment. 2010;9(1):7–22.
6. Loeb GE, Ghez C. The motor unit and muscle action. In: Kandel ER, Schwartz JH, Jessell TM, editors. Principles of neural science. New York: McGraw-Hill; 2000. p. 674–693.
7. Fayemi PO. Perceptions on ante-mortem welfare, quantitation of pain and pregnancy biomarkers, muscular fibre architecture and quality of Dohne Merino offal [PhD thesis]. Alice: University of Fort Hare; 2013.
8. Albrecht E, Teuscher F, Ender K, Wegner J. Growth- and breed-related changes of muscle bundle structure in cattle. J Anim Sci. 2014;84:2959–2964. http://dx.doi.org/10.2527/jas.2006-345
9. Xia JJ, Berg EP, Lee JW, Yao G. Characterizing beef muscles with optical scattering and absorption coefficients in VIS-NIR region. Meat Sci. 2007;75:78–83. http://dx.doi.org/10.1016/j.meatsci.2006.07.002
10. Albrecht E, Lembcke C, Wegner J, Maak S. Prenatal muscle fiber development and bundle structure in beef and dairy cattle. J Anim Sci. 2013;91:3666–3673. http://dx.doi.org/10.2527/jas.2013-6258
11. Mapiye C, Chimonyo M, Marufu MC, Muchenje V. Stress reactivity and its relationship to beef quality in Nguni steers supplemented with Acacia Karroo leaves. Animal. 2011;5(9):1361–1369. http://dx.doi.org/10.1017/S1751731111000395
12. Strydom PE, Naude RT, Smith MF, Scholtz MM, Van Wyk JB. Characterisation of indigenous African cattle breeds in relation to meat quality traits. Meat Sci. 2000;55:79–88. http://dx.doi.org/10.1016/S0309-1740(99)00128-X
13. Muchenje V, Dzama K, Chimonyo M, Raats JG, Strydom PE. Meat quality of Nguni, Bonsmara and Aberdeen Angus steers raised on natural pasture in the Eastern Cape, South Africa. Meat Sci. 2008;79:20–28. http://dx.doi.org/10.1016/j.meatsci.2007.07.026
14. Terlouw C. Stress reactions at slaughter and meat quality in pigs: genetic background and prior experience. A brief review of recent findings. Livestock Prod Sci. 2005;94:125–135. http://dx.doi.org/10.1016/j.livprodsci.2004.11.032
15. Terlouw EMC, Rybarczyk P. Explaining and predicting differences in meat quality through stress reactions at slaughter: The case of Large White and Duroc pigs. Meat Sci. 2008;79:795–805. http://dx.doi.org/10.1016/j.meatsci.2007.11.013
16. Robinson DL, Ferguson DM, Oddy VH, Perry D, Thompson J. Genetic and environmental influences on beef tenderness. Aust J Exp Agric. 2001;41:997–1003. http://dx.doi.org/10.1071/EA00039
17. Swatland HJ. Fibre-optic spectrophotometry of beef relative to sarcomere length. Arch Tier Dummerst. 2003;46(1):31–34.
18. Damez JL, Clerjon S. Meat quality assessment using biophysical methods related to meat structure. Meat Sci. 2008;80:132–149. http://dx.doi.org/10.1016/j.meatsci.2008.05.039
19. Greeff JC, Safari E, Fogarty NM, Hopkins DL, Brien FD, Atkins KD, et al. Genetic parameters for carcass and meat quality traits and their relationships to live weight and wool production in hogget Merino rams. J Anim Breed Genet. 2008;125:205–215. http://dx.doi.org/10.1111/j.1439-0388.2007.00711.x
20. White A, O'Sullivan A, Troy DJ, O'Neill EE. Effects of electrical stimulation, chilling temperature and hot-boning on the tenderness of bovine muscles. Meat Sci. 2006;73:196–203. http://dx.doi.org/10.1016/j.meatsci.2005.11.020

21. Marino R, Albenzio M, Della Malva A, Caroprese M, Santillo A, Sevi A. Changes in meat quality traits and sarcoplasmic proteins during aging in three different cattle breeds. Meat Sci. 2014;98:178–186. http://dx.doi.org/10.1016/j.meatsci.2014.05.024

22. Meat Safety Act, Act no. 40 of 2000, South Africa. Red Meat Regulations, 2004 Part V: Humane treatment of animals and slaughter process [Section 11(1) (b)], 64. Requirements for vehicles.

23. South African Meat Industry Company (SAMIC). Classification of South African beef – A key to consumer satisfaction. Pretoria: SAMIC; 2006.

24. Statistical Analysis System (SAS). SAS/STAT Guide to personal computers, version 6. Cary, NC: SAS Institute; 2003.

25. Statistical Analysis System (SAS). JMP 9: Discovering JMP. Cary: SAS Institute; 2010.

26. Ferguson DM, Warner RD. Have we underestimated the impact of pre-slaughter stress on meat quality? Meat Sci. 2008;80:12–19. http://dx.doi.org/10.1016/j.meatsci.2008.05.004

27. Allen DG, Lamb GD, Westerblad H. Skeletal muscle fatigue: Cellular mechanisms. Physiol Rev. 2008;88:287–332. http://dx.doi.org/10.1152/physrev.00015.2007

28. Dalla Costa OA, Faucitano L, Coldebella A, Ludke JV, Peloso JV, Dalla Roza D, et al. Effects of the season of the year, truck type and location on truck on skin bruises and meat quality in pigs. Livestock Sci. 2007;107:29–36. http://dx.doi.org/10.1016/j.livsci.2006.08.015

29. Hoffman LC, Lühl L. Causes of cattle bruising during handling and transport in Namibia. Meat Sci. 2012;92:115–124. http://dx.doi.org/10.1016/j.meatsci.2012.04.021

30. Van der Wal PG, Engel B, Reimert HGM. The effect of stress, applied immediately before stunning on pork quality. Meat Sci. 1999;53:101–106. http://dx.doi.org/10.1016/S0309-1740(99)00039-X

31. Silva CCG, Rego OA, Simões ERE, Rosa HJD. Consumption of high energy maize diets is associated with increased soluble collagen in muscle of Holstein bulls. Meat Sci. 2010;86:753–757. http://dx.doi.org/10.1016/j.meatsci.2010.06.017

32. Hulof F, Ouhayoun J. Muscular pH and related traits in rabbits: A review. World Rabbit Sci. 1999;7(1):15–36.

33. Sañudo C, Macie ES, Olleta JL, Villarroel M, Panea B, Albertí P. The effects of slaughter weight, breed type and ageing time on beef meat quality using two different texture devices. Meat Sci. 2004;66:925–932. http://dx.doi.org/10.1016/j.meatsci.2003.08.005

34. Hanzelková Š, Simeonovová J, Hampel D, Dufek A, Šubrt J. The effect of breed, sex and aging time on tenderness of beef meat. Acta Vet Brno. 2011;80:191–196. http://dx.doi.org/10.2754/avb201180020191

35. Ryu JH, Kim SH, Lee HY, Bai JY, Yam YD, Bae JW, et al. Innate immune homeostasis by the homeobox gene *Caudal* and commensal gut mutualism in *Drosophila*. Science. 2008;319(5864):777–782. http://dx.doi.org/10.1126/science.1149357

36. Martínez-Cerezo S, Sañudo C, Panea B, Medel I, Delfa R, Sierra I, et al. Breed, slaughter weight and ageing time effects on physico-chemical characteristics of lamb meat. Meat Sci. 2005;69:325–333. http://dx.doi.org/10.1016/j.meatsci.2004.08.002

37. Sañudo C, Olleta JL, Campo MM, Panea B, Renand G, Turin F, et al. Meat quality of ten cattle breeds of the southwest of Europe. FAIR1 CT95 0702 – Final Report. c2014 [cited 2014 Nov 26]. p. 190–232. Available from: http://www.recercat.net/bitstream/handle/2072/4685/informefinalcarne.pdf.

38. Monsón F, Sañudo C, Sierra I. Influence of cattle breed and ageing time on textural meat quality. Meat Sci. 2004;68:595–602. http://dx.doi.org/10.1016/j.meatsci.2004.05.011

39. Węglarz A. Quality of beef from semi-intensively fattened heifers and bulls. Anim Sci Pap Rep. 2010;28(3):207–218.

40. Schiavon S, De Marchi M, Tagliapietra F, Bailoni L, Cecchinato A, Bittante G. Effect of high or low protein ration combined or not with rumen protected conjugated linoleic acid (CLA) on meat CLA content and quality traits of double-muscled Pietmontese bulls. Meat Sci. 2011;89:133–142. http://dx.doi.org/10.1016/j.meatsci.2011.03.025

41. Li C, Zhou G, Xu X, Zhang J, Xu S, Ji Y. Effects of marbling on meat quality characteristics and intramuscular connective tissue of beef longissimus muscle. Asian Australas J Anim Sci. 2006;19(12):1799–1808. http://dx.doi.org/10.5713/ajas.2006.1799

42. Christensen M, Ertbjerg P, Failla S, Sañudo C, Richardson RI, Nute GR, et al. Relationship between collagen characteristics, lipid content and raw and cooked texture of meat from young bulls of fifteen European breeds. Meat Sci. 2011;87(1):61–65. http://dx.doi.org/10.1016/j.meatsci.2010.09.003

43. Koohmaraie M, Geesink GH. Contribution of post-mortem muscle biochemistry to the delivery of consistent meat quality with particular focus on the calpain system. Meat Sci. 2006;74:34–43. http://dx.doi.org/10.1016/j.meatsci.2006.04.025

44. Ahnström ML. Influence of pelvic suspension on beef meat quality [PhD thesis]. Uppsala: Swedish University of Agricultural Sciences; 2008.

45. Webb EC, O'Neill HA. The animal fat paradox and meat quality. Meat Sci. 2008;80:28–36. http://dx.doi.org/10.1016/j.meatsci.2008.05.029

46. Petracci M, Cavani C. Muscle growth and poultry meat quality issues. Nutrients. 2012;4:1–12. http://dx.doi.org/10.3390/nu4010001

47. Muchenje V, Dzama K, Chimonyo M, Strydom PE, Hugo A, Raats JG. Some biochemical aspects pertaining to beef eating quality and consumer health: A review. Food Chem. 2009;112:279–289. http://dx.doi.org/10.1016/j.foodchem.2008.05.103

48. Dubost A, Micol D, Picard B, Lethias C, Andueza D, Bauchart D, et al. Structural and biochemical characteristics of bovine intramuscular connective tissue and beef quality. Meat Sci. 2013;95:555–561. http://dx.doi.org/10.1016/j.meatsci.2013.05.040

49. Łukasiewicz M, Niemiec J, Wnuk A, Mroczek-Sosnowska N. Meat quality and the histological structure of breast and leg muscles in Ayam Cemani chickens, Ayam Cemani x Sussex hybrids and slow-growing Hubbard JA 957 chickens. J Sci Food Agric. 2014;95(8):1730–1750. http://dx.doi.org/10.1002/jsfa.6883

50. Lind V, Berg J, Eilertsen SM, Hersleth M, Eik LO. Effect of gender on meat quality in lamb from extensive and intensive grazing systems when slaughtered at the end of the growing season. Meat Sci. 2011;88:305–310. http://dx.doi.org/10.1016/j.meatsci.2011.01.008

51. Guerrero A, Velandia Valero M, Campo MM, Sanudo C. Some factors that affect ruminant meat quality: From the farm to the fork: Review. Acta Sci Anim Sci. 2013;35(4):335–347.

52. Zuckerman H, Bowker BC, Eastridge JS, Solomon MB. Microstructure alterations in beef intramuscular connective tissue caused by hydrodynamic pressure processing. Meat Sci. 2013;95:603–607. http://dx.doi.org/10.1016/j.meatsci.2013.05.041

53. Li YX, Cabling MM, Kang HS, Kim TS, Yeom SC, Sohn YG, et al. Comparison and correlation analysis of different swine breeds meat quality. Asian Australas J Anim Sci. 2013;26(7):905–910. http://dx.doi.org/10.5713/ajas.2012.12622.

54. Kopuzlu S, Onenc A, Bilgin OC, Esenbuga N. Determination of meat quality through principal components analysis. J Anim Plant Sci. 2011;21(2):151–156.

55. Vimiso P, Muchenje V. A survey on the effect of transport method on bruises, pH and colour of meat from cattle slaughtered at a South African commercial abattoir. S Afr J Anim Sci. 2013;43:105–111. http://dx.doi.org/10.4314/sajas.v43i1.13

56. Byrne CE, Troy DJ, Buckley DJ. Post-mortem changes in muscle electrical properties of bovine *M. longissimus dorsi* and their relationship to meat quality attributes and pH fall. Meat Sci. 2000;54:23–34. http://dx.doi.org/10.1016/S0309-1740(99)00055-8

57. Purchas RW. An assessment of the role of pH differences in determining the relative tenderness of meat from bulls and steers. Meat Sci. 1990;27:129–140. http://dx.doi.org/10.1016/0309-1740(90)90061-A

58. Kušec G, Kralik G, Petričević A, Gutznirtl H, Grgurić D. Meat quality indicators and their correlation in two crosses of pigs. Agric Conspec Sci. 2003;86(2):115–119.

59. Hoffman LC, Kroucamp M, Manley M. Meat quality characteristics of springbok (*Antidorcas marsupialis*). 4: Sensory meat evaluation as influenced by age, gender and production region. Meat Sci. 2007;76:774–778. http://dx.doi.org/10.1016/j.meatsci.2007.02.020

60. Gajana CS, Nkukwana TT, Marume U, Muchenje V. Effects of transportation time, distance, stocking density, temperature and lairage time on incidences of pale, exudative (PSE) and the physicochemical characteristics of pork. Meat Sci. 2013;95:520–525. http://dx.doi.org/10.1016/j.meatsci.2013.05.028

61. Andrés-Bello A, Barreto-Palacios V, García-Segovia P, Mir-Bel J, Martínez-Monzó J. Effect of pH on color and texture of food products. Food Engineer Rev. 2013;5:158–170. http://dx.doi.org/10.1007/s12393-013-9067-2

62. Dewi SH, Gurnadi E, Priyanto R, Manalu W. Sucrose supplementation, insulin injection and resting period prior to slaughtering on meat physical characteristics in sheep exposed to stressful transportation. J Agric Sci Technol. 2011;1:1162–1165.

63. Hambrecht E, Eissen JJ, Nooijen RIJ, Ducro BJ, Smits CHM, Den Hartog LA, et al. Pre-slaughter stress and muscle energy largely determine pork quality at two commercial processing plants. J Anim Sci. 2004;82:1401–1409.

64. Lin HS, Lee JY, Ke BC. Effect of cooking method and storage on the quality of minced pork. J Agric Sci Technol. 2011;1:1249–1256.

65. Singh P, Wani AA, Saengerlaub S, Langowski H-C. Understanding critical factors for the quality and shelf-life of map fresh meat: A review. Crit Rev Food Sci Nutr. 2015;51(2):146–177. http://dx.doi.org/10.1080/10408390903531384

66. Mancini RA, Hunt MC. Current research in meat color. Meat Sci. 2005;71:100–121. http://dx.doi.org/10.1016/j.meatsci.2005.03.003

67. Hudson NJ. Mitochondrial treason: A driver of pH decline rate in post-mortem muscle? Anim Prod Sci. 2012;52:1107–1110. http://dx.doi.org/10.1071/AN12171

68. Scholtz MM. Improvement of meat quality in commercial beef and pork production systems. S Afr J Anim Sci. 2007;8:31–34.

69. Kim YH, Lonergan SM, Huff-Lonergan E. Protein denaturing conditions in beef deep semimembranosus muscle results in limited μ-calpain activation and protein degradation. Meat Sci. 2010;86:883–887. http://dx.doi.org/10.1016/j.meatsci.2010.06.002

Inward foreign direct investment and transfer of environmentally sound technology in Angola

AUTHOR:
Albert E. Manyuchi[1,2]

AFFILIATIONS:
[1]Department of Political Sciences, University of Pretoria, Pretoria, South Africa
[2]Anova Health Institute, UCSF–CDC National Surveillance, Johannesburg, South Africa

CORRESPONDENCE TO:
Albert E. Manyuchi

EMAIL:
manyuchi@anovahealth.co.za

POSTAL ADDRESS:
Department of Political Sciences, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

DATES:
Received: 09 Oct. 2015
Revised: 05 Mar. 2016
Accepted: 04 Apr. 2016

KEYWORDS:
energy sector; greenfield; brownfield; environment; sovereignty principles; constitutional provisions

HOW TO CITE:
Manyuchi AE. Inward foreign direct investment and transfer of environmentally sound technologies in Angola. S Afr J Sci. 2016;112(7/8), Art. #2015-0391, 7 pages. http://dx.doi.org/10.17159/sajs.2016/20150391

Many developing countries have relied on foreign direct investment as a primary means to acquire technologies. However, there has been inadequate empirical research on the nexus between foreign direct investment and the transfer of environmentally sound technology (EST), specifically focused on African countries. In this paper I explore whether inward foreign direct investment in Angola's energy sector has indeed transferred ESTs. My study encompasses illustrative case studies specifically related to energy firms, and the data were drawn from literature and in-depth individual interviews. The results indicate that Angola has used its national policy framework and institutions to promote inward foreign direct investment, and has harnessed appropriate international regimes to acquire ESTs. Countries may therefore invoke sovereignty principles enshrined in constitutional provisions, or may utilise international regimes to attract ESTs through foreign direct investment. I recommend that further studies be conducted to explore this subject area, drawing examples from other African countries and differing economic sectors.

Introduction

Foreign direct investment (FDI) is known to be important and necessary for the economic growth and industrial development of many countries, especially developing countries. FDI provides capital, facilitates the transfer and upgrading of technologies, enhances marketing skills of domestic firms, and diffuses organisational and managerial know-how for economic growth.[1] The perceived and actual benefits of FDI have caused many developing countries to implement a regime to attract foreign investors, including aspects such as investment liberalisation and a variety of national and supra-national policies.[2] Such policies are aimed at reassuring, incentivising, protecting, building trust and confidence, and generally attracting the interest of foreign investors.[2] Zarsky[3] observes that 'from sub-Saharan Africa to East Asia, Russia to Latin America, [the] hunger for FDI has exploded', primarily motivated by the desire to realise economic growth.

Although other channels do exist, the FDI activities of transnational companies (TNCs) are at the forefront of transferring technologies among countries.[4] This transfer includes environmentally sound technology (EST) and the sharing of knowledge, such as environmentally sound practices.[5] Thus, ESTs and practices are part-and-parcel of the assets of TNCs. These assets may be deployed through FDI to assist developing host countries to adapt to and mitigate the effects of climate change, to enhance the management of natural resources, and ultimately to achieve sustainable development.

However, African policymakers do not necessarily understand the nexus between FDI and sustainable development. Whilst most African countries want to receive FDI, they still – almost inadvertently – reduce FDI to financial capital flows, and rather parochially view FDI as a potent ingredient to achieve economic growth. However, FDI should be viewed holistically as a bundle of resources that affect economies in a multi-dimensional manner. Also, the role of FDI in the management of environmental resources has tended to be somewhat peripheral to the FDI agenda of African countries. Therefore, in the African context, there is a striking lack of sound evidence for the role of FDI in the transfer of – or FDI-carrying – ESTs and practices. There is general concern about the ability of FDI to carry and transmit ESTs, and specifically whether inward FDI (IFDI) to African countries actually does transfer ESTs.

My analysis focuses on whether IFDI does transfer ESTs within the African context. I used a number of firm-specific case studies from the energy sector in Angola. My two main research questions were: (1) Are increased FDI flows to Angola's energy sector carrying ESTs or associated with the transfer of ESTs? (2) Has FDI actually resulted in such transfers? The results of my study can provide African policymakers with empirical evidence to enable them to engage more effectively in international negotiations, and to develop high-quality, useful FDI policies.

The paper is organised as follows. In the next section I present the conceptual framework that guided my analysis, followed by a description of the methodology used. The fourth section elaborates on the findings of the research, followed by a discussion. In the final section I offer some concluding remarks.

