The quality of the early hominin fossil record : implications for evolutionary analyses

THE QUALITY OF THE EARLY HOMININ FOSSIL RECORD

IMPLICATIONS FOR EVOLUTIONARY ANALYSES

Simon Joseph Maxwell

Thesis submitted to the
UCL-Birkbeck Institute of Earth and Planetary Sciences
for the degree of Doctor of Philosophy

2018

This thesis is dedicated to the memory of Sarah Delf.

Declaration

Declaration of originality

I, Simon Joseph Maxwell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Copyright declaration

Doctoral committee

Dr. Philip J. Hopley
Department of Earth and Planetary Sciences
Birkbeck, University of London

Professor Paul Upchurch
Department of Earth Sciences
University College London

Dr. Christophe Soligo
Department of Anthropology
University College London

Thesis examiners

Professor Fred Spoor
Department of Earth Sciences
Natural History Museum

Professor Roger Benson
Department of Earth Sciences
University of Oxford

Funding

Supported in full by the Natural Environment Research Council [NE/L002485/1].

Acknowledgements

There are quite a few people without whom this project would never have become a reality, and quite a few more who just made the journey a whole load of fun. My biggest thanks go to my three supervisors, Phil Hopley (Earth and Planetary Sciences, Birkbeck), Paul Upchurch (Earth Sciences, UCL), and Christophe Soligo (Anthropology, UCL), for their support, guidance, and enthusiasm from start to finish. I also thank Christophe for facilitating my teaching at UCL.

Many individuals replied to my incessant, and very wanting, emails regarding hominin fossil material and phylogenetic data. Their kind replies helped fill numerous gaps in the early hominin fossil database I compiled during my first year, and this formed the primary resource for much of my work. In order of correspondence, I would like to thank Scott Simpson, Tim White, Carol Ward, John Harris, Yohannes Haile-Selassie, Frederick Manthi, Meave Leakey, Lucas Delezene, Bernard Wood, Frederick Grine, Darryl de Ruiter, Chris Stringer, John Hawkes, Mana Dembo, and David Strait for their valuable, and in some cases, unpublished information[*]. I would especially like to thank Bernard Wood for showing interest in my work, and providing advice and encouragement on multiple occasions during my PhD. For their help writing vital R code I thank Juan Cantalapiedra[†] and Graeme Lloyd[‡]. My biggest thanks, however, go to those I didn't have to ask—those that made their data and R code freely available.

I would especially like to thank my family for their support. In particular, my ma and pa for their financial aid and for offering safe refuge to write. I would also like to thank my friends for their encouragement and oft-needed escapism: there's nothing better than a trip to Yorkshire or the Lake District to break the monotony.

Starting this journey as part of the first (2014) cohort of the London NERC DTP made it a shared experience unlike any other. So for that I thank the London NERC DTP committee and managerial boards.

Second to lastly, a huge thanks go to the folks—past and present—of Room 611 (Earth and Planetary Sciences, Birkbeck). Sorry for finishing *in absentia* but I wouldn't have it any other way.

Lastly, I would like to thank Charlie Underwood (Earth and Planetary Sciences, Birkbeck) for his feedback during my upgrade viva, the six anonymous reviewers, one associate editor, and two editors who read various parts of this thesis and whose helpful comments greatly improved it, and my two examiners, Fred Spoor (Natural History Museum) and Roger Benson (University of Oxford).

Done.

[*] Email correspondence 02.04.2015–04.01.2016.

[†] The Use of Phylogenies in the Study of Macroevolution (Transmitting Science), September 19–23 2016, Bercelona, Spain.

[‡] Tools and Methods for Constructing the Tree of Life (University of York), December 19–20 2016, York, UK.

Abstract

Hominins (the clade including modern humans and their fossil ancestors) were a taxonomically and morphologically diverse group during the Plio-Pleistocene, and their evolution documents the only known transition to obligate bipedalism in primates. However, many aspects of their shared evolutionary history remain frustratingly unclear due to uncertainty about whether change in the fossil record reflects genuine evolutionary change or variation in our sampling of the rock record. Here, a comprehensive assessment of the quality of the early African hominin fossil record is presented. A specimen database of all early African hominin fossils (>5000) has been compiled including taxonomic, geological, anatomical, and bibliographic information. Using a range of sampling metrics (fossil-bearing formations, collection effort, sampled area, and ghost lineage diversity), it is shown that the pulsed-like pattern of uncorrected (taxic) hominin diversity is almost entirely controlled by rock availability. By contrasting taxic with phylogenetically corrected diversity, hominin diversification appears unconstrained through the late Miocene and Pliocene, with diversity constantly increasing until a single peak is reached in the early Pleistocene. Phylogenetically corrected diversity shows no discernible link with sampling metrics and there is no direct evidence that shifts in climatic conditions drove diversification. A study of specimen completeness through geological time shows that while sampling metrics (specifically sustained collection effort at rich deposits) have a major influence on patterns of specimen completeness, specimen completeness has only a moderate influence on diversity patterns. It also shows that specimen completeness is poorest during the period most pertinent to human origins, the estimated *Pan-Homo* divergence date, in large part due to under-sampling (<4% of Africa by sampled area). In combination, this work illustrates that the hominin fossil record is by no means an unbiased depiction of evolutionary events, and therefore its quality and incompleteness should be fully understood before any interpretation of macroevolutionary patterns.

Contents

List of Figures

List of Tables

List of Abbreviations

AICc	Second-order Akaike Information Criterion
cc	cubic centimetres (cm^3)
CC	Common-cause hypothesis
CCM	Character Completeness Metric
CI	Confidence interval
EARS	East African Rift System
FAD	First Appearance Datum
FDR	False Discovery Rate
Fm.	(geological) formation
GD	Generalised differencing
GDE	Ghost lineage diversity estimate
HBC	Hominin-bearing collections
HBF	Hominin-bearing formations
HBL	Hominin-bearing localities
km	kilometres
LAD	Last Appearance Datum
LVI	Lake variability index
Ma	Mega-annum or millions of years ago (refers to a date)
MBF	Mammal-bearing formations
mm	millimetres
Myr	million years (refers to an amount of time)
NOO	Number of opportunities to observe
OTU	Operational taxonomic unit
p	Significance value
PBF	Primate-bearing formations
PDE	Phylogenetic diversity estimate
PIU	Palaeontological Interest Unit
r	Pearson's product-moment correlation coefficient
RED	Redundancy hypothesis
RRB	Rock record bias hypothesis
R^2	Coefficient of determination
SCM	Skeletal Completeness Metric
TDE	Taxic diversity estimate
w_i	Akaike weight
ρ	Spearman's rank correlation coefficient
τ	Kendall tau rank correlation coefficient

List of Hominin Fossil Abbreviations

AL	Afar Locality, Hadar, Ethiopia
ALA-VP	Alayla-Vertebrate Paleontology, Middle Awash, Ethiopia
ARA-VP	Aramis-Vertebrate Paleontology, Middle Awash, Ethiopia
ABD	Asbole Dora, Gona, Ethiopia
ASI-VP	Asa Issie-Vertebrate Paleontology, Middle Awash, Ethiopia
BAR	prefix for the fossils recovered at Lukeino, Tugen Hills, Baringo District, Kenya (2000–)
BEL-VP	Belohdelie-Vertebrate Paleontology, Woranso-Mille, Ethiopia
BOU-VP	Bouri-Vertebrate Paleontology, Middle Awash, Ethiopia
BRT-VP	Burtele-Vertebrate Paleontology, Woranso-Mille, Ethiopia
BSN	Busidima Formation, Afar, Ethiopia
DNH	Drimolen hominin, Drimolen, South Africa
FJ	Fejej, Ethiopia
KNM-BC	Kenya National Museums followed by the site code for the Baringo District, Kenya, for fossils recovered from the Chemeron Formation
KNM-ER	Kenya National Museums followed by the site code for East Rudolf (now referred to as Koobi Fora), Kenya
KNM-KP	Kenya National Museums followed by the site code for Kanapoi, Kenya
KNM-LT	Kenya National Museums followed by the site code for Lothagam, Kenya
KNM-TH	Kenya National Museums followed by the site code for Tugen Hills, Kenya (pre-2000)
KNM-WT	Kenya National Museums followed by the site code for West Turkana, Kenya
KT	Koro Toro, Chad
L	prefix for fossils collected by the American-led group from the Shungura Formation, Omo Basin, Ethiopia
LH	Laetoli hominid, Tanzania
LD	Ledi-Geraru, Ethiopia
MH	Malapa hominin, South Africa
MLD	Makapansgat Limeworks Deposits, South Africa
OH	Olduvai hominid, Tanzania
Omo	prefix for fossils collected by the French-led group from the Shungura Formation, Omo Basin, Ethiopia
SK	prefix for fossils recovered from Swartkrans, South Africa (1948–1953)
SKW	prefix for fossils recovered from Swartkrans, South Africa (1968–1979)
SKX	prefix for fossils recovered from Swartkrans, South Africa (1979–1986)
Sts	fossil hominins recovered from the Sterkfontein type site, South Africa (1947–1949)
StW	Sterkfontein Witwatersrand, South Africa
Taung	Taung Child Type Site, Buxton-Norlim Limeworks, Taung, South Africa
TM	Transvaal Museum, South Africa
TM	Toros-Menalla, Chad
UR	Uraha, Malawi

Chapter 1

General introduction

1.1 Studying hominin origins

A major issue relating to the origin and evolution of hominins (the clade including modern humans and their fossil ancestors) concerns the quality of their fossil record (Maxwell *et al.*, 2018, *in preparation*). The quality of the hominin fossil record directly informs our ability to accurately determine their time of origin, and to determine whether patterns of speciation, extinction, and diversity in the fossil record are a genuine depiction of their evolutionary history or an artefact of our incomplete sampling of the rock record (Benton, 2003).

The timing of hominin origins can be constrained by (1) the *Pan-Homo* (chimpanzee-modern human) divergence date based on molecular data, and (2) the date of the earliest hominin fossil. The first date represents a maximum estimate and the second date represents a minimum estimate for the origin of the hominin clade (Soligo *et al.*, 2007). Molecular estimates suggest a divergence date of 8.0–4.0 million years ago (Ma) for hominins and panins (e.g., Glazko & Nei, 2003; Patterson *et al.*, 2006; Steiper & Young, 2006, 2009; Stone *et al.*, 2010; Perelman *et al.*, 2011; Langergraber *et al.*, 2012; Prüfer *et al.*, 2012; Scally *et al.*, 2012; Springer *et al.*, 2012; Pozzi *et al.*, 2014), indicating that morphologically diagnosable hominin characteristics started to accumulate during the late Miocene and early Pliocene. However, the minimum (earliest fossil) estimates for the origin of hominins range from 7.0–4.4 Ma depending on the taxonomic composition of the clade (Simpson, 2013). While the earliest undisputed hominin is known from 4.2-million year old (Myr) deposits (Leakey *et al.*, 1995, 1998), three highly fragmentary, possible hominin genera pre-date this taxon (Senut *et al.*, 2001; Brunet *et al.*, 2002; White *et al.*, 2009), however, the case for each of them being a hominin is weak (Wood & Boyle, 2016, 2017).

The second issue concerns the quality of the early African fossil record and its influence on evolutionary pattern recognition and evolutionary process inference. The fossil record of any organism is an incomplete representation of evolutionary history (Raup, 1972, 1976*b*; Martin, 1993; Benton *et al.*, 2011). This follows naturally from the fact that not all members of a taxon are fossilised and ultimately sampled. Indeed, while reflecting on this concern, Darwin once lamented: "I look at the natural geological record as history of the world imperfectly kept, and written in a changing dialect; of this history we possess the last volume alone, relating only to two or three countries. Of this volume, only here and there a short chapter has been preserved, and of each page, only here and there a few lines…" (Darwin, 1859:310–311).

However, concerns about the poor quality of the hominin fossil record and the sporadic nature of our sampling of the rock record have remained untested in palaeoanthropology, being of apparently little interest to the study of hominin macroevolution despite the potential to mislead meaningful interpretations (Smith & Wood, 2017). A long-standing assumption based on a literal (face-value) reading of the fossil record is that all major events in hominin evolution were a direct result of Plio-Pleistocene climatic change and variability (e.g., Dart, 1925; Vrba, 1988; deMenocal, 1995, 2004; Potts, 1996, 1998; Kingston, 2007; Maslin & Trauth, 2009; Grove, 2011, 2012; Shultz & Maslin, 2013). However, the finding that poor fossil record quality severely limits the degree to which climate shifts and major evolutionary events in hominin evolution can be meaningfully compared, questions such an approach (Hopley, 2018, *in review*).

This thesis addresses some key issues in hominin palaeobiology and macroevolution related to the quality of the early African hominin fossil record (Maxwell *et al.*, 2018).

1.2 Introduction to the Hominini

Hominins are members of the tribe Hominini, the clade including modern humans (*Homo sapiens*) and all taxa more closely related to them than to modern chimpanzees and bonobos (*Pan*), which are members of the tribe Panini or panins (Wood & Harrison, 2011). Hominins belong to the superfamily Hominoidea (ape and human clade), family Hominidae (great ape and human clade), and subfamily Homininae (African great ape and human clade; Fig. 1.1). Molecular estimates for the hominin-panin last common ancestor generally agree on a late Miocene to early Pliocene divergence date of 8.0–4.0 million years ago (Ma) (e.g., Patterson *et al.*, 2006; Perelman *et al.*, 2011; Springer *et al.*, 2012; Pozzi *et al.*, 2014), and this is supported by statistical analyses of the primate fossil record that explicitly incorporate gaps in fossil sampling during divergence date estimation (e.g., Wilkinson *et al.*, 2011). However, a more precise divergence date within this time window is currently unclear. A taxonomy of Hominoidea can be found in Appendix A (Wood & Harrison, 2011), however, it must be stressed that the *Pan* + *Homo* clade has no distinct taxonomic nomen in this taxonomy. A composite cladogram is shown in Fig. 1.2 (Strait *et al.*, 2007). In this cladogram the earliest hominins are collapsed into a polytomy because they are all yet to be included in a single cladistic analysis due to specimen incompleteness (Strait *et al.*, 2007). The gracile australopiths (= *Australopithecus*) then branch off the tree in succession (= in a pectinate fashion), and it is clear that they are not a natural (= monophyletic) group insofar as they are not all more closely related to each other than they are to other hominins (Strait, 2010). The robust australopiths (= *Paranthropus*) appear to be more closely related to *Homo* (with support for both being monophyletic reasonably strong), although their relationship to *Kenyanthropus* is currently unclear. Many aspects of hominin systematics are contested (see Strait *et al.*, 2007 for a review), and the cladogram depicted in Fig. 1.2 is a phylogenetic hypothesis that specifically emphasises uncertainty.

Hominins appear to occupy a search-intensive terrestrial feeding niche similar to that of an Old World monkey (e.g., *Papio*) (White *et al.*, 2015). They represented a morphologically and ecologically disparate group throughout the Plio-Pleistocene and their evolution includes the transition to obligate bipedalism, reduced sexual dimorphism, grossly enlarged brains, extended life history, increased carnivory, and tool manufacture and use (Fleagle, 2013). Hominini is currently (according to a highly

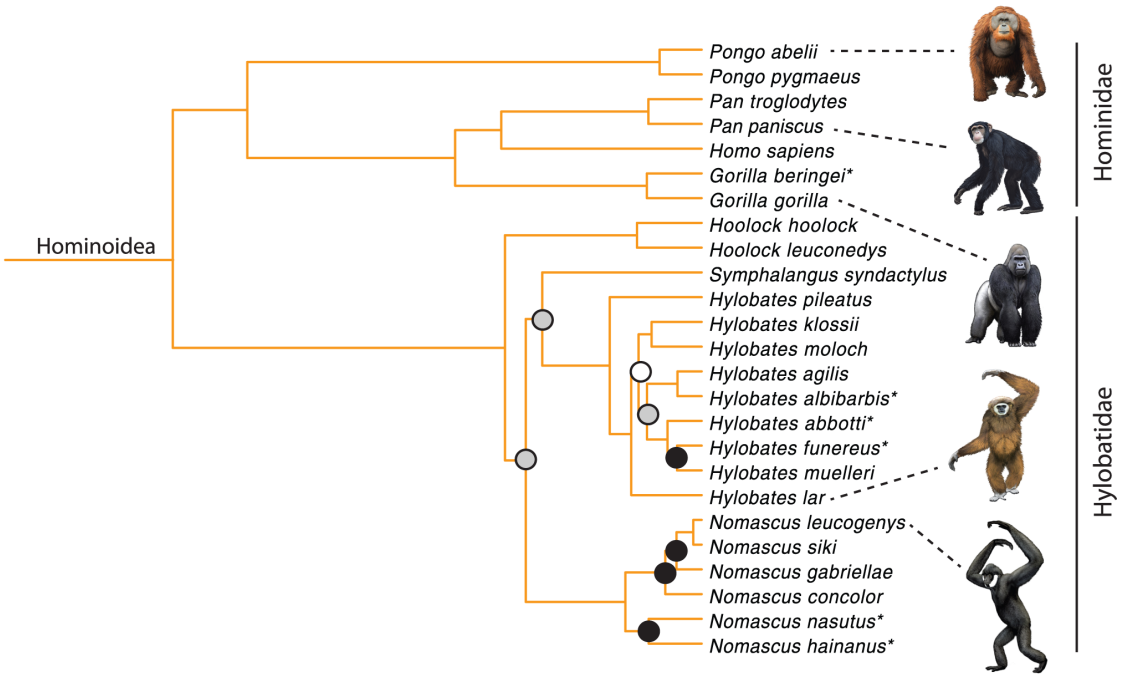

Figure 1.1: Molecular phylogeny of Hominoidea. Bootstrap support for each node is shown by no circle (>95%), black circle (70%–95%), grey circle (50%–70%), or white circle (<50%). Hominini are shown within Hominidae and represented by the branch leading to modern humans (*Homo sapiens*) after their divergence from modern chimpanzees (*Pan*). Modified from Springer *et al.* (2012).

speciose [splitting] interpretation) composed of at least 22 species (e.g., Wood & Richmond, 2000; Wood & Lonergan, 2008; Wood & Boyle, 2016) (Fig. 1.3). In contrast, according to a less speciose (lumping) taxonomy, Hominini is composed of 8 species (e.g., White, 2003). Because of the ambiguity and debate surrounding the phylogenetic placement of many fossil hominins, it is far more common to see an informal taxonomy in which hominins are grouped by evolutionary *grade* (Huxley, 1958, 1959)—and not by *clade*—into possible and probable early hominins, archaic hominins, megadont archaic hominins, transitional hominins, pre-modern humans, and anatomically modern humans (Wood & Lonergan, 2008). While the hominins within each grade may not cluster phylogenetically, they are united by a similar ecological situation, or adaptive strategy.

The earliest undisputed hominin *Australopithecus anamensis* is known from 4.2–3.9-Myr deposits in Kenya (Leakey *et al.*, 1995, 1998) and Ethiopia (White *et al.*, 2006) (Fig. 1.4). However, there are currently three generally recognised but disputed genera that pre-date the australopiths and which are known from 7.0–4.4 Ma (Simpson, 2013). In this section, the defining characteristics of Hominini are outlined and the earliest fossil evidence for hominin evolution is described (in roughly chronological order), with specific reference to their morphological similarity and dissimilarity to *Gorilla*, *Pan*, and other fossil hominins.

1.2.1 Defining Hominini and distinguishing human ancestors

The assignment of a fossil specimen or taxon to the tribe Hominini is contingent upon that specimen or taxon possessing a precise set of unique characteristics (synapomorphies) shared by hominins

and no other clade. So, in order to confidently distinguish a human ancestor from a closely related non-hominin taxon, the key question to ask is what unique characteristics constitute specialisations of the hominin lineage after its divergence from the panin lineage (Andrews & Harrison, 2005). Such characteristics need to be present in modern humans but demonstrably different from those of fossil hominoids, and distinct from characteristics present in *Pan* and fossil panins (if they are known). Candidate synapomorphies include facial shortening, encephalisation, smaller and more vertically implanted incisors, reduction in the size and degree of sexual dimorphism of the canines, modification of the lower third premolar associated with a reduced honing function of the upper canine, thick enamel, postcanine megadontia, and also specialised features of the trunk, hip, knee, and foot associated with adaptations to upright posture (orthogrady) and terrestrial bipedalism (Wood & Harrison, 2011:350). However, encephalisation, thick enamel, and postcanine megadontia can be discounted immediately. First, the dramatic increase in absolute brain and postcanine tooth size both occurred much later in hominin evolution (shortly before or after 3.0 Ma) and are, therefore, information redundant with respect to the unique characteristics which distinguish the earliest hominins from the earliest panins or any other non-hominin clade. Second, thick enamel on the cheek teeth also occurs in other late Miocene fossil hominoids (e.g., *Griphopithecus*, *Kenyapithecus*, and *Sivapithecus*), presumably a result of parallel shifts in dietary behaviour in response to changing ecological conditions (Begun, 2007; Wood & Harrison, 2011).

The remaining characteristics used to distinguish a stem hominin from a stem panin, a stem hominine, or closely related hominid include:

1. ***Canine reduction and loss of the C/P_3 honing complex.*** In modern and fossil catarrhines (hominoids + cercopithecoids), a triangular, projecting upper canine is continuously sharpened by occlusion against the lower third premolar. In contrast, hominins are characterised by a shift to apically-dominated canine wear, suggesting a limited role for the canine in social organisation and reduced male-male competition (Fleagle, 2013). It is important to recognise, however, that a number of late Miocene Eurasian hominoids (e.g., *Oreopithecus*, *Ouranopithecus* and *Gigantopithecus*) also display canine reduction in conjunction with the partial loss of the C/P_3 honing complex (Wood & Harrison, 2011). Canine reduction can also been seen in bonobos, *Pan paniscus* (Kelley, 2001).

2. ***Position and orientation of the foramen magnum.*** Modern humans display a foramen magnum that is more anteriorly positioned than any other primate (Russo & Kirk, 2013), highlighting a major reorganisation of the basicranium. It is commonly associated with bipedalism and, therefore, routinely accepted as a defining characteristic of Hominini (e.g., Brunet *et al.*, 2002; Zollikofer *et al.*, 2005; Suwa *et al.*, 2009*a*; Kimbel *et al.*, 2014). However, a more anteriorly positioned and horizontally oriented foramen magnum is broadly associated with head carriage, facial length, and brain size in hylobatids, short-faced monkeys (e.g., *Saimiri*), and indriids (Lieberman *et al.*, 2000; Strait, 2001; Ruth *et al.*, 2016), rather than uniquely with an upright posture and terrestrial bipedalism. Moreover, the basicranium is unknown in many late Miocene fossil hominoids, meaning the polarity of this character transformation is unclear.

3. ***Modifications of the trunk, hip, knee, and foot associated with obligate bipedalism.*** Bipedalism

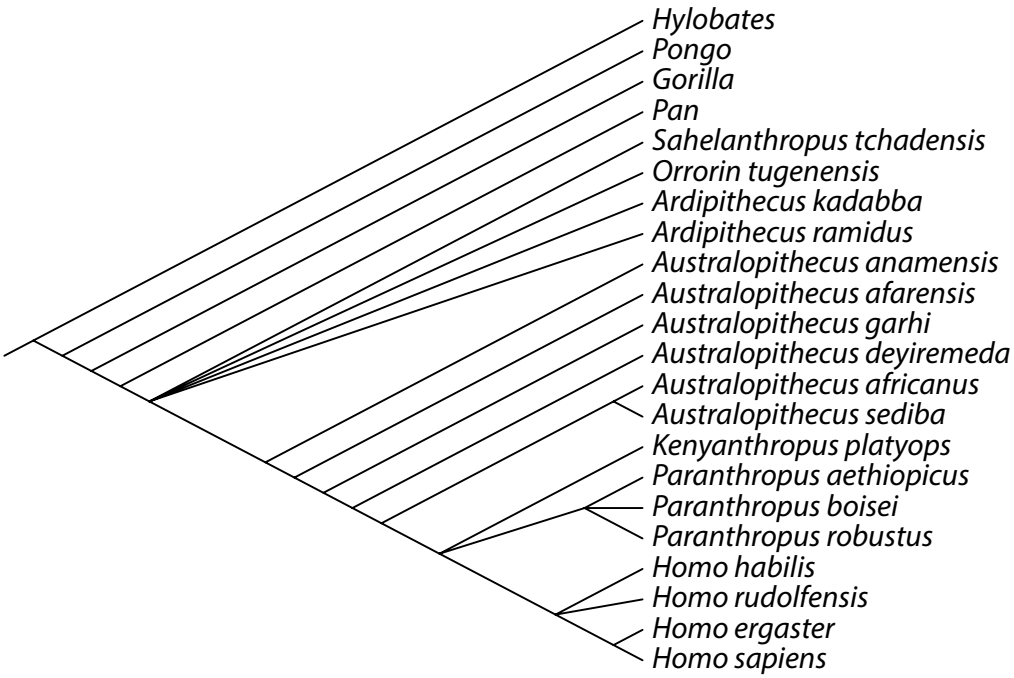

Figure 1.2: Composite cladogram of Hominini within the superfamily Hominoidea. The cladogram is reproduced from Strait & Grine (2004; Fig. 10), with the earliest purported hominins collapsed into a polytomy to reflect phylogenetic uncertainty, and *Australopithecus sediba* placed as a sister-taxon to *Australopithecus africanus* (Irish *et al.*, 2013; Prang, 2016; Kimbel & Rak, 2017).

is a highly specialised and unusual form of locomotion that is found today in only one primate: modern humans (Fleagle, 2013). Fortunately, anatomical specialisations for obligate bipedalism require modification to multiple parts of the musculo-skeletal system, all of which relate to maintaining balance and the trunk's center of gravity as close to the middle of the body as possible (Harcourt-Smith, 2007). These include a shorter and broader ilium, an infero-superiorly short pubic symphysis, a well-developed anterior inferior iliac spine, a large ischial spine, a discrete greater sciatic notch, a long femoral neck with the greater trochanter low in relation to the superior border of the neck, medial condyle of the distal femur similar in size to the lateral condyle, femora that converge distally (valgus knee), the presence of a bicondylar angle, a rigid mid-foot with a longitudinal and transverse arch, enlargement of the calcaneal tuberosity, and an adducted hallux (Richmond *et al.*, 2001; Harcourt-Smith & Aiello, 2004; Crompton *et al.*, 2008). While there is general agreement that terrestrial bipedalism is a synapomorphy of Hominini, the precise anatomical characteristics that signature this locomotor pattern are debated (e.g., Ruth *et al.*, 2016; Russo & Kirk, 2017).

It is this suite of unique characteristics that forms the basis for recognising human ancestors in palaeoanthropology. Therefore, in order to confidently identify a stem hominin in the fossil record it is necessary that at least one of these phylogenetically diagnostic skeletal regions is known and of suf-

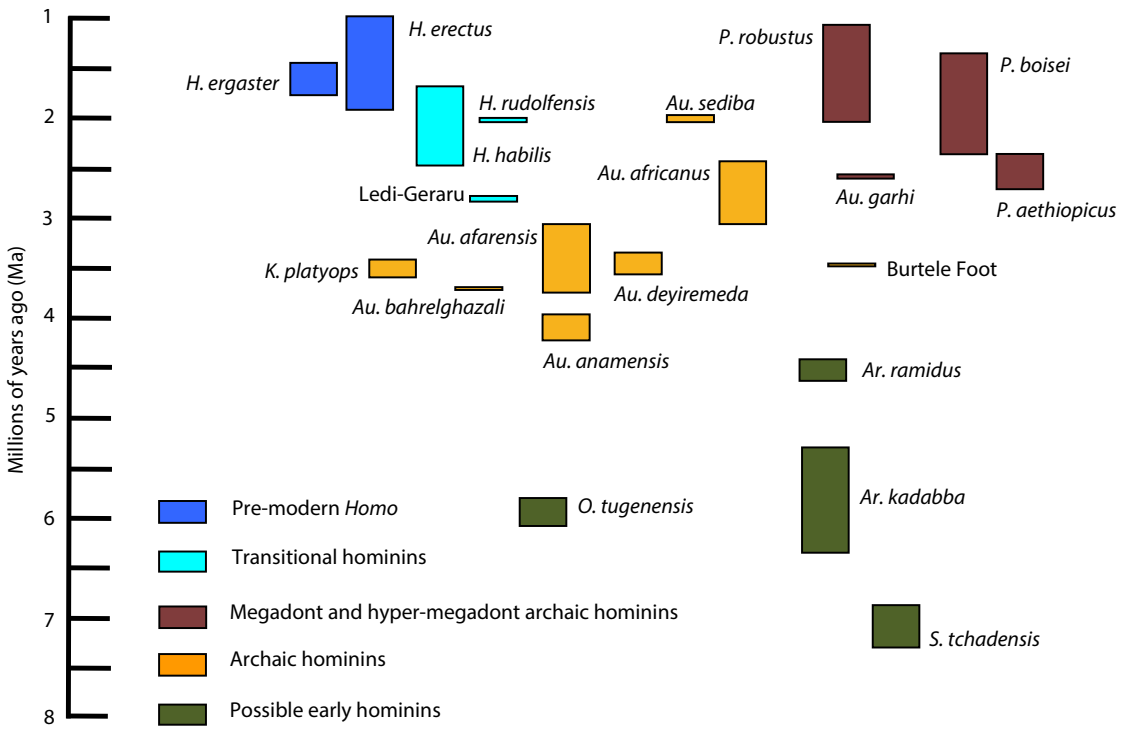

Figure 1.3: Stratigraphic range of early African hominins. The bottom of each bar represents the first appearance datum (FAD) while the top represents the last appearance datum (LAD). The height of each bar is therefore equal to the stratigraphic range of each taxon. The colour scheme represents the grade-based taxonomy proposed by Wood & Lonergan (2008). Only those early African hominins included in later analyses are shown. For an estimate of the stratigraphic range of early African hominins including dating error see Fig. 2 in Wood & Boyle (2016). Modified version of Fig. 1 in Wood & Boyle (2016).

ficient quality to permit a thorough assessment (Maxwell *et al.*, 2018, *in preparation*). Because of the questionable usefulness of canine reduction and foramen magnum position as defining characteristics of hominins, obligate bipedalism is the most widely recognised hominin synapomorphy (MacLatchy *et al.*, 2010; Wood & Harrison, 2011; Simpson, 2013). However, despite a recent meta-analysis showing that postcrania do contain a useful phylogenetic signal (Mounce *et al.*, 2016) and compelling evidence that teeth are particularly poor at reconstructing phylogenetic relationships (Sansom *et al.*, 2017), no cladistic analysis of hominins has to date included postcranial characters (e.g., Chamberlain & Wood, 1987; Wood, 1988, 1991, 1992*b*; Skelton & McHenry, 1992; Lieberman *et al.*, 1996; Strait *et al.*, 1997; Strait & Grine, 2001, 2004; Cameron, 2003; Kimbel *et al.*, 2004; Irish *et al.*, 2013; Dembo *et al.*, 2015, 2016; Mongle *et al.*, 2018, *in review*). Cladistic analyses of hominoids do tend to include postcranial characters (e.g., Begun, 2001; Finarelli & Clyde, 2004; Nengo *et al.*, 2017), however, they also tend to exclude hominins. The reassessment of hominoid phylogeny by Finarelli & Clyde (2004), on the other hand, did include hominins by combining *Australopithecus* + *Homo* into a single Operational Taxonomic Unit (OTU). The cranio-dental synapomorphies identified by Strait & Grine (2004) and the postcranial synapomorphies identified by Finarelli & Clyde (2004) are shown in Table 1.1. In spite of the wealth of evidence to suggest that these widely recognised synapomorphies are probably homoplasies (Wood & Harrison, 2011), palaeoanthropologists continue to adopt the view that these characteristics evolved

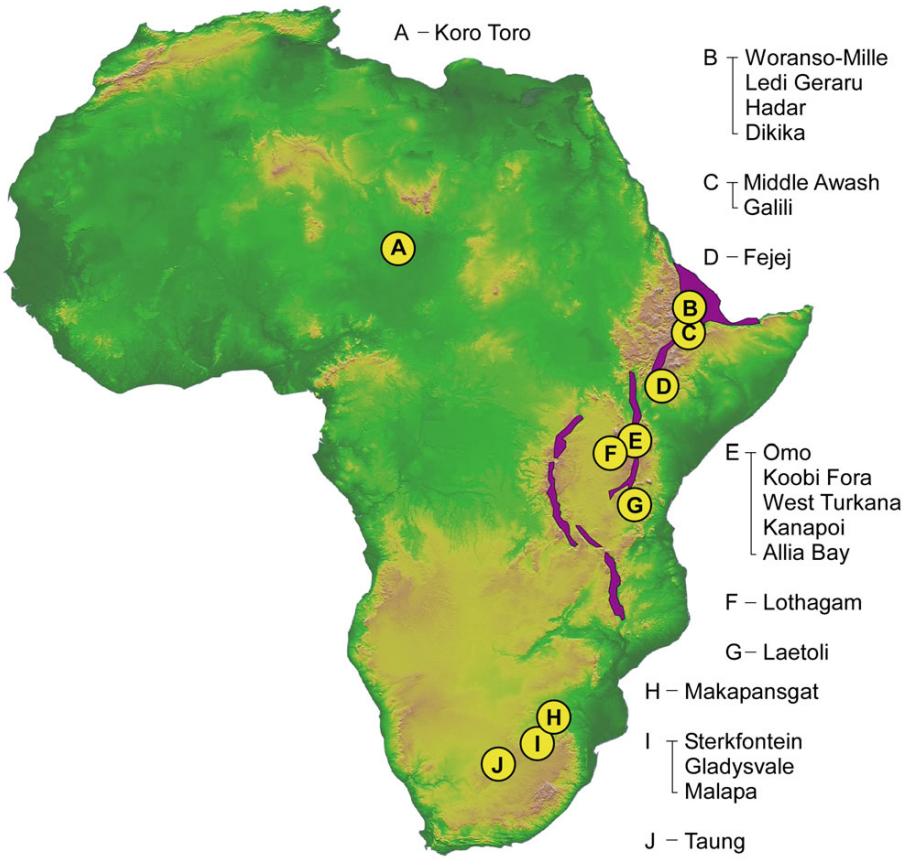

Figure 1.4: Key early African hominin sites. Site "A" also includes the near-by type locality of *Sahelanthropus tchadensis* (TM 266). Not shown are the hominin-bearing deposits 700 km south of site "G" in Malawi (Malema and Uraha). It is clear from this map that the entirety of the early African hominin fossil record samples only three major areas: the Djurab Desert (Chad), the East African Rift System (EARS), and the South African cave and karst deposits. Map from Behrensmeyer & Reed (2013).

only once, disregarding the evidence that many hominin synapomorphies are in whole or in part shared with middle and late Miocene fossil hominoids (Andrews & Harrison, 2005:112). While it is possible that all of the currently recognised synapomorphies did not occur in the earliest hominins, it is clear that a thorough reassessment of hominin systematics inclusive of late Miocene fossil hominoids is desperately needed, particularly in light of our absence of evidence for the evolutionary history of our closest living relative, the chimpanzee (the earliest fossils confidently assigned to *Pan* are around half a million years old; McBrearty & Jablonski, 2005).

It has recently been suggested that the Late Miocene European hominid *Graecopithecus freybergi* is a possible hominin based on its P_4 root configuration (Fuss *et al.*, 2017). However, the case for P_4 root fusion being a hominin synapomorphy is weak and, as such, there is little evidence hominins originated in Europe ca., 7.2 Ma (Harrison, 2010*a*).

Sahelanthropus

The earliest purported hominin to appear in the fossil record is the 7.0-Ma *Sahelanthropus tchadensis* from the late Miocene of Chad (Brunet *et al.*, 2002, 2005). The *Sahelanthropus* hypodigm includes a

near complete but distorted cranium (TM 266-01-060-1), three mandibular fragments, and four isolated teeth from three localities in the Djurab Desert (Fig. 1.2). *Sahelanthropus* is assigned to Hominini based on three characteristics in the basicranium and dentition. First, a more anteriorly positioned foramen magnum compared to *Gorilla* and *Pan*, which may reflect an upright posture and, therefore, a transition to obligate bipedalism (Zollikofer *et al.*, 2005). Second, a flat, long, more horizontally oriented nuchal plane and a large nuchal crest similar to the condition in later hominins and unlike that seen in *Pan* (Zollikofer *et al.*, 2005). These characteristics are indicative of large neck musculature presumably used to support the head and maintain a horizontal gaze. Third, small canines that lack a C/P_3 honing complex and absence of a canine diastema (Brunet *et al.*, 2002, 2005). Geometric morphometric methods also support a clustering with later hominins, particularly *Australopithecus africanus* (Fleagle *et al.*, 2016), a pattern also found by Guy *et al.* (2005). Unfortunately, no late Miocene fossil hominoids were included in either study.

Sahelanthropus differs from *Gorilla* and *Pan* in its short, less prognathic subnasal region and short basioccipital (Brunet *et al.*, 2002). It further differs from *Gorilla* (which it has been claimed to resemble; Wolpoff *et al.*, 2002) in its lack of a supratoral groove (Brunet *et al.*, 2002). Its status as a distinct taxon is uncontested (MacLatchy *et al.*, 2010) based on autapomorphic features such as its remarkably thick supraorbital ridge which is outside the range of male *Gorilla* (Brunet *et al.*, 2002). However, the hominin status of *Sahelanthropus* is hotly contested (e.g., Wolpoff *et al.*, 2002, 2006; Begun, 2004), based on the probably homoplastic suite of hominin characteristics of the basicranium and dentition (Wood & Harrison, 2011), and characteristics shared with *Gorilla* and *Pan* in the neurocranium. For example, *Sahelanthropus* is similar to *Gorilla* and *Pan* in having a small, ape-like brain size (360–370 cc), a low, long superior contour of the neurocranium, and pronounced postorbital constriction (Guy *et al.*, 2005). In addition, the shape and shallowness of the palate resemble the conditions found in *Gorilla* and *Pan* (Guy *et al.*, 2005), and the number of P_4 roots (2) is ape-like and not one-rooted as in modern humans (Emonet *et al.*, 2014). The lack of postcranial remains means there is no direct evidence that *Sahelanthropus* is a biped (Richmond & Hatala, 2013), and only evidence of a highly orthograde posture.

The hominin status of *Sahelanthropus* is well supported phylogenetically (Strait *et al.*, 2007). However, no cladistic analysis of Hominini has included any late Miocene African hominoids. A recent hierarchal Bayesian tip-dating analysis including 13 extinct hominoids and 100 non-coding genomic loci placed *Sahelanthropus* as the sister-taxon of a clade containing *Gorilla* + *Pan* + *Homo* (Matzke *et al.*, 2013), making it a stem hominine. It is also claimed that at 7.2–6.8 Ma (Lebatard *et al.*, 2008) *Sahelanthropus* is too old to be a hominin (Wolpoff *et al.*, 2006). However, if the 8.0-Ma *Chororapithecus* is a member of the *Gorilla* clade (Suwa *et al.*, 2007), then this would support a hominin-panin divergence date of approximately 7.0 Ma.

Orrorin

Orrorin tugenensis is the earliest purported hominin to include cranial and postcranial remains (Senut *et al.*, 2001). The *Orrorin* hypodigm includes three proximal femur fragments, a humeral shaft fragment, a proximal manual phalanx, a fragmentary mandible, and isolated teeth from four fossiliferous localities in the Lukeino Formation, Baringo County, Kenya (Senut *et al.*, 2001), all of which are dated 6.1–5.7 Ma (Pickford & Senut, 2001). In terms of the dento-gnathic evidence, *Orrorin* differs from *Go-*

rilla, *Pan*, and *Ardipithecus* in its thicker enamel (Senut *et al.*, 2001), however, it is reported to have thick enamel by some (e.g., Pickford, 2012) and thin enamel by others (e.g., White *et al.*, 2009). It is also characterised by its smaller postcanine teeth compared to australopiths, a large C with a distinct mesial groove, and no molar cingulum (Senut *et al.*, 2001). It is important to remember the absence of a distinct medial groove on the C is a hominin synapomorphy (Table 1.1). The C is similar in size to *Pan* but with evidence of apical wear (Pickford, 2012), and lacks the elevated crown shoulders found in *Sahelanthropus*, *Ardipithecus*, and later hominins. Senut *et al.* (2001) added the lower molar, KNM-LU 335 (Pickford, 1975), to the *Orrorin* hypodigm, which is similar to *Pan* in its cusp morphology (McHenry & Corrucchini, 1980) and australopiths in terms of buccal flare (Ungar *et al.*, 1994). It has been suggested that pronounced lower molar flare, present in KNM-LU 335, is a hominin synapomorphy (Singleton, 2003), which would support its hominin status. However, there is no significant evidence of lower molar flare in *Sahelanthropus* and *Ardipithecus* (Singleton, 2003). One dental specimen originally assigned to *Orrorin* (BAR 1001'00) has since been re-assigned to a non-hominin hominoid (Pickford & Senut, 2004), as have an upper and lower molar with purported similarities to *Gorilla* (Senut, 2007).