Conceptual framework

The transfer of ESTs has been at the centre of international discussion and debate for several decades.[6] The acquisition and diffusion of technologies, including ESTs, through FDI is a public policy matter that concerns sovereign states and entities – such as governments and legislatures. Sovereign states are in a position to exercise sovereign principles of autonomy, control and international legal recognition in order to regulate FDI.[7] They have the necessary resources to deploy structural power[8] to actively direct FDI to stimulate the transfer of ESTs.

Generally, there are three main reasons why states become involved in processes of transfer, acquisition and diffusion of ESTs. First, the introduction of ESTs in an economy requires the creation of appropriate regulatory frameworks aimed at protecting the environment. Such rules encompass regulations about waste management, wastewater treatment, conservation, poisonous gas emissions and climate change. Governments usually create a raft of policies and legislation that provide incentives for adherence to, and sanctions for non-compliance with, these rules. Second, the development, commercialisation and subsequent transfer of ESTs rely more on

public funds than on private investment.[9] Finally, as stated by Less and McMillan, 'unlike other types of technology, ESTs often necessitates public seed funds as incentives for companies to initiate EST-related research and development'.[10] Hence in many situations, ESTs are developed and commercialised by small and medium enterprises.[11] These enterprises depend heavily on government funding and support to penetrate domestic and international markets.

There are four methods by which countries can obtain technologies – including ESTs – from TNCs. These are FDI; joint ventures; purchase of technology in contractual form; and reverse engineering, copying and imitation.[12] In this paper I focus exclusively on IFDI. TNCs involved in IFDI hold at least 10% of the equity shares in an investment they make in a host country. IFDI activities of TNCs often take the form of setting up local production facilities, known as greenfield investments; or by purchasing existing businesses, known as brownfield investment, through mergers and acquisitions.[13]

Various theories have been put forward to explain the motivation, magnitude and spatial distribution of FDI, including the investment development model[14], the product life cycle model[15] and location-specific advantage theory.[16] Dunning's eclectic paradigm is the most widely used model.[17] The investment development model posits that IFDI to a country is systematically related to the stages and structure of that country's economic development. Therefore, a country can move sequentially through five stages, from being a net IFDI recipient to becoming a net exporter of outward FDI.[18] Using this theory, Angola's energy sector can be viewed as being in an initial phase of the development path, as it still depends considerably on inward flows of FDI.

According to the product life cycle model, products are initially made and sold in the local markets of developed countries. However, as the products mature and production becomes standardised, they are exported to external developing country markets. When the cost of producing a product in the domestic market outweighs the costs in other countries, its production is relocated abroad, mainly through the channel of FDI.[19] The product life cycle model also attempts to explain FDI by third-world TNCs. According to this theory, third-world TNCs have generally bought technology from the developed world. Because such technology is suitable for an area with a large market, firms that import such technology will export their products once local demand has been met. As the products become more familiar to foreign markets and as the markets for such products gradually become established, the third-world firms show a preference for setting up subsidiaries abroad rather than exporting or other channels. Hence, the product life cycle theory provides an explanation for the involvement of TNCs from emerging countries – such as Brazil, China and Argentina – in Angola's energy sector.

The eclectic paradigm explains that TNCs choose FDI over other investment methods in order to obtain ownership, internalisation and location-specific advantages.[20] Ownership advantages derive from firm-specific resources and capabilities, including superior intangible and tangible assets and skills – such as multinational corporations' experience, firm size, and ability to develop product differentiation. Internalisation advantages are related to reduction in transaction costs. Location-specific advantages refer to host country characteristics such as policies. The eclectic paradigm explains IFDI to Angola's energy sector through exposing the country's locational advantages and highlighting the ownership advantages inherent in the investing firms.

The eclectic paradigm further clarifies the four motives underlying direct investments: resource-seeking, market-seeking, efficiency-seeking, and strategic asset or capability-seeking.[21] Resource-seeking IFDI arises from the desire of TNCs to acquire particular types of resources that are not available at home, such as natural resources (raw materials) or resources offered at a lower cost, such as unskilled labour. Market-seeking IFDI involves investments aimed at exploiting the possibilities granted by markets of much greater dimensions, or when TNCs follow suppliers or customers that have built foreign production facilities. This enables the TNC to adapt its goods to local needs or tastes and to save on costs associated with serving other markets from a distance. TNCs motivated by efficiency-seeking IFDI take advantage of various factor endowments, cultures, institutional arrangements, economic systems and policies, and market structures that are amenable to efficient production methods. Strategic asset-seeking IFDI occurs when a foreign direct investor acquires a foreign entity, not for immediate profit or gain but for long-term strategic purposes. Recent literature indicates that developing countries, including African countries, attract primarily resource-seeking and market-seeking IFDI.[22] Resource-seeking IFDI usually targets the extractive economic sectors, such as mining, which are environmentally deleterious.

IFDI may transfer technologies such as ESTs through different channels, including vertical and horizontal linkages, as well as spillovers.[23] On the one hand, the transfer of technologies, in particular ESTs, through vertical linkages encompasses backward and forward linkages. Whereas backward linkages refer to relations with suppliers of parts, components, materials and services, forward linkages refer to relations with buyers – either consumers or other companies that utilise the intermediate products of TNCs in their own processes. On the other hand, the transfer of technologies through horizontal linkages occurs through demonstration, competition and labour migration.[24]

Transferrable ESTs through FDI may include 'hardware and software elements',[25] comprising '(1) capital goods and equipment, (2) skills and know-how for operations and maintenance, and knowledge and expertise for innovation'.[26] Mansfield[27], for example, distinguishes between material transfer, design transfer and capacity transfer, whereas Bell[28] identifies three distinct flows of transferable technology. Bell's categories range from Flow A to Flow C, all generally highlighting the transfer of capital, blueprints and expertise. For these different technologies to constitute ESTs, they should facilitate cleaner production and technological leapfrogging, and abate pollution, thereby creating 'pollution halos' when applied in a given context.[29]

Methodology

To examine the transfer of ESTs through FDI inflows to Angola's energy sector, I selected a number of illustrative case studies of specifically energy-oriented firms. This case study approach is based on the work of Yin[30]. Angola was chosen as the FDI host country because it has an extensive energy sector to which there are high FDI inflows, and there is strong government control of this sector. The methodology I used encompassed a thorough review of relevant documents, and face-to-face interviews with key informants.

The documents I reviewed included official reports on FDI and energy, as well as academic literature on a wide range of interdisciplinary fields – including the transfer of technology and ESTs, the nature of FDI, sustainable development in Africa and FDI, and international relations issues related to FDI and the transfer of ESTs. Most of the reports and other documents I reviewed (such as business plans, project plans and project designs) were accessed through internet searches or by contacting various stakeholders (by phone or email). The people I contacted to obtain such documents included government officials, representatives of TNCs and their subsidiaries, and civil society and intergovernmental organisations.

In addition to this documentary review, I collected primary data through structured discussions with 50 interviewees. The participants were accessed through a snowball sampling technique. This technique was chosen because Angola's energy sector is highly securitised, and it is therefore quite difficult for employees to openly share information. The focused interviews were conducted with government officials, members of intergovernmental organisations, non-governmental organisations, and domestic and international firms. To be eligible for interview, participants were required to be knowledgeable about IFDI to Angola, Angola's energy sector, FDI inflows to the energy sector, energy sector projects, and technology transfer, including the diffusion of ESTs. Some interviewees corroborated the information I had collected from various secondary literature sources and reports.

As far as data analysis is concerned, in terms of secondary data I asked reviewers to analyse theme-based summary sheets of triangulated data. The reviewers identified gaps in the data, conflicting data, or areas where

they believed saturation point was not achieved. For gaps and areas where data were perceived not to be saturated, I probed these topics during the interviews. Where secondary data conflicted substantially, I did not analyse the discrepancies and effectively discarded the data in question.

Interview data were collected by note-taking during discussions and information extracted from email correspondence, which was subsequently classified according to subject and summarised. The data were then thematically presented on Microsoft Excel spreadsheets. These theme-based spreadsheets – containing both secondary and primary data – were merged, and the various theme-based findings clearly presented in a case-study format.

Findings

Inward FDI flows to Angola's energy sector

The main research finding was that since 2004 there have been increased inflows of FDI into Angola's energy sector. This has comprised the subsectors of oil and gas, renewables and non-renewables, alternative energy sources, power-generation (electricity) and transmission. Although the oil and gas subsector has historically received the lion's share of IFDI, there have also been direct investments into other sub-sectors, as shown in Table 1.

Generally, IFDI to the energy sector in Angola can be viewed from the vantage point of two different periods: before and after 2004. In an interview on 14 July 2014, an official from the Angola National Private Investment Agency stated:

> *The civil war ended in 2002. The government quickly moved to focus on development, and among the key sectors to be modernised was the energy sector. Several actors were approached between 2002 and 2003. A breakthrough came in late 2004 by way of a Chinese loan and direct investments. We can, therefore, see the era before 2004 as an era where not much was done to attract IFDI to the energy sector, and the period post-2004 as a period of sudden increases in direct investments.*

The key differences in terms of FDI inflows to the energy sector in Angola during the pre-2004 and post-2004 periods, respectively, are presented in Table 2.

Since 2004, Angola's energy sector has been receiving direct investment flows. These investments have typically targeted new (greenfield) projects rather than existing (brownfield) enterprises. Furthermore, such investments have mainly been resource-seeking and market-seeking.

State's role in the transfer of ESTs in the energy sector through IFDI

One of my key research findings was that through the active involvement of the government of Angola, the inflows of FDI to Angola's energy sector have indeed transferred ESTs. The Angolan government influences the transfer of ESTs into the energy sector through IFDI using a three-pronged approach. This was explained by an official from the Ministry of Energy and Water in an interview on 14 July 2014:

> *For foreign investments to bring to a country desired technologies such as those that protect the environment, deliberate actions by the*

Table 1: Main inward foreign direct investment into non-oil and gas energy subsectors

Operator	Energy sub-sector	Partners	Location	Project name
BIOCOM	Bio-energy	Sonangol, Odebrecht, Damer Industria SA	Malanje Province	Biocom Project
Hydro Chikapa 1	Hydro-Power	Alrosa, ENE	Lunda Sul Province	Hydro Chikapa 1
Angola Liquefied Natural Gas (LNG) Project	Bio-Energy	Chevron, Sonangol, Total, Eni	Zaire Province	Angola LNG Project
Fortune CP	Solar	Government of Angola	Huíla Province	Solar projects for health centres, community centres, and educational facilities
Proef	Bio-Energy	Sonangol	Zaire Province	Bio-Energy Project

Source: Author's compilation

Table 2: Differences in inward foreign direct investment to the energy sector, before and after 2004

Key variables	IFDI flows to Angola's energy sector, pre-2004	IFDI flows to Angola's energy sector, post-2004
Sources of IFDI	Mainly from developed countries, such as US, UK, France, Spain, Italy, the Netherlands.	From both developed and developing countries, such as China, Argentina, Brazil.
Concentration of IFDI in the sector	Direct investments were mainly into oil and gas.	Diversifying from oil and gas to include renewables, alternative energy sources, and non-renewables.
Policy framework	State-centred delivery of energy to consumers through the national grid, and very little participation of independent power producers.	Reforms now underway to open up the space to independent power producers, and the promotion of off-grid decentralised energy solutions.
Role of the state in projects	Owner of all energy projects through equity shares.	Creates the framework for various stakeholders to operate, and only gets involved in some projects of strategic importance to the country.

Source: Author's compilation

Key: IFDI – inward foreign direct investment

> *government are necessary. In our case, we created institutions, crafted some policies, and also engaged with various players at international level, and these combined actions have benefited us through increased FDI inflows, some technologies, and new projects.*

First, and following from this approach, I found that the Angolan government has created state-owned enterprises that partner with foreign companies in any project arising from IFDI. In the oil and gas subsector, Sonangol is the main state-owned enterprise. For example, in the Angola Liquefied Natural Gas (LNG) Project, Chevron and Sonangol hold 36.4% and 22.8% equity shares respectively, whilst Total, BP and Eni each have 13.6% equity shares. Thus, Chevron and Sonangol are the project core leaders. Similarly, in the bio-energy subsector, the Angolan government has permitted investors to develop projects subject to a partnership with Sonangol. For example, in the Biocom bio-energy project, a 40% equity share is owned by the Brazilian firm Odebrecht, a 40% equity share by the Angolan company Damer Industria SA, and 20% by the state-run petroleum company Sonangol Holdings EP. In an interview on 14 July 2014, an official from the Angola National Private Investment Agency stated that 'Sonangol's involvement in this investment is aimed to ensure that national interests are protected.'

The Angolan state has also established sector-specific state-owned enterprise for the power subsector. These include the power utility company *Empresa Nacional de Electricidade* (ENE), which manages the transmission network and operates over 80% of power-generation facilities and distribution systems outside of Luanda, and *Empresa de Distribuição de Electricidade* (EDEL), which manages power-generation and distribution within Luanda. A third power SEO is *Gabinete de Aproveitamento do Médio Kwanza* (GAMEK), which facilitates the design and development of large hydro-power projects in the Kwanza River Basin. The ENE is a mandatory partner whenever a foreign investor wants to invest in power generation in Angola. For example, ENE owns 45% equity shares in Hydro Chikapa 1, an IFDI project that includes the Alrosa group of Russia, which owns a 55% equity share in the power-generating company Hydrochikapa SARL.

Second, the Angolan government has facilitated FDI to carry or transfer ESTs through explicit policies and legislation that provide incentives for direct investors. The literature[31] shows that the Angola National Private Investment Agency uses the following criteria in determining tax and duties incentives, or reductions for a given investment: (1) type and value of the investment, (2) contribution towards the realisation of Angola's economic development strategy, (3) views on direct and indirect capital gains, (4) complexity of the investment, (5) estimated time required for a return on capital, (6) type of technology to be utilised, (7) commitment to reinvestment of profits, (8) volume of goods or services to be produced, and (9) the creation of production lines. Using these criteria, the Angola National Private Investment Agency may offer an extraordinary tax incentive for investments perceived as highly relevant for the country's strategic development, creating at least 500 jobs, contributing to a major boost in technological innovation and scientific research, and with exports that could exceed USD50 million and inputs valued at above USD50 million. The actual incentives available include import rights, deferral of tax payments, accelerated amortisation and depreciation, tax payment deductions, exemptions and credits.

Most of the projects in the energy sector have been greenfield investments; hence, these have been offered a variety of incentives. In the cases of the Biocom, Angola LNG and Hydro Chikapa 1 projects, the government deliberately promoted cleaner technologies in the form of hardware and machinery through exempting all equipment for constructing these plants from taxes and important duties. Most importantly, in the case of Angola LNG project, tax and import duty exemptions of all plant material facilitated pollution abatement, and assisted in creating 'pollution halos' by halting the rampant pollution that resulted from gas-flaring.

Third, the government harmonises its domestic laws with international laws, and domesticates international regimes aimed at facilitating international transfer of technology, including ESTs. Angola is a party to the United Nations Framework Convention on Climate Change (UNFCCC) of 1992[32] and ratified the Kyoto Protocol of 1997 in May 2007. In terms of these agreements, Angola has qualified for securing carbon credits from the Clean Development Mechanism.

As a result, two key studies were conducted in 2006 and 2007, assessing the eligibility of the Angola LNG Project for carbon credits. In December 2007, Sonagas (a subsidiary of Sonangol) presented the project during the proceedings of COP13 in Bali, Indonesia. In November 2008, *Banco Espirito Santo Angola* confirmed its interest in buying carbon credits that result from this project, through a letter of intention. In 2010 Angola established a Designated National Authority through Decree No. 2/10. To adhere to the Clean Development Mechanism requirements, at the end of August 2011 all supporting project documents were lodged with the Clean Development Mechanism, and the validation process commenced on 10 November 2011 with the publication of the Angola LNG Project details on the UNFCCC website.

In addition, the Angolan government has deployed its international relations apparatus to engage with international finance institutions, such as the World Bank and the African Development Bank, to facilitate IFDI carrying ESTs. Angola is a member of the World Bank's Global Gas-Flaring Reduction Partnership (GGFRP), and its interactions with this institution were of prime importance in designing the Angola LNG Project. In October 2005, October 2006 and February 2007, the GGFRP funded several conferences and presentations to raise awareness about the Angola LNG Project. Thus through deploying international relations, the GGFRP managed to finance activities that led to the realisation of the Angola LNG Project – a project that deals with the huge problem of pollution resulting from gas-flaring. Because Angola has good relations with and is eligible for support from the World Bank, the country also voluntarily made use of the World Bank and the International Finance Corporation's policies and guidelines to conduct environmental impact assessments for the Angola LNG project. As explained by an Angola LNG Project official, assimilating these 'international environmental guidelines on environmental impact assessments helped to enhance the project's appeal to international stakeholders' (interview with Angola LNG Official, 21 July 2014).

IFDI to the energy sector and the types of EST transferred

Another key research finding is that the three-pronged approach of the Angolan government has directed IFDI to transfer of ESTs mainly in the form of hardware, in particular machinery and equipment necessary for production. As an official from the Ministry of Petroleum said in an interview on 14 July 2014:

> *Our greatest benefit from these direct investments comes from the machinery and equipment that we receive from investors … which enables us to immediately exploit the energy sources and also deliver energy services to consumers. This machinery should be viewed as very important, because it is key to production.*

Indeed, most of the energy projects carried out in the energy sector have been greenfield investments that demand investments in new machinery and equipment for plants. Because Angola does not have the technological capabilities to manufacture energy sector machinery and equipment locally, the country has been dependent on technology transfer from abroad. In an interview on 14 July 2014, an official from the Ministry of Petroleum stated:

> *We, like many developing countries, are still importing or buying basic things, because we do not have the industry to manufacture products locally. We buy and import electric bulbs at the moment; think of a time when we will be able to manufacture components for a power plant or an oil-rig.*

My research showed that new machinery and equipment constituting the plant for Hydro Chikapa 1, Angola LNG and the Biocom project were transferred from Russia, Europe and America, and Brazil respectively.

Generally, hardware transferred to Angola through FDI inflows can be considered to be ESTs because it is new and therefore enables cleaner production. Furthermore, the plant for the Angola LNG Project reduced pollution that arose from gas-flaring, hence facilitating the creation of 'pollution halos'. The transfer of this new technology, never used in Angola before, has enabled the country to leapfrog to new frontiers of LNG and bio-energy technologies.

The transfer of hardware, especially machinery and equipment to harness renewable energy, is another novel area where ESTs are being transferred through inflow of FDI. This hardware includes machinery and equipment for wind and solar energy. As an official from the Ministry of Energy and Water stated in an interview on 14 July 2014:

> *We have announced a public-private partnership to construct Angola's first 100MW wind park at Tômbwa in the Namibe Province. Our hope is that we will get partners who have modern technology and are willing to share the technologies with us. So far, we are convinced that wind and solar [energy] are the way to go forward, but without the appropriate technology we can't follow this desired route.*

Recently the government awarded a tender to develop the wind–energy and solar-radiation resource map of Angola to a Spanish firm, Ereda. Ereda is also obliged to transmit soft technologies to local Angolans during this mapping phase. Likewise, direct investments into photovoltaic energy have been on the increase, especially in the form of off-grid small-scale solutions that target rural public facilities such as clinics and hospitals. The UK-based Fortune CP, for example, has been involved in implementing solar-power systems in collaboration with the Ministry of Health in ten rural clinics in Huila Province.

Besides transfer of ESTs in the form of hardware, IFDI to Angola's energy sector has also transferred ESTs in the form of know-how, especially people-embodied know-how. ESTs in the form of knowledge on safety, health and enhanced environmental practices were passed on to local staff who participated in the environmental impact assessment process for the Angola LNG and Biocom projects. Similarly, training given by Brazilian experts to Angolans employed in the Biocom bio-energy project transferred environmental best practices for sugarcane production and environmentally-friendly agricultural activities. In the Angola LNG Project, locals have been trained in methods to conserve biodiversity, especially marine animals such as sea turtles, which otherwise would have been adversely affected by the construction of gas plant and pipelines. The transfer of ESTs through training in environmental best practices also indirectly leads to the transfer of paper-embodied ESTs. However, the capacity of IFDI to transfer paper-embodied ESTs is limited because Angola is a Portuguese-speaking country, whereas some investors in the energy sector speak English, Spanish or Mandarin.

Although I found evidence for transfer of ESTs in the form of hardware and know-how, I could not find evidence that any capacity transfer has taken place. In other words, I found no proof that the transfer of ESTs in the form of 'know-why' technology has taken place. As an official from the Ministry of Energy and Water said in an interview on 14 July 2014:

> *Our main problem has been to ensure that we have qualified Angolans at the helm of these energy sector projects in contrast to the current situation where most of the senior managers and engineers are foreign expatriates. Educating our people is important, but we need to institute and enforce the policies that promote foreigners to train locals. This is a rather difficult task.*

This means that the transfer of ESTs to Angola has not yet built the capacities and capabilities of Angolans to independently create and design new and locally adapted environmentally friendly best practices. The local capacity to adapt, adopt and transform the acquired hardware technologies remains limited.

FDI inflow to the energy sector and the transfer of ESTs to the broader economy

Another research finding is that whilst FDI inflows to the energy sector have transferred ESTs, especially hardware in greenfield investments, such transfers have been limited to this sector and have not had a wider impact on the overall Angolan economy. This was explained in an interview on 14 July 2014 by a Ministry of Environment official:

> *Whilst we are making great strides in transforming our economic activities so that they do not harm the environment, some sectors have been very successful whilst others have not. For example, I would say all new projects in the energy sector have had the benefit of the best existing technologies. However, old projects remain unchanged and polluting. Furthermore, we still have environmentally-harmful technologies and practices in the agricultural and other sectors. Therefore, we need to analyse these issues much more holistically.*

From these comments it is clear that IFDI to Angola's energy sector has not comprehensively transformed either the energy sector or other sectors of the Angolan economy, in terms of the deployment of ESTs and practices. What is apparent is that there are pockets of environmentally-sound projects that coexist with environmentally-degrading enterprises within the energy sector, and uncertainty about what is transpiring in other sectors. Currently, the amendment of outdated legislation and statutes, such as those governing the power subsector, seems to be the Angolan government's main plan for forward movement.

Discussion

The objective of this research was to explore whether increased FDI inflows to Angola's energy sector have successfully transferred ESTs. I also wanted to examine the types of technology transferred. My main findings were that Angola has attracted increased IFDI to its energy sector, especially since 2004, and these inflows have indeed transferred ESTs. Pivotal to promoting the transfer of ESTs through IFDI are actions by the government, which has created institutional structures and policy frameworks that promote EST-carrying FDI.

My findings illustrate four types of institutions that essentially direct IFDI to transfer ESTs. First, certain state-owned enterprises partner with foreign investors and advance national interests in these investments, for example by ensuring adherence with sustainable development imperatives. Second, law-making institutions – such as the National Assembly – draw up domestic legislation, and harmonise and domesticate international regimes. Third, government institutions provide oversight for policymakers, the legislature and other relevant bodies. In Angola's energy sector, this role is fulfilled by line ministries such as the Ministry of Petroleum and the Ministry of Energy and Water. Fourth, certain state institutions build relations and interact with international actors, with the aim of attracting FDI inflow and promoting the transfer of ESTs. The primary feature that is apparent from analysing this institutional architecture is that the transfer of ESTs through IFDI does not occur automatically, but can be achieved through deliberate intervention by government. Furthermore, like other studies[33], my research highlights the importance of institutions and policies. My approach departs from other studies by stressing the central role of governments in promoting EST-carrying IFDI through the deployment of international-relations apparatuses, and through the domestication of global regimes.

Based on the findings of this study, it transpires that there are two types of policy that essentially direct FDI inflows to transmit EST. The first type is explicit policies: specific policies and provisions for IFDI that carries ESTs, such as environmental legislation and EST-incentivised programmes. The second type is implicit policies: these policies do not make direct provision for FDI or ESTs, but their implementation directly affects the transfer of ESTs through IFDI.

Table 3: Types of environmentally sound technologies transferred through inward foreign direct investment to Angola's energy sector

Capital goods and equipment	Skills and know-How	Knowledge and expertise ('know-why')
These take the form of hardware, especially machinery and equipment transferred in all greenfield energy-sector projects. These are new, clean, less-polluting technologies, and they also facilitate technology leapfrogging.	This occurs especially through people-embodied practices, transferred in the training of Angolans inside and outside the country, as well as new knowledge being transferred in renewables and alternative energy sources. However, transfer of paper-embodied ESTs and practices is limited by language barriers.	The limited availability or lack of Angolans at managerial and technical levels constrains the transfer of know-why ESTs. Also, the lack of manufacturing capability limits the country's capacity to transform acquired technologies to a level where they can be environmentally friendly.

Source: Author's compilation

Key: EST = environmentally sound technology

Implicit policies include human resource issues, such as policies on training of locals – the implementation of which results in transfer of ESTs in the form of people and paper-embodied know-how through IFDI. Through these policies, the state has managed to create a favourable or conducive environment for IFDI that carries ESTs.

My research findings furthermore show that although IFDI does transfer ESTs in Angola's energy sector, the inflows do not uniformly transfer all types of EST. Some ESTs are easily transferrable and are indeed transferred, whilst others are not, as shown in Table 3.

The main type of ESTs transferred take the form of capital goods. Most of the machinery and equipment being transferred are essential for production. Achanda and Gosch[34] state that 'the flow of capital goods and services adds to the production capacity of the transferee'. However, because the incoming machinery and equipment are modern, the whole country benefits from much more efficient and cleaner production.

The transfer of skills and know-how for environmental management has also occurred, especially people-embodied know-how, which seems to be more easily transmitted than paper-embodied know-how. The problem of passing on ESTs in the form of paper-embodied know-how can be attributed to language barriers that exist between Angolans and foreign investors. Moreover, the transfer of knowledge and expertise for innovation or know-why is rather limited. My findings on transferrable technologies that are easily diffused through IFDI are similar to those of other studies[19].

The limited transfer of know-why ESTs seems closely linked to Angola's technological capability. Generally, the country's lack of endogenous technological capability hinders its ability to modify systems and to adapt, adopt and modify externally acquired technologies. With low-level technological capacity, it even becomes more difficult to modify the acquired technologies to be environmentally friendly. Thus the country needs to build its endogenous technological capability, in particular manufacturing capabilities that would enhance its capacity to manufacture or transform transferred technologies. The limited impact of EST-carrying IFDI in brownfield investments in the energy sector, as well as in the broader economy, might suggest that policies on FDI as well as sector-specific and national policies should be strengthened. I recommend there should be closer integration of FDI policies and institutions with the national system of innovation and educational framework.

Conclusion

In this study, I examined whether IFDI into Angola's energy sector has transferred ESTs. I concluded that the active involvement of the Angolan government in creating policy frameworks and institutional structures, and in deploying international relations, has facilitated IFDI carrying ESTs to the country's energy sector. This finding suggests that countries can generally utilise sovereignty principles to attract FDI that transfers ESTs. Furthermore, FDI recipient countries should build their endogenous technological capability, which would enable those countries to adopt, adapt and transform externally acquired technologies to be environmentally friendly. This research is among the few Africa-focused studies that provide empirical evidence to policymakers and negotiators on FDI inflows and the transfer of EST. I recommend that further research be conducted to explore this subject area, perhaps using different countries and economic sectors as case studies.

References

1. Diyamett B, Mutambla M. Foreign direct investment and local technological capabilities in least developed countries: Some evidence from the Tanzanian manufacturing sector. Afr J Sci Technol Innov Dev. 2014;6(5):401–414. http://dx.doi.org/10.1080/20421338.2014.983305
2. Mugabe JO. Governing foreign direct investment in Sub-Saharan Africa: Policies and practices reconsidered. In Zarsky L, editor. International investment for sustainable development: Balancing rights and rewards. London: Earthscan; 2005. p. 74–96
3. Zarsky L, editor. International investment for sustainable development: Balancing rights and rewards. London: EarthScan; 2005.
4. Lall S, Narula R. Foreign direct investment and its role in economic development: Do we need a new agenda? EJDR. 2004;16(3):447–464. http://dx.doi.org/10.1080/0957881042000266589
5. Chudnovsky D, Lopez A. TNCs and the diffusion of environmentally-friendly technologies to developing countries (Occasional Paper No. 9). In: Cross-border environmental management in transnational corporations. New York: UNCTAD; 1999.
6. Mugabe J, Clark N. Technology transfer and the Convention on Biological Diversity: Emerging policy and institutional issues. Biopolicy International Series Number 18. Nairobi: African Centre for Technology Studies; 1996.
7. Krasner SD. Abiding sovereignty. Int Polit Sci Rev. 2001;22(3):229–251. http://dx.doi.org/10.1177/0192512101223002
8. Strange S. States and markets. London: Pinter Publishers; 1988.
9. Intergovernmental Panel on Climate Change. Methodological and technological issues in technology transfer. Geneva: United Nations Environment Programme; 2000.
10. Less CT, McMillan S. Achieving the successful transfer of environmentally sound technologies: Trade-related aspects. Trade and Environment Working Paper No. 2005-02. Paris: OECD; 2005.
11. United Nations Conference on Trade and Development. The role of publicly funded research and publicly owned technologies in the transfer and diffusion of environmentally sound technologies. New York: UNCTAD; 2000.
12. United Nations Conference on Trade and Development. Foreign direct investment, the transfer and diffusion of technology, and sustainable development. New York: UNCTAD; 2010.
13. Dunning JH, Lundan MS. Multinational enterprises and the global economy. London: Edward Elgar Publishing; 2008.
14. Dunning JH. Explaining the international direct investment position of countries: Towards a dynamic or developmental approach. Weltwirtschaftliches Archiv. 1981;117:30–64
15. Tarzi S. Host countries and foreign direct investment from the emerging markets. IntRelat.1999;14(4):15–32,http://dx.doi.org/10.1177/004711789901400402

16. Yeung HWC. Introduction: Competing in the global economy. In: Yeung HWC (editor). The globalisation of business firms from emerging economies. Cheltenham: Edward Elgar. 1999; p. 1–15. http://dx.doi.org/10.1142/9781848160286_0001

17. Dunning JH. The eclectic paradigm as an envelope for economic and business theories of MNE activity. Int Bus Rev. 2000;(9)2:163–190. http://dx.doi.org/10.1016/S0969-5931(99)00035-9

18. Miguel F. The investment development path hypothesis: A panel data approach for Portugal and the cohesion countries. The Business Review. 2009(2); p. 170–178. Cambridge.

19. Yean TS. Outward foreign direct investment from Malaysia: An exploratory study. Sudostasien Aktuell. 2007;5:44–72.

20. Dunning JH. The eclectic (OLI) paradigm of international production: Past, present and future. Int J Econ Bus. 2000;(8)2:173–190.

21. Ajala BO. Gearing FDI towards sustainable development in Nigeria: The role of the WTO TRIMS agreement. [LLM Dissertation], University of Pretoria, Pretoria]; 2010

22. Narula R, Dunning JH. Industrial development, globalization and multinational enterprises: New realities for LDCs. Oxford: Oxford Development Studies; 2000. http://dx.doi.org/10.1080/713688313

23. Farole T, Winkler D, editors. Making foreign direct investment work for Sub-Saharan Africa: Local spillovers and competitiveness in global value chains. Washington, DC: World Bank; 2014. http://dx.doi.org/10.1596/9781464801266

24. Saggi K. Trade, foreign direct investment, and international technology transfer: A survey. Working Paper Series in International Economics No. 2349. Washington, DC: World Bank; 2000.

25. United Nations Conference on Trade and Development. Positive measures to promote sustainable development, particularly in meeting the objectives of multilateral environmental agreements. Geneva: UNCTAD; 1997.

26. Urban F, Siciliano G, Sour K, Lonn PD, Tan-Mullins M, Mang G. South-South technology transfer of low-carbon innovation: Large Chinese hydropower dams in Cambodia, Sustain Dev. 2015;23:232–244. http://dx.doi.org/10.1002/sd.1590

27. Mansfield E. Technology transfer, innovation and public policy. Technology, Productivity and Economic Policy. New York: W.W. Norton; 1982.

28. Bell M. Continuing industrialisation, climate change and international technology transfer. Report prepared in collaboration with the Resource Policy Group, Oslo, Norway. Brighton: University of Sussex, Science Policy Research Unit; 1990.

29. Gallagher KS, Zarsky L. FDI spill-overs and sustainable industrial development: Evidence from US firms in Mexico's Silicon Valley. [Article on the Internet]. Discussion Paper No. 18: Working Group on Development and the Environment in the Americas; 2008. [Cited 2013 October 14] https://ase.tufts.edu/gdae/Pubs/rp/DP18Zarsky_GallagherApr08.pdf

30. Yin RK. Case study research: Design and methods. California / London: Sage Inc; 2009. http://dx.doi.org/10.4135/9781483348858.n8

31. ANIP / USA. How to invest in Angola? Washington D.C.: Angola National Private Investment Agency; 2009

32. United Nations Framework Convention on Climate Change. [Document on the Internet]. 1992 [Cited 2013 October 16]. http://unfccc.int/kyoto_protocol/status_of_ratification/items/2613.php

33. Asiedu E. Foreign direct investment in Africa: The role of natural resources, market size, government policy, institutions and political instability. World Econ. 2006;29(1):63–67. http://dx.doi.org/10.1111/j.1467-9701.2006.00758.x

34. Achanda A, Gosch P. Technology transfer and the environment. In: Juma C, Sanchez V, editors. Biodiplomacy: Genetic resources and international relations. Nairobi: African Centre for Technology Studies. 1994; p. 56–76

Universities are becoming major players in the national system of innovation

AUTHORS:
David Walwyn[1]
Laurens Cloete[2]

AFFILIATIONS:
[1]Graduate School of Technology Management, University of Pretoria, Pretoria, South Africa
[2]Council for Scientific and Industrial Research, Pretoria, South Africa

CORRESPONDENCE TO:
David Walwyn

EMAIL:
david.walwyn@up.ac.za

POSTAL ADDRESS:
Graduate School of Technology Management, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

DATES:
Received: 10 Sep. 2015
Revised: 12 Dec. 2015
Accepted: 04 Apr. 2016

KEYWORDS:
R&D survey; labour market; university R&D; business R&D; business confidence

HOW TO CITE:
Walwyn D, Cloete L. Universities are becoming major players in the national system of innovation. S Afr J Sci. 2016;112(7/8), Art. #2015-0358, 8 pages. http://dx.doi.org/10.17159/sajs.2016/20150358

Based on data from South Africa's research and development (R&D) surveys, the country's R&D expenditure has grown in real terms by 52% over the period 2001 to 2012. This growth has been driven by government funding, which rose from 34% of the total funding in 2003 to 45% by 2012. Much of the additional funding has been granted to universities, with government support of R&D in this sector rising 450% in nominal terms, or 250% in real terms, over the same period. This funding focus, indicative of a growing role for universities as R&D performers within the national system of innovation, follows a pattern set earlier in many developed countries and reflects a revision in the state's steering of knowledge creation. The R&D Survey also revealed a decline in the average cost of research, as expressed by expenditure per full-time equivalent researcher. This finding suggests that the researcher labour market is being better supplied and the constraints identified by earlier reviews are slowly being overcome. Both trends are highly positive for the research system. However, the 34% decline in business R&D expenditure since its peak in 2008 is a matter of concern and needs to be addressed. In particular, the level of state-industry embeddedness must be increased to encourage private investment and to overcome South Africa's present growth constraints in respect of developing competitive medium- to high-technology sectors.

Introduction

In South Africa, the Centre for Science, Technology and Innovation Indicators has undertaken and published a research and development (R&D) survey annually since 2003; other agencies undertook similar surveys at mostly biennial intervals over the period 1983 to 2001, apart from 1995, when no data were published.[1,2] Such surveys are rich in data that can inform policy and improve the overall performance of the R&D system. However, the link between survey data and policy appears to be made only sporadically, despite a number of recent reviews on science and technology policy.[3-5]

We attempt to stimulate a more regular policy discussion based on the results of the survey, especially the role of public-funded R&D in supporting the national system of innovation. We concentrate on three issues, namely the source and beneficiaries of the growth in South Africa's expenditure on R&D, changes in the R&D labour market, and the decline in business R&D expenditure. The trends are discussed within the context of previous studies on the R&D role of the state, including the need for a more focused and transformative innovation policy[3,4,6], constraints in R&D human resources[6-9], and business R&D performance[10]. The extent to which and the manner in which the state involves itself in R&D are important questions that can have significant consequences on overall economic performance. Facing multiple options and instruments, the state must act in a focused yet balanced fashion to address the key challenges, without disfavouring any particular sector.

Background literature

Research, development and economic growth

There has been a distinct, if not deliberate, change in the profile of public (government) funding for R&D in South Africa since 2003. This change reflects an international trend in the role of governments with regard to their support for R&D. To examine the significance of this change, an initial discussion of the theoretical basis for public funding of R&D is required.

Governments worldwide are major funders of R&D. This fact is justified initially on the basis that research is characterised by a market failure, in which business enterprises and private investors fail to invest at an optimal level in R&D because of the inherent uncertainty, indivisibility and inappropriability of the research sector.[11,12] Accordingly the role of the state is to fill this void and to supplement private R&D investment, thereby ensuring that its benefits are maximised. However, this perspective is by no means unchallenged. On the one hand, its neo-liberal opponents argue for reduced state involvement in the economy, based on the belief that state intervention is economically inefficient and therefore socially undesirable.[13] On the other hand, some scholars hold the view that the state must act not as a stopgap but as a leader working proactively.[14] The latter perspective recognises a more prominent role for public-funded R&D, in which funding creates opportunities for subsequent innovation, rather than simply attempting to support private-funded R&D or prevent the dissipation of previous investment by business enterprises.

Over the period 1980 to 2000, the neo-liberal position was popular, with the result that several developed countries decreased their levels of public R&D funding. In more recent years there has been a recovery in support for public R&D, based on new evidence from country-level studies. Beginning with Chalmers Johnson[15] who studied the history of modern Japanese industrial policy, followed by various other scholars in the developmental state school, new theories on the role of the state in enabling economic development have been proposed[16-21]. In a recent contribution that builds on these theories, Breznitz[22] offers a model of how developing countries can achieve economic transformation through state-led interventions that promote rapid innovation-based industries. This

model is based on detailed case studies from Israel, Taiwan and Ireland. Breznitz[18] argues that there are three critical factors for productive interventions by the state, namely:

- a clear framework for how the state will acquire the necessary knowledge and skills to support such industries (including R&D)
- a set of policies to solve the inherent market failure of industrial R&D, including strategies to encourage private investment
- significant effort to nurture local competence and embed local firms in global production networks.

Breznitz[22] further noted that a plurality of solutions existed to the problem of developing rapid innovation-based industries in developing countries, within a context of globalisation and fragmented supply chains. However, the requirement for a close partnership between the state and business enterprise – referred to as 'embeddedness' – was universal.[18] Although Breznitz's work did not cover South Africa, it did cover countries of a similar size and stage of development in the 1970s and for which the source of economic growth had significant overlap. Furthermore the core recommendations of his work, namely that partnership, learning and skills development are crucial to successful transitions, accord with other studies on middle-income countries including Latin America. Raising levels of skills and technology transfer, nourishing local competitive firms, developing medium- and high-technology exports, and articulating a clear, simple industrial policy all underpin such transitions to a developed economy.[23,24]

Part of the confusion in the literature can be ascribed to the non-recursive nature of the variables. For example, economic growth is both caused by, and can in turn cause, an increase in knowledge production. Some authors have sought to examine these relationships in more detail. The nature and direction of causality between the accumulation of knowledge (principally through R&D) and economic growth has been studied using scientometric data and a bootstrap causality analysis.[25] The study concluded that the link between R&D and economic development is weak in developed countries that invest too little in basic research. A similar conclusion was reached in a separate study in the United Kingdom, in which the researchers argued that excessively short-term economic interests had led to the closure of the country's research laboratories, resulting in lower levels of technological innovation and ultimately less economic growth.[26] In South Africa, a positive correlation between academic research output and economic growth has been observed, indicating that such knowledge production has indeed benefited the economy.[27]

Nevertheless knowledge production, learning and skills development are not the only important factors; it is also critical that policies should identify areas of specialisation and incentivise specific sectors in which a country can establish its comparative advantage[28]. This imperative imposes on R&D policy and hence public R&D funding the need to facilitate not only human resource development, but also product or service development in sectors that can lead to significant economic growth.