The most convincing evidence for hominin status can be found in the proximal femur BAR 1002'00, which includes the head, neck, lesser trochanter and approximately 2/3 of the shaft (Senut *et al.*, 2001). The greater trochanter and distal end are missing (Senut *et al.*, 2001). The femoral head is spherical and rotated anteriorly with an intertrochanteric groove, the femoral neck is long and anteroposteriorly compressed, the lesser trochanter projects medially (as in modern humans and chimpanzees, and unlike the posteriorly projecting lesser trochanter of australopiths), and the gluteal tuberosity is well-developed (Pickford *et al.*, 2002; Galik *et al.*, 2004). The presence of an intertrochanteric groove is common in modern humans and rare or absent in other primates (Lovejoy *et al.*, 2002; DeSilva *et al.*, 2006), and is suggestive of the kind of full hip extension associated with bipedalism. However, because of its presence in *Pongo* and some atelines and pitheciines (Stern & Susman, 1983), it may not relate directly to a modern human-like form of bipedalism (Crompton *et al.*, 2008). However, the cortical bone is thicker inferiorly than superiorly on the femoral neck (Ohman *et al.*, 2005), differing from the approximately equal cortical thicknesses found in modern hominines. Together, this lead Senut *et al.* (2001) to suggest that *Orrorin* is a sister-taxon to *Homo*, however, there is very little support for such a phylogenetic hypothesis (Strait *et al.*, 2007). Both the humeral shaft fragment (BAR 1004'00) and proximal manual phalanx (BAR 349'00) are ape-like, with the latter showing a degree of curvature similar to *Pan* (Senut *et al.*, 2001; Richmond & Jungers, 2008), and suggesting *Orrorin* maintained some arboreal adaptations for climbing behaviour (Almécija *et al.*, 2013).

Ardipithecus

Ardipithecus is the earliest multi-specific hominin genus. While some regard the earlier *Ardipithecus kadabba* and later *Ardipithecus ramidus* as a single anagenetic lineage (White *et al.*, 2009), others argue that they actually represent distinct genera due to a purported lack of synapomorphies uniting them (e.g., Begun, 2004). However, their precise phylogenetic position, and relation to one another, is unclear, because only the latter has been included in cladistic analyses (Strait *et al.*, 2007).

Ardipithecus kadabba is a poorly-known taxon from the late Miocene and early Pliocene of Ethiopia (Haile-Selassie, 2001; Haile-Selassie *et al.*, 2004). The evidence for *Ardipithecus kadabba* being a ho-

minin is the weakest of any claim. It is known from a right mandibular fragment with M_3 and isolated left mandibular dentition (ALA-VP-2/10), 16 isolated teeth, a humeral mid-shaft and proximal ulna, a fragmentary clavicle, an intermediate manual phalanx, and a pedal proximal phalanx (Haile-Selassie, 2001; Haile-Selassie *et al.*, 2004, 2009). Most fossils are dated 5.8–5.4 Ma, with a P^4 at 6.3 Ma and a pedal phalanx at 5.2 Ma (WoldeGabriel *et al.*, 2001; Simpson *et al.*, 2015). It is assigned to Hominini based on the proximal pedal phalanx (AME-VP-1/71), which has a dorsally canted proximal articular surface like *Australopithecus afarensis* and unlike *Pan* (Haile-Selassie, 2001), yet phalanx curvature is ape-like though less than *Pan* (Haile-Selassie *et al.*, 2009). This characteristic loading at the metatarsophalangeal (MTP) joint in bipeds is reflected in the dorsal orientation of the basal articular surface of the proximal phalanx. Hominoids, whose feet are adapted to grasping, have proximal pedal phalanges that do not routinely experience dorsiflexion at the MTP joint (Simpson, 2013). The geographic and stratigraphic separation of AMA-VP-1/71 from the bulk of the *Ardipithecus kadabba* hypodigm also raises doubt about its assignment to this taxon. The deep, steep-sided olecranon fossa of the distal humeral (ASK-VP-3/78) differs from later hominins, which have more elliptical and shallower fossae, however, the clavicle (STD-VP-2/893) is modern human-like (robust) with a strongly marked deltoid insertion (Haile-Selassie, 2001).

The $\underline{C}$ has a medial groove as in hominoids and *Orrorin*, and there is some indication of a honing complex. The holotype canine has a posteriorly oriented wear facet, which is present in hominoids with a C/P_3 honing complex (Haile-Selassie, 2001). However, the facet is worn horizontally, not diagonally, suggesting the absence of a fully functioning C/P_3 honing complex (Haile-Selassie *et al.*, 2004). This morphology is similar to the presumed most recent common ancestor of *Pan* and *Homo* (MacLatchy *et al.*, 2010). Similarly, the right upper canine (ASK-VP-3/400) also has little apical wear, and in this regard is similar to *Pan* and unlike *Sahelanthropus* and australopiths. The upper and lower canines are projecting and interlocking as in male gorillas and chimpanzees. *Ardipithecus kadabba* differs from *Ardipithecus ramidus* in that the apical crests on the $\underline{C}$ is longer and the P_3 crown outline is asymmetrical (Haile-Selassie, 2001). Moreover, it differs from *Orrorin tugenensis* in $\underline{C}$ crown shape and size; it is similar to *Sahelanthropus* and *Ardipithecus ramidus* in its intermediate enamel thickness (which is less thicker than hominins, but much thicker than hominines), and from the thick enamel of *Orrorin* and australopiths.

Ardipithecus ramidus (4.5–4.3 Ma) is the earliest Pliocene hominin and the first with a hypodigm to sample a substantial portion of the skeleton (White *et al.*, 1994, 1995, 2009; Semaw *et al.*, 2005). The majority of the fossil material is from the Middle Awash (Ethiopia), with the possibility of additional material from Lothagam (Kenya). The holotype is ARA-VP-6/1, an associated set of 10 upper and lower teeth (White *et al.*, 1994), however, the most significant fossil is the partial skeleton ARA-VP-6/500 (White *et al.*, 2009). *Ardipithecus ramidus* can be distinguished from *Gorilla* by its more incisiform canine morphology, and in the smaller absolute size of its dentition and limbs (*Ardipithecus ramidus* weighed around 50 kg compared to a Western gorilla's 160 kg). It differs from *Pan* in the reduction in I^1 size, and elongate and relatively larger M_3 and less crenulated molars. It also differs from other extinct hominoids in its relatively broader lower molars (White *et al.*, 1994).

Ardipithecus ramidus is assigned to Hominini based on characters of the cranium and dentition, including relatively small P_3 without a functional canine-premolar honing complex; reduced canine

size dimorphism; and a more anteriorly positioned foramen magnum (White *et al.*, 2009; Kimbel *et al.*, 2014). Suwa *et al.* (2009*a*) report that the upper canine is shorter than the lower canine, a condition not seen in any anthropoid (Delezene, 2015). Leakey *et al.* (1995, 1998), Kimbel *et al.* (2006), and White *et al.* (2006) have noted that *Ardipithecus kadabba–Ardipithecus ramidus–Australopithecus anamensis–Australopithecus afarensis* form a continuum from ape-like to human-like canine-premolar anatomy, implying these taxa represent an ancestor-descendant sequence. In the partial basicranium ARA-VP-1/500, the anterior border of the foramen magnum is almost in line with the carotid canal (MacLatchy *et al.*, 2010). The basioccipital region is also shorter than in *Gorilla* and *Pan*, which is linked to a more habitually orthograde posture or neural reorganisation (White *et al.*, 1994; Suwa *et al.*, 2009*a*). In addition, the partial temporal ARA-VP-1/125 displays marked pneumatization of the temporal squama and the tympanic is tubular (both of which link it with extant and extinct hominines and the australopiths). However, some of the characteristics—including thin molar enamel (which match with dental microwear and isotopic evidence of a generalised frigivore-omnivore diet; Suwa *et al.*, 2009*b*; Grine *et al.*, 2013), asymmetrical upper and lower third molars, and the size relationships between the canines and cheek teeth—place *Ardipithecus ramidus* closer to *Pan* than to any late Miocene hominin (MacLatchy *et al.*, 2010).

Support for hominin status can also be found in the postcranium, including characters inferred to be indicative of substantial bouts of bipedality, such as the presence of a greater sciatic notch, anterior inferior iliac spine, and dorsal canting of the pedal phalanx (Lovejoy *et al.*, 2009*a*, 2009*b*, 2009*c*). However, postcranially, *Ardipithecus ramidus* is arguably the most unusual hominin. It lacks the elongation of the metacarpals that characterises modern hominoids (though the manual phalanges are elongated and curved as in suspensory hominoids), and the dorsal surface of the proximal metacarpals do not possess ridges or expanded heads (Lovejoy *et al.*, 2009*a*). However, *Ardipithecus ramidus* displays an abducted (opposable) hallux, absence of longitudinal arch in the foot, relatively equal fore- to hind-limb lengths, an African ape-like ischium with a large ischial tuberosity, and pedal phalanges that are curved and similar in length to those in *Gorilla*.

Phylogenetically, *Ardipithecus ramidus* has been reconstructed as a sister-taxon to the australopiths (Strait *et al.*, 2007). Irrespective of whether *Ardipithecus ramidus* is phylogenetically a hominin, it seems that ecologically it is most similar to an ape (Andrews, 1995). However, no single characteristic of *Ardipithecus kadabba* and *Ardipithecus ramidus* definitively demonstrate that either or both are members of the modern human-African ape lineage, or a hominin (Begun, 2004). Moreover, descriptions of its postcranial anatomy being more primitive that any other modern or extinct ape except *Proconsul* are difficult to reconcile with a phylogeny based on cranio-dental characters (Fleagle, 2013).

The placement of *Ardipithecus ramidus* in any part of hominoid phylogeny results in remarkably high levels of homoplasy (Wood & Harrison, 2011). If *Ardipithecus ramidus* is not a hominin then it would require the parallel evolution of a number of shared specialisations with later hominins in the basicranium, dentition, and ilium. On the contrary, if *Ardipithecus ramidus* is a hominin, then it would require remarkably high levels of homoplasy in modern hominoids (Wood & Harrison, 2011). White and co-workers (2009; Lovejoy *et al.*, 2009*d*) argue that the plesiomorphic characteristics of *Ardipithecus ramidus* indicate that the last common ancestor of *Gorilla*, *Pan*, and *Homo* (that is, hominines) lacked the locomotor and positional adaptations common to all hominoids. Shared hominoid char-

acteristics relating to forelimb suspensory behaviour, orthogrady, and vertical climbing are, therefore, argued to have arisen independently in hylobatids and each great ape lineage (White *et al.*, 2009).

Australopithecus

Australopithecus is the earliest undisputed group of hominins. Shared australopith characteristics include (Kimbel, 2007):

1. Brain size approximately equal to an ape (range ca., 380–550 cc).
2. Inferosuperiorly short, vertical mid-face with massive zygomaticomaxillary region and strong subnasal prognathism.
3. Large (in relation to body size) premolars and molars capped by variably thick enamel.
4. Transversely thick mandibular body and tall ascending rami.

Australopithecus anamensis is the earliest known australopith and found in 4.2–3.9-Ma deposits in Kenya and Ethiopia (Leakey *et al.*, 1995, 1998). It is known from mostly cranio-dental material including the holotype KNM-KP 29281 (a mandible with complete dentition but lacking the rami). Postcranial fossils are known including a fragmentary tibia (KNM-KP 29285), distal humerus (KNM-KP 271), radial fragment (KNM-ER 20419), capitate, and manual phalanx (Ward *et al.*, 2001, 2013, 2017, *in press*).

Australopithecus afarensis (Johanson *et al.*, 1978) is known from Laetoli (Tanzania), Dikika, Hadar, Maka, and Woranso-Mille (Ethiopia), and possibly East Turkana (Kenya) and Bahr el Ghazal (Chad). Its stratigraphic range spans from 3.7–3.0 Ma, however, there are specimens of uncertain affiliation at 3.9 Ma (BEL-VP-1/1). *Australopithecus afarensis* is known from nearly 400 specimens from the Hadar Formation alone, including the partial skeleton AL 288-1 (Johanson *et al.*, 1982), the near-complete skull AL 444-2 (Kimbel & Rak, 2010), and the AL 333 assemblage of at least 13 individuals (known as the first family).

Australopithecus bahrelghazali (Brunet *et al.*, 1995, 1996) is named to accommodate a mandibular symphyseal fragment, an isolated upper third premolar, and a maxilla fragment from Bahr el Ghazal (Chad). Cosmogenic nuclide dating yields an age of 3.58 Ma (Lebatard *et al.*, 2008). It has been argued that the material is of insufficient quality to make an accurate taxonomic assignment (White, 2002), and that the supposed apomorphies are represented in *Australopithecus afarensis* material from Laetoli, Hadar, and Maka (Kimbel, 2007). For example, LH 24 has a three-rooted premolar (White *et al.*, 2000) and AL 444-2 has a vertical symphyseal cross section (Kimbel *et al.*, 2004).

Australopithecus deyiremeda is a recently described hominin from Burtele and Waytaleyta in the Woranso-Mille study area (Ethiopia), dated to 3.5–3.3 Ma (Haile-Selassie *et al.*, 2015). It is known from a fragmentary maxilla (BRT-VP-3/1), and 12 other cranio-dental fossils. It differs in maxillary shape from *Australopithecus afarensis* and *Kenyanthropus platyops* (Spoor *et al.*, 2016), supporting claims for a new hominin taxon.

Australopithecus africanus (Dart, 1925) is known from cave deposits at Makapansgat, Taung, Sterkfontein, and Gladysvale (South Africa), and is dated 3.0–2.4 Ma. However, the dating of these hominin-bearing deposits is only known with considerable uncertainty (Pickering *et al.*, 2011). The hypodigm of

Table 1.1: List of hominin synapomorphies.

Character (state)	Source
Projection of nasal bones above frontomaxillary suture (projected)	Strait & Grine (2004)
Infraorbital foramen location (variable)	Strait & Grine (2004)
Occipital marginal (O-M) sinus present in high frequency (yes)	Strait & Grine (2004)
External cranial base flexion (moderate)	Strait & Grine (2004)
Inclination nuchal plane (intermediate)	Strait & Grine (2004)
Position of foramen magnum relative to bi-tympanic line (anterior margin at the bi-tympanic line)	Strait & Grine (2004)
Canine reduction (somewhat)	Strait & Grine (2004)
Positions of cusps on mandibular teeth (lingual cusps approximate margin; buccal cusps slightly lingual to margin)	Strait & Grine (2004)
Mesiobuccal protrusion of P_3 crown base (moderate)	Strait & Grine (2004)
Extensive mesial groove on $\underline{C}$ (no)	Strait & Grine (2004)
Insertion of genioglossus (above inferior transverse torus)	Strait & Grine (2004)
Condylar canal (absent/infrequent)	Strait & Grine (2004)
Ethmo-lacrimal contact (full: 100%)	Strait & Grine (2004)
Ethmo-sphenoid contact (present: 100%)	Strait & Grine (2004)
Second metacarpal facet on capitate (continuous)	Finarelli & Clyde (2004)
Metacarpal heads broadest (palmarly)	Finarelli & Clyde (2004)
Lower iliac height (short)	Finarelli & Clyde (2004)
Trochanteric fossa (unique)	Finarelli & Clyde (2004)
Distal tibia facet (square)	Finarelli & Clyde (2004)
Lateral malleolus (small)	Finarelli & Clyde (2004)
Flexor hallucis longus groove (intermediate)	Finarelli & Clyde (2004)
Cuboid length (long)	Finarelli & Clyde (2004)
Ectocuneiform/first metatarsal joint (distal)	Finarelli & Clyde (2004)
Cuneiform length (long)	Finarelli & Clyde (2004)
Size of first metatarsal sesamoid grooves (large)	Finarelli & Clyde (2004)
First metatarsal length (long)	Finarelli & Clyde (2004)
Transverse arch (present)	Finarelli & Clyde (2004)
Ectocuneiform facet on navicular (distal)	Finarelli & Clyde (2004)
Pedal phalangeal curvature (straight)	Finarelli & Clyde (2004)

Australopithecus africanus is numerically the best of any hominin, and although poorly catalogued there are more than 500 specimens (the vast majority of which are from Sterkfontein). Cranio-dental material is by far the most abundant (including the near-complete crania Sts 5 and Taung 1), but there is at least one of each long bone, and much of the vertebral column is represented in the partial skeleton StW 431 (Toussaint *et al.*, 2003). It remains to be seen whether the associated skeleton StW 573 from Sterkfontein Member 2 and 12 hominin fossils recovered from the Jakovec Cavern (Partridge *et al.*, 2003) belong to *Australopithecus africanus* or a different taxon (Clarke, 2008).

Australopithecus garhi (Asfaw *et al.*, 1999) is known from the partial cranium BOU-VP-12/130 and four other cranio-dental specimens found in the Middle Awash (Ethiopia). *Australopithecus garhi* combines *Paranthropus*-like postcanine megadontia with large incisors and canines and enamel that lacks the extreme thickness seen in *Paranthropus* (Asfaw *et al.*, 1999). For this reason it is often linked ecologically with *Paranthropus* (Wood & Lonergan, 2008). A partial skeleton including a long femur and forearm is also known from nearby deposits but are not assigned to *Australopithecus garhi* (Asfaw *et al.*, 1999). The humerofemoral index of this specimen is *Homo*-like in proportions. However, the forearm displays *Pongo*-like brachial proportions well outside the range of other hominins (Richmond *et al.*, 2002).

Australopithecus sediba (Berger *et al.*, 2010) is named to accommodate more than 200 specimens from Malapa (South Africa), including the partial skeletons MH1, a sub-adult presumed male, and MH2, an adult presumed female. The phylogenetic position of *Australopithecus sediba* is hotly debated, mainly because of the juvenile status of the holotype (MH1). However, the weight of evidence suggests that *Australopithecus sediba* and *Australopithecus africanus* are sister-taxa based on the anatomy of their dentition (Irish *et al.*, 2013), feet (Prang, 2016), and cranium (Kimbel & Rak, 2017).

Kenyanthropus

Kenyanthropus platyops (Leakey *et al.*, 2001) is found in 3.5–3.3-Ma deposits at Lomekwi, West Turkana (Kenya). It is known from a relatively complete but crushed cranium (KNM-WT 40000), a paratype maxilla, and 34 other cranio-dental specimens including three mandible fragments, a maxilla fragment, and isolated teeth (Leakey *et al.*, 2001). *Kenyanthropus platyops* is distinguished from other australopiths by its short, flat face, anteriorly-situated zygomatic root, and flatter and more vertically orientated malar region (Spoor *et al.*, 2010). The phylogenetic placement of *Kenyanthropus platyops* is uncertain (Strait *et al.*, 2007), having been reconstructed as either a sister-taxon to *Paranthropus* or a *Paranthropus* + *Homo* clade.

Paranthropus

Paranthropus aethiopicus (Arambourg & Coppens, 1968) is the earliest megadont hominin and is known from 2.7–2.3-Ma deposits in the Omo-Turkana Basin (Kenya). The hypodigm is composed of 68 specimens, 67 of which are cranio-dental, including the adult cranium from Lomeckwi (KNM-WT 17000), a partial mandible (KNM-WT 16005), and isolated teeth from the Shungura Formation. The taxon is defined by a massive face, pronounced subnasal prognathism, and very large sagittal and nuchal crests. A tibial fragment (EP 1000/98) from the Upper Ndolanya Beds, Laetoli (Tanzania), is the only postcra-

nial specimen. Most view *Paranthropus aethiopicus* and *Paranthropus boisei* as an ancestor-descendant pair (Wood & Schroer, 2013).

Paranthropus boisei (Leakey, 1959) is a hyper-megadont hominin known from deposits at Olduvai Gorge and Peninj (Tanzania), the Omo and Konso (Ethiopia), Malema (Malwai), Chesowanja, Koobi Fora, and West Turkana (Kenya), and is securely dated between 2.3–1.3 Ma. Compared with other members of *Paranthropus*, it has smaller anterior teeth, absolutely larger cheek teeth, a robust mandible, and pronounced sagittal and nuchal crests. It is unusual among hominins in that isotopic evidence indicates a diet made up almost entirely of C_4 foods (Sponheimer *et al.*, 2013).