Prior reviews of South Africa's R&D policy

The previous section briefly reviewed the variable nature of returns to R&D expenditure, how such returns can be measured, and the state's role in supporting R&D and ultimately facilitating economic development. We accepted the notion that the goal should be firm-level innovation, and that this outcome is linked at least partly to public expenditure on R&D; and we noted that South Africa has been characterised by a static level of innovation and economic activity in the important high-technology sectors[29]. Given these observations, the country's slow response to higher levels of public R&D funding has already been noted and studied. Prior reviews have considered this important issue and have identified several constraints. The first main constraint is the lack of focus or specialisation in R&D, and the second is a failure to provide sufficient R&D human resources.

Focus of South Africa's R&D

Concerns about a lack of focus within South Africa's R&D system and the small number of researchers spread over multiple projects have been raised in a number of reviews.[3,6] These concerns arose partly in response to growing evidence that specialisation is important in countries such as South Africa that are seeking to escape the middle-income trap.[30] Although a balanced and diversified R&D system might be the end goal for a developing country, the transition strategy requires a strong focus on specific sectors, which are then able to catch up with developed countries and build competitive local industries.

An attempt was made to address the issue of focus with the adoption in 2008 of the Ten-Year Innovation Plan.[27] This plan proposed five 'bold interventions in critical areas', labelled as grand challenges and covering the bio-economy, space science and technology, energy security, global change science with a focus on climate change, and human and social dynamics.[31] The Department of Science and Technology – and more broadly government in general – implemented the plan by directing at least a portion of the additional funding from National Treasury to large projects with close alignment to the grand challenges. These projects included the Karoo Array Telescope[32], the pebble bed modular reactor[33], the electric car (Joule)[34], and the development of a HIV microbicide (START trial)[35].

Unfortunately this focus has yet to yield measurable economic dividends. With the notable exception of Karoo Array Telescope, these projects were terminated without having reached their goals. Although failure is intrinsic to R&D, the high rate of failures is concerning. Furthermore it is claimed that the state's investment in innovation has been overly focused on big science projects, with insufficient attention to the priorities of business and social development.[36]

The debate about diversity versus focus within government-funded R&D is particularly an issue for small systems with limited resources. Options favouring diverse funding instruments and hence potential outcomes have the advantage of at least some measure of success, whereas overly focused programmes may all fail because of the inherently risky and unpredictable nature of R&D. This tension is at the core of portfolio management theory and the risk mitigation properties of a balanced portfolio.[37] However, small systems may fail to achieve critical mass in any particular area because of the inevitable dilution of available resources that accompanies diversity.

In this respect, criticism of resource allocation can always be made; only hindsight will identify the most perfect solution. Nevertheless there are suggestions that South Africa's choices for specialisation have been misguided. The present challenges are overly focused on long-cycle science-based sectors, instead of the recommended short-cycle technology-based opportunities. The latter category is more likely to yield the desired economic outcomes.[30]

Human resources for R&D

The severe shortage of human resources for R&D has been identified as a fundamental constraint to economic growth and R&D in every review since 1994. For example, the 2007 report by the Organisation for Economic Cooperation and Development (OECD) identified two looming crises: the engineering gap (deficit between supply of engineering and management skills or capabilities), and the limited supply of university graduates capable of undertaking research.[6] Similarly Kaplan[9] noted that any further expansion of the system will be 'predicated on a significant expansion in supply of skills' and that more resources devoted to R&D will have the effect of driving up unit costs and lowering productivity, rather than expanding output.[5] Likewise the 2012 Ministerial Review of Science, Technology and Innovation Landscape of South Africa noted that the 'biggest constraints are the stuttering pipeline of trained and knowledgeable people at all levels'.[36]

Seekings and Nattrass[38] highlight the deficiencies in the South African labour market as the key factor in the dual issues of employment and inequality. The market is characterised by a shrinking of low-wage formal employment opportunities and higher earnings or better working

conditions for white- and blue-collar workers, as a result of an enduring skills shortage. The same diagnosis is reached by a separate study by the Reserve Bank, which concluded that the country needs to triple the growth rate of its skilled labour, thereby significantly increasing the pool of skilled workers and reducing their cost, so that firms can expand their skills base without bidding-up wages.[39] Based on an economic model, the latter study predicts that relieving the skills constraint would raise potential growth to 6.7% by 2025, but that this result would require long-term reform across the education and training spectrum.

The divergence between labour market conditions for unskilled versus skilled labour is an enduring feature of apartheid policies, which specifically sought to establish and preserve high wages and protected incomes for a minority group.[40] Policies such as trade liberalisation and the promotion of a high-productivity growth path have favoured capital-intensive firms over labour-intensive firms, resulting in rising real wages for employed people in 2014 compared with 1994.[40,41] As a result, the benefits of economic growth since 1994 have largely been experienced by the lower and upper middle classes, with the poor and unemployed being excluded.

The Ministerial Task Team summarised the problem in the 'stuttering pipeline' as being inadequate schooling and training systems, low university participation (admission) rates, high drop-out rates, minimal enrolments for advanced postgraduate study, an ageing research cadre, and high barriers to the expansion of the postdoctoral sector.[36] Proposed solutions included quadrupling the number of technical colleges, curriculum reform, improved functioning of post-graduate training programmes, and a new cohort of research institutes to undertake multi-focus high-level research.[36]

Given rising public expenditure on education and the long-standing nature of the problem, it is important to consider what progress has been achieved and whether the constraints are being eased by the various interventions. Unfortunately a number of surveys indicate little progress, especially at primary and secondary school level.[42,43] In terms of undergraduate education, the Council for Higher Education reported that only one in four students complete their degrees in regulation time, that 55% of students never complete their undergraduate studies, and that access/success rates are still racially skewed. The net result of the disparity is that less than 5% of African and coloured youth succeed in any form of higher education.[44] Participation, let alone achievement, remains a big challenge for higher education. The Department of Higher Education and Training plans to increase university enrolments to 1.5 million by 2030, and to increase the role and alignment with industry needs of technical and vocational education and training colleges. The department also plans to introduce foundation programmes at these colleges for additional instruction in mathematics and science, to prepare students for university. Further insight on changes in the human resource pipeline can be extracted from the R&D Survey.

Results of R&D Survey

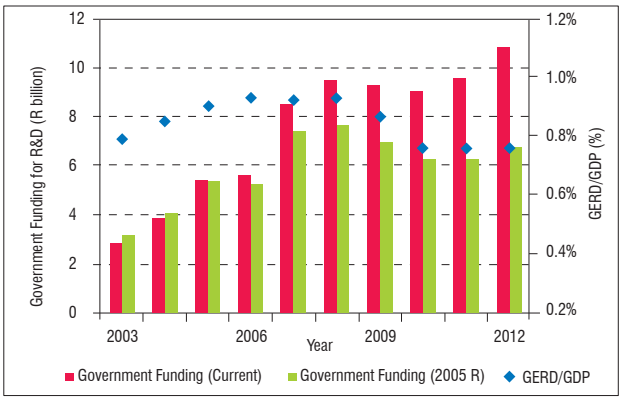
The report on the 2012/2013 R&D Survey covers a broad range of questions, including transformation, socio-economic objectives, R&D in key technology missions such as biotechnology, and regional disparities in R&D effort.[45] However, in this review of the results, only three themes will be covered as follows:

- increasing government budget appropriations or outlays for R&D, with much of the increases being allocated to universities
- increasing numbers of full-time equivalent (FTE) researchers, and declining expenditure per FTE
- declining business performance of R&D (BERD), measured as both performance itself and funding.

Government R&D expenditure and performance

Government is spending more on R&D than it did 10 years ago, with the increase in government budget appropriations or outlays for R&D over this period being 214% in real terms and 385% in nominal terms (Figure 1). Public funding of R&D has risen from 28% to 45% of gross domestic expenditure on R&D, and is now the dominant source of funds.

Much of the additional funding has been allocated to universities, whose R&D performance – as measured by higher education expenditure on R&D – has risen from R3.6 billion in 2007 to R7.3 billion in 2012. This marks an increase of 202% in 5 years (Figure 2). Universities now account for 34% of the total R&D performance, up from 19% in 2007.



Key: GERD – gross domestic expenditure on R&D

Figure 1: Government funding of R&D (2003 to 2012).

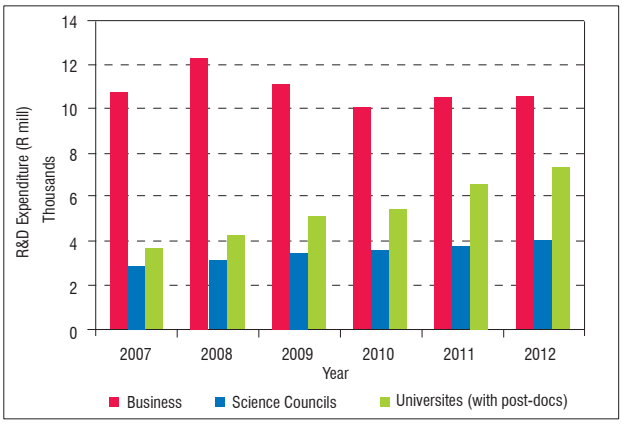
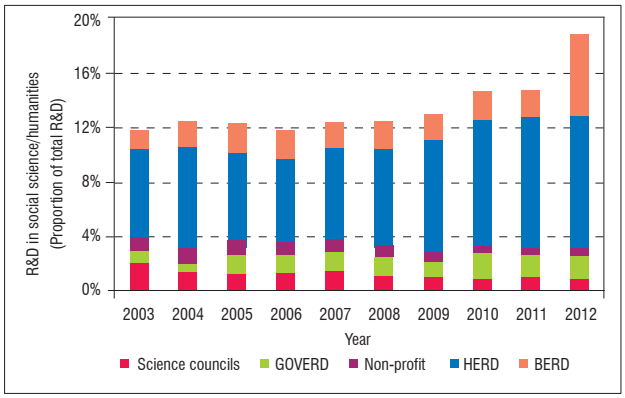


Figure 2: R&D performance by business, science and universities (2007 to 2012).

The growing role of universities in the overall R&D landscape reflects an international trend and aligns with the recommendations of the Ministerial Committee.[36] Universities in general have become more important actors within national systems, mainly as a consequence of the perceived lower social return from public research institutes.[46] Total public funding of universities has reached about 25% of gross domestic expenditure on R&D in OECD countries, or 0.43% of gross domestic product (GDP). Similarly, intramural expenditure on R&D by government, which measures R&D performance by government agencies, is declining in OECD as a percentage of GDP, and has now dropped to 0.27% of GDP.[47]

In terms of expenditure in different research fields, Kahn[48] previously pointed to the high proportion of South Africa's higher education expenditure on R&D devoted to the social sciences and humanities. Between 2002 and 2012 this proportion remained constant at 31%. However, the consequence of the growth in higher education expenditure on R&D as a proportion of total R&D expenditure, and certain structural changes to BERD, has been an increase in the overall ratio of spending on the social sciences and humanities (Figure 3). This proportion rose from 12% in 2002 to 19% in 2012. A corollary result was a decline in the proportion of experimental development from 38% to 28% over the same period. The latter is the Frascati[49] category, which records expenditure of development as opposed to research and is typically undertaken in the natural or engineering sciences on close-to-market R&D projects. The structural changes in BERD mentioned earlier are a decrease of R905 million in R&D relating to information, computer and communication technologies; and an increase of R972 million in R&D in the social sciences.



Key: GOVERD – government expenditure on R&D, HERD – higher education expenditure on R&D, BERD – business expenditure on R&D

Figure 3: Rising R&D expenditure in social sciences and humanities.

Disappointingly for the Department of Science and Technology, the ratio of gross domestic expenditure on R&D to GDP has not responded to the growth in the economy or the increase in government funding. It has instead remained at 2003 levels, as shown in Figure 1. The Department of Science and Technology's ten-year innovation plan set a target of 2% for this ratio, to be achieved by 2018, but South Africa is unlikely to achieve a figure of even half this value. Furthermore, limited economic growth, rising fiscal pressure and the low level of confidence within the private sector in the country's economic future suggest that the shortfall is unlikely to be met by either the public or the private sector. In the absence of significant re-allocations from other portfolios, the only alternatives for the Department of Science and Technology are to be more selective in its investments, to seek ways of achieving productivity increases and to more closely monitor the progress of its existing projects. Questions of specialisation and focus have already been discussed; it is likely that such questions will become even more relevant in the coming years.

Adjustments in the R&D labour market

Given the increase in funding, it is not unexpected that universities in South Africa have expanded in terms of the number of active researchers, with FTEs rising from 10 000 in 2007 to 13 744 in 2012 (Figure 4). However, the decline in the number of science council and business researchers, mainly because of the loss of 475 FTEs in science councils and 1 256 FTEs in business enterprises in a single year – 2010 – is of major concern.

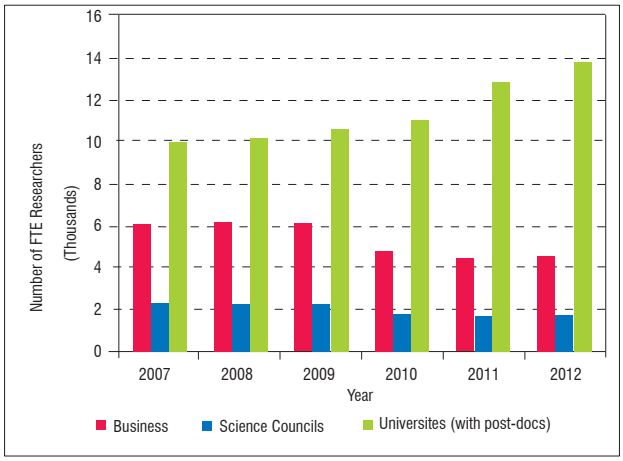


Figure 4: Trends in FTE researchers in the three R&D performance sectors (2007 to 2012).

The data indicate a recovery in the researcher labour market, despite concerns expressed in earlier reviews about labour market shortages and rising unit labour costs. Although the number of FTE researchers in universities has risen substantially, the unit costs per FTE have decreased over the same period, indicating that the R&D labour market conditions have eased at least for universities. Although R&D expenditure per FTE is only a proxy indicator for labour costs, it does provide some indication of what is happening in the market, because labour costs account for about 70% to 80% of total R&D expenditure. Notwithstanding this relaxation of labour market conditions, relative to their international peers and on a purchasing power parity basis, the average R&D spend per South African researcher is still higher than in some developed countries, and further adjustments in the labour market are necessary. The values in the graph shown in Figure 5 were calculated initially in USD purchasing power parity, and were then changed to 2005 constant ZAR using the purchasing power parity rate for 2005.

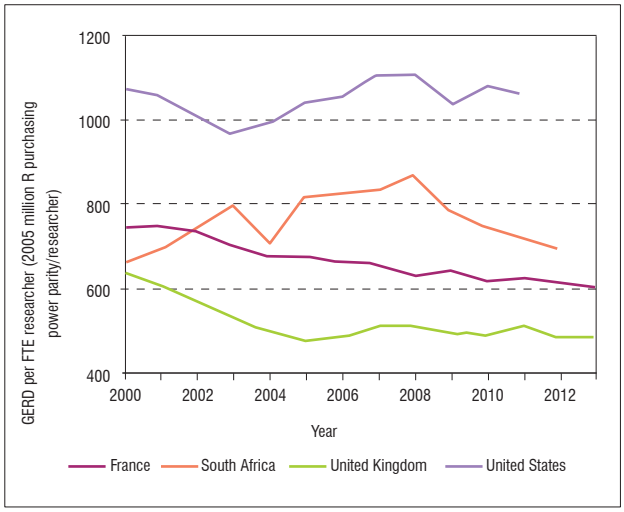


Figure 5: R&D expenditure per FTE for South Africa and comparator countries.

As predicted by Walwyn and Scholes[50], the cost of R&D in South Africa's public research institutions (known as science councils) has risen. Expenditure per FTE, which increased by 36% in real terms over the period 2007 to 2012, exceeded the value from the business sector for the first time in 2012 (Figure 6). These data reflect a growing equality in the market between R&D employment conditions in the public and private sectors, but could be indicative of the need for an adjustment to public research institutions' cost structures if these organisations are to compete more successfully against universities.

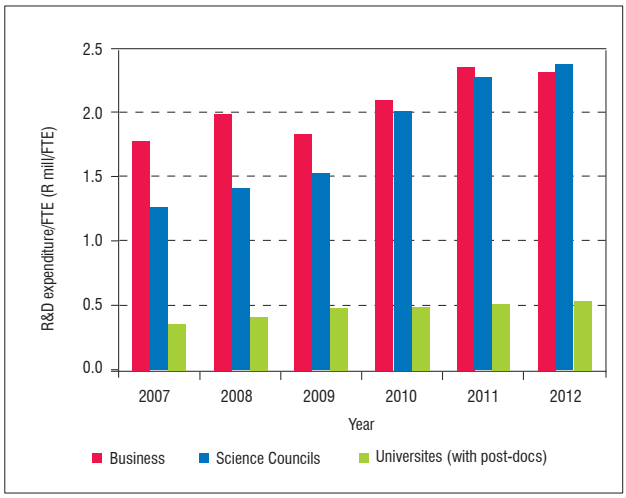


Figure 6: Trends in South Africa's R&D expenditure per FTE.

Business R&D

Of all the R&D metrics factors discussed in this article, the question of business funding for R&D is perhaps the greatest concern. (Note that

business funding is different from business performance of R&D, where the latter is referred to as BERD within the R&D Survey's nomenclature, irrespective of source of funding). As shown in Figure 7, it is clear that both funding and performance of R&D in this sector has declined in real terms since its peak during 2006 to 2008 (a drop of 34% in real terms).[51] The data for performance are confounded by the funding for the pebble bed modular reactor, in which government invested R8.8 billion between 1999 and 2010.[52]

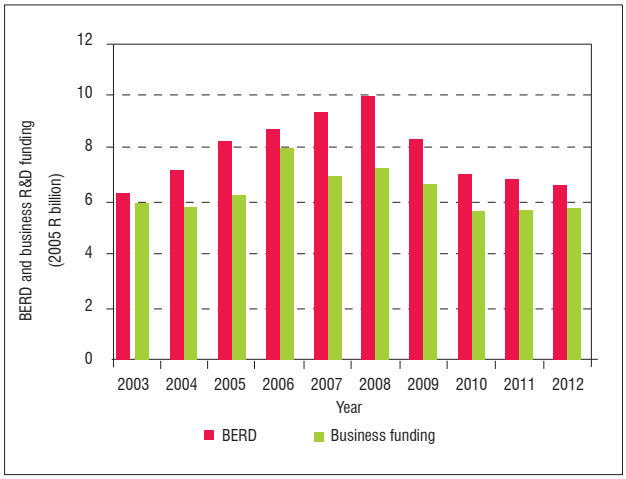


Figure 7: BERD and business R&D funding in constant 2005 R billion (2003 to 2012).