Paranthropus robustus (Broom, 1938) is found in cave deposits at Swartkrans, Kromdraii, Drimolen, Gondolin, and Cooper's Cave (South Africa). It is known from over 400 specimens of which the majority are isolated teeth. *Paranthropus robustus* can be distinguished from the contemporary and roughly sympatric *Australopithecus africanus* by a larger brain, wider face, and postcanine megadontia. Moreover, while the anterior pillars of *Australopithecus africanus* are a hollow column of cortical bone, in *Paranthropus robustus* they are a column of dense trabecular bone (Villmoare & Kimbel, 2011). *Paranthropus robustus* differs from *Paranthropus boisei* in its dietary breadth, which includes substantially more C_3 foods (Sponheimer *et al.*, 2013).

Early *Homo*

Homo habilis (Leakey *et al.*, 1964) is known from 2.3–1.6 Ma deposits at Olduvai Gorge (Tanzania) and Koobi Fora (Kenya), and possibly also Chemeron and West Turkana (Kenya), the Omo and Hadar (Ethiopia), Sterkfontein, Drimolen, and Swartkrans (South Africa). It is known from mostly craniodental material (e.g., KNM-ER 1813) with few postcranial remains (Johanson *et al.*, 1987) that can be assigned with confidence. *Homo habilis* can be distinguished from the australopiths by reduced subnasal prognathism, relatively thin molar enamel, and narrower premolars and molars (Kimbel *et al.*, 1997).

Homo rudolfensis (Alexeev, 1986; *sensu* Wood, 1992*a*) is known from the Turkana Basin of northern Kenya and possibly Uraha (Malawi; Schrenk *et al.*, 1993). It ranged from ca., 2.0–1.8 Ma, however, if the Malawian specimen UR 501 is indeed *Homo rudolfensis* then its first appearance date will be 2.4 Ma (Wood & Boyle, 2016). *Homo rudolfensis* is defined by a larger brain (730 cc in KNM-ER 1470), flatter, broader face, and larger postcanine teeth with thicker enamel compared to *Homo habilis* (Wood, 1991; Leakey *et al.*, 2012). The face of *Homo rudolfensis* is widest in its mid-part compared to the face of *Homo habilis* which is widest superiorly. Spoor *et al.* (2015) also report that the dental arcade of *Homo rudolfensis* is different (e.g., more divergent tooth rows, flatter anterior dental arch) from *Homo habilis*.

African *Homo erectus* is known from Koobi Fora and West Turkana (Kenya), the Middle Awash, Gona, Garba, Gambore (Ethiopia), Olduvai Gorge and Makuyuni (Tanzania), Ain Maarouf (Morocco), Tighenif (Algeria), Buia (Eritrea), Yayo (Chad), and Sterkfontein and Swartkrans (South Africa). The nomen *Homo ergaster* (Groves & Mazák, 1975) is used to distinguish specimens that are more primitive (e.g., KNM- ER 730 and 992) than Asian *Homo erectus* (Antón, 2003), though Spoor *et al.* (2007) argue that many of the cranial differences between *Homo ergaster* and *Homo erectus* are size related and not specific differences. *Homo erectus* is most well known from the remarkably complete skeleton KNM-WT 15000, the first specimen to show modern human-like dental and limb proportions (Walker & Leakey, 1993). The nomen *Homo erectus* is used throughout this thesis to refer to *Homo ergaster* and all

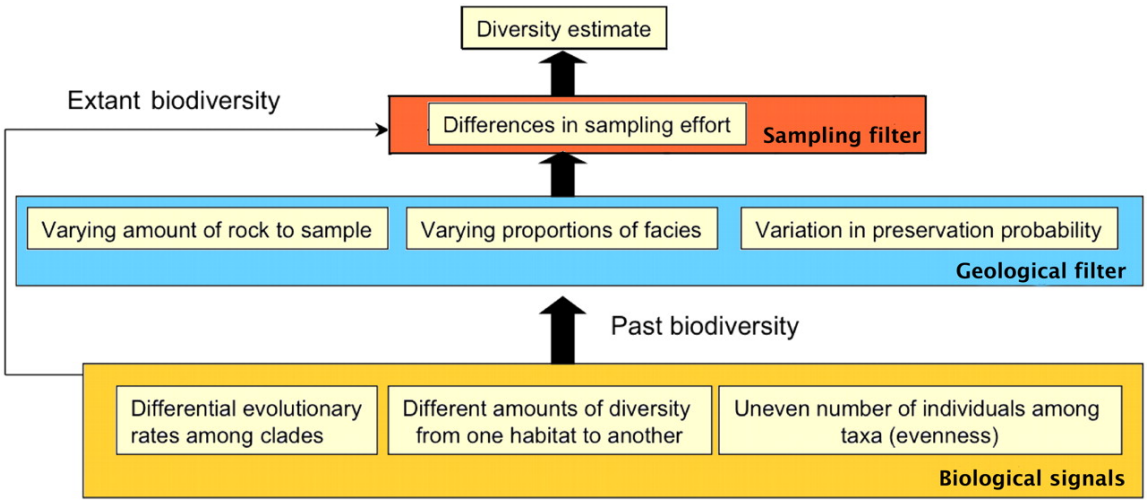

Figure 1.5: Factors influencing palaeodiversity estimation. Schematic flow chart showing how the diversity estimate that palaeoanthropologists frequently use to understand hominin diversification represents a filtered version of the original biological diversity present in the geological past. From Smith & McGowan (2011).

African fossils assigned to *Homo erectus*. Such a designation is often referred to as *Homo erectus sensu lato* (= in the broad sense).

1.3 Fossil record quality through space and time

In a groundbreaking paper, Raup (1972) outlined the principle biases that effect the fossil record at the species level which introduce error into diversity patterns. This list is outlined below and summarised in Fig. 1.5.

(1) *Range charts.* The earliest diversity estimates were based on compendia of stratigraphic range data rather than on fossil occurrence data (e.g., Valentine, 1969; Sepkoski *et al.*, 1981). For example, if a taxon first occurred in the Miocene and last occurred in the Pleistocene its stratigraphic range will span the entire Pliocene. Range-through diversity estimates have the benefit of requiring minimal information and inferring diversity in time bins that do not contain fossil-bearing rocks. For example, a time bin (e.g., the Pliocene) could completely lack fossil-bearing rock but be credited with yielding considerable diversity. Under such a scenario, low diversity would be more simply explained by poor fossil sampling and not a genuine feature of a clade's evolutionary history. Range charts can lead to phenomena known as edge effects. Range charts are incomplete in that true stratigraphic ranges are unknown, meaning they will underestimate diversity at either end of a taxon's range (Raup, 1972). However, range charts will have a higher probability of range truncation at the older end (first appearance) because older rocks have a greater chance of non-exposure or destruction by erosion and metamorphism (Raup, 1972). Mass extinction events can also produce a specific type of edge effect (Signor & Lipps, 1982): during a mass extinction many taxa will die out in a single event but, due to range truncation, not all stratigraphic ranges will end at the extinction event. In fact, many taxa will appear to go extinct before the event and the mass extinction will appear gradual: this effect is named the Signor-Lipps effect (Signor & Lipps, 1982).

(2) *Influence of extant records.* Since our understanding of extant taxa (neontology) is far better

than that of the fossil record, fossil taxa with extant members will probably have their stratigraphic range extended to the present (Raup, 1972). This leads to a specific type of edge effect known as the Pull of the Recent (Raup, 1979; Sahney & Benton, 2017). By extending the range of extinct taxa with extant members to the present day, younger rocks are biased toward higher diversity and lower extinction compared to older rocks.

(3) *Duration of geological time bins.* The time bins employed in diversity estimation can also distort diversity patterns. For example, longer time bins will show higher diversity and shorter time bins will show lower diversity (Raup, 1972). One would expect that during longer time bins more taxa will speciate and go extinct, raising that time bin's diversity (Foote, 1994). Moreover, during longer time bins there will also be more sedimentation, and a higher probability of fossilisation (Miller & Foote, 1996). However, time bin duration and raw taxic diversity repeatedly show no correlation (e.g., Benson *et al.*, 2010; Mannion *et al.*, 2011, 2015; Bennett *et al.*, 2018). Raup (1972) also noted that geological time units are commonly delineated based on biostratigraphy, and therefore the duration of these units are not independent of turnover through geological time.

(4) *Monographic effects.* Raup (1972) suggested that the level of interest in a particular group or geographic area will affect apparent diversity, as will the quality of the taxonomic research into a group. Interest in a particular clade, either for reasons of popularity (e.g., dinosaurs and hominids) or usefulness (e.g., foraminifera), will also lead to substantially more work done on these clades and potentially more taxa being named (Raup, 1972). The tendency of workers to examine particular geographic areas is also well documented (e.g., Hill, 2007; Uhen & Pyenson, 2007; Brocklehurst *et al.*, 2012), with palaeoanthropologists showing a preference for known hominin-bearing deposits in the East African Rift Valley and South African cave networks. The amount and areal extent of collection effort is of critical importance in shaping our knowledge of the fossil record (Raup, 1972, 1976*b*; Barnosky *et al.*, 2005). Regions where surface exploration is active will, statistically, yield more specimens and a higher species richness. This may seem rather trivial: of course, fossils are unlikely to be found in regions where they are not being actively searched for, but variation in the amount of collection effort and study interest (geographically and stratigraphically) can have a major impact on apparent diversity patterns. Raup (1972) also noted a time-dependent aspect of monographic effects: if a clade has extant members, morphological information is better closer to the Recent, in turn affecting taxonomic assignments.

(5) *Lagerstätten.* Lagerstätten deposits are those which contain remarkably abundant or complete fossils (e.g., Messel Pit, Germany, and the La Brea Tar Pits, California). The quantity of fossil material will result in higher apparent diversity, while the quality of fossil material will increase the amount of diagnostic information available for taxonomic assignment. Lagerstätten deposits have been shown to correlate with peaks in diversity and specimen completeness (Brocklehurst *et al.*, 2012; Friedman & Sallan, 2012; Dean *et al.*, 2016). The distribution of Lagerstätten throughout the rock record is not systematic but they appear more common in younger rocks (Raup, 1972). The greatest effect, however, is to add noise to the diversity data in much in the same way that monographic bursts produce artificial peaks in diversity.

(6) *Area-diversity relationships.* The greater the sampled area the higher a diversity estimate is likely to be (Raup, 1972). When a new geographic area is explored, the rate of taxon discovery increases rapidly. This is due in part to increased sampling, but also a result of the fact taxa are geographically restricted

due to either climatic factors or barriers to dispersal. The larger the area of habitat the higher the diversity found in that area. However, apparent diversity depends not only on the amount of exposed rock, but the manner in which exposed rock is distributed (Raup, 1972).

(7) *Thickness, area, and volume of exposed rock.* Raup (1972) first reported that stratigraphic intervals with the greatest exposed rock yield more collecting sites, more specimens, and a higher species richness. The relationship—if any—between Neogene rock exposure and African primate diversity is unknown, as temporally fine-grained information on exposed sedimentary rock is not available at the continental level, as yet. Preliminary tests have been performed with the scant data that is currently available, but a relationship remains unclear. In the Turkana Basin (Kenya), for example, a positive—albeit non-significant—correlation has been found between the thickness of each formation (Bowen, 1974; Harris *et al.*, 1998a, 1988b; McDougall & Feibel, 1999; Feibel, 2003; Gathogo & Brown, 2006) and the number of vertebrate specimens recorded in the Turkana Public Database (Bobe *et al.*, 2011) (r = 0.796, n = 5). In contrast, preliminary work by Thompson *et al.* (2017) at three collection areas in the Koobi Fora Formation, East Turkana (Kenya), demonstrated that there is no link between primate (cercopithecid + hominin) fossil abundance and outcrop or collection area.

Large-scale studies comparing palaeodiversity to the rock record include those that estimate the surface outcrop area of sedimentary rocks from geological maps and their accompanying monographs (Raup, 1972, 1976*a*; Smith, 2001; Crampton *et al.*, 2003; Smith & McGowan, 2007; McGowan & Smith, 2008; Wall *et al.*, 2009, 2011; Fröbisch, 2013), the area of exposed rock (Dunhill, 2011, 2012; Dunhill *et al.*, 2012; Walker *et al.*, 2017), the number of gap- or hiatus-bound rock packages (Peters, 2005, 2006; Peters & Heim, 2010), or counts of the number of named formations as a proxy for the amount of exposed sedimentary rock (Peters & Foote, 2001; Crampton *et al.*, 2003; Fröbisch, 2008; Barrett *et al.*, 2009; Butler *et al.*, 2009; Benson *et al.*, 2010; Marx & Uhen, 2010; Benson & Butler, 2011; Benton *et al.*, 2011; Mannion *et al.*, 2011; Upchurch *et al.*, 2011; Benson & Upchurch, 2013; Newham *et al.*, 2014; Dunhill & Wills, 2015; Dean *et al.*, 2016). In a surprisingly large number of these studies a statistically significant positive correlation has been found linking the rock record and apparent diversity, both in the marine and terrestrial realms. The strength and ubiquity of these correlations implies that short-term fluctuations in apparent diversity predominantly reflect fluctuations in the sedimentary rock record. Geological processes such as erosion, subduction, and the deposition of other layers continuously destroy rocks, rendering some periods of time inaccessible to study. Moreover, rates of sedimentation have fluctuated through Earth history, and a higher rate of sedimentation increases the probability of preservation (Raup, 1972).

Since Raup (1972), many other factors have been shown to influence the quality of the fossil record and in particular diversity patterns.

Errors in radiometric dating. Range-through diversity relies on accurate first and last appearance dates. This is particularly problematic when combining occurrence data from the East African Rift System (EARS) and South African palaeocave deposits. Deposits in the EARS are securely dated due to the presence of volcanic ash layers (tuffs) that are interspersed between fossil-bearing deposits. Because of their regularity and the ability to date them precisely, fossil-bearing deposits in the EARS are frequently constrained to within tens of thousands of years. In contrast, the South African palaeocave deposits are often constrained within hundreds of thousands of years (Wood & Boyle, 2016).

Size of an organism. The body size of an organism has also been shown to be important. The larger

the organism the more robust the bones and the greater the probability of fossilisation (Behrensmeyer *et al.*, 2000). Larger fossils may be collected more often due to a desire for large articulated specimens for museum displays (Brown *et al.*, 2013). Large mammals (> 180 kg) have been shown to be overrepresented relative to modern faunas at many hominin-bearing sites (Soligo & Andrews, 2005). Similar levels of non-equivalence with respect to modern faunas have also been found for small mammals (< 1 kg), which are underrepresented (Soligo & Andrews, 2005).

Specimen completeness. Temporal trends in specimen completeness have been shown to influence diversity patterns. However, the precise relationship depends heavily on the clade of interest. Mesozoic birds and Late Cretaceous sauropodomorph dinosaurs display higher apparent diversity when specimens are more complete (Mannion & Upchurch, 2010; Brocklehurst *et al.*, 2012). In contrast, early synapsids display lower apparent diversity when specimen are more complete, suggesting that much of their species richness may be an artefact of fragmentary fossils (Brocklehurst & Fröbisch, 2014). The relationship between specimen completeness and body size is more complex. Small- and large-bodied ichthyosaurs are less complete than ichthyosaurs of intermediate size (Cleary *et al.*, 2015), and small birds, though more likely to be destroyed, are easier to bury rapidly and preserve whole (Brocklehurst *et al.*, 2012).

Time-averaging. Taphonomic time-averaging occurs when the turnover rate of a population is considerably higher than the net rate of sedimentation (Behrensmeyer *et al.*, 2000). This produces a single stratum containing taxa of successive, non-contemporary populations admixed together. Erosion can also lead to time-averaging: many fossils in the East African Rift System (EARS), for example, are surface finds. These may be lag deposits combining eroded layers of unknown thickness and time. Time-averaging is closely related to fossil record completeness (Kowalewski, 1996). A fossil record is incomplete whenever a time bin does not contain any fossil material; time-averaging on the other hand can produce an over-complete fossil record by mixing older and/or younger material together (note that in this scenario the strata in question are over-complete, whereas adjacent strata would be incomplete). Time-averaging can, therefore, give the appearance of sympatry and artificially inflate both phenotypic variation and diversity estimates. Moreover, time-averaging also has the effect of combining taxa from different phases of a climate cycle, a phenomenon named climate-averaging, which is particularly problematic when relating evolutionary events with climatic conditions (Hopley & Maslin, 2010).

This large—and growing—body of work demonstrates that the quality of the fossil record differs substantially through geological time and between clades, and that the quality of the fossil record should be studied individually for each clade if they are to be used to derive macroevolutionary patterns' (Tarver *et al.*, 2011). The hominin fossil record ought not to be exempt from such scrutiny.

1.4 Defining fossil record quality

Quality refers to the adequacy and fidelity of the fossil record as an archive of evolutionary events (Paul, 1991). The adequacy of the fossil record refers not to whether the fossil record is complete (complete knowledge in science is unattainable) but to whether the fossil record is *sufficiently* complete to answer questions about an organism's evolution and palaeobiology. The fidelity of the fossil record refers to whether the biological signal contained within the rock record is a genuine depiction of evolutionary

events, and if not, what are the factors biasing it.

Bias refers to any biological, geological, or anthropogenic factor that can introduce error into interpretations of data from the fossil record (Benton *et al.*, 2011). A sampling metric seeks to quantify these biases through geological time.

Completeness refers to the proportion of all life on Earth that is known from identifiable fossils or the proportion of time actually represented by fossiliferous sediments. In Chapter 3, completeness refers to the amount of time represented by fossiliferous sediments relative to the amount of time represented by gap, whether due to a lack of fossiliferous sediment, collection effort, or both. This is the concept of completeness that Darwin (1859) himself emphasised (Hunt, 2010). However, in Chapter 4 completeness refers to the amount of skeletal representation or the number of scorable phylogenetic characters in the fossils themselves.

1.5 Compilation of the Hominin Fossil Database

The Hominin Fossil Database (HFDB) constructed for this thesis (using Microsoft Excel) contains over 5000 entries (each representing a single fossil element) and includes taxonomic, geographic, stratigraphic, anatomical, and bibliographic information for each specimen plus additional notes on, for example, specimen preservation, percentage completeness, and taxonomic uncertainty. It includes all African hominin fossils from the late Miocene, Pliocene, and early Pleistocene. It also includes dubious specimens referred but not formally assigned to a specific genus or species (i.e., those labelled cf. or aff.). HFDB is the product of an exhaustive literature survey. Its compilation has benefitted from several comprehensive and detailed compendia, namely the *Catalogue of Fossil Hominids. Part I: Africa* (Oakley *et al.*, 1977), *Wiley-Blackwell Encyclopedia of Human Evolution* (W-BEHE; Wood, 2011), *Cenozoic Mammals of Africa* (Werdelin & Sanders, 2010), and the Turkana Public Database (Bobe *et al.*, 2011). The assignment of specimens to genera and species follows current conventional taxonomic assignment in W-BEHE. If a specimen does not have an entry in W-BEHE then its current conventional taxonomic assignment is based on the consensus in the literature. In spite of the continued debate surrounding the taxonomic status of *Sahelanthropus tchadensis*, *Orrorin tugenensis*, and *Ardipithecus kadabba*—a debate exacerbated by the small number of fossil specimens in their hypodigms, the lack of overlap in the parts of the skeleton represented (Wood & Boyle, 2016), and a meagre comparative sample of early to late Miocene African hominoids (Begun, 2010)—these taxa are conventionally reported as hominins and are therefore included (see § 1.2 for further information).

1.6 Outline of thesis

Despite the quickening pace of discovery and re-analysis of current fossil evidence, the hominin fossil record remains, at key points, frustratingly incomplete, and there has been a systemic lack of interest in its quality at the macroevolutionary scale. Consequently, the majority of research into early hominin macroevolution has been based on a direct reading of the fossil record. It has been suggested that one reason for the absence of a debate on this subject may be the perception that hominin fossils are so scarce that they are unlikely to pass a rigorous analysis of their completeness (Hopley, 2018, *in review*).