The decline in business funding for R&D would be even more severe if one were to exclude Sasol, whose funding for R&D rose from R376 million in 2002–2003 to R1.26 billion in 2012[53] and by 2012 constituted 12% of BERD in South Africa. A target or desirable value for BERD has been much debated over a long period, with proposed values ranging from 0.7% to 1.8% of GDP, or at a company level from 4% to 25% of revenue. Little agreement has been reached on the value with the highest return – or indeed even the methodology by which this value should be calculated.[54] In earlier work it has been suggested that South Africa's BERD should be about 0.9% of GDP (or about 3 times the present value of 0.34%); this figure was based on an analysis of the country's industry structure and benchmark values for each industry, as derived from an international comparison.[55]

A more recent approach to determining optimal R&D expenditure has followed the formulation of the Cobb-Douglas production function[56-58], as shown in Equation 1. The critical parameter for this discussion is the exponent *y*, which can also be expressed in the form of the Research Quotient (RQ). The latter is a normalised value calculated using the following equation:

$$RQ = y^{*}181 + 81 \qquad \text{Equation 1}$$

Although the basis for the normalisation is not explained in the original paper[56], an RQ of 100 is understood to represent a breakeven point for companies. A value of less than 100 corresponds with *y* of less than 0.107, and implies that the firm (or country) destroys rather than creates value with its R&D efforts. We undertook a limited analysis on selected South African companies using this methodology. Encouragingly, the results showed at least one of the companies, Sasol, has a coefficient of 0.412 and hence an RQ of about 155. This figure is on par with the leaders in the sector (Table 1) and confirms the high added value of Sasol's R&D efforts.

Table 1: A comparison of Research Quotient (RQ) for Sasol and other leading international companies

Company	Sector	RQ
Sasol	Fuels and chemicals	155
Usec Inc	Chemicals	136
Medicines Co	Drugs	130
China Petroleum and Chemicals	Oil and gas	124
Amazon.Com Inc	Online retail	123
Salix Pharmaceuticals Ltd	Drugs	119

Source: http://www.amkanalytics.com/Pages/rq50 *for all companies except Sasol (RQ for Sasol was calculated in our study)*

The Sasol RQ result indicates the successful implementation of a strategy built on in-house R&D and technology management – an approach that has enabled the company to retain high levels of profitability. Sasol's revenue reached R181 billion in 2012, with an operating profit (gross earnings) consistently above 20% of revenue (Figure 8). Understandably these returns have been under pressure following the collapse of the crude oil price, although the most recent data on the company's performance are not yet publically available.

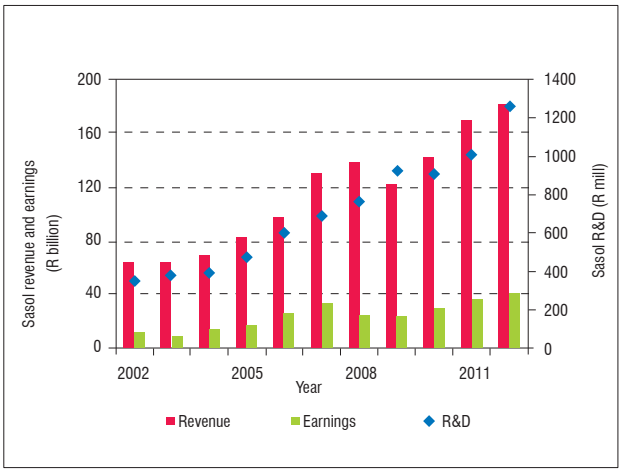


Figure 8: Sasol's revenue, earnings and R&D expenditure (2002 to 2012).

The initial results suggest that South African companies receive acceptable returns from their R&D expenditure. However, a perception remains that increased R&D is not judicious within the present business environment, and levels of BERD have continued to decline. The reasons for this decline might be complex, but appear to correlate with falling business confidence in the economy, which manifests in many ways – including reluctance to invest in R&D.

The economic upswing between September 1999 and November 2007 was the longest in South Africa's history[59] and was associated with growing business confidence, as indicated in Figure 9. During this period BERD also grew substantially, reaching a high of R10 billion (in constant 2005 rands) in 2008, before declining significantly between 2009 and 2012. The data indicate that BERD appears to track business confidence with a lag, not unexpectedly, of about a year. The implications of this correlation are discussed in the next section.

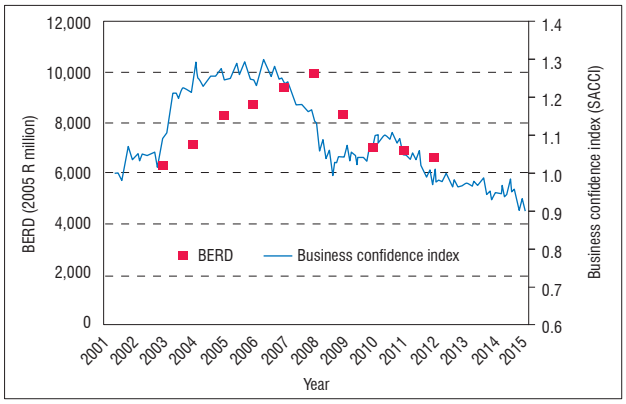


Figure 9: Trends in business confidence and business expenditure on R&D (2001 to 2015).

Discussion

Trends in R&D reveal interesting patterns in response to changes in the business environment, the policy context, social priorities, and advances in knowledge and technology. In this respect R&D is both a reflection of the past and a forecast of the future. The 2012/2013 R&D Survey illustrates several such trends, including an increasing role of universities in public-funded research, which follows an international trend of the movement of funds from public research institutions to universities. The preference towards universities or higher education institutions as performers of public-funded R&D has different drivers depending on specific national contexts. In South Africa, the trend can be ascribed to a number of overlapping and simultaneous adjustments within the system.

Firstly, universities have increased their activities in applied research. They have done so by establishing closer links with the private sector, setting up technology transfer offices, pursuing the registration and licensing of intellectual property arising from their R&D, and adopting the commercialisation of knowledge within institutions as a significant component of their mandates (in addition to teaching and research). Public research institutions, which have traditionally performed the role of adapting new knowledge for the development of novel products and services, are facing strong competition from universities in terms of access to competitive funding, networks with industry, sponsorship of specific research programmes and training of personnel in new technologies. In this competitive environment, universities have a strong cost advantage because of a rather loosely-applied notion of full cost and a generally lower cost per FTE researcher (Figure 6). The latter can be explained by the observation that much university R&D is conducted by doctoral or post-doctoral students, who are not remunerated by the universities except in unusual situations.

The second important aspect influencing the issue of universities and public research institutions is the implementation of new public management (NPM) and its effect within both institutions. NPM refers to the introduction of, firstly, a business-type managerialism adapted from the private sector. This management approach includes performance agreements, fixed-term contracts for senior managers, business planning, new financial techniques, full-cost accounting and greater autonomy for line managers. Secondly, NPM has introduced new institutional economics based on greater use of market mechanisms – including privatisation, removal of government subsidies, public choice, competition, quasi-markets, citizens as clients and customer satisfaction.[60] Although NPM was applied equally in both environments (universities and public research institutions), universities benefitted more in a financial sense from the change. For instance, an important consequence of NPM was the shift from block funding to competitive funding[47], with public research institutions having to compete for public funds alongside the universities. It has already been noted that universities have an inherent cost advantage in this contest for funding because of their use of indirectly-funded postgraduate students as researchers.

Rising levels of public-funded R&D within universities has the additional benefit of producing the necessary human resources to directly support the economy's transition from a resource-based to a knowledge-intensive structure. However, this link assumes requisite capacity within business enterprises to absorb these additional resources through intensification of business-supported R&D. As already noted, South Africa's development of this capacity has been slow and BERD has declined since 2008.

Furthermore, based on the trends in the South African Chamber of Commerce and Industry (SACCI) business confidence index, we predict that this decline has persisted since 2012/2013 (the most recent year for which R&D expenditure data are available) to the present, and will continue in the near future. In December 2015 the index reached a new low of 79.6, and the average in 2015 of 86.4 was the lowest since 1993 (when it was 81.3).[61] Assuming that this index is indeed predictive of business expenditure on R&D, as suggested by Figure 9, this trend does not auger well for BERD results in future years.

The situation presents an ongoing problem for government and requires a novel response. Its gravity can be appreciated even more profoundly if one considers the extent to which public policy has been reformed to support business R&D. For instance, government introduced a 150% R&D tax credit system in 2006; it provides financial support for technology entrepreneurs through the Technology Innovation Agency and the Support Programme for Industrial Innovation; it encouraged university-industry linkages through the Technology for Human Resources Programme; and through the Department of Science and Technology, government provides financial support for a wide range of R&D projects and collaborations. Given the disappointing outcome of these incentives, the South African government should consider other possibilities for productive interventions. These might include the three critical factors identified by Breznitz[22], the specialisation strategy proposed by Lee[62], and improved coordination between the Department of Science and Technology and other government departments to ensure more effective application of the various instruments. In particular, government needs to define a clear framework for how the state will acquire the necessary knowledge and skills to support new industries, how it will nurture local competence, and more importantly how it will embed local firms in global production networks.

The Department of Science and Technology also needs to clarify the mandates of its respective performance agencies, including the science councils, given the changes introduced by NPM and the new approaches to the commercialisation of university-based intellectual property. The role of science councils such as the Human Science Research Council and the Africa Institute of South Africa, which perform only basic and applied research (as defined by the R&D Survey), should be examined. There may be cost advantages associated with universities performing this type of research instead, and concomitant benefits of associated human capital development. In addition, the potential duplication of facilities and capabilities required for experimental development – and which already exist within public research institutions – should be avoided. Instead, government should seek to improve relationships between universities and science councils through a targeted instrument to support meaningful collaboration, especially in projects and areas that do not attract private sector interest.

The state's allocation of resources to support R&D is subject to various pressures, including the need to directly support business R&D, to grant more funding to universities and to sustain key infrastructure in the science councils. The final allocations must reflect at least a consideration of these priorities and a rational attempt to justify government decisions, based on a conscious logic model that articulates the relationship between funding and desired outcomes. Unfortunately such a model is complex and difficult to define. Many of the variables are non-recursive (bidirectional) and the relationship between them varies depending on the context. Assuming that South Africa is a middle-income country struggling to escape the proverbial trap of such countries, the need to pay more attention to learning and specialisation is appropriate, as it would in turn lead to gains in total factor productivity rather than factor accumulation. As a result, it could be argued that an increase

in public funding for universities within the present context of overall human resource constraints is justifiable. Evidence for such a shift, as noted in our study, is aligned with innovation policy studies and could over time lead to the economic growth so urgently desired. However, this attention by the state to the human resource pipeline needs to be combined with strengthening the science councils to deliver on short-cycle technologies, as well as a large effort to ignite business R&D. Business R&D has fallen to low levels and this is becoming a major impediment to the potential for future economic growth in South Africa.

Conclusion

Investment in R&D is an important decision for all countries, but especially in countries with a appreciable set of structural and economic problems, such as South Africa. Returns from R&D are by no means guaranteed, and several studies have shown variable outcomes and weak causality. It is therefore critical that the country's R&D Survey results are critically analysed on a regular basis and research policy is continually adjusted based on the insights from such analysis.

In the most recent survey, it is apparent that South Africa's expenditure on R&D has grown in real terms by 52% between 2001 and 2012. This increase was driven by government funding, which rose from 34% of total R&D funding in 2003 to 45% by 2012. Much of the additional funding has been granted to universities, with government support of R&D in this sector rising 450% in nominal terms and 250% in real terms over the same period. The survey also reveals a decline in the average cost of research, as expressed by expenditure per full time equivalent (FTE) researcher. Although the latter is a crude measure, the decrease suggests that the researcher labour market is better supplied and the constraints identified by earlier reviews are being overcome slowly. Both trends (rising government funding and declining unit costs) are highly positive for the system. However, the 34% decline in business R&D expenditure since its peak in 2008 is a matter of concern and needs to be addressed. In particular, the levels of specialisation, state-industry embeddedness and effective incentivisation of BERD must all be increased if South Africa is to overcome its present growth constraints and develop a competitive high-technology sector.

Authors' contributions

The authors contributed equally to this paper.

References

1. Department of Science and Technology. South African national survey of research and experimental development: Statistical report 2013/14. Pretoria: Statistics South Africa, 2016.
2. Department of Science and Technology. South African national survey of research and experimental development: Statistical report 2013/14. Pretoria: Statistics South Africa, 2015.
3. Kaplan D. Science and technology policy in South Africa past performance and proposals for the future. Science, Technology & Society. 2008;13(1):95–122. http://dx.doi.org/10.1177/097172180701300104
4. Kahn MJ. Rhetoric and change in innovation policy: The case of South Africa. Science, Technology & Society. 2013;18(2):189–211. http://dx.doi.org/10.1177/0971721813489447
5. Kaplan D. South Africa's national research and development strategy: A review. Science Technology & Society. 2004;9(2):273–294. http://dx.doi.org/10.1177/097172180400900204
6. OECD. Review of South Africa's innovation policy. Paris: Organisation for Economic Cooperation and Development, 2007.
7. Blankley W, Kahn M. The history of research and experimental development measurement in South Africa and some current perspectives. S Afr J Sci. 2005;101(3/4):151–156.
8. Kahn M. The South African national system of innovation: From constructed crisis to constructed advantage? Sci Public Policy. 2006;33(2):125–136. http://dx.doi.org/10.3152/147154306781779109
9. Kaplan DE. The constraints and institutional challenges facing industrial policy in South Africa: A way forward. Transformation: Critical Perspectives on Southern Africa. 2007;64(1):91–111. http://dx.doi.org/10.1353/trn.2008.0000
10. Mustapha N, Blankley W, Makelane H, Mololotja N. Trends in research and development expenditure in South Africa (2010–2013): Policy implications. Pretoria: HSRC, 2015
11. Arrow K. Economic welfare and the allocation of resources for invention. In: Rosenberg N, editor. The rate and direction of inventive activity: Economic and social factors. Princeton: Princeton University Press; 1962. p. 609–626. http://dx.doi.org/10.1515/9781400879762-024
12. Joseph RA, Johnston R. Market failure and government support for science and technology: Economic theory versus political practice. Prometheus. 1985;3(1):138–155. http://dx.doi.org/10.1080/08109028508628976
13. Biersteker TJ. Reducing the role of the state in the economy: A conceptual exploration of IMF and World Bank prescriptions. Int Stud Quart. 1990;34(4):477–492.
14. Mazzucato M. The entrepreneurial state: Debunking public vs. private sector myths. London: Anthem Press; 2013.
15. Johnson CA. MITI and the Japanese miracle: The growth of industrial policy: 1925–1975: Stanford University Press; 1982.
16. Evans PB. Embedded autonomy: States and industrial transformation: Cambridge Univ Press; 1995.
17. Amsden AH, Chu W-w. Beyond late development: Taiwan's upgrading policies. MIT Press Books. 2003;p. 1.
18. Amsden AH. Asia's next giant: South Korea and late industrialization. New York: Oxford University Press; 1989.
19. O'Riain S. The politics of high tech growth: Developmental network states in the global economy: Cambridge University Press; 2004. http://dx.doi.org/10.1017/CBO9780511499609.003
20. O'Riain S. The flexible developmental state: Globalization, information technology and the 'Celtic Tiger'. Polit Soc. 2000;28(2):157–193. http://dx.doi.org/10.1177/0032329200028002002
21. Ansell C. The networked polity: Regional development in Western Europe. Governance. 2000;13(2):279–291. http://dx.doi.org/10.1111/0952-1895.00136
22. Breznitz D. Innovation and the state: Political choice and strategies for growth in Israel, Taiwan, and Ireland. New Haven: Yale University Press; 2007. http://dx.doi.org/10.12987/yale/9780300120189.001.0001
23. Cantwell J. Innovation and competitiveness. The Oxford handbook of innovation. Oxford: Oxford University Press; 2005. p. 543–567.
24. Kim YK, Lee K. Different impacts of scientific and technological knowledge on economic growth: Contrasting science and technology policy in East Asia and Latin America. Asian Economic Policy Review. 2015;10(1):43–66. http://dx.doi.org/10.1111/aepr.12081
25. Ntuli H, Inglesi-Lotz R, Chang T, Pouris A. Does research output cause economic growth or vice versa? Evidence from 34 OECD countries. Journal of the Association for Information Science and Technology. 2015;66:1709–1716. http://dx.doi.org/10.1002/asi.23285
26. Jones R. The UK's innovation deficit & how to repair it. SPERI Paper No 6. Sheffield. 2013.
27. Inglesi-Lotz R, Pouris A. The influence of scientific research output of academics on economic growth in South Africa: An autoregressive distributed lag (ARDL) application. Scientometrics. 2013;95(1):129–139. http://dx.doi.org/10.1007/s11192-012-0817-3
28. Lee H. Cycle time of technologies and criteria for specialization. Personal communication. 2015.
29. National Advisory Council on Innovation. South African science, technology and innovation indicators 2014. Pretoria: Department of Science and Technology, 2015.
30. Lee K. Schumpeterian analysis of economic catch-up: Knowledge, path-creation, and the middle-income trap. Cambridge: Cambridge University Press; 2013. http://dx.doi.org/10.1017/CBO9781107337244

31. Department of Science and Technology. Ten year innovation plan: Innovation towards a knowledge-based economy 2008–2018. Pretoria: Department of Science and Technology; 2008.

32. Davidson DB. Meerkat and SKA phase 1. Antennas, Propagation & EM Theory (ISAPE). 10th International Symposium: IEEE. 2012. http://dx.doi.org/10.1109/ISAPE.2012.6409014

33. Thomas S. The pebble bed modular reactor: An obituary. Energy Policy. 2011;39(5):2431–2440. http://dx.doi.org/10.1016/j.enpol.2011.01.066

34. Swart G. Innovation lessons learned from the Joule EV development. IAMOT 2015; Cape Town: IAMOT; 2015.

35. Karim QA, Karim SSA, Frohlich JA, Grobler AC, Baxter C, Mansoor LE, et al. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. Science. 2010;329(5996):1168–1174. http://dx.doi.org/10.1126/science.1193748

36. Department of Science and Technology. Ministerial review committee on science, technology and innovation landscape in South Africa: Final report. Pretoria: Department of Science and Technology; 2012.

37. Cooper RG, Edgett SJ, Kleinschmidt EJ. New product portfolio management: Practices and performance. Journal of Product Innovation Management. 1999;16(4):333–351. http://dx.doi.org/10.1111/1540-5885.1640333

38. Seekings J, Nattrass N. Policy, politics and poverty in South Africa: Palgrave Macmillan; 2015.

39. Faulkner D, Loewald C, Makrelov K. Achieving higher growth and employment: Policy options for South Africa. Pretoria: National Treasury, 2013.

40. Seekings J. State-building, market regulation and citizenship in South Africa. European Journal of Social Theory. 2016;19(2):191–209.

41. Nattrass N, Seekings J. Should and can labour-surplus, middle-income economies pursue labour-intensive growth? The South African challenge. Centre for Social Science Research, Working Paper No 351, 2015.

42. Spaull N. South Africa's education crisis: The quality of education in South Africa 1994–2011. Johannesburg: Centre for Development and Enterprise, 2013.

43. Schwab K, Sala-i-Martin X. The global competitiveness report 2014–2015. Geneva: 2014.

44. Council on Higher Education. Higher education participation 2011. Pretoria: Council on Higher Education, 2013.

45. Department of Science and Technology. 2012/13 national survey of research and experimental development. [Executive summary.] Pretoria: Department of Science and Technology; 2014.

46. Guellec D, Van Pottelsberghe de la Potterie B. From R&D to productivity growth: Do the institutional settings and the source of funds of R&D matter? Oxford Bulletin of Economics and Statistics. 2004;66(3):353–378. http://dx.doi.org/10.1111/j.1468-0084.2004.00083.x

47. OECD. Public research institutions: Mapping sector trends. Paris: 2011.

48. Kahn M. Internationalization of R&D: Where does South Africa stand? S Afr J Sci. 2007;103(1/2):7–12.

49. OECD. Frascati manual 2002: Proposed standard practice for surveys on research and experimental development. Paris: OECD; 2002.

50. Walwyn D, Scholes RJ. The impact of a mixed income model on the South African CSIR: A recipe for success or disaster? S Afr J Sci. 2006;102(5/6):239–243.

51. Department of Science and Technology. South African national survey of research and experimental development main analysis report: 2011/12. Pretoria: Department of Science and Technology; 2014.

52. McKune C. Pebble bed modular reactor demonstration plant is funded but not constructed. S Afr J Sci. 2010;106(5/6):2–4. http://dx.doi.org/10.4102/sajs.v106i5/6.287

53. Sasol. Annual financial statements 2012/3. Johannesburg: Sasol, 2013.

54. Hartmann GC, Myers MB, Rosenbloom RS. Planning your firm's R&D investment. Res Tech Man. 2006;49(2):25–36.

55. Walwyn D. A target for South Africa's business expenditure on research and development based on industry structure. S Afr J Sci. 2008;104(9/10):340–344.

56. Knott AM. The trillion-dollar R&D fix. Harvard Business Review. 2012;76.

57. Gwynne P. Research note: Measuring r&d productivity. Res Tech Man. 2015;58(1):19–23. http://dx.doi.org/10.5437/08956308X5801008

58. Euchner J. From the editor: How much is enough? Res Tech Man. 2015;58(1):9–11. http://dx.doi.org/10.5437/08956308X5801002

59. Venter JC. Business cycles in South Africa during the period 1999 to 2007. Pretoria: South African Reserve Bank, 2009.

60. Cameron R. New public management reforms in the South African public service: 1999–2009. Journal of Public Administration. 2009;44(4.1):910–942.

61. South African Chamber of Commerce and Industry (SACCI). Business confidence index August 2015. South African Chamber of Commerce and Industry, 2015.

62. Lee C, Narjoko D. Escaping the middle-income trap in southeast Asia: Micro evidence on innovation, productivity, and globalization. Asian Economic Policy Review. 2015;10(1):124–127. http://dx.doi.org/10.1111/aepr.12090

A unique fingerprint? Factors influencing attitudes towards science and technology in South Africa

AUTHORS:
Lars Guenther[1]
Peter Weingart[1]

AFFILIATION:
[1]Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, South Africa

CORRESPONDENCE TO:
Lars Guenther

EMAIL:
larsguenther@sun.ac.za

POSTAL ADDRESS:
Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Ryneveld Street, Stellenbosch 7602, South Africa

DATES:
Received: 23 Mar. 2016
Revised: 19 May 2016
Accepted: 23 May 2016

KEYWORDS:
perceptions of science and technology; scientific literacy; South Africa; sources of scientific information

HOW TO CITE:
Guenther L, Weingart P. A unique fingerprint? Factors influencing attitudes towards science and technology in South Africa. S Afr J Sci. 2016;112(7/8), Art. #2016-0093, 4 pages. http://dx.doi.org/10.17159/sajs.2016/20160093

From an international perspective, research in the field of public attitudes towards science and technology has been conducted since the 1970s. A frequently articulated – and empirically supported – assumption is that strong interest in and knowledge about science in a society is associated with more favourable attitudes towards science. This positive attitude in turn affects support for public funding of science. However, this research field is not without controversy, and for the South African population many questions remain unanswered. Initial research has not explored the factors that shape attitudes towards science and technology in detail. We re-analysed data from the Human Sciences Research Council to explore the above assumption. Interestingly, for the South African population, higher levels of scientific literacy and use of information sources are associated with more promises but also more reservations towards science and technology. This is especially true for relatively young and educated survey respondents. In international comparison, South Africa shows a unique fingerprint to some extent, but also shares characteristics with industrially developing countries of Europe (such as Greece or Portugal). To understand the correlations better, future research should aim to examine the overall picture when investigating the diverse South African population more extensively.

Introduction

Starting in the 1970s, research on public perceptions of science and technology (S&T) was driven by the idea that national success is largely dependent on innovation in S&T, and that innovation requires a supportive public.[1] However, research programmes were initiated mainly because of a rising public scepticism towards science in Western countries.[2] Researchers were afraid that this scepticism would result in funding cuts for scientific programmes. As a result, surveys measuring public perceptions turned into a regular activity in many countries.[3] The idea behind this research field is that a combination of interest in and knowledge about science shapes attitudes towards science. In turn, these attitudes affect outcomes such as support for public funding among the voting population.[1]

For South Africa, certain initial findings have led researchers to conclude that this country has a *unique fingerprint* with regard to public attitudes towards S&T.[4] South Africans generally show a mix of positive and negative attitudes about S&T. In this paper, 'promises' refer to people's positive expectations and beliefs about the benefits of S&T, whereas 'reservations' refer to negative views about S&T. A comparison of recent data with earlier data[5] showed that for South Africans, scientific promises have dropped slightly whereas some reservations have increased significantly. Age and education seem to influence the promise–reservation ratio in South Africa.[4] However, the claim that South Africa has a *unique* fingerprint requires further empirical support.

Specifically, the influence of scientific literacy and the use of information sources on 'promises' and 'reservations' regarding S&T has not been investigated in enough detail. According to the literature, these two factors play a crucial role in public perceptions of S&T.

Short theoretical review: Public perceptions of science and technology

Surveys conducted in other countries found that the more scientifically literate the public, the more favourable is their attitude towards S&T.[6] A meta-analysis showed a weak but positive correlation.[2] If there is a positive perception of science, then there is also more support for its public funding.[7] In addition, favourable attitudes towards S&T and higher literacy both seem to be influenced by sociodemographics, with gender (male), age (younger), and education level (higher) being among the important variables.[6,8]

In general, the correlation between scientific literacy and attitudes towards S&T has long been debated.[1,9] For some time, researchers have adhered to the so-called *deficit model*, believing that if the public were only more scientifically literate they would hold more favourable attitudes towards science, and scepticism would vanish.[6] That is why many educational initiatives have been carried out in Western countries.[10] However, the success of such initiatives as well as the whole terminology have been questioned. In addition, cross-cultural studies in Europe have shown that in industrially developing countries, a strong positive correlation exists between scientific literacy and attitudes towards S&T, whereas in post-industrial countries this correlation is weaker.[10] The reasoning is that in industrial societies, only a small elite is really knowledgeable (high socioeconomic stratification[11]), and greater knowledge leads to a more positive attitude.

Moving along the continuum of economic development, in a post-industrial society one finds that knowledge is widely distributed, mainly because education is more accessible but also because developed countries make greater demands on their citizens in terms of scientific and technical skills.[12] Positive attitudes, then, become less predictable by scientific knowledge as public attitudes commonly become more sceptical.[13] That is why, in general, basic assumptions of the deficit model (i.e. greater scientific literacy leads to more favourable attitudes to science) are mainly confirmed in industrially developing countries. However, in these countries, overall, an ambivalent picture emerges: large parts of such societies also show strong pessimism towards science, as was found in some European countries.[11]

Interestingly, for South Africa, we have no reports answering whether and to what extent factors such as scientific literacy and use of information sources differentially influence attitudes towards S&T. Addressing this research gap was the central goal of our investigation. Our main research question was: *What are the strongest predictors influencing public attitudes towards S&T in South Africa?*

Method

Research design and sample

Our investigation was a secondary analysis of data from the Human Sciences Research Council 2010 wave of the South African Social Attitudes Survey (SASAS 2010). The number of participants (*n*) was 3183. To obtain a representative sample of the South African population, three stratification variables were used: province, geographic type and majority population group, and the data were weighted accordingly.[4] Data collection took place between November and December 2010, using face-to-face interviews. The SASAS 2010 included, among others, items to measure attitudes towards S&T, scientific literacy and sources of scientific information.[4]

Measurement

Within SASAS 2010, seven items measured respondents' attitudes to S&T (α=0.80) using a 5-point rating scale (1=strongly agree; 5=strongly disagree).[4] Our confirmatory factor analysis using principal component analysis and Varimax rotation (KMO=0.84) resulted in two factors. The first factor (4 items; α=0.79) was 'promises towards S&T' (Eigenvalue=3.25, explaining 46% of the variance). The second factor (3 items; α=0.65) was 'reservations towards S&T' (Eigenvalue=1.06, explaining 15% of the variance).

Hence, two additive indexes served as dependent variables in our study: promises (mean=2.52; *s.d.*=1.387) and reservations (2.77±1.225) towards S&T. These factors are congruent with the literature.[8,14] 'Promises' refer to positive expectations and beliefs in the benefits of S&T (sample item: 'S&T are making our lives healthier, easier, and more comfortable'), whereas 'reservations' refer to negative consequences and predispositions concerning S&T (sample item: 'Science makes our way of life change too fast').[15]

Table 1: Claims to measure scientific literacy and responses

Scientific claim	True	False	Don't know
The centre of the Earth is very hot.	79%†	9%	12%
Electrons are smaller than atoms.	39%†	33%	28%
Antibiotics kill viruses as well as bacteria.	43%	38%†	19%
Human beings developed from earlier species of animals.	34%†	46%	20%
The sun rotates around the Earth.	47%	42%†	10%
The oxygen we breathe comes from plants.	82%†	8%	9%

† = *correct answer*

We set the independent variables to be 'scientific literacy', 'sources of scientific information' and sociodemographic data. Scientific literacy had been measured using six items asking respondents to indicate whether a scientific claim is true or false, or that they did not know (α=0.84). The results of the analysis are shown in Table 1. Based on the number of correct answers, an additive index for scientific literacy (3.14±1.445) was created. Sources of scientific information (α=0.87) had been assessed on a 5-point rating scale (1=very often; 5=never) asking how often several sources were used by respondents to access information about S&T. The mean scores (with *s.d.*) were as follows: television 2.57±1.494; radio 2.86±1.396; newspapers 3.30±1.387; books/magazines 3.56±1.341; the Internet 4.19±1.274; other people 3.21±1.330; and public spaces 4.04±1.219.

Relevant sociodemographic variables for this investigation were gender (52% female), age (37.03±20 years), level of education (low 18%, medium 67%, high 15%), and social class (lower 42%, working 24%, middle 27%, upper middle 4%, upper 1%).

Findings

To answer the research question, hierarchical regressions for both dependent variables were tested (Table 2). Using 'promises towards S&T' as a dependent variable (F=683051.17; $d.f.$=12; p<.001), the findings showed that the more scientifically literate respondents were, the higher the promises towards S&T. Scientific literacy was the strongest predictor. Other meaningful findings were that the more respondents used television and books/magazines as sources of scientific information, the greater their perception of promises towards S&T. In addition, two sociodemographic variables had an important influence. In this sample, the younger the participants and the higher their level of education, the more favourable their attitude towards S&T.

For 'reservations towards S&T' as a dependent variable (F=338065.82; $d.f.$=12; p<.001), the findings showed that the more scientifically literate the respondents, the higher their reservations towards S&T. Again, scientific literacy was the strongest predictor. In addition, for the sample we studied, reservations towards S&T increased the more respondents used television, radio, newspapers, or public spaces as sources of scientific information. However, reservations decreased when books/magazines or the Internet were used as a source. Three sociodemographic variables were among the strongest predictors: reservations were stronger for younger people, more educated respondents, and people from lower social classes.

Conclusion

In terms of the South African population, our study showed that knowledge (greater knowledge), age (younger) and education (more educated) were associated with more favourable attitudes to S&T. In an international comparison, South Africa shares characteristics with industrially developing countries of Europe, such as Greece and Portugal.[11] In post-industrial European countries, the correlation between knowledge and positive attitudes is weaker.[10] In the USA, promises and reservations are negatively correlated.[14] For Europe, researchers found that lower social classes of the population in less advanced countries hold stronger reservations towards S&T than higher social classes.[15] In South Africa, however, scientific literacy increased both promises and reservations towards S&T; hence, South Africa indeed has a *unique fingerprint*.

A rationale for this finding could be that more knowledgeable, educated and younger South Africans see S&T as the route to progress; however, they also have reservations that science changes their lives too much.[11] In a recent international study, the South African sample was the one that agreed most strongly that people believe too often in science and not enough in feelings and faith. However, South Africans also believed most strongly that science is able to solve problems[16] – perfectly representing this ambivalence. One finding of our study, namely that a lower social class also shares more reservations, shows that future investigations need to explore in greater detail the specific correlations between the variables.

Despite these illuminating findings, our investigation was merely a secondary analysis of existing data. We thus had no control over the type of constructs and items tested, and as a result can present only a partial picture. Further research questions should include whether this attitudinal ambivalence among relatively more literate and educated South Africans leads to greater support of public funding. In addition, the role of trust in science and scientific institutions, or interest in science, needs to be further explored, because both are central when measuring public perceptions of science.[12] To understand the role of information sources fully, a qualitative design could explore what kinds of sources shape public attitudes. Because South Africa is on many levels a highly diverse country, it would also be interesting to see to what extent the life-worlds of different segments of the public influence people's perceptions of science. We propose to answer these research questions in our future investigations.

Table 2: Results of hierarchical regressions

	Model 1		Model 2		Model 3	
Variables	***β***	***t***	***β***	***t***	***β***	***t***
Dependent variable: Promises towards science and technology						
Scientific literacy	-.384	-2322.37	-.350	-2077.90	**-.328**	-1905.79
Television as information source			.189	793.49	**.175**	735.21
Radio as information source			.001	5.94	.019	79.53
Newspapers as information source			-.017	-65.67	-.030	-116.30
Books/magazines as information source			.070	274.17	**.054**	210.21
Internet as information source			-.018	-87.01	-.034	-163.08
Other people as information source			-.012	-58.69	-.007	-34.31
Public spaces as information source			.019	89.99	.013	62.11
Gender					.013	80.14
Age					**.078**	467.59
Level of education					**-.080**	-435.49
Social class					-.012	-69.46
Adjusted R^2	14.8		19.5		20.8	
Dependent variable: Reservations towards science and technology						
Scientific literacy	-.230	-1316.08	-.227	-1273.38	**-.221**	-1210.62
Television as information source			.131	518.98	**.128**	505.87
Radio as information source			.049	188.56	**.053**	204.91
Newspapers as information source			.078	288.14	**.079**	290.11
Books/magazines as information source			-.063	231.78	**-.074**	-269.24
Internet as information source			-.096	443.43	**-.092**	-418.72
Other people as information source			.050	229.10	.047	218.51
Public spaces as information source			.067	302.54	**.067**	305.04
Gender					.039	225.01
Age					**.051**	287.12
Level of education					**-.050**	-257.40
Social class					**.054**	292.43
Adjusted R^2	5.3		10.7		11.6	

Note. All results are significant ($p<.001$). Bold numbers in Model 3 are the strongest predictors.

Acknowledgements

This work is based on research supported by the South African Research Chairs Initiative of the Department of Science and Technology together with the National Research Foundation (NRF) of South Africa (Grant Number 93097). The opinions, findings, conclusions and recommendations expressed in this paper are those of the authors, and the NRF does not accept any liability in this regard. The data used in this study were obtained from the South African Human Science Research Council. For discussions on conceptualising the study, we wish to thank Michael Gastrow, Marina Joubert and Kalayvan Nadar. We also wish to thank our proofreader Jackie Viljoen.

Authors' contributions

L.G. analysed the data and was the main author. P.W. was the project manager and was involved in conceptualising the study and writing.

References

1. Besley JC. The state of public opinion research on attitudes and understanding of science and technology. B Sci Technol Soc. 2013;33(1/2):12–20. http://dx.doi.org/10.1177/0270467613496723
2. Allum N, Sturgis P, Tabourazi D, Brunton-Smith I. Science knowledge and attitudes across cultures: A meta-analysis. Public Underst Sci. 2008;17:35–54. http://dx.doi.org/10.1177/0963662506070159

3. Raza G, Singh S, Shukla R. Relative cultural distance and public understanding of science. Science, Technology & Society. 2009;14(2):269–287. http://dx.doi.org/10.1177/097172180901400204

4. Reddy V, Gastrow M, Juan A, Roberts B. Public attitudes to science in South Africa. S Afr J Sci. 2013;109(1/2):1–8. http://dx.doi.org/10.1590/sajs.2013/1200

5. Pouris A. Interests, public attitudes and sources of scientific information in South Africa. 2001. Available from www.visits.web.cern.ch/visits/pcst2001/proc/Pouris.doc

6. Sturgis P, Allum N. Science in society: Re-evaluating the deficit model of public attitudes. Public Underst Sci. 2004;13:55–74. http://dx.doi.org/10.1177/0963662504042690

7. Muñoz A, Moreno C, Luján JL. Who is willing to pay for science? On the relationship between public perception of science and the attitude to public funding of science. Public Underst Sci. 2012;21(2):242–253. http://dx.doi.org/10.1177/0963662510373813

8. Miller JD. Public understanding of, and attitudes toward, scientific research: What we know and what we need to know. Public Underst Sci. 2004;13:273–294. http://dx.doi.org/10.1177/0963662504044908

9. Allum N, Sibley E, Sturgis P, Stoneman P. Religious beliefs, knowledge about science and attitudes towards medical genetics. Public Underst Sci. 2014;23(7):833–849. http://dx.doi.org/10.1177/0963662513492485

10. Bauer MW, Allum N, Miller S. What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. Public Underst Sci. 2007;16:79–95. http://dx.doi.org/10.1177/0963662506071287

11. Durant J, Bauer MW, Gaskell G, Midden CJH, Liakopoulos M, Scholten LM. Two cultures of public understanding of science and technology in Europe. In: Dierkes M, Von Grote C, editors. Between understanding and trust: The public, science and technology. London: Routledge; 2000. p. 131–156.

12. Bauer MW, Durant J, Evans G. European public perceptions of science. Int J Public Opin R. 1994;6(2):163–186. http://dx.doi.org/10.1093/ijpor/6.2.163

13. Bauer MW. The evolution of public understanding of science: Discourse and comparative evidence. Science, Technology & Society. 2009;14(2):221–240. http://dx.doi.org/10.1177/097172180901400202

14. Nisbet MC, Scheufele DA, Shanahan J, Moy P, Brossard D, Lewenstein B. Knowledge, reservations, or promise? A media effects model for public perceptions of science and technology. Commun Res. 2002;29(5):584–608. http://dx.doi.org/10.1177/009365002236196

15. Pardo R, Calvo F. Mapping perceptions of science in end-of-century Europe. Sci Commun. 2006;28(1):3–46. http://dx.doi.org/10.1177/1075547006291895

16. National Science Foundation. Science and Engineering Indicators 2014. [Document on the Internet]. c 2014 [Cited 2016 March 14]. Available from http://www.nsf.gov/statistics/seind14/content/chapter-7/chapter-7.pdf

Two dung beetle species that disperse mimetic seeds both feed on eland dung

AUTHORS:
Jeremy J. Midgley[1]
Joseph D.M. White[1]

AFFILIATION:
[1]Department of Biological Sciences, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Jeremy Midgley

EMAIL:
jeremy.midgley@uct.ac.za

POSTAL ADDRESS:
Department of Biological Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa

DATES:
Received: 11 Apr. 2016
Revised: 26 May 2016
Accepted: 28 May 2016

KEYWORDS:
Scarabaeus spretus; *Epirinus flagellatus;* seed dispersal; mimicry; deception; eland; *Ceratocaryum argenteum*

HOW TO CITE:
Midgley JJ, White JDM. Two dung beetle species that disperse mimetic seeds both feed on eland dung. S Afr J Sci. 2016;112(7/8), Art. #2016-0114, 3 pages. http://dx.doi.org/10.17159/sajs.2016/20160114

Scarabaeus spretus zur Strassen was observed to roll and bury *Ceratocaryum argenteum* (Restionaceae) seeds in the sandplain fynbos of the Potberg area of the De Hoop Nature Reserve, South Africa. This species is the second dung beetle species found to be deceived by the faecal mimicry of *C. argenteum* seeds – the first species being *Epirinus flagellatus.* An isotopic analysis suggests that both these dung beetle species most likely feed on eland (*Taurotragus oryx*), not bontebok (*Damaliscus pygargus pygargus*), dung. Thus the model in this mimicry is eland dung; this interaction suggests large herbivores are an integral part of this fynbos.