Alternatively, researchers may assume that bias is randomly distributed in the hominin fossil record and so any evolutionary signal is dampened, not distorted (Raup, 1991); or that the fossil record is a completely un-biased archive depicting only real evolutionary events, so any consideration of its quality is unnecessary (e.g., Potts & Faith, 2015:13). Though either of the latter points may be true, the tendency of previous studies to disregard the architecture of the sedimentary rock record, the geographic and stratigraphic heterogeneity of collection effort, and other limitations inherent to fossil data, could significantly hamper our understanding of hominin origins and evolution (Smith & Wood, 2017; Wood & Boyle, 2017; Hopley, 2018, *in review*)—with the greatest danger being the totality of our knowledge having been read literally from the fossil record.

This thesis aims to remedy this by providing an up-to-date examination of the diversification of early African hominins, and the first detailed examination of the quality of their fossil record. It is structured as a series of semi-autonomous article-chapters book-ended by this general introduction and a conclusion. Because this thesis includes separate article-chapters, each with its own introduction, detailed accounts of background information specific to each chapter are not included in this general introduction. Likewise, this thesis does not include a separate chapter dedicated to the entire methodology as each article-chapter contains its own methodology section.

Chapter 1 (the current chapter) presents an overview of the early African fossil evidence for hominin evolution and outlines the major biases affecting fossil occurrence data.

Chapter 2 presents and compares taxic and phylogenetically corrected estimates of early African hominin diversity, and assesses the confidence in the first and last appearance dates used to create palaeodiversity curves.

Chapter 3 assesses whether the diversification patterns in Chapter 2 are a genuine biotic signal or instead driven by the non-random distribution of available rock and collection effort through geological time.

Chapter 4 assesses the specimen and taxon completeness of the early African hominin fossil record, and whether there is a relationship between completeness, diversity, and sampling metrics.

Chapter 5 presents a summary of the key findings, a discussion of the limitations, and proposes several new avenues for future research in hominin palaeobiology.

Chapter 2

Phylogenetic and taxic perspectives on early hominin diversity

This chapter is an extended version of parts of the following publication: Maxwell SJ, Hopley PJ, Upchurch P, Soligo C, 2018. Sporadic sampling not climatic forcing drives early hominin taxic diversity. Proceedings of the National Academy of Sciences, ***115****(19), 4891–4896.*

2.1 Introduction

Understanding fluctuations in palaeodiversity is key to elucidating diversification dynamics in deep time. Peaks and troughs in palaeodiversity are driven by temporal variation in speciation and extinction rate (Marshall's (2017) fourth law of palaeobiology), and an accurate reconstruction of their magnitude and sequence enables palaeobiologists to deduce major events or transitions in the history of a clade (e.g., adaptive radiations and mass extinctions). Moreover, diversification is of relevance to broader macroevolutionary questions such as the processes underlying adaptive radiation, the relative importance of intrinsic *versus* extrinsic factors, co-evolution, and inter- and intra-species competition. It is, therefore, unsurprising that the study of palaeodiversity, and the factors that influence its estimation, have received considerable attention in palaeobiology (see McGowan & Smith (2011) and references therein).

In recent years there have been a number of publications on hominin palaeodiversity (Grove, 2011; Shultz & Maslin, 2013; Foley, 2016; Wood & Boyle, 2016, 2017), with many more describing the frequency of their first and last appearance in the fossil record (e.g., Vrba, 1988; Foley, 1994; deMenocal, 1995, 2004; Kimbel, 1995; Grove, 2012; Potts, 2013). In all of these publications, each taxon's first appearance date (FAD) in the fossil record is treated as a speciation event and each taxon's last appearance date (LAD) is treated as an extinction event. The number of taxa through geological time are then counted based on the assumption that FADs and LADs accurately describe the timing and sequence of evolutionary events. The fossil record offers the only direct window into deep time diversification (Marshall, 2017). However, it is incomplete, patchy, and biased (Raup, 1972, 1976*b*; Benton *et al.*, 2011). The use of such an approach to estimating palaeodiversity has been shown to produce grossly inaccurate diversity patterns and should therefore be treated with great caution. Many attempts have been made to correct

the distortion using phylogenetic (Norell, 1992; Norell & Novacek, 1992; Smith, 1994), modelling (Smith & McGowan, 2007; Lloyd, 2012), and subsampling approaches (Raup, 1975; Alroy, 2010). However, no attempts have been made to estimate hominin palaeodiversity using sampling-corrected approaches. Therefore, any palaeodiversity estimate based solely on first and last appearance in the fossil record must be viewed with caution.

The main aims of this chapter are (1) to present and compare taxic and phylogenetically corrected estimates of early African hominin palaeodiversity, and (2) to assess the confidence in the first and last appearance dates used to create these palaeodiversity curves. (For brevity, the terms palaeodiversity and diversity are henceforth synonymous.)

2.1.1 Previous studies of hominin diversity

Previous studies of hominin diversity can be divided into three groups: first, those that provide a summary of hominin taxonomy (e.g., Wood & Lonergan, 2008; Wood & Boyle, 2016); second, those that question the number of taxa in the hominin fossil record (e.g., Foley, 1991; Fleagle, 1995; Begun, 2004); and third, those that assess the factors controlling diversity (e.g., Foley, 2005; Hopley, 2018, *in review*). Within the latter group, the vast majority emphasise change in climatic conditions as the primary—and often single—driver of diversification (e.g., Dart, 1925; Vrba, 1988, 1995; deMenocal, 1995, 2004; Foley, 1994; Kimbel, 1995; Potts, 1996, 1998; Kingston, 2007; Grove, 2012; Shultz & Maslin, 2013).

The first testable hypothesis linking climate and diversity combined macroevolutionary theories of punctuated equilibrium and species selection (e.g., Eldredge & Gould, 1972; Vrba, 1984) to produce the turnover-pulse hypothesis (Vrba, 1985*a*). The central tenet of the turnover-pulse hypothesis is that species are habitat-specific and persist under a specific range of environmental conditions (e.g., temperature, rainfall, vegetation; Vrba, 1993). Stenotopic (specialist) taxa are more affected by climatic change than eurytopic (generalist) taxa due to their narrower range of habitable conditions and, therefore, are more likely to go extinct during periods of climatic change. Vrba argues that the apparent pulses of bovid turnover (speciation, extinction, and migration; Vrba, 1985*b*) at 3.6, 2.7–2.5, and 1.8 Ma, can be explained by pulses of climatic change caused by shifts in orbital (Milankovitch) cyclicity, global cooling and the intensification of Northern Hemisphere glaciation, African aridification, and the development of the Walker Circulation (see Maslin *et al.*, 2014 for a review).

Behrensmeyer *et al.* (1997) provided an anthropogenic explanation for this trend: they argued that these apparent turnover pulses are the result of temporal heterogeneity in collection effort as, in the case of bovids, diversity and the number of bovid-bearing localities correlate significantly at East Turkana (Spearman's $r = 0.891$), West Turkana ($r = 0.769$), and the Omo ($r = 0.636$; Behrensmeyer *et al.*, 1997). The appearance of pulsed turnover in the bovid fossil record is thus explained as variation in the number of sampled localities that preserve bovid fossils through time (Behrensmeyer *et al.*, 1997). Other shortcomings of the turnover-pulse hypothesis include: (1) both empirical data and simulations demonstrate that the appearance of pulsed turnover is an artefact of an incomplete fossil record (McKee, 1996, 2001); (2) turnover in the fossil record is generally gradual after sampling standardisation (Bibi & Kiessling, 2015); (3) turnover pulses are not concurrent in multiple mammalian lineages (White, 1995; Werdelin & Lewis, 2005; Frost, 2007); and, (4) turnover pulses can arise from entirely stochastic evolutionary processes (Barr, 2017).

Subsequent hypotheses focus not on directional shifts in climatic conditions but climatic variability. Potts (1996, 1998) and other proponents (e.g., Bobe *et al.*, 2002; Bobe & Behrensmeyer, 2004; Bonnefille *et al.*, 2004; Owen *et al.*, 2008; Grove, 2011) argued that climatic variability had a greater impact on hominin diversity during the Plio-Pleistocene. Recently, Potts & Faith (2015) linked periods of pronounced variability identified in regional aridity records—each ~0.2 Myr in duration—to diversification in the hominin lineage. They found that the first appearance datum (FAD) of most (8 out of 9) eastern African hominins coincides with periods of high climatic variability, and this overlap is significantly more likely than chance. They did not find the same pattern for each last appearance datum (LAD), mirroring the earlier findings of Grove (2012) who suggested global-scale extrinsic factors drive speciation, but local-scale extrinsic factors or subsequent inter-species competition drive extinction (*contra* Foley, 1994).

Building on this, Maslin & Trauth (2009) and Trauth *et al.* (2010) found that acute orbital forcing made the East African Rift System (EARS) sensitive to precessional or half-precessional (~10–20 ka) pulses of wetter and more variable climate, and that episodes of extreme humidity punctuated the gradual drying trend of the Plio-Pleistocene. In response, local climate became substantially less arid and ephemeral deep lakes occurred throughout the EARS during these pulses (Trauth *et al.*, 2010). Shultz & Maslin (2013) have since argued that the appearance and disappearance of deep lakes and peaks and troughs in hominin diversity are coincident, and that lake levels drove hominin diversification and migration by population vicariance and allopatric speciation.

Each of the above hypotheses suffers the same limitations. First, if an event in a palaeoclimate record and the hominin fossil record are coincident, a hypothesis is proposed to explain the possible mechanism for this relationship. However, concurrence does not prove consequence (*cum hoc ergo propter hoc*). Second, in all previous studies, fluctuations in the number of hominin taxa are accepted as genuine changes in diversity, even though it is possible that such fluctuations reflect variation in the quality of the fossil record. None of these previous studies considered the effect of sampling on diversity patterns or alternative non-climatic explanations. This has resulted in an abundance of Court Jester (Barnosky, 2001) or climate-driven hypotheses for hominin evolution but no significant improvement in our understanding or meaningful consensus on the link between climate and evolution (Smith & Wood, 2017; Hopley, 2018, *in review*). Some authors have gone so far as to say that because hominin FAD and LAD occur in periods of high and low climatic variability with equal probability, the hominin fossil record is an un-biased record of evolutionary events (Potts & Faith, 2015:13). Others still have ignored the issue of sampling entirely (e.g., Maslin & Trauth, 2009; Shultz & Maslin, 2013). Either these authors are ignorant of the limitations inherent to fossil data or they deem a consideration of fossil record quality and alternative diversity estimates unnecessary.

2.1.2 Current hominin diversity estimates and their problems

Current hominin diversity estimates use raw (empirical, face-value, uncorrected) counts of the number of observed taxa through time. If a taxon is observed in the stratigraphic record at a particular time, and later at another, its range is assumed to span the gap, and the number of taxa present are then summed to produce a depiction of overall diversity through time. This technique is based on the observed stratigraphic range of taxa and has been termed the *taxic* approach or taxic diversity estimate (TDE; Levinton, 1988). Whilst this method is simple and requires minimal information, it has been shown to be biased

by sampling heterogeneity and other sources of error in the sedimentary rock record, and may provide grossly inaccurate estimates of diversity (Raup, 1972, 1976*b*; Benton *et al.*, 2011).

Smith (1994) proposed that, in principle, these problems can be partially corrected if an understanding of the phylogenetic relationships of taxa are taken into account. If, for example, a lineage is not observed in the stratigraphic record but its presence can be inferred from a phylogeny, based on the assumption that two taxa must have split from their common ancestor at the same time (Norell, 1992), a cladistically-implied, as-yet unsampled *ghost lineage* (Norell, 1993) is incorporated into the diversity estimate. The number of taxa are then estimated based on the number of lineages (observed and inferred) in each time bin. This technique has been termed the *phylogenetic* approach or phylogenetic diversity estimate (PDE; Smith, 1994).

The PDE approach, however, is not entirely free from criticism (Foote, 1996*a*; Wagner, 2000*a*, 2000*b*; Wagner & Sidor, 2000).

First, many cladistic analyses do not sample all of the taxa within a particular clade. This can be problematic if the taxa excluded from an analysis are non-randomly distributed in the stratigraphic record (e.g., the earliest members of a clade), and could increase the probability of errors in tree topology.

Second, the sampling correction is uni-directional: a PDE corrects the first appearance of taxa by extending origination times backwards but does not offer a corresponding correction that extends extinction times forward (Foote, 1996*a*; Wagner, 2000*a*), creating higher diversity earlier in time.

Third, polytomies in a phylogeny will also create error as the ghost lineage of all taxa in the polytomy will be extended to the FAD of the oldest taxon (Upchurch & Barrett, 2005).

Fourth, most PDEs are generated using the basic method (Laurin, 2004), where the oldest fossil of a clade is used to represent the node age (Norell, 1992; Smith, 1994). However, this will have the negative effect of producing zero-length branches (ZLBs) when the youngest members of a clade are nested relatively basally within the cladogram (Bell & Lloyd, 2015). Branch-sharing methods were developed to eliminate ZLBs. They include either (1) adding a share of the first directly ancestral positive branch length proportional to the amount of evolutionary change along that branch (Ruta *et al.*, 2006), or (2) sharing the first directly ancestral positive branch length equally among branches (Brusatte *et al.*, 2008). More recently, probabilistic methods have become increasingly popular (Bapst & Hopkins, 2017). For example, the *cal3* time-scaling method constrains each node between the age of the previous node and the FAD of the oldest taxon (Bapst, 2013). Branch lengths are then stochastically sampled from a distribution based on the most likely amount of missing evolutionary history, determined by the rate of speciation, extinction, and sampling (Bapst, 2013). Alternatively, the Hedman method is a Bayesian approach where the age of each node is sampled from a uniform distribution based on the age of the earliest representative of the sister-taxon (Hedman, 2010; Lloyd *et al.*, 2016).

Finally, current cladistic methods produce cladograms under the assumption that speciation is bifurcating: one parent taxon produces two daughter taxa and, in the process, becomes extinct (Norell, 1992; Smith, 1994; Wagner & Erwin, 1995). If ancestors are included in a cladogram, they will be interpreted as the sister-taxon to their descendants, leading to the inference of an incorrect ghost lineage and a higher PDE (Lane *et al.*, 2005). The probability of sampling a direct ancestor was claimed to be low enough to be negligible (e.g., Norell, 1993). However, it has been shown that the probability of sampling

Table 2.1: Stratigraphic range and confidence interval data for early hominin taxa. First and last appearance dates were taken from Wood & Boyle (2016) and updated using the Hominin Fossil Database (Appendix C). Upper and lower 95% confidence intervals (CIs) were calculated using equation (2) in Marshall (1990) and an updated database of horizon counts (H) from Hopley (2018, *in review*). Those taxa for which a meaningful CI cannot be calculated (i.e., $H \leq 2$) are represented by a hyphen (–).

Taxon	FAD (Ma)	LAD (Ma)	Stratigraphic range (Ma)	H	CI
Sahelanthropus tchadensis	7.00	7.00	0.00	1	–
Orrorrin tugenensis	6.10	5.70	0.40	4	0.92
Ardipithecus kadabba	6.30	5.20	1.10	7	0.92
Ardipithecus ramidus	4.51	4.30	0.21	3	1.01
Kenyanthropus platyops	3.54	3.35	0.19	4	0.44
Australopithecus anamensis	4.20	3.90	0.30	8	0.20
Australopithecus afarensis	3.70	3.00	0.70	22	0.13
Australopithecus bahrelghazali	3.58	3.58	0.00	1	–
Australopithecus deyiremeda	3.50	3.30	0.20	2	–
Australopithecus africanus	3.00	2.40	0.60	4	1.38
Australopithecus garhi	2.50	2.50	0.00	1	–
Australopithecus sediba	1.98	1.98	0.00	1	–
Paranthropus aethiopicus	2.66	2.30	0.36	13	0.13
Paranthropus boisei	2.30	1.30	1.00	31	0.13
Paranthropus robustus	2.00	1.00	1.00	8	0.68
Homo habilis	2.33	1.65	0.68	13	0.24
Homo rudolfensis	2.00	1.78	0.22	4	0.51
African *Homo erectus*	1.90	0.78	1.12	26	0.18

ancestor-descendant pairs each at least once is far from negligible (e.g., Foote, 1996*b*). Sophisticated methods for inferring relationships while simultaneously assessing the likelihood of ancestors (support for budding cladogenesis or anagenesis) (e.g., Gavryushkina *et al.*, 2017), and *a posteriori* methods that time scale a cladogram while simultaneously assessing the likelihood of ancestors (Bapst, 2013), have recently emerged.

Despite the limitations of the PDE approach, it has been shown to capture more real diversity than TDE in model simulations (Lane *et al.*, 2005), and consistently outperforms TDE up to a tree topology error rate of 50% (Brocklehurst, 2015). No method of estimating diversity is entirely un-biased, so here a pluralistic approach is taken and both diversity estimates, multiple phylogenies, and a probabilistic time-scaling method are used.

2.2 Methodology

2.2.1 Diversity metrics

Taxic diversity estimate. Taxic methods assess diversity by counting the number of observed taxa in a series of time bins based on their stratigraphic range. The conservative FAD and LAD data set (n = 18) in Wood & Boyle (2016:40) formed the basis of the TDE (Table 2.1). 0.25-Myr time bins between 7.0 and 1.0 Ma were used to compile TDE (producing 24 time bins). If a FAD or LAD falls on the boundary of

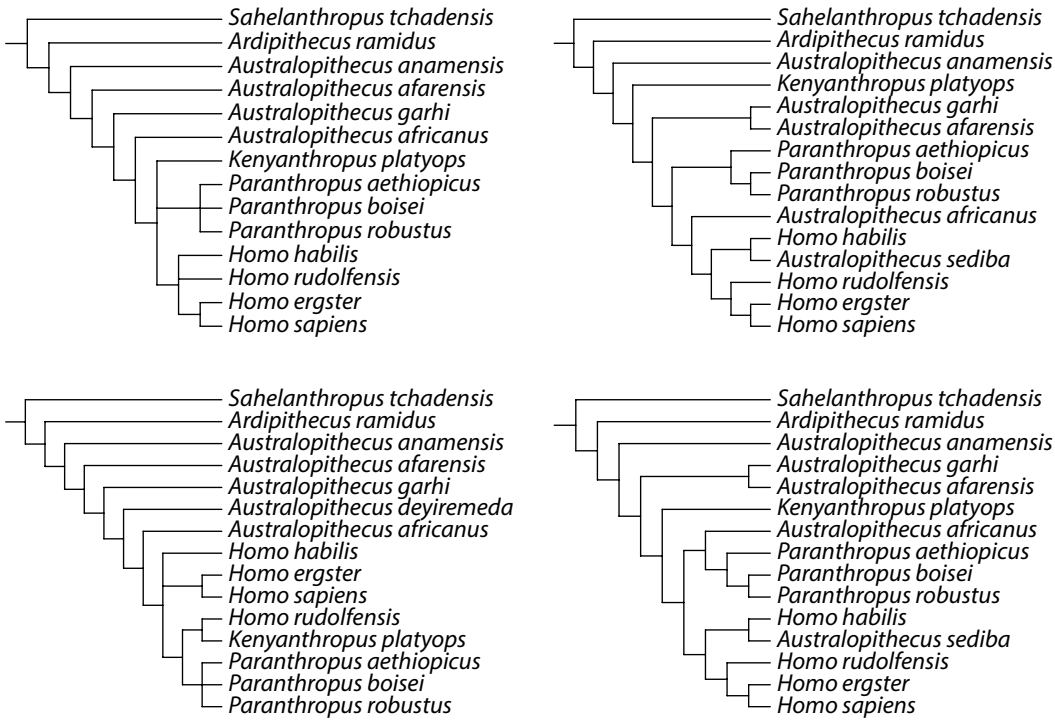

Figure 2.1: Early hominin cladistic relationships according to the four most recent studies. (**a**) Strait & Grine (2004). (**b**) Dembo *et al.* (2015). (**c**) Haile-Selassie *et al.* (2015). (**d**) Dembo *et al.* (2016).

a time bin (e.g., 2 Ma), that taxon is deemed present in the younger bin only (in this case, 2–1.75 Ma).

Phylogenetic diversity estimate. Phylogenetic methods assess diversity by counting the number of lineages (observed and inferred) in a series of time bins using a dated (time scaled) phylogeny. PDEs were generated using equivalent time bins and the cladograms of Strait & Grine (2004), Dembo *et al.* (2015), Haile-Selassie *et al.* (2015), and Dembo *et al.* (2016), as these cladograms represent the most recent phylogenetic analyses and include the largest sample of early African hominin taxa (Fig. 2.1). Polytomies in the strict consensus (Strait & Grine, 2004) and majority-rule (Haile-Selassie *et al.*, 2015) cladograms were resolved based on the order of first appearance using the paleotree function timeLadderTree (Bapst, 2012). This had little effect on each PDE compared to randomly resolving polytomies and mitigated the fact that most bifurcating resolutions of a consensus tree are not themselves among the optimal topologies used to create that consensus (see Fig. 4 in Bell & Lloyd, 2015). The Dembo *et al.* (2015, 2016) cladograms were fully resolved so required no further treatment. In order to maximise comparability between diversity estimates, Eurasian taxa and taxa younger than 1 Ma were pruned from each cladogram (after time-scaling), as the focus here is early African hominin diversity dynamics.

The phylogenetic method requires that branch lengths are proportional to time. To do this, each tree was time-scaled using taxon duration data and the three-rate-calibrated time-scaling (*cal3*) method (Bapst, 2013). The *cal3* method constrains the age of each node between the date of the previous node (except for the root) and the FAD of daughter lineages. The age of each node is then calculated by the probability density of the amount of unobserved evolutionary history implied by each node age, a

probability dependent on rates of speciation, extinction, and sampling in the fossil record (Bapst, 2013). These densities are then used to stochastically sample the possible ages for each node (Bapst, 2013). Speciation, extinction, and sampling rate were first determined empirically using the durationFreq function in paleotree (Bapst, 2012). This function applies a maximum likelihood optimisation to the distribution of taxon durations and returns the best fitting sampling probability and extinction rate to explain the distribution (Foote, 1997). Speciation and extinction rate are assumed equal given the tight relationship observed in the fossil record (Stanley, 1979). Results presented for the calculation of speciation, extinction, and sampling rate are based on the taxon durations shown in Table 2.1, as the main interest here is the sampling and diversification of early hominin taxa. This method, however, produced an estimated sampling rate that differed markedly from previous estimates for mammals (Bapst & Hopkins, 2017). Moreover, the frequency-ratio or freqRat method (Foote & Raup, 1996) did not provide a meaningful estimate of sampling because the frequency distribution of taxon durations violated model assumptions (the equations of Foote & Raup (1996) require the frequency distribution of the log of taxon durations is linear). To combat this, the sampling rate Tavaré *et al.* (2002) reported for primates (0.023 per lineage Ma) was used plus a maximum root age of 8 Ma (reviewed in Bradley, 2008). Because node ages are stochastically picked from a distribution defined by the probability of different amounts of unobserved evolutionary history, no single time-scaled tree is correct. Therefore, to account for uncertainty in the age of each node and improve analytical rigour, 1000 time-scaled trees were generated. The median diversity across all 1000 trees was calculated along with confidence intervals based on two-tailed 95% upper and lower quantiles (Bapst, 2012). It is this median PDE which is used in the statistical tests. Interestingly, the *cal3* method produced median node ages that correlate strongly with the node ages produced by Dembo *et al.* (2015) ($r \geq 0.98$, $p < 0.001$) and Dembo *et al.* (2016) ($r \geq 0.93$, $p < 0.001$) in their Bayesian tip-dating analyses. Tip-dating methods tend to produce node ages that are several million years—sometimes tens of millions of years—older than the minimum (i.e., fossil) divergence date (Bapst *et al.*, 2016), whereas *cal3* node age distributions tend to be similar to the minimum divergence date (Bapst & Hopkins, 2017). The agreement between tip-dating and *cal3* is, therefore, likely a result of the range of possible node ages being tightly constrained by the FAD and input topologies.

Several studies have argued that *Australopithecus anamensis* and *Australopithecus afarensis* (Kimbel *et al.*, 2006), and *Paranthropus aethiopicus* and *Paranthropus boisei* (Wood & Schroer, 2013), represent anagenetically evolving lineages (or chronospecies). It has also been suggested that *Ardipithecus kadabba* and *Ardipithecus ramidus* are probable chronospecies (White *et al.*, 2009:78), however, this has not been formally tested and, as a result, is not generally accepted (Wood & Grabowski, 2015). To test whether the treatment of chronospecies as sister-taxa distorts PDEs, those trees that support hypotheses of ancestry (i.e., that reconstruct these taxa as time-successive sister-taxa) were re-drawn to represent the chronospecies as single lineages. These trees were time scaled as above and the median diversity calculated across all 1000 trees.

2.2.2 Statistical tests

Two statistical tests were used to compare each diversity estimate. Spearman's rank correlation coefficient (ρ) is a non-parametric measure of the correlation between two rank-ordered variables, and the Kendall tau rank correlation coefficient (τ) is a non-parametric measure of the extent to which two vari-

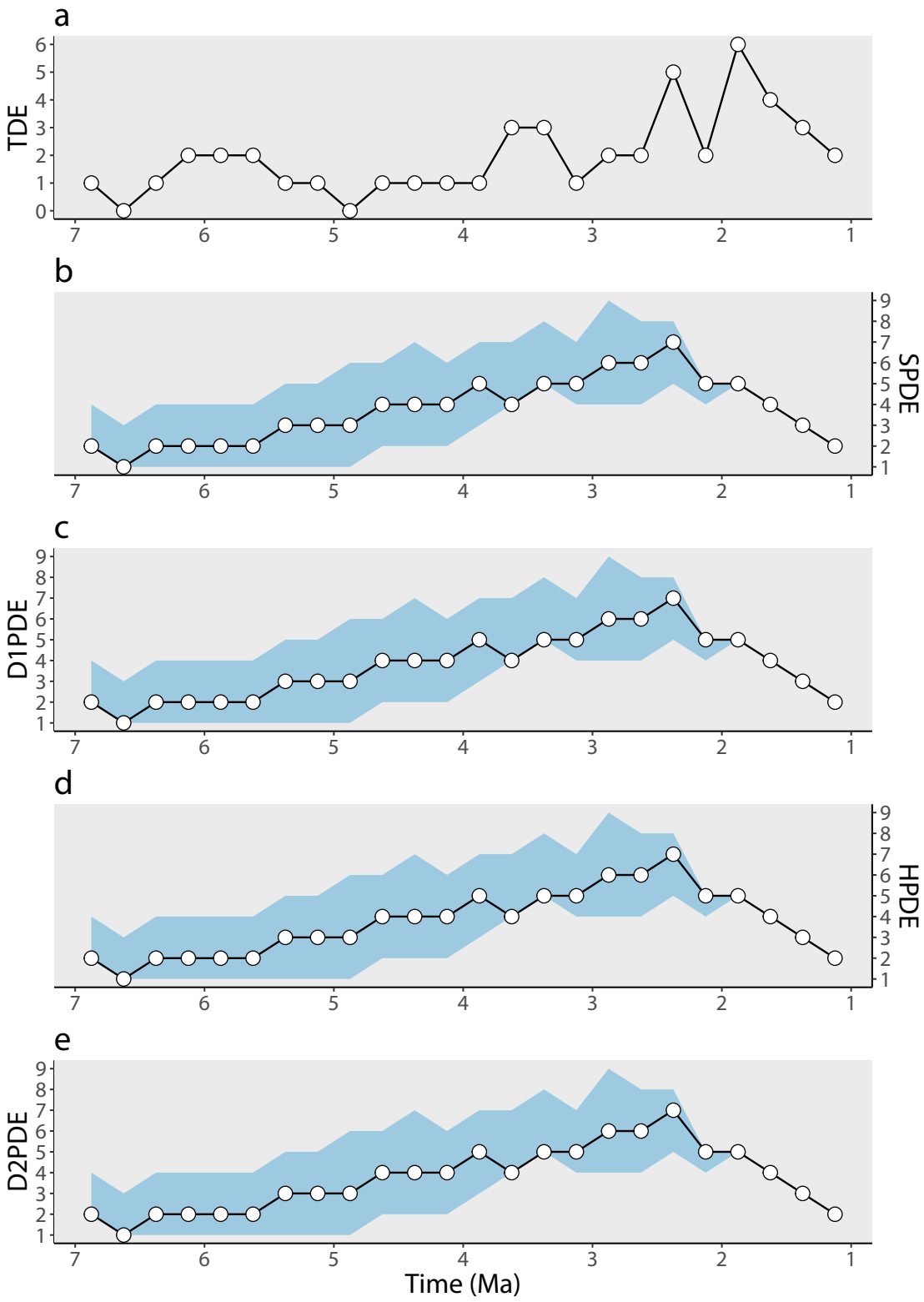

Figure 2.2: Early hominin diversity estimates through geological time. (**a**) Taxic diversity estimate (TDE). (**b**) Strait & Grine (2004) phylogenetic diversity estimate (SPDE). (**c**) Dembo *et al.* (2015) phylogenetic diversity estimate (D1PDE). (**d**) Haile-Selassie *et al.* (2015) phylogenetic diversity estimate (HPDE). (**e**) Dembo *et al.* (2016) phylogenetic diversity estimate (D2PDE). Total bin diversity (including range-through taxa) is plotted at the midpoint age of each time bin. The blue envelopes (**b**–**e**) represent upper and lower 95% confidence intervals based on 1000 time-scaling replicates.

ables change synchronously (Hammer & Harper, 2006). Time series were de-trended and corrected for temporal autocorrelation by generalised differencing (GD) prior to regression (McKinney, 1990; R code from G.T. Lloyd: http://www.graemetlloyd.com/methgd.html). Long-term trends and autocorrelation (i.e., the effects of the statistical overprinting of an interval's diversity onto the interval proceeding it) tend to result in spurious detection of correlation between time series (McKinney, 1990; Alroy, 2000; Benson & Butler, 2011) and must be removed prior to performing statistical tests. First, the presence of a long-term trend is assessed by regressing the time series against the midpoint age of each time bin (in Ma). If the regression line provides a significant fit to the data ($p < 0.05$), residuals from the regression represent a statistically de-trended time series. Second, autocorrelation is evaluated by regressing each point in the time series against the point from the interval directly preceding it (i.e., a lag of 1). If the regression line is non-significant ($p > 0.05$) then no statistical autocorrelation is present and no further action is required. If autocorrelation is present then GD is implemented using Equation (2.1): a generalised differenced time series (t_{GD}) is produced by multiplying each point in the time series at a lag of 1 (t_{i-1}) by the slope of the regression line or autocorrelation coefficient (a), and subtracting this value from the original (or de-trended) time series (t_i).

$$t_{GD} = t_i - at_{i-1} \tag{2.1}$$

The significance of correlations were evaluated based on original p-values and p-values adjusted for the implementation of multiple tests using the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995). The problem of Type I statistical errors (the incorrect rejection of a true null hypothesis) is a risk for multiple pairwise comparisons where the sheer number of tests means that correlations might be recognised erroneously (false positives). To perform the Benjamini-Hochberg procedure p-values are first sorted in ascending order, and then each observed p-value is divided by its percentile rank (Noble, 2009). GD and FDR correction are standard practice in palaeobiology (see McGowan & Smith (2011) and references therein).

Linear and exponential models were fit to the net positive diversification component of each PDE to identify whether diversification is best described by an additive or exponential model (Benton, 2007). Non-linear regression models were constructed using the nlsfit function in R package easynls, with the Akaike Information Criterion (AIC) and Akaike weights (w_i) used to quantify the likelihood of each model. The diversity analyses were carried out in R 3.5.0 (R Development Core Team, 2018) and an R script for Chapter 2 can be found in Appendix B.

2.2.3 Confidence intervals on stratigraphic ranges

Confidence intervals (CIs) are typically calculated using the FAD and LAD of a taxon and the number of fossil-bearing horizons from which that taxon is known (e.g., Strauss & Sadler, 1989; Marshall, 1990, 1994, 1997). They do not indicate a taxon's true stratigraphic range but provide an estimate of the statistical confidence that a taxon's absence is not simply the result of a failure to sample that taxon in appropriate facies (Smith, 1994). To determine the influence of sampling on the known stratigraphic range of early hominins, 95% CIs were calculated on their entire range according to Equation (2) in Marshall (1990). The difference between a taxon's lower CI and FAD, and upper CI and LAD, can be

regarded as a speciation window (SW) and extinction window (EW), respectively (a period in which one can be confident the taxon of interest genuinely appeared or disappeared given the sampling of their fossil record). Marshall's (1990) model requires that fossils: (1) are randomly distributed through time; (2) are statistically independent of one another; and, (3) have a constant recovery potential (Marshall, 1990). These assumptions have not been tested for the hominin fossil record so Wald-Wolfowitz runs tests were performed (in R) to investigate the null hypothesis of randomness and independence in horizon counts through geological time (Hammer & Harper, 2006). In spite of these assumptions, CIs serve as a valuable check against reading a pattern of first and last appearances in the fossil record as the literal dates of speciation and extinction events. CIs were calculated in PAST 3.18 (Hammer *et al.*, 2001) using an up-to-date database of horizon counts (Hopley, 2018, *in review*).

2.3 Results

2.3.1 Observed hominin taxic diversity

The updated TDE (Fig. 2.2*a*) is similar to the diversity curve of previous analyses (e.g., Foley (2016), calculated using 0.5 Ma time bins), displaying three peaks: first, in the middle Pliocene (3.6 Ma); second, in the early Pleistocene (2.4 Ma); and third, in the early-middle Pleistocene (1.9 Ma) (dates refer to the midpoint age of each time bin rounded to one decimal place). Peak TDE ($n = 6$) occurs during the early-middle Pleistocene. These peaks are separated by troughs ca., 3.0 Ma and 2.0 Ma, and low but fluctuating diversity during the late Miocene and early Pliocene (7.0–4.0 Ma). Two time bins in the late Miocene and early Pliocene (~6.8–6.5 Ma and ~5.0–4.8 Ma, respectively) do not contain any identifiable hominin fossils and therefore have zero taxic diversity. Overall, TDE displays a "classic spiky" curve indicative of a genuine signal of speciation and extinction overlain by major fluctuations in sampling (Benton *et al.*, 2013:23). This reinforces the need to assess early hominin diversity using alternative, sampling-corrected estimates.

2.3.2 Phylogenetic diversity

The Strait & Grine (2004) PDE (SPDE), Dembo *et al.* (2015) PDE (D1PDE), Haile-Selassie *et al.* (2015) PDE (HPDE), and Dembo *et al.* (2016) PDE (D2PDE) are shown in Fig. 2.2*b–e*. Each PDE converges on a diversification pattern qualitatively different to the TDE. SPDE, D1PDE, and D2PDE display a long-term increase in diversity from the late Miocene to the early Pleistocene, each reaching peak diversity ($n = 7$, 8, and 8, respectively) at 2.4 Ma. HPDE differs from the other curves in two aspects. First, although it also displays a long-term increase from the late Miocene onwards, peak diversity occurs during the middle Pliocene (~3.4 Ma). Second, where diversity peaks and then begins to decline in SPDE, D1PDE, and D2PDE, HPDE remains high from 3.4 to 2.4 Ma, after which diversity then begins to decline (Fig. 2.2). Finally, each PDE generally implies 1 or 2 more taxa per bin than the TDE (Fig. 2.2) and thus no time bins have zero diversity.

2.3.2.1 Statistical comparisons between taxic and phylogenetic diversity

There is no statistically significant correlation between TDE and any PDE (Fig. 2.3 and Table 2.2).

Table 2.2: Results of the statistical analyses comparing taxic and phylogenetic diversity estimates. See text for an explanation of the abbreviations of each diversity estimate. Statistically significant correlations are shown in bold. *Significant at $p \leq 0.05$. **Significant at $p \leq 0.05$ after false discovery rate (FDR) correction (Benjamini & Hochberg, 1995)

Comparison	Spearman's ρ	Kendall's τ
TDE *versus* SPDE	0.286	0.225
TDE *versus* D1PDE	0.298	0.202
TDE *versus* HPDE	0.299	0.217
TDE *versus* D2PDE	0.323	0.233
SPDE *versus* D1PDE	**0.792****	**0.660****
SPDE *versus* HPDE	**0.672****	**0.486****
SPDE *versus* D2PDE	**0.767****	**0.613****
D1PDE *versus* HPDE	**0.789****	**0.605****
D1PDE *versus* D2PDE	**0.967****	**0.905****
HPDE *versus* D2PDE	**0.745****	**0.542****
SPDE *versus* $SPDE_{AN}$	**0.568***	**0.486****
HPDE *versus* $HPDE_{AN}$	**0.686****	**0.557****

2.3.2.2 Statistical comparisons between different phylogenies

Each PDE is strongly correlated with every other PDE, both before and after false discovery rate (FDR) correction (Fig. 2.4 and Table 2.2). Further, SPDE and HPDE both correlate significantly with their equivalent PDE under an anagenetic scenario ($SPDE_{AN}$ and $HPDE_{AN}$). However, SPDE and $SPDE_{AN}$ do not correlate significantly after FDR correction (Table 2.2).

2.3.2.3 Best-fit diversification model

There is strong support for an exponential increase in hominin diversification for each PDE (Table 2.3), corresponding to a mean net diversification rate of 0.39 species/Myr.

2.3.3 Comparing confidence intervals on stratigraphic ranges

Per-taxon horizon counts are shown in Fig. 2.5 and through geological time in Fig. 2.6 (alongside the frequency of first and last appearance dates). Runs tests demonstrate that horizon counts through geological time depart significantly from randomness ($p < 0.001$). Hopley (2018, *in review*) reported a temporal coincidence between TDE and the number of hominin-bearing horizons (HBH) and a statistically significant positive correlation if also found here ($\rho = 0.779$, $p < 0.001$).

CIs on early hominin stratigraphic ranges are shown in Table 2.1. The mean CI is 0.53±0.12 (standard error) Ma ($n = 13$), comparable to the mean stratigraphic range (0.45 ± 0.09 Ma, $n = 18$). However, the mean duration of African Neogene large mammals is 1.3–1.4 Ma (Vrba, 2000; Bibi & Kiessling, 2015), triple that of an early hominin. In the Turkana Basin, Kenya, species stratigraphic ranges are statistically indistinguishable among bovids, cercopithecids, equids, felids, hippopotamids, hyaenids, and suids (Bibi & Kiessling, 2015). Longer durations do occur in antilopins, elephantids, and giraffids and shorter durations in hominins (Bibi & Kiessling, 2015). However, the influence of sampling on the longevity of hominins remains poorly understood (but see Hopley, 2018, *in review*).

Australopithecus afarensis, *Paranthropus aethiopicus*, and *Paranthropus boisei* have the greatest statistical confidence on their stratigraphic range (0.13 Ma). *Australopithecus anamensis*, *Homo habilis*, and African *Homo erectus* are represented by 8, 13, and 26 horizons, corresponding to CIs of 0.20, 0.24, and 0.19 Ma, respectively. Most taxa, however, are poorly documented, with CIs ranging from 1.38 Ma for *Australopithecus africanus*, 1.01 Ma for *Ardipithecus ramidus*, and 0.91 Ma for *Orrorin tugenensis* and *Ardipithecus kadabba* (Table 2.1). The uncertainty in the stratigraphic range of many early hominins is exemplified by the fact that *Sahelanthropus tchadensis*, *Australopithecus bahrelghazali*, *Australopithecus garhi*, and *Australopithecus sediba* are singletons (taxa known only from a single horizon; Fig. 2.5) and therefore cannot be assigned a stratigraphic range or meaningful CIs. This is also the case for the recently described *Australopithecus deyiremeda* ($H = 2$). There is no statistically significant correlation between CI and FAD ($r = 0.478$, $p = 0.098$, $n = 13$), implying that confidence in the known stratigraphic range of early hominins is independent of geological time.

2.4 Discussion

2.4.1 Early hominin diversity patterns: is diversification pulsed?

The taxic and phylogenetic diversity curves in Fig. 2.2 do not agree on many aspects of early hominin evolutionary history. In particular, TDE and PDE differ in both the timing and magnitude of peaks and troughs in early hominin diversity. TDE and PDE do agree that diversity reached a peak at ~2.4 Ma, implying that hominins may have genuinely experienced maximum taxonomic diversity at this time. However, the discrepancy between TDE and PDE for the majority of geological time potentially represents problems inherent to the sampling of the fossil record.

Late Miocene. The earliest hominin fossils currently occur in the late Miocene. This includes *Sahelanthropus* (Brunet *et al.*, 2002)—repeatedly reconstructed as a hominin despite possessing several hominine (particularly *Gorilla*) characteristics (Strait *et al.*, 2007)—and *Orrorin* (Senut *et al.*, 2001). Recently, the description of a mandibular premolar from the Adu-Asa Formation, Ethiopia, extends the first appearance of *Ardipithecus* to 6.3 Ma (Simpson *et al.*, 2015). The fossil record apparently indicates relatively low but fluctuating diversity during the late Miocene. Each PDE also supports low diversity,

Table 2.3: Competing models for early hominin diversification. For the net positive diversification component of each phylogenetic diversity estimate (PDE), the Akaike weight (w_i) shows the relative likelihood for each diversification model. R^2 shows the variance explained by each model. The slope corresponds to the diversification rate (in species/Myr) for each model. For the SPDE, D1PDE, and D2PDE net positive diversification spans from 7.0–2.4 Ma, while for the HPDE net positive diversification spans from 7.0–3.4 Ma (Fig. 2.2).