Introduction

The dung beetle *Epirinus flagellatus* was observed to roll and bury seeds of the Cape plant *Ceratocaryum argenteum* (Restionaceae) at a site in the Potberg part of the De Hoop Nature Reserve in South Africa.[1] This primary dispersal of seeds involves chemical and visual mimicry because neither the dung beetle nor its larvae can eat these hard seeds. Chemically the seeds have characteristics of the dung of both of the most common large herbivores in the reserve: the eland (*Taurotragus oryx*) and the bontebok (*Damaliscus pygargus pygargus*).[1] However, the seeds are more similar in shape and size to the smaller faeces of the bontebok, which is then the possible visual model that *C. argenteum* mimics. At the same Potberg site, during February 2016, we observed similar seed dispersal of *C. argenteum* seeds by another dung beetle, *Scarabaeus spretus* zur Strassen. The aim of this paper is to document this new burial behaviour and to investigate both dung beetle species to determine whether the faeces of the bontebok or the eland is the likely model of the mimic.

The bontebok is a short grass grazer whereas the eland is a mixed feeder.[2,3] The two dominant grassland/renosterveld grass species at Potberg are *Cymbopogon popschilli* (Andropogoneae) and *Cynodon dactylon* (Chloridoideae),[3] to which can be added the relatively widespread *Themeda triandra* (Andropogoneae). All three species utilise the C4 photosynthetic pathway rather than the C3 pathway.[4] This pathway is common in tropical grasses whereas the C3 system is more common in woody plants and temperate grasses. The enzymes of these two different photosynthetic pathways produce different carbon $\delta^{13}C$ signatures in their photosynthetic products. The relatively rare stable isotope of carbon,^{13}C, is slightly heavier than the more common ^{12}C, which affects the ratios of these isotopes in different plants depending, for example, on enzyme preferences for the lighter isotopes. Fractionation is the process which reflects changes in relative proportions of isotopes, such as ^{13}C:^{12}C during C3 photosynthesis. Fractionation can also occur in ^{15}N because, as it is heavier than ^{14}N, it may increase in tissues depending on factors such as levels of metabolism, catabolism and excretion. Thus animals are typically enriched by +3–5‰ in $\delta^{15}N$ compared to their diet, although typically they are less than +1‰ enriched in $\delta^{13}C$.[5] During metamorphosis, larval tissue is broken down and then used to form new adult tissue and thus metamorphosis is also known to increase both $\delta^{13}C$ and $\delta^{15}N$ in the adult tissue in much the same way as would happen in adult tissue with an increase in trophic level.[6] Thus the hypothesis that bontebok dung is the likely model *C. argenteum* mimics can be tested using the isotopic method for diagnosing animal diets, including those of dung beetles.[7]

Methods

The study took place in the Potberg area of the De Hoop Nature Reserve (34.374420 S, 20.533060 E) in the sand plain vegetation type in which *C. argenteum* grows. During 3 days in early February 2016, we placed out 5 to 10 piles of *C. argenteum* seeds, with each pile comprising 10–20 seeds. Piles of seeds were 10 m apart on the edge of a 100-m stretch of a sand road through natural vegetation. We monitored the seed piles in the early morning for approximately 2 h (starting at about 08:00). This experiment took place after a 24-h rain event.

Dung samples were taken in various vegetation types in the Potberg reserve. These types were renosterveld (dominated by the shrub *Elytopappus rhinocerotis* (Asteraceae)), grassland (dominated by *Cynodon dactylon*), salt marshes (dominated by Chenopodiaceae), valley bottom fynbos (dominated by the Proteaceae shrubs *Leucadendron linifolium*/ *L. coniferum),* sand plain fynbos *(*dominated by *Leucadendron laureolum*, where *C. argenteum* occurs) and limestone fynbos (dominated by *Leucadendron meridianum*). Eland dung was found at all six sites whereas bontebok dung was found at all but the last two fynbos sites (sand plain fynbos and limestone fynbos). Previously, Radloff et al.[3] noted that bontebok avoid fynbos whereas eland are found throughout fynbos, including limestone fynbos. To reduce chances of pseudoreplication, we sampled only a single pellet of bontebok or eland dung from a dung pile; only dung piles greater than 5 m apart were sampled and, as judged by colour, only relatively fresh samples were collected until a total of 10 pellets had been sampled from within each vegetation type. Dung pellets and dung beetle exoskeletons were dried and analysed for $\delta^{13}C$ and $\delta^{15}N$, in ‰, using standard techniques at the Archaeometry Lab at the University of Cape Town. Dung beetle larvae have chewing teeth and are vigorous detritivores that depend on the plant remains that constitute the dung ball, rather than being dependent on microbiota associated with the ball.[8] Although female beetles do select small fragments of plant remains from dung to constitute brood balls,[8] this is not likely to significantly affect the isotopic signature of these balls nor the signature of the exoskeletons of adults that emerge from these balls.

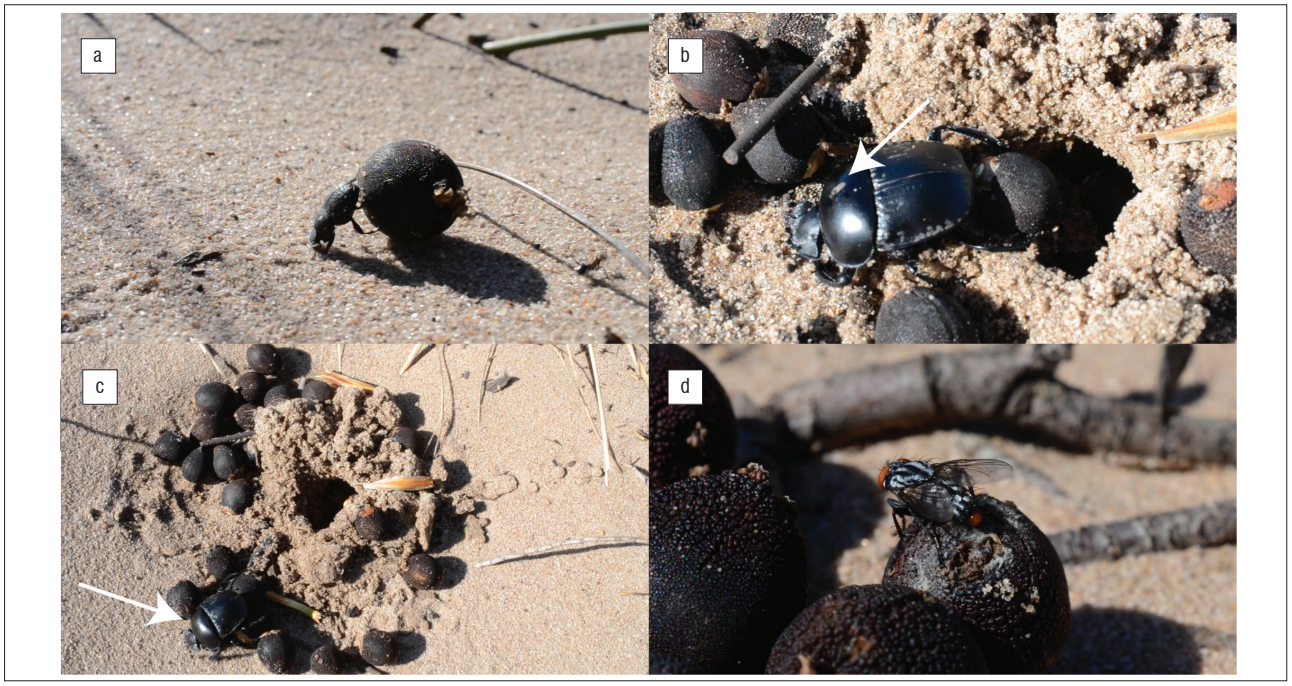


Figure 1: (a) *Epirinus flagellatus* rolling a *Ceratocaryum argenteum* seed; (b) *Scarabaeus spretus* rolling a seed (the arrow indicates a sphaerocerid lesser dung fly); (c) the large hole made by *S. spretus* for burying several seeds (the arrow indicates the location of the dung beetle); and (d) a female sarcophagid fly on a seed.

Results

Dung beetles arrived at seed stations within a few minutes of placing seeds out; thus within 2 h each day, more than 10 beetles had arrived at seed piles along our short 100-m transect and had started burying seeds (Table 1 and Figure 1a–c). *E. flagellatus* crawled out of the vegetation towards seed piles, with only an occasional individual flying in, whereas all *S. spretus* individuals flew towards the seed piles. It was clear, based on the direct flight or crawling paths of both species to the seeds, that the attraction is strongly chemical. A *S. spretus* beetle even flew into a paper bag containing seeds. All cases of *S. spretus* burial involved limited movement of seeds (<0.25 m) from seed piles, whereas *E. flagellatus* moved seeds up to 2 m. *S. spretus* beetles were observed to frantically bury up to three seeds ($n=2$) and often five or more seeds ($n=4$) per excavated hole (see the video in the supplementary material online). *E. flagellatus* was observed to only bury seeds individually, similarly to observations by Midgley et al.[1] Flies of the Sarcophagidae were frequently observed to settle on *C. argenteum* seeds (Figure 1d), indicating that they too are deceived by the scent of the seeds. These 'flesh flies' are typically attracted to dung, carrion or rotting vegetation.[9] Lesser dung flies (Sphaeroceridae) were observed on *S. spretus* (Figure 1b).

Samples of the grasses *Cynodon dactylon* and *Themeda triandra* from Potberg have a typical C4 isotopic signal ($n=2$ for each species, mean δ^{13}C of -13.61‰ and -14.29‰, respectively).

Table 1: Observations of dung beetles at *Ceratocaryum argenteum* seed stations at Potberg

Date	Duration (min)	Number of *Epirinus flagellatus* individuals	Number of *Scarabaeus spretus* individuals
4 February 2016	120	10	3
5 February 2016	135	8	7
6 February 2016	90	7	3
Total	345	25	13

Dung of eland and bontebok are significantly different in both δ^{13}C (U=52, $p<0.001$, Mann-Whitney test) and δ^{15}N (U=309, $p<0.0001$, Mann-Whitney test) (Table 2). Bontebok graze a fairly equal mixture of C4 and C3 plants to create a mean δ^{13}C value of -20.10‰, whereas eland are mostly eating C3 plants (Figure 2 and Table 2).

Table 2: Isotope analyses of dung beetles and dung from Potberg

	n	δ^{13}C	δ^{15}N
Epirinus flagellatus	8	-25.76 (0.54)	4.48 (1.24)
Scarabaeus spretus	9	-26.45 (0.37)	6.92 (0.97)
Bontebok dung	40	-20.1 (2.44)	2.6 (0.89)
Eland dung	60	-26.71 (1.33)	1.23 (0.78)

Values shown are mean (s.d.) in ‰.

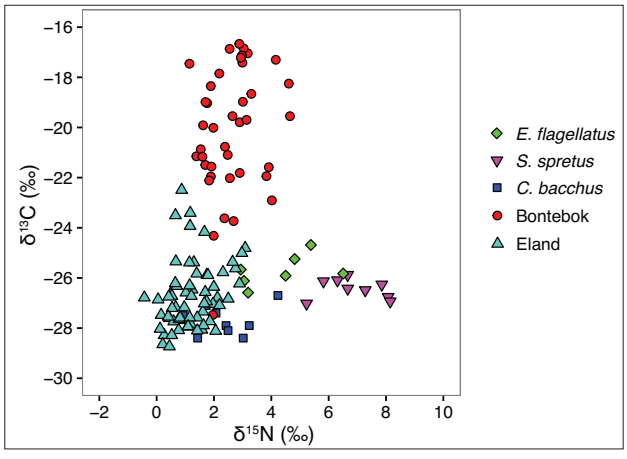


Figure 2: The distribution of isotopes (δ^{13}C and δ^{15}N) of the dung of bontebok and eland as well as of three dung beetle species.

Discussion

The increase in δ^{15}N as a result of metamorphosis is in the range of +3‰ and +5‰ for a selection of insects ranging from Diptera to Coleoptera to Lepidoptera[6] and the increase in δ^{13}C is about +1‰[5]. Dung beetle adults whose larvae fed on bontebok dung should thus have values of slightly more than -20‰ δ^{13}C but up to 7.6‰ δ^{15}N. The mean δ^{13}C values of both *E. flagellatus* and *S. spretus* clearly indicate a C3 dung diet and are thus much closer to that of the eland dung (Table 2 and Figure 2). The δ^{15}N values are 3.25‰ and 5.7‰ above eland dung but only 1.9‰ and 4.32‰ above bontebok dung. The evidence from δ^{15}N of the beetles is less equivocal about the larval food source because of the small 1‰ difference in dung between the two herbivore species. Overall the isotope results are compatible with eland being the main larval source of dung. Also bontebok dung is rare in the lowland fynbos habitat of *C. argenteum*. As these two dung beetle species are from different genera and are both attracted to *C. argenteum* seeds, these results indicate that the seed chemistry and deception by *C. argenteum* is not dung beetle species-specific. The deception appears most likely to be modelled on the chemistry of eland, rather than bontebok, dung.

Not much is known of the feeding biology of *Scarabaeus* or *Epirinus* beetles.[10] For both species, we observed diurnal activity and no pair formation at seed burial sites. *C. argenteum* seeds are about the same size as bontebok droppings and about half the size of eland droppings. That many seeds were buried per site for *S. spretus* suggests that several pellets of eland dung are typically used for feeding or egg laying, whereas *E. flagellatus* only buried a single seed per burial event. This distinctive burial behaviour of the two species likely results in differential recruitment patterns for *C. argenteum* seedlings. Single *E. flagellatus* burials would lead to lower intraspecific competition between seedlings compared with the multiple burials by *S. spretus*.

Many other dung beetles occur at Potberg, for example, the millipede-eating *Sceliages adamastor*.[10] The δ^{15}N dung beetle values presented in Table 1 provide a framework to interpret those of *S. adamastor* to determine whether this beetle is an obligate insectivore. For example, δ^{15}N values of an obligate millipede-eater should be a trophic level above herbivorous dung beetles such as *S. spretus* and *E. flagellatus* (i.e. they should have δ^{15}N values greater than 7‰). The very large Addo flightless dung beetle (*Circellium bacchus*) also occurs at Potberg[10]; elsewhere it feeds on elephant dung[10] but as there are no elephants at Potberg, its diet there is unknown and could too be clarified using the isotopic method. Being flightless, *C. bacchus* is often killed on roads and our analysis of nine roadkill individuals (mean δ^{13}C of -27.70‰ and δ^{15}N of 2.39‰) indicates eland dung is also its major larval food source.

The fact that *Ceratocaryum argenteum* is an element of deep sand fynbos,[11] implies that sufficient quantities of large herbivore dung, such as that of eland, occurred in this vegetation. This would maintain the associated dung beetle species and the deceptive relationship between *C. argenteum* and these species. There is some debate as to whether large herbivores were once more common in fynbos and in this area of the Cape.[3] Our observation that *C. bacchus*, *E. flagellatus* and *S. spretus* utilise eland dung suggests that the eland is, and has been, a key species in this system and should be carefully managed as such. Finally, we suggest that there are now sufficient examples of seed dispersal by beetles for use of the term coleopterochory. This term would include primary dispersal such as that described above, as well as examples of beetle endozoochory[12] in which small seeds are swallowed, as well as incidental or secondary dispersal in dung or with fruit[1].

Acknowledgements

We thank the National Research Foundation for Incentive Funding to J.J.M. and a PhD bursary to J.D.M.W.

Authors' contributions

Both authors participated in field work, analysis and writing.

References

1. Midgley JJ, White JDM, Johnson SD, Bronner GN. Faecal mimicry by seeds ensures dispersal by dung beetles. Nat Plants. 2015;1, Art. #15141. http://dx.doi.org/10.1038/nplants.2015.141
2. Skinner JD, Chimimba CT. Mammals of the southern African subregion. Cape Town: Cambridge University Press; 2005. http://dx.doi.org/10.1017/CBO9781107340992
3. Radloff FG, Mucina L, Snyman D. The impact of native large herbivores and fire on the vegetation dynamics in the Cape renosterveld shrublands of South Africa: Insights from a six-yr field experiment. Appl Veg Sci. 2014;17:456–469. http://dx.doi.org/10.1111/avsc.12086
4. Osborne CP, Salomaa A, Kluyver TA, Visser V, Kellogg EA , Morrone O, et al. A global database of C4 photosynthesis in grasses. New Phytol. 2014;204:441–446. http://dx.doi.org/10.1111/nph.12942
5. Peterson BJ, Fry B. Stable isotopes in ecosystem studies. Ann Rev Ecol Syst. 1987;18:293–320. http://dx.doi.org/10.1146/annurev.es.18.110187.001453
6. Tibbets TM, Wheeless LA, Martinez del Rio C. Isotopic enrichment without change in diet: An ontogenetic shift in δ15N during insect metamorphosis. Funct Ecol. 2008;22:109–113. http://dx.doi.org/10.1111/j.1365-2435.2007.01342.x
7. Stavert JR, Gaskett AC, Scott DJ, Beggs JR. Dung beetles in an avian dominated island ecosystem: Feeding and trophic ecology. Oecologia. 2014;176:259–271. http://dx.doi.org/10.1007/s00442-014-3001-z
8. Byrne MJ, Watkins B, Bouwer G. Do dung beetle larvae need microbial symbionts from their parents to feed on dung? Ecol Entomol. 2013;38:250–257. http://dx.doi.org/10.1111/een.12011
9. Picker M, Griffiths CL, Weaving A. A field guide to the insects of South Africa. Cape Town: Struik; 2004.
10. Davis ALV, Frolov AV, Scholtz CH. The African dung beetle genera. Pretoria: Protea Book House; 2008. p. 272.
11. Linder HP. Two new species of *Ceratocaryum* (Restionaceae). Kew Bull. 2001;56:465–477. http://dx.doi.org/10.2307/4110968
12. De Vega C, Montserrat Arista M, Ortiz PL, Herrera CM, Talavera S. Endozoochory by beetles: A novel seed dispersal mechanism. Ann Bot. 2011;107:629–637. http://dx.doi.org/10.1093/aob/mcr013

Note: This article is accompanied by supplementary material.

Anaerobic digestion of donkey dung for biogas production

AUTHORS:
Patrick Mukumba[1]
Golden Makaka[1]
Sampson Mamphweli[2]

AFFILIATIONS:
[1]Department of Physics, University of Fort Hare, Alice, South Africa
[2]Fort Hare Institute of Technology, University of Fort Hare, Alice, South Africa

CORRESPONDENCE TO:
Patrick Mukumba

EMAIL:
pmukumba@ufh.ac.za

POSTAL ADDRESS:
Department of Physics, University of Fort Hare, Private Bag X1314, Alice 5700, South Africa

DATES:
Received: 16 Jan. 2016
Revised: 24 Apr. 2016
Accepted: 30 May 2016

KEYWORDS:
substrate; hydrolysis; fermentation; acetogenesis; methanogenesis

HOW TO CITE:
Mukumba P, Makaka G, Mamphweli S. Anaerobic digestion of donkey dung for biogas production. S Afr J Sci. 2016;112(9/10), Art. #2016-0013, 4 pages. http://dx.doi.org/10.17159/sajs.2016/20160013

Biogas can provide a solution to some of South Africa's energy needs, especially in rural areas of the Eastern Cape province that have plentiful biogas substrates from donkeys, goats, sheep, cattle and chicken. We investigated the effectiveness of donkey dung for biogas production using a designed and constructed cylindrical field batch biogas digester. The donkey dung was collected from the University of Fort Hare's Honeydale Farm and was analysed for total solids, volatile solids, total alkalinity, calorific value, pH, chemical oxygen demand and ammonium nitrogen. The biogas composition was analysed using a gas analyser. We found that donkey dung produced biogas with an average methane yield of 55% without co-digesting it with other wastes. The results show that donkey dung is an effective substrate for biogas production.