	Linear model				Exponential model			
	w_i	R^2	slope	p (slope)	w_i	R^2	slope	p (slope)
SPDE	0.17	0.93	−1.14	< 0.001	0.83	0.94	−0.31	< 0.001
D1PDE	0.00	0.86	−1.38	< 0.001	1.00	0.95	−0.40	< 0.001
HPDE	0.16	0.92	−1.91	< 0.001	0.84	0.97	−0.47	< 0.001
D2PDE	0.00	0.86	−1.19	< 0.001	1.00	0.94	−0.37	< 0.001

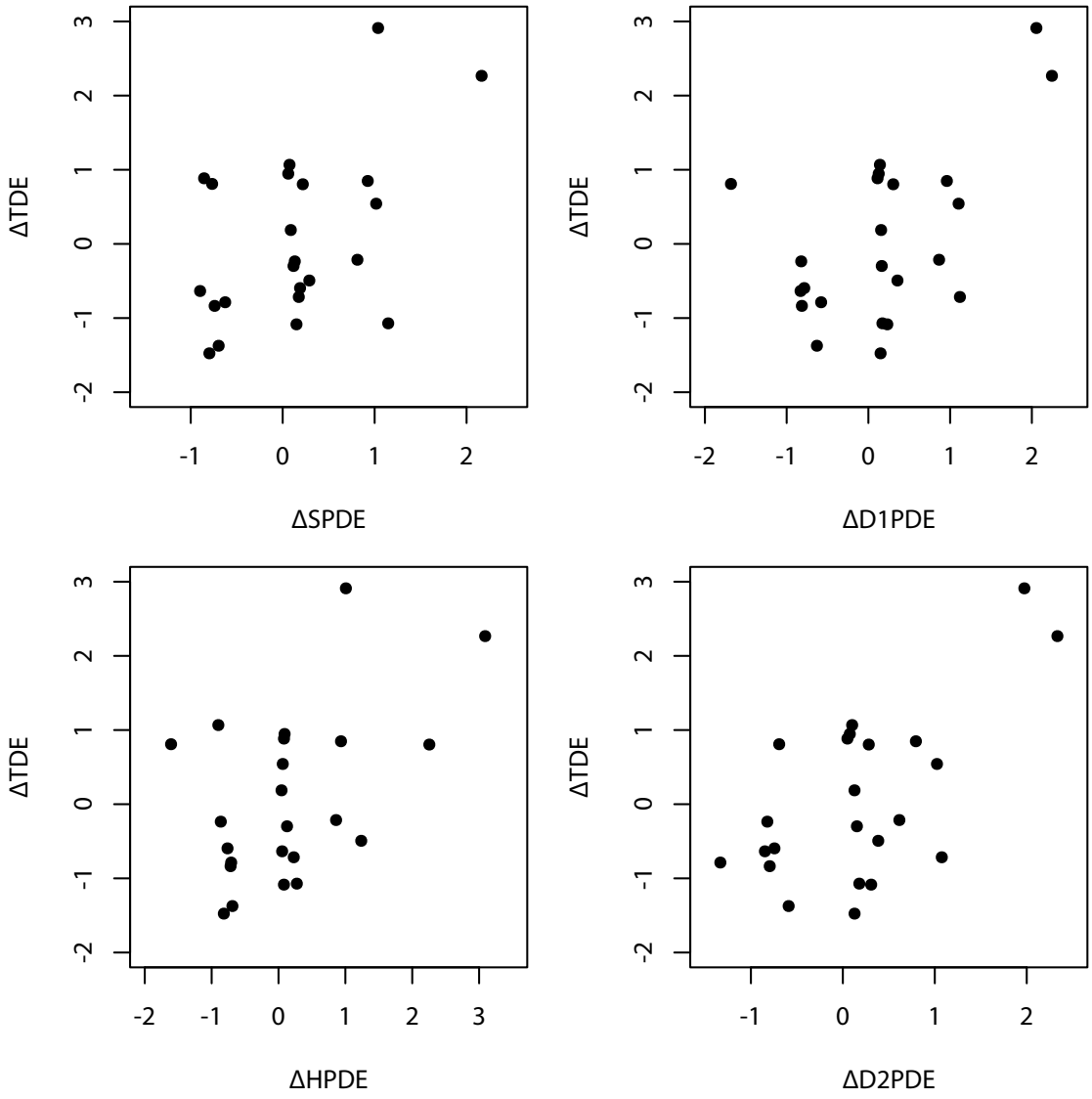

Figure 2.3: Scatter plots showing the relationships between the taxic diversity estimate (TDE) and each phylogenetic diversity estimate (PDE). (**a**) TDE against SPDE. (**b**) TDE against D1PDE. (**c**) TDE against HPDE. (**d**) TDE and D2PDE. See Table 2.2 for correlation coefficients and *p*-values.

however, short-term fluctuations are not supported by ghost lineage data. This suggests that taxic diversity patterns are influenced by the number of opportunities to observe fossils (see Chapter 3), and that failure to sample rather than genuine absence is responsible for many short-term fluctuations in TDE (Fig. 2.2*a*). Phylogenetic estimates, on the other hand, imply near-stasis during the late Miocene and, as a result, 1 to 2 lineages in the 6.8–6.5 Ma time bin (where TDE equals zero). Taxon sampling is known to influence PDEs, particularly if those taxa excluded from a cladistic analysis are rogue taxa (their presence/absence substantially alters tree topology) or non-randomly distributed in the stratigraphic record (Upchurch & Barrett, 2005). Both *Orrorin* and *Ardipithecus kadaba* are excluded in each of the phylogenies analysed in the current study. Given the uncertainty in the range of these taxa (0.91 Ma) and their unknown placement in hominin phylogeny, PDEs during this period are definitely an under-

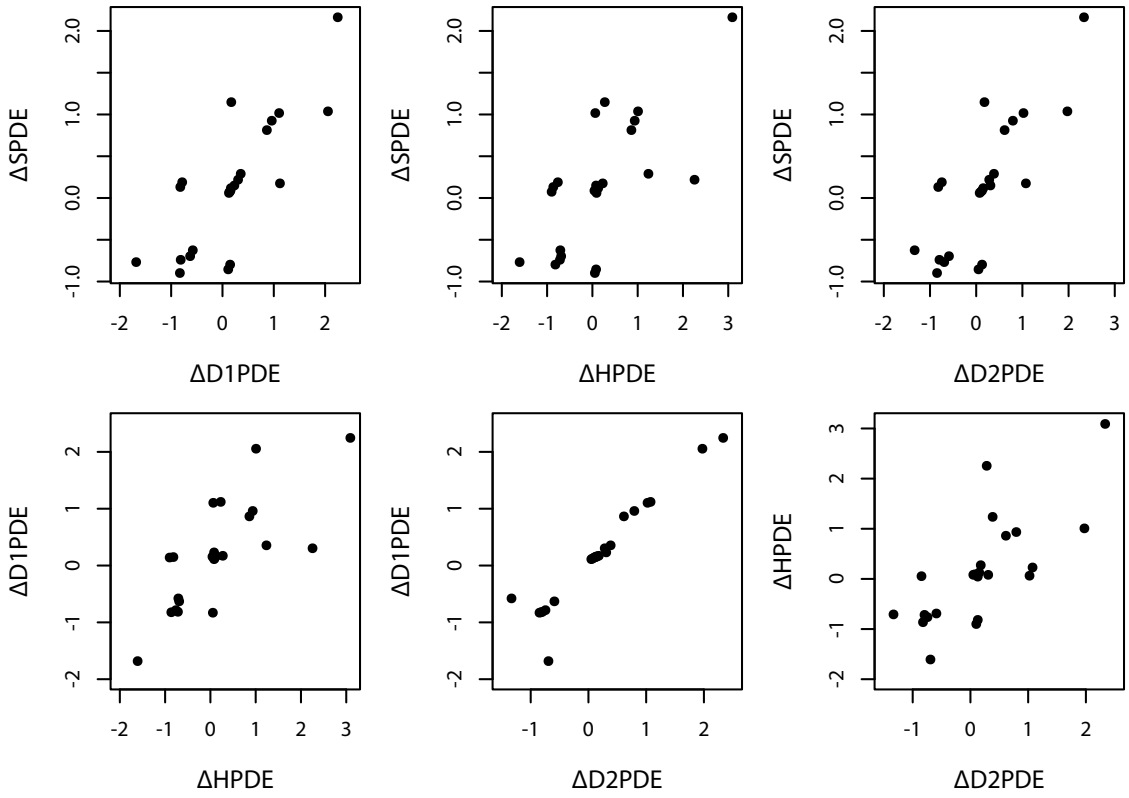

Figure 2.4: Scatter plots showing the relationships between each phylogenetic diversity estimate (PDE). Δ indicates that the time series has undergone generalised differencing prior to statistical testing(**a**) SPDE against D1PDE. (**b**) SPDE against HPDE. (**c**) SPDE against D2PDE. (**d**) D1PDE and HPDE. (**e**). D1PDE against HPDE (**f**) HPDE against D2PDE. See Table 2.2 for correlation coefficients and *p*-values.

estimate of true diversity—assuming these taxa are indeed hominins (Begun, 2004). This has lead some to suggest Late Miocene hominins may have been as diverse as late Miocene apes (Begun, 2010).

Pliocene. The first half of the Pliocene (5.3–4.0 Ma) is characterised by low-standing taxic diversity. Excluding a TDE of zero in the 5.0–4.8 Ma time bin (once again a failure to sample), only one taxon is known at any one time. Phylogenetic methods also imply little change in diversity but indicate the presence of 3 to 4 additional lineages in each time bin relative to the TDE.

The second half of the Pliocene (4.0–2.6 Ma) shows a rapid three-fold increase in TDE at 3.6 Ma followed by an equally rapid three-fold decrease at 3.4 Ma (Fig. 2.2*a*). TDE then remains low until the early Pleistocene. The discovery of *Kenyanthropus platyops*, *Australopithecus bahrelghazali*, *Australopithecus deyiremeda*, and the Burtele foot (BRT-VP-2/73) alongside *Australopithecus afarensis* at 3.4–3.6 Ma has led to the suggestion that middle Pliocene hominins were as speciose as later hominins (Spoor, 2015; Haile-Selassie *et al.*, 2016). The magnitude of this apparent peak is recovered in each PDE, however, it forms part of a near-linear increase in diversity and not a discrete turnover event (*contra* Vrba, 1995). While SPDE, D1PDE, and D2PDE agree closely during this period, HPDE implies higher diversity throughout, peaking at approximately 1.0 Ma before other estimates. PDEs are known to overestimate diversity earlier in a clade's history (Foote, 1996*a*; Wagner, 2000*a*), and the inclusion of *Australopithecus deyiremeda* in HPDE does increase Pliocene diversity by extending all subtending branches to at

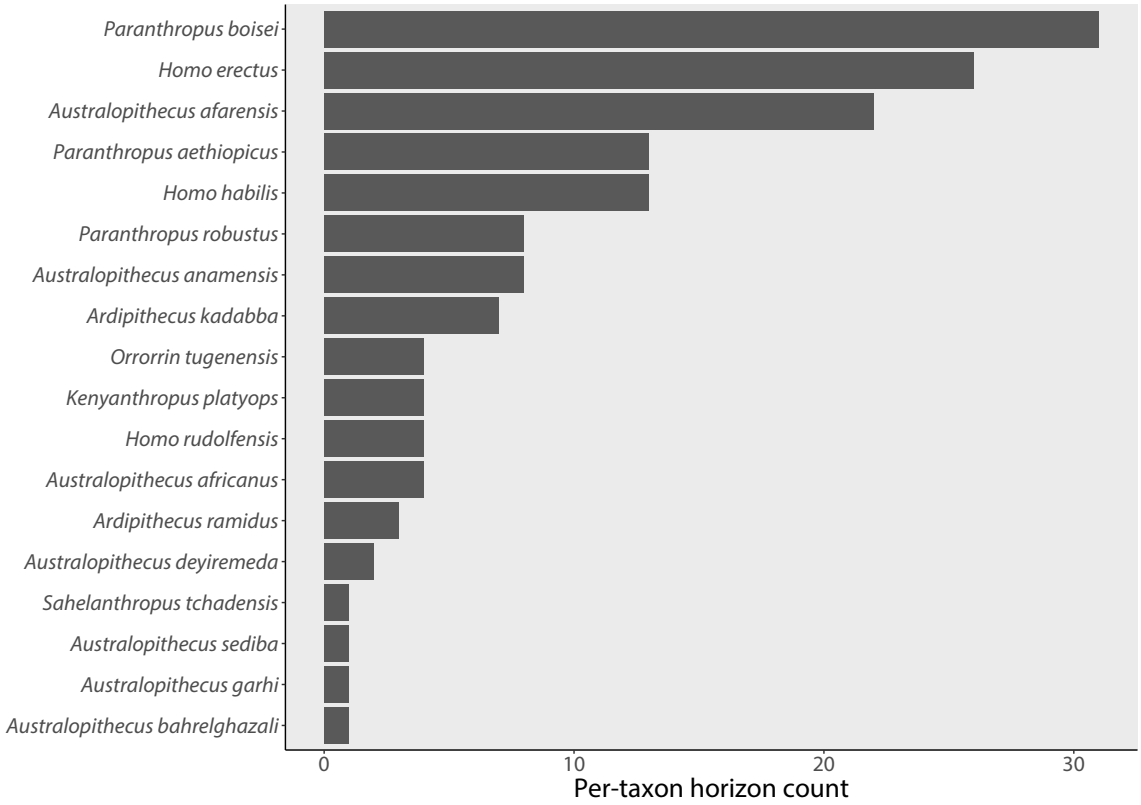

Figure 2.5: Per-taxon horizon counts for each early hominin. Per-taxon horizon counts were taken from Hopley (2018, *in review*) and systematically updated using the Hominin Fossil Database (HFDB). Raw data can be found in Appendix C.

least 3.6 Ma (Fig. 2.2). Hominin diversity, therefore, may have been much higher throughout the entire Pliocene than has been proposed previously. The greatest discrepancy between TDE and PDE occurs between 3.1–2.6 Ma, with each PDE implying 5–6 more lineages than TDE (Fig. 2.2). The abundance of cladistically-implied lineages relative to sampled taxa suggests that this period, despite being central to the origin of *Homo* and *Paranthropus*, is one of the most poorly sampled in hominin evolutionary history (Kimbel, 2009).

Early Pleistocene. The early Pleistocene has long been thought of as a major transition in hominin evolution. It includes the earliest appearance of increased body size, human-like limb proportions, increased brain size, extended life history, high-quality diet including increased carnivory, increased cultural complexity, and advanced social organisation (Hublin *et al.*, 2015). Taxic estimates (e.g., Shultz & Maslin, 2013; Foley, 2016) have long characterised this period as one of maximal richness, and this is found here with peaks at 2.4 Ma (n = 5) and 1.9 Ma (n = 6). Generally, PDEs agree that early hominin diversity reached a pinnacle in the early Pleistocene, with 8 taxa inferred at 2.4 Ma and 6 taxa at 1.9 Ma (Fig. 2.2). However, they provide no evidence for multiple peaks in diversity separated by major troughs (= pulsed turnover) (e.g., Shultz & Maslin, 2013). Once again, these highs are recovered in each PDE, however, they form part of a gradual rise culminating in a single peak in the early Pleistocene. This suggests that highs in observed diversity more accurately represent true diversity and that the greatest departure from true diversity occurs during troughs (see Chapter 3).

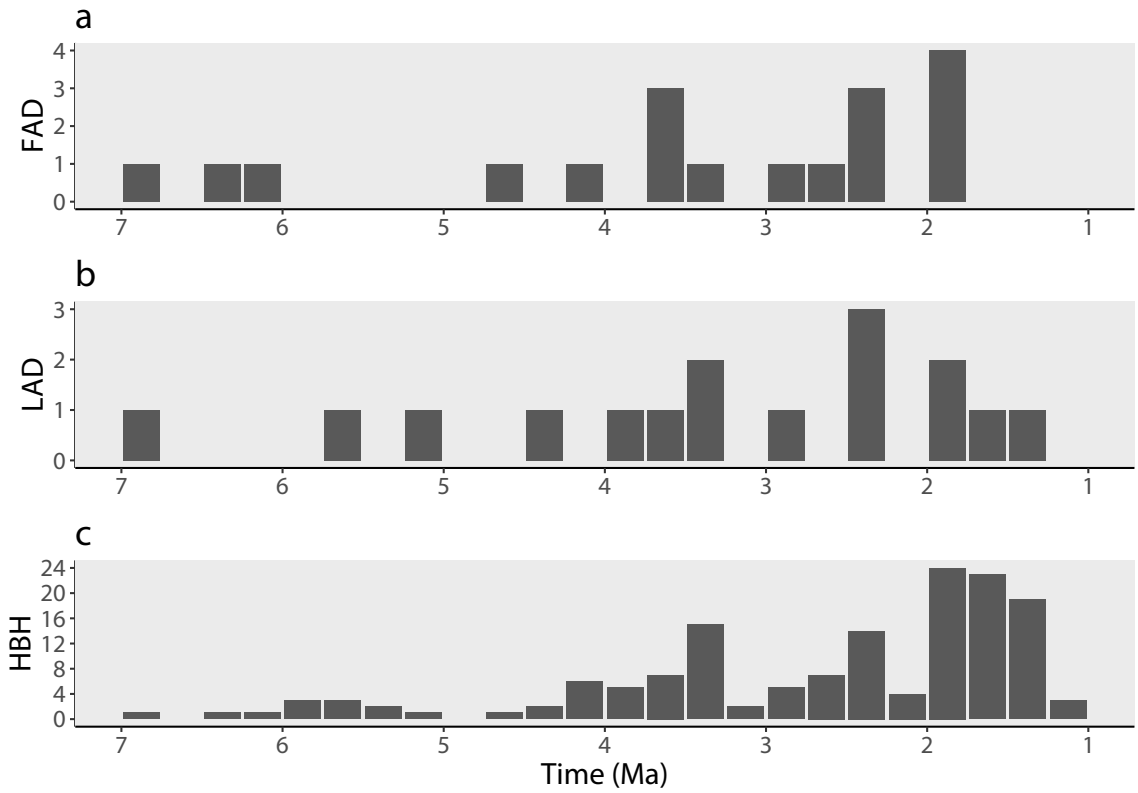

Figure 2.6: Frequency of first and last appearance dates and counts of the number of hominin-bearing horizons. (**a**) Frequency of first appearance dates (FAD). (**b**) Frequency of last appearance dates (LAD). (**c**) Counts of the number of hominin-bearing horizons (HBH) through geological time. FADs and LADs were taken from Wood & Boyle (2016) and the horizon counts were taken from an updated version of the database originally compiled by Hopley (2018, *in review*).

From 2.4 Ma onwards, TDE drops from 6 to 2 and PDE drops from 8 to 2. This represents a decrease of 67%–75%, returning early hominin diversity to that seen during the late Miocene-early Pliocene. The post-2.0 Ma drop in both TDE and PDE (Fig. 2.2) may represent a sequence of gradual or coordinated extinctions, or a mass extinction event artificially truncated by the Signor-Lipps effect (Signor & Lipps, 1982; Lane *et al.*, 2005). The probability of sampling a taxon decreases as one approaches a mass extinction and this is particularly problematic for poorly preserved/sampled taxa (Signor & Lipps, 1982). During a mass extinction, many taxa become extinct simultaneously, resulting in an increase in *zombie lineages* (un-sampled portion of a taxon's range occurring after its last appearance in the fossil record) and the appearance of a prolonged extinction (Lane *et al.*, 2005). PDEs cannot account for zombie lineages (note the lack of confidence intervals on each PDE after 1.9 Ma), and the fact that both TDE and PDE depict this pattern does not add support to either a gradual or mass extinction scenario. Given the finding that TDE is grossly distorted by artificial range truncation, it is equally probable that this decrease is the result of poor sampling.

Gould & Eldredge (1977) suggested in their appraisal of punctuated equilibrium in human evolution, that gaps in the hominin fossil record should be treated as data (i.e., genuine absence) and not artefacts of differential preservation/poor sampling. The mode of evolution in the hominin lineage has remained a hot topic (e.g., Eldredge & Tattersall, 1975; Cronin *et al.*, 1981; Foley, 2016; Kimbel & Vill-

moare, 2016), yet the assumption that the hominin fossil record is complete—or, at least, adequate for evolutionary inquiry—has been routinely accepted but never formally tested (Hopley, 2018, *in review*). This explanation for gaps in the fossil record has remained unchallenged and, as a result, hominin diversification is often described as successive adaptive radiations (e.g., Foley, 2002). However, the extension of first appearances back to their most probable divergence date based on estimates of speciation, extinction, and sampling in the fossil record (*cal3*), and the inclusion of this inferred portion of the fossil record in diversity estimates, produces a radically different pattern of diversification, and one that finds little evidence of episodic or pulsed change.

2.4.2 Does the choice of phylogeny make a difference?

Each phylogeny produces a similar diversity curve, however, there are two key differences between the phylogenies that may effect PDEs: first, each phylogeny includes a different number of taxa; and second, the relationships between taxa (topology) differs in each phylogeny.

Different numbers of taxa. SPDE matches D1PDE and D2PDE closely throughout much of the time period under study (Fig. 2.2), although D1PDE and D2PDE estimate a higher number of lineages between 3.0 and 2.0 Ma. This reflects the greater number of early hominins (14) incorporated into the Dembo *et al.* (2015, 2016) analyses compared to the Strait & Grine (2004) analysis (13). The additional taxon, in this case the 1.98-Ma *Australopithecus sediba*, is repeatedly placed as the sister-taxon to the 2.33-Ma *Homo habilis*. The difference in peak diversity between SPDE (7) and D1PDE and D2PDE (8) is thus caused by the inclusion of *Australopithecus sediba*. Similarly, the Haile-Selassie *et al.* (2015) analysis also includes 14 taxa, but the additional taxon is the 3.3–3.5-Ma *Australopithecus deyiremeda*. The latter is consistently placed as the sister-taxon of a clade containing *Australopithecus africanus*, *Paranthropus*, and *Homo* (including *Kenyanthropus platyops*). The difference in the timing of peak diversity in HPDE is thus caused by the inclusion of *Australopithecus deyiremeda* (Fig. 2.2).

Different tree topologies. The phylogenetic relationships of early hominins are hotly debated (see Strait *et al.*, 2007), yet there are arguably only two unstable taxa: *Kenyanthropus platyops* and *Australopithecus africanus* (Strait *et al.*, 2007). The Strait & Grine (2004) analysis places *Kenyanthropus platyops* as the sister-taxon of either *Paranthropus* or a *Paranthropus* + *Homo* clade. Dembo *et al.* (2015) place *Kenyanthropus platyops* as the sister-taxon of all hominins except *Sahelanthropus tchadensis*, *Ardipithecus ramidus*, and *Australopithecus anamensis*, whereas Dembo *et al.* (2016) place *Kenyanthropus platyops* as the sister-taxon of all hominins except *Sahelanthropus tchadensis*, *Ardipithecus ramidus*, *Australopithecus anamensis*, *Australopithecus afarensis*, and *Australopithecus garhi.* The phylogenetic relationship of *Kenyanthropus platyops* to other hominins thus has little effect on SPDE, D1PDE, and D2PDE (Fig. 2.2 and Table 2.2). In contrast, Haile-Selassie *et al.* (2015) place *Kenyanthropus platyops* as the sister-taxon of *Homo rudolfensis*, and *Kenyanthropus platyops* + *Homo rudolfensis* as the sister-taxon of *Paranthropus*. This implies a ghost lineage for *Homo rudolfensis*, and *Homo habilis* + *Homo erectus*, extending back to ca., 4.0 Ma. Despite this, HPDE does correlate significantly with SPDE, D1PDE, and D2PDE (Table 2.2).

Strait & Grine (2004) place *Australopithecus africanus* as the sister-taxon of *Kenyanthropus platyops*, *Paranthropus*, and *Homo*. Dembo *et al.* (2015) place *Australopithecus africanus* as the sister-taxon of a clade containing *Australopithecus sediba* + *Homo*, whereas Dembo *et al.* (2016) place *Australopithecus*

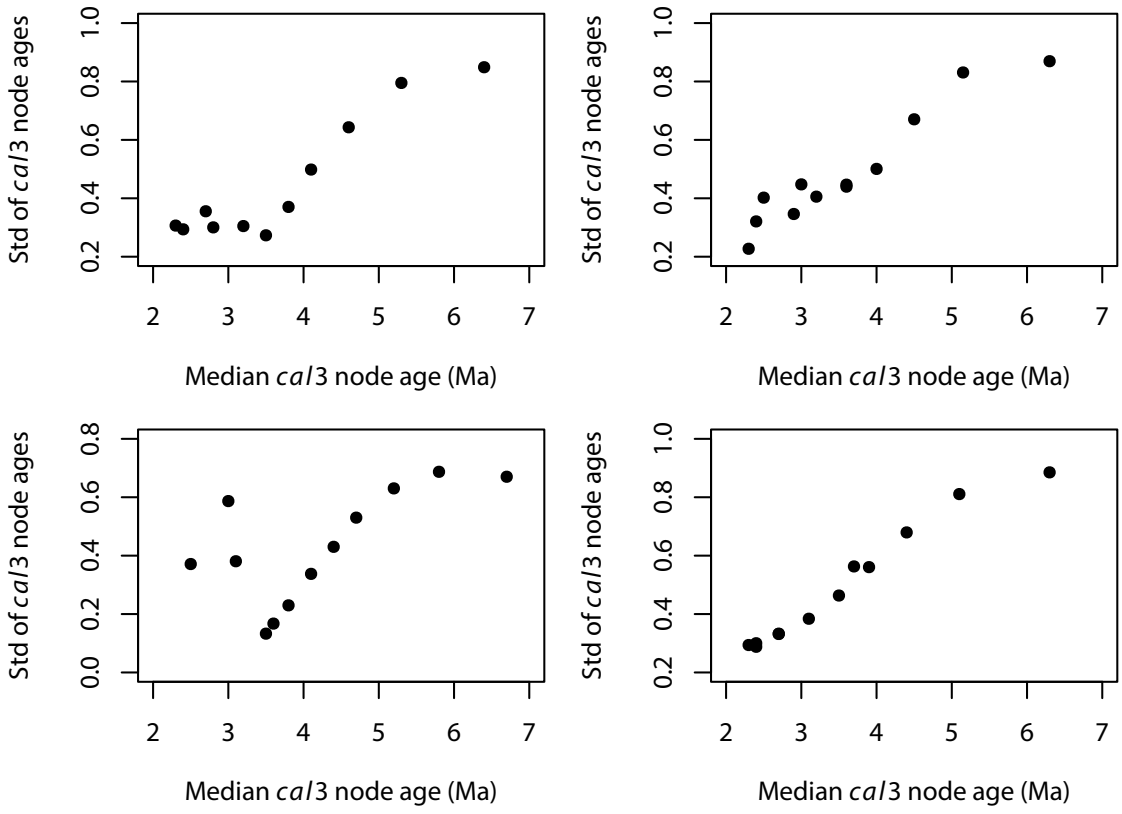

Figure 2.7: Scatter plots showing the relationships between standard deviation (std) of *cal3* node age estimates and stratigraphic age for each phylogeny. (**a**) SPDE. (**b**) D1PDE. (**c**) HPDE. (**d**) D2PDE. The age of the root is omitted from each regression because it is constrained between 8.0 Ma (the upper estimate of the *Pan-Homo* divergence date) and the age of the oldest first appearance date (in this case the age of *Sahelanthropus tchadensis* = 7.0 Ma), and so uncertainty in its estimation is anomalously small.

africanus in a robust australopith (*Paranthropus*) clade. Despite the major systematic implications of each topology, the different placements of these taxa have little effect on the resultant PDE, probably because the clustering of nodes in each phylogeny tightly constrains the possible divergence dates of each species, irrespective of the exact topology. Plotting the standard deviation of *cal3* node ages against median *cal3* node age for each phylogeny demonstrates that most nodes are very tightly constrained, and that node age uncertainty (standard deviation) increases linearly toward the root (Fig. 2.7).

Reconstructing chronospecies as anagenetically evolving lineages. It is often assumed that the chance of sampling an ancestor in the fossil record is negligible (e.g., Norell, 1993; but see Foote, 1996*b*). However, there is strong evidence for anagenesis in the early hominin fossil record (Kimbel *et al.*, 2006; Wood & Schroer, 2013; see also White *et al.*, 2009). If ancestors are included in a cladogram, they will be interpreted as the sister-taxa to their descendants, leading to the inference of an incorrect ghost lineage and inflated PDE (Lane *et al.*, 2005). However, for those studies that support hypotheses of ancestry (Strait & Grine, 2004; Haile-Selassie *et al.*, 2015), there are no major differences in the pattern of their diversity curves (Table 2.2). Although there are fewer lineages inferred under an anagenetic scenario, the overall trend in phylogenetic diversity is similar, suggesting that (for early hominins at least) the inclusion of ancestors has no significant effect on PDEs (Table 2.2).

Overall, each PDE displays a very similar curve, and any differences between estimates can be readily accounted for by taxon selection, tree topology, and tree size.

2.4.3 Diversification models

Fitting different mathematical models to the net positive diversification component of each PDE shows that an exponential model of diversification is best supported for early African hominins (Table 2.3). In terms of an evolutionary branching model, an exponential increase in diversity is equivalent to a birth-death model in which speciation and extinction are constant through time (Nee, 2006). Under such a scenario, a steady rate of diversification within each lineage will scale up to a doubling of diversity from any one time period to the next, since total diversity is ever-increasing (Benton, 2007). Unbounded diversification would imply that ecological constraints were either non-existent or not particularly important drivers of early hominin diversification (Benton & Emerson, 2007; Rabosky, 2013), and there are three possible explanations for such a pattern:

1. Competitive exclusion, predation, and other ecological interactions were rare among hominins, or their effects were weakened by niche partitioning.
2. Climate remained relatively stable from the late Miocene to the Plio-Pleistocene boundary.
3. Speciation and extinction were independent of abiotic factors.

First, many early hominins were sympatric (Wood & Boyle, 2016), and it has been posited that the most probable explanation for diversification within any sympatric group of primate taxa is niche partitioning (Robinson, 1963; Jolly, 1970). For example, dietary adaptations are one of the key drivers of hominin evolution and are important indicators of resource exploitation and niche partitioning (Teaford & Ungar, 2000). However, the precise mechanisms underlying niche partitioning in hominins are unclear (Ackermann & Smith, 2007). While there is strong evidence for isotopic niche differentiation (Sponheimer *et al.*, 2013), there is little understanding of how multiple sympatric hominins occupied a generalist feeder niche for millions of years.

Second, climate-driven models of hominin diversification invoke change and variability in the physical environment to explain both positive and negative net diversification (e.g., Vrba, 1988; Foley, 1994; Kimbel, 1995; Kingston, 2007; Grove, 2011, 2012; Potts, 2013; Shultz & Maslin, 2013); stable climatic conditions are invoked when net diversification nears zero. In the absence of discrete climatic change or during low climatic variability, stable climatic conditions could free hominins from constrains imposed by their physical environment and lead to unbounded diversification (until, of course, the carrying capacity of the landscape is reached).

The most difficult to reconcile with human evolutionary theory is the third explanation. The finding that speciation and extinction were roughly constant rejects any role for discrete climatic events as causal agents in hominin diversification. Moreover, these findings underscore that the relative importance of biotic *versus* abiotic factors in hominin macroevolution is by no means clear, and that Red Queen hypotheses may offer more viable explanations despite the wide range of Court Jester hypotheses described in the literature today. Ultimately, if coordinated diversification shifts become readily apparent with the discovery of new hominins, and can be linked to climatic, tectonic, and geographic drivers,

then Court Jester hypotheses can continue to provide an explanatory framework in human evolutionary theory. However, if diversification shifts occur gradually, randomly, or are unique to a particular clade, Red Queen hypotheses offer the simplest explanation (Benton, 2009).

Further work linking phenotype to niche, identifying the adaptive fitness of a trait, and on the relative importance of directional selection *versus* neutral genetic drift in driving morphological change must now become key research foci if we are to fully understand hominin diversification.

2.4.4 Confidence intervals and stratigraphic uncertainty

Palaeodiversity estimates require accurate first and last appearances, yet it is clear that, for several taxa, current FAD and LAD provide a gross underestimate of their true stratigraphic range. For example, four early hominins have no known stratigraphic range. For those hominins that do, the confidence with which we can be consider a FAD to represent a speciation event and, similarly, a LAD to represent an extinction event, is very poor. In combination, this raises problems for their use in time-series analyses of palaeodiversity.

Down-weighting. One method of compensating for stratigraphic uncertainty (namely dating error) is to down-weight the contribution of taxa in bins for which their placement is uncertain. Proportional-weighting assigns a fractional contribution to each bin based on the number of bins that fall in a taxon's stratigraphic uncertainty (Hunt *et al.*, 1994). Equal-weighting, on the other hand, counts one taxa present in each bin spanning a taxon's stratigraphic uncertainty (Upchurch & Barrett, 2005). For example, *Australopithecus africanus* has CIs of 1.4 Ma (corresponding to six 0.25-Ma bins): under a proportional-weighting strategy it would contribute 1/6 to the six bins preceding its FAD and six bins succeeding its LAD, whereas under an equal-weighting strategy it would contribute 1 to each of these bins. Despite the uncertainty in early hominin stratigraphic ranges, down-weighting is not employed in the current study for two reasons.

First, suppose a taxon's CIs imply it originated in one of two consecutive time bins, X and Y, but that, in reality, it originated in X. Proportional-weighting would overestimate diversity in Y and underestimate diversity in X. In the same scenario, equal-weighting would overestimate diversity in Y but correctly estimate diversity in X. Each strategy, therefore, adds additional error to diversity estimates. Second, CIs do not provide any information on the most probable date that a speciation or extinction event occurred within a speciation window (SW) or extinction window (EW), respectively (Smith, 1994). Therefore, any attempt to down-weight diversity based on CIs is arbitrary and greatly affected by time bin duration.

Taxonomic uncertainty. CIs rely on accurate and up-to-date horizon counts, and the exclusion of disputed (cf., aff., or sp.) specimens may artificially inflate CIs. For example, if the partial mandible from Tabarin, Kenya (KNM-TH 13150), represents *Ardipithecus ramidus* at 4.42 Ma (e.g., Wood, 2011; Kissel & Hawks, 2015) then its CIs half from 1.01 to 0.48 Ma—a decrease of 48%. If the isolated dentition from Fejej, Ethiopia (FJ-4-SB-1 and -2), represent *Australopithecus anamensis* (e.g., White, 2002) then its CIs decrease from 0.20 to 0.17 Ma. Lastly, if the ~3.89 Ma partial cranium from Belohdelie, Ethiopia (BEL-VP-1/1; Asfaw, 1987), the partial skeleton from South Turkwell, Kenya (KNM-WT 22944; Ward *et al.*, 1999*a*), and the fragmentary dentition from Lothagam (e.g., KNM-LT 23181; Leakey & Walker, 2003), represent *Australopithecus afarensis* then both the stratigraphic range and CIs of this taxon change. The

stratigraphic range increases from 0.7 Ma to ~0.9 Ma and number of horizons from 22 to 25, hence the CIs increase slightly from 0.13 to 0.15 Ma. This indicates that the debated assignments of several specimens can have a large effect on CIs. The 2.8-Ma *Homo* partial mandible LD 350-1 (Villmoare *et al.*, 2015) is also likely to have a dramatic effect on our confidence in the time of origin of *Homo* (see Bobe & Leakey, 2009). Indeed, if assigned to *Homo habilis*, its CI increases from 0.24 to 0.37, placing the origin of *Homo* firmly in the middle Pliocene (3.2–2.8 Ma).

2.5 Conclusion

Trends in early African hominin diversity are frequently linked to change in their physical environment, however, our understanding of hominin diversity dynamics is muddied by a tendency to read the fossil record literally. Taxic (= literal, face-value) diversity suggests stasis punctuated by three major turnover events. First, the australopith radiation in the middle Pliocene (~3.6 Ma), possibly an ecological response to increased grasslands, characterised by increased sympatry and highly variable (e.g., *Kenyanthropus*) and atypical diets for mammals (Sponheimer *et al.*, 2013). Second, the origin of *Paranthropus* and *Homo* in the early Pleistocene (~2.4 Ma), representing the first major divergence in morphological and ecological (dietary) specialisation in the hominin lineage (Cerling *et al.*, 2013) and invasion of new morphospace in primate evolution (Fleagle *et al.*, 2016). Third, the radiation of *Paranthropus* and *Homo* and later demise of *Australopithecus* by ca., 1.9 Ma. However, phylogenetic diversity suggests gradual, constant-rate lineage turnover from the late Miocene to the Plio-Pleistocene boundary, implying that sampling bias is a major control of TDE and that TDE does not provide a reliable depiction of early hominin diversification dynamics. Debate surrounding the quality of the early hominin fossil record has focussed exclusively on site-specific taphonomic processes effecting assemblage formation and taxon abundances, and not on the large-scale geological and anthropogenic factors influencing macroevolutionary patterns. The difference between uncorrected (taxic) and partially corrected (phylogenetic) diversity estimates reported here indicates that the fossil record is, at present, inadequate for studying diversity patterns without correction for sampling heterogeneity.

Summary

To summarise, the results of this chapter have shown:

1. There is considerable statistical uncertainty that the dates of first and last appearance reflect speciation and extinction, respectively. Moreover, this uncertainty (mean = 0.53 Ma; maximum = 1.38 Ma) is comparable in duration to the stratigraphic range (mean = 0.45 Ma) of early hominins.

2. Irrespective of dating error (the primary concern in past studies of diversification), the sampling of the early hominin fossil record indicates that speciation (extinction) is likely to have occurred substantially earlier (later) than first (last) occurrences imply.

3. Raw taxic diversity supports low-standing diversity during the first half of hominin evolution, punctuated by three rapid turnover pulses in the Plio-Pleistocene.

4. The four phylogenies used to estimate phylogenetically corrected (= lineage) diversity each include different taxa and support different phylogenetic hypotheses, yet all imply gradual change in species richness and provide no evidence of episodic or pulsed diversification.

5. Each PDE supports an exponential increase in net diversification for the first four to five million years of hominin evolution. This suggests constant-rate diversification with speciation rate greater than extinction rate, and implies that ecological constants were either non-existent or not particularly important drivers of early hominin diversification.

Chapter 3

Geological and anthropogenic controls on the sampling of the early hominin fossil record

This chapter is an extended version of the following publication: Maxwell SJ, Hopley PJ, Upchurch P, Soligo C, 2018. Sporadic sampling not climatic forcing drives early hominin taxic diversity. Proceedings of the National Academy of Sciences, ***115****(19), 4891–4896.*

3.1 Introduction

Many recent publications have reported a significant positive correlation between palaeodiversity and a range of metrics for fossil sampling (Raup, 1972; Sheehan, 1977; Peters & Foote, 2001, 2002; Smith, 2001; Crampton *et al.*, 2003; Smith & McGowan, 2007; McGowan & Smith, 2008; Fröbisch, 2008, 2013; Barrett *et al.*, 2009; Butler *et al.*, 2009, 2011, 2013; Wall *et al.*, 2009, 2011; Benson *et al.*, 2010; Benson & Butler, 2011; Benton *et al.*, 2011; Mannion *et al.*, 2011; Upchurch *et al.*, 2011; Lloyd *et al.*, 2012*a*; Benson & Upchurch, 2013; Brocklehurst *et al.*, 2013; Newham *et al.*, 2014; Tennant *et al.*, 2016). The co-variation between sampling metrics and palaeodiversity can be explained by three hypotheses: (1) the rock record bias (RRB) hypothesis, that variable collection effort and its underlying driver, the amount of sedimentary rock available to sample, controls apparent palaeodiversity (Raup, 1972, 1976*b*; Smith, 2001); (2) the common-cause (CC) hypothesis, that both genuine diversity and the rock/fossil records are both independently driven by a third environmental factor (Peters, 2005, 2006; Hannisdal & Peters, 2011); or (3) the redundancy (RED) hypothesis, that palaeodiversity and sampling metrics are information redundant (non-independent) with respect to each other (Benton *et al.*, 2011, 2013; Dunhill