Introduction

Anaerobic digestion is a biological process that naturally occurs when bacteria decompose organic matter, producing mainly methane (CH_4) and carbon dioxide (CO_2) in an oxygen-free environment.[1] The anaerobic digestion process is divided into four steps as follows: hydrolysis, fermentation (acidogenesis), anaerobic oxidation (acetogenesis) and methanisation.[2-4]

Hydrolysis

This is an enzyme-mediated stage, where insoluble organic compounds such as proteins, fats, lipids and carbohydrates are converted into soluble organic components, such as amino acids, fatty acids and monosaccharides.[5]

Fermentation

Acetate is the main end product of this step. Volatile fatty acids are also produced, as are carbon dioxide and hydrogen. Table 1 shows the major acids produced through fermentation processes in anaerobic digesters.[6]

Table 1: Major acids produced through fermentation processes in anaerobic digesters

Name	Formula
Acetate	CH_3COOH
Butanol	$CH_3(CH_2)_2CH_2OH$
Formate	$HCOOH$
Ethanol	CH_3CH_2OH
Lactate	$CH_3CHOHCOOH$
Methanol	CH_3OH
Propanol	$CH_3CH_2CH_2OH$
Succinate	$HOOCCH_2CH_2COOH$

Source: Based on information in Gerardi MH, 2003

Acetogenesis

In this step, the volatile acids are broken down into acetate and hydrogen.[7] This process is represented by equations 1 to 3:

$$CH_3CH_2COOH + 2H_2O \rightarrow CH_3COOH + CO_2 + 3H_2 \quad \text{Equation 1}$$

$$CH_3CH_2CH_2COOH + 2H_2O \rightarrow 2CH_3COOH + 2H_2 \quad \text{Equation 2}$$

$$2CO_2 + 4H_2 \rightarrow CH_3COOH + 2H_2O \quad \text{Equation 3}$$

Methanogenesis

In this step, acetate, formaldehyde, hydrogen and carbon dioxide are converted to methane and water. Table 2 shows substrates used by methane-forming bacteria.[6]

Table 2: Substrates used by methane-forming bacteria

Substrate	Chemical formula
Acetate	CH_3COOH
Carbon dioxide	CO_2
Carbon monoxide	CO
Formate	HCOOH
Hydrogen	H_2
Methanol	CH_3OH
Methylamine	CH_3NH_2

Source: Based on information in Gerardi MH, 2003[6]

Several parameters within an anaerobic digester play key roles in the physical environment and efficiency of digestion and biogas production potential. These parameters include pH-value, temperature, concentration of solids, hydraulic retention time, redox potential, volatile solids (VS) and loading rate, inocula, carbon–nitrogen ratio, toxicity, ammonium (NH_4), particle size, water content, agitation frequency, organic loading rate and volatile fatty acids.

All metabolic processes in bacteria are brought about by enzymes. There is a temperature range within which these microbes thrive. When the temperature is not favourable, the enzymes may be denatured, which hampers their digestion process. In this regard, bacteria are classified according to their preferred temperature. Psychrophilic bacteria work best between 10 °C and 20 °C, mesophilic bacteria between 20 °C and 35°C, and thermophilic bacteria between 45 °C and 60 °C.[2] Anaerobic digestion is very efficient in the thermophilic range. However, rural type digesters that lack external heating use mesophilic bacteria, as temperatures higher than 35 °C are very hard to obtain. For mesophilic bacteria, optimal digestion occurs at about 35 °C, whereas for thermophilic bacteria the optimum is 55 °C.[8]

Different substrates have different biogas yields. Swine manure has a better biogas output in terms of volume compared with cow dung, poultry manure, sheep manure, algae and night-soil. In addition, cow dung has a lower biogas yield than sheep manure.[9] We were unable to find data on the biogas yield of donkey dung. The aim of our study was therefore to investigate the effectiveness and performance characteristics of anaerobic digestion of donkey dung for biogas production, using a cylindrical field batch biogas digester.

Methodology

Source of substrate

Donkey dung was collected from Honeydale Farm, which belongs to the University of Fort Hare. Before water was added to the substrates, total solids (TS) of the prepared sample was determined to find out the amount of water to be added to the substrates before feeding into the batch digester. The most favourable TS value for biogas production is 8%.[10]

Determination of total solids

Total solids is the weight of dry material remaining after drying; it is also called dry weight. This is the portion of a substrate remaining after the elimination of moisture. Different samples of substrates were weighed using digital weighing scales. The weighed samples were placed in an oven at 105 °C for 24 hours. After heating the samples were reweighed, and TS was calculated using equation 4.[11]

$$TS(mg/L) = \frac{(W_1 - W_2)}{(W_3 - W_2)} \times 1{,}000{,}000 \qquad \text{Equation 4}$$

Where: W_1 = weight of dried residue and dish (g)

W_2 = weight of dish (g)

W_3 = weight of wet sample and dish (g)

Determination of volatile solids

The dried residue was weighed and heated in a crucible inside a furnace for 2 hours at 550 °C to 600 °C. The residue was cooled in a desiccator and was then weighed. The ignition, cooling, desiccating and weighing steps were repeated until the weight change was 50 mg. The final weight was recorded. Volatile matter was determined using equation 5.[11]

$$VS(mg/L) = \frac{(W_1 - W_2)}{(W_1 - W_3)} \times 1{,}000{,}000 \qquad \text{Equation 5}$$

Where: W_1 = weight of solids + weight of dish before ignition at 550°C (g)

W_2 = weight of solids + weight of dish after ignition at 550°C (g)

W_3 = weight of empty dish

Substrate parameters

The following parameters in substrate were determined: pH, chemical oxygen demand (COD), TS, VS, ammonia-nitrogen (NH_4-N), total alkalinity (T_A), and temperature (T). All the analytical determinations were performed according to the standard methods for examination of water and wastewater.[11] The slurry temperature was measured using the type-K thermocouples, and the digital pH meter measured the feed, digester slurry and effluent pH. Figure 1 is a photo of a Crison pH meter that was used to measure pH values. Figure 2 shows the calorimeter (CaL2K-ECO) that was used to measure caloric values of donkey dung (with diagrammatic captions added).

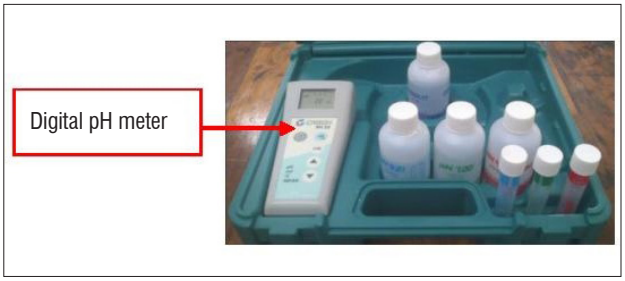


Figure 1: Digital pH meter

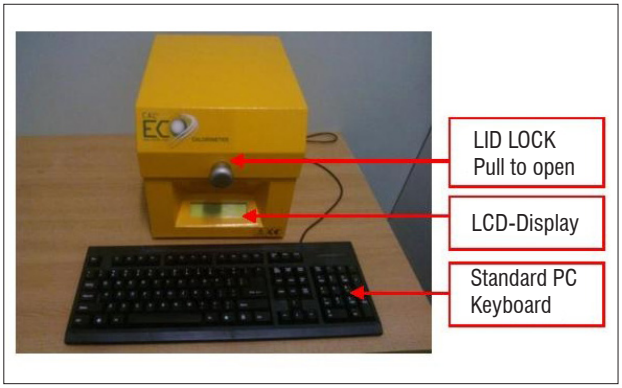


Figure 2: Calorimeter (CAL2K-ECO)

Biogas analysis

The data acquisition system consisted of a palladium–nickel (Pd/Ni) sensor and a non-dispersive infrared sensor. Biogas composition was analysed using a biogas analyser (a non-dispersive infrared sensor for

sensing methane and carbon dioxide, and a palladium–nickel sensor for sensing hydrogen and hydrogen sulphide).

Data on biogas composition were recorded by a CR1000 data logger at intervals of 2 minutes. The biogas analyser and CR1000 data logger were powered by a 12 V DC battery that was connected to a 20 W photovoltaic module. The slurry and ambient temperatures were measured using type-K thermocouples connected to the same CR1000 data logger as the biogas sensors. The data logger was interfaced to a computer. Figure 3 shows the complete data acquisition system with biogas and temperature sensors.

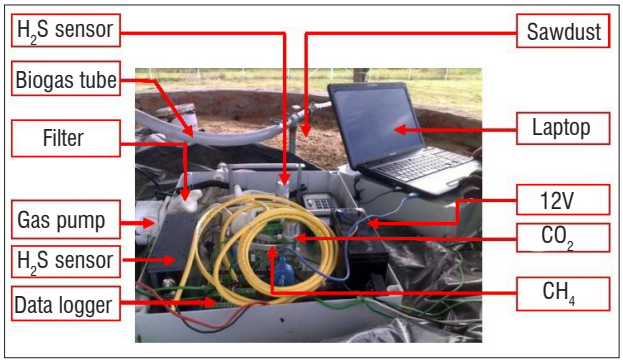


Figure 3: Data acquisition system

Results and discussion

Table 3 shows characteristics of donkey dung. The higher concentrations of ammonium nitrogen between 940 mg/L and 1223 mg/L were beneficial to the anaerobic digestion micro-organisms, and this enhanced biogas production.

Table 3: Substrate characteristics for donkey dung

Parameter	Unit	Donkey dung
Total solids	mg/L	198778.83
Volatile solids	mg/L	144189.99
Total alkalinity	mg/L	6276–6343
Ammonium-nitrogen	mg/L	940–1223
Calorific value	MJ/g	29.83
Volatile solid /total solid %	%	72.54

Figure 4 shows biogas production for donkey dung. We observed that biogas production was low at the beginning and end periods of the anaerobic digestion process. This is because biogas production rate is directly proportional to the specific growth rate of methanogens.

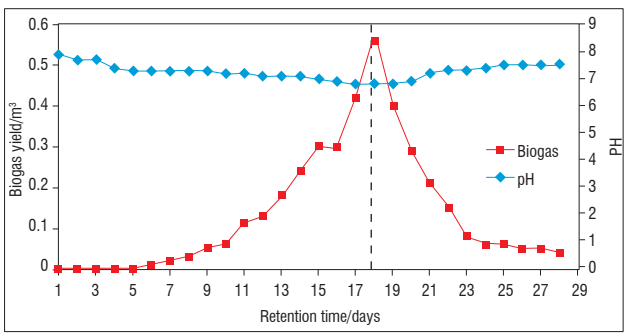


Figure 4: Biogas yield and pH values for donkey dung for a retention time of 28 days

With reference to Figure 4, the donkey dung biogas production peaked at 0.56 m^3 on Day 18. The total biogas yield of donkey dung was 3.8 m^3 over a retention time of 28 days. The maximum biogas yield for donkey dung (63.4% or 2.41 m^3) was attained in the first 18 days of the batch experiment, before the biogas yield declined. However, the biogas yield from Day 18 to Day 28 was 36.6% (1.39 m^3) of the total biogas production. The biogas production (Y) of donkey dung from Day 4 to Day 18 is represented by the following equation:

$$Y = 0.02t^2 - 0.06t + 0.047 \quad \text{for } 4 \leq t \leq 18 \quad \text{Equation 6}$$

The biogas yield started to decline from Day 18 to Day 28. The decay equation is given by:

$$Y = 65.039e^{-0.27t} \quad \text{for } 18 \leq t \leq 28 \quad \text{Equation 7}$$

As shown in Figure 4, the initial pH of donkey dung was 7.9. The pH was observed to decline with time, attaining a minimum value of 6.8 where an optimum biogas production of 0.56 m^3 was achieved. The decline is due to the conversion of substrate to acids during the stages of acidogenesis and acetogenesis of methane production. As from Day 18, the pH began to increase as the acids produced were converted to methane by methanogens. The pH (R) equation for the best fit is represented as follows:

$$R = 0.041t^2 - 0.1278t + 7.9795 \quad \text{Equation 8}$$

Where R represents pH values and t represents retention time.

Figure 5 shows COD values during the anaerobic digestion of donkey dung. As shown in Figure 5, concentration COD of donkey dung entering the batch was 40 110 mg/L. The higher biogas yield from donkey dung was attributed to this higher concentration of COD. The concentration of COD decreased from 40 110 mg/L to 25 110 mg/L because of the conversion of substrate into biogas. Optimum biogas production occurred at a COD profile of 26 436 mg/L. The concentration of COD declined as of Day 20 and by Day 28 was very low, indicating a low biogas yield (a higher COD destruction means a high biogas yield). The COD (S) best fit can be approximated by the following equation:

$$S = 3.4277t^3 - 139.09t^2 + 737.13t + 39.3601 \quad \text{Equation 9}$$

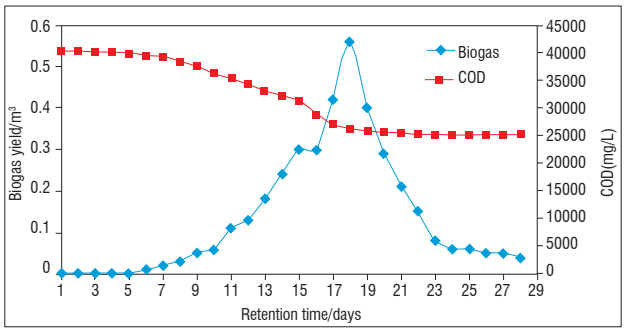


Figure 5: Relationship between gas production and COD for donkey dung

The relationship between biogas production and total alkalinity (T_A) is shown in Figure 6. The effluent total alkalinity for donkey dung ranged from 6235 mg/L to 6343 mg/L. Therefore, alkalinity values were within the normal range, above the threshold alkalinity of 500 mg/L suggested for anaerobic digestion.

The high alkalinity levels imply that there was a high buffering capability that increased the biogas yield. The higher the alkalinity, the greater the buffering capacity in the anaerobic digestion process; this in turn promoted a stable pH value and resulted in an increased biogas yield.

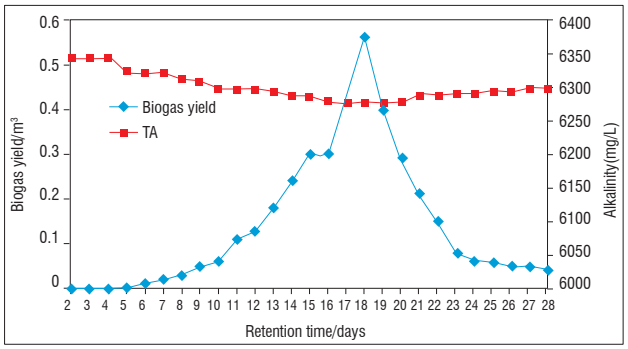


Figure 6: Relationship between biogas yield and total alkalinity for donkey dung

The relationship between biogas yield and NH_4-N for donkey dung is shown in Figure 7. The NH_4-N concentrations for donkey dung ranged from 920 mg/L to 1223 mg/L. Therefore, donkey dung had appropriate NH_4-N levels to maintain stable anaerobic digestion performance. The higher NH_4-N final values corresponded with lower biogas production. The lowest NH_4-N concentration was 920 mg/L. We observed that the lowest NH_4-N concentrations corresponded to the highest biogas yield, and no inhibition occurred because the NH_4-N concentrations fell within the desired optimum range (below 1500 mg/L) for better biogas production by the methanogens. Previous studies have reported that inhibiting concentrations of NH_4-N are higher than 1500 mg/L.[12-14]

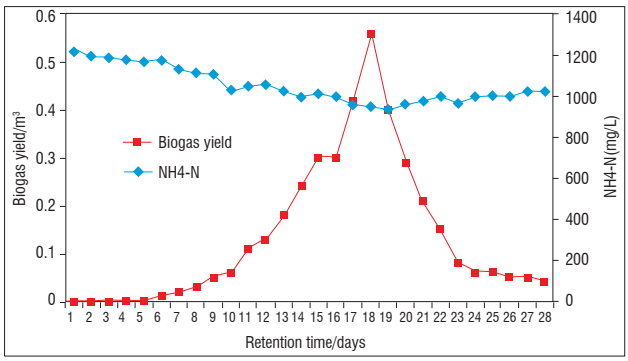


Figure 7: Relationship between biogas yield and NH4-N for donkey dung

Table 4 shows the biogas composition for donkey dung. The biogas composition of donkey dung consisted of methane, carbon dioxide and hydrogen. The composition of hydrogen and other gases in the biogas was 5%, whereas the composition of carbon dioxide was 40%. In addition, the methane content in the biogas was 55%. We observed that biogas from donkey dung did not contain carbon monoxide.

Table 4: Biogas composition of donkey dung

Gases	Composition (%)
Methane (CH_4)	55
Carbon dioxide (CO_2)	40
Carbon monoxide (CO)	0
Hydrogen (H_2) and other gases	5

The concentration of COD entering the digester for donkey dung was high, and this contributed to a higher methane yield of 55% in donkey dung. Donkey dung had high values of CV, VS and COD and hence high methane yield.

Conclusion

We investigated the potency of donkey dung for biogas production. We analysed, and here present, the performance characteristics of anaerobic digestion in a cylindrical field batch biogas digester. The results showed that the biogas from donkey dung has an average methane yield of 55%. In addition, we observed that biogas production from donkey dung was low at the beginning and end of the retention period of the anaerobic digestion process, because biogas production rate is directly proportional to the specific growth rate of methanogens. From the results of the study, we conclude that donkey dung is a potential substrate in biogas digesters for rural digesters in the Eastern Cape Province of South Africa. This conclusion is based on the methane yield in donkey dung.

References

1. Arsova L. Anaerobic digestion of food waste: Current status, problems and an alternative product. [MSc thesis], USA: Columbia University; 2010.
2. Davidsson Å. Increase of biogas production at waste water treatment plant: Addition of urban organic waste and pre-treatment of sludge. [PhD thesis]. Water and Environmental Engineering, Department of Chemical Engineering, Sweden: Lund University; 2007.
3. Leksell N. Present and future digestion capacity of Käppala waste treatment plant: A study in a laboratory scale. Sweden; 2005.
4. Polprasert C. Organic waste recycling – technology and management. London: IWA, Publishers; 2007.
5. Yadvika SS, Sreekrishman TR, Kohli S, Rana V. Enhancement of biogas production from solid substances using different techniques: A Review. Bioresour Technol. 2004;(95):1–10. http://dx.doi.org/10.1016/j.biortech.2004.02.010
6. Gerardi MH. The microbiology of anaerobic digester. New Jersey: John Wiley & Sons; 2003. http://dx.doi.org/10.1002/0471468967
7. Wiese J, König R. Monitoring of digesters in biogas plants: Laboratory analysis and process analysis: Biogas plant monitoring. Germany: EnerCess GmbH & Hach Lange GmbH; 2009.
8. Marchaim U. Biogas processes for sustainable development. FAO Agricultural Services. Bulletin 95; 1992.
9. Nijaguna BT. Biogas technology. New Delhi: New Age International (P) Limited, Publishers; 2002.
10. Ituen EE, John NM, Bassey BE. Biogas production from organic waste in Akwa Ibom State of Nigeria. In: Yanful EK, editor. Appropriate technology for environmental protection in the developing world. Selected papers from ERTEP 2007, July 17–19 2007, Ghana, Africa. [E-book publication] Springer Science & Business Media; 2009. p. 93–99
11. American Public Health Association (APHA). Standard methods for examination of water and waste water. Washington DC: WEF; 2005.
12. Calli B, Mertoglu B, Inac B, Yenigun O. Effects of free ammonia concentrations on the performance of anaerobic bioreactors. Process Biochem. 2005;(40):1285–1292. http://dx.doi.org/10.1016/j.procbio.2004.05.008
13. Chen Y, Cheng JJ, Creamer KS. Inhibition of anaerobic digestion process: A review. Bioresour Technol. 2008;(9):4044–4064.
14. Fricke K, Santen H, Wallman R, Hütter A, Dichtl N. Operating problems in anaerobic digestion plants resulting from nitrogen in MSW. Waste Managt. 2007;(27):30–43. http://dx.doi.org/10.1016/j.wasman.2006.03.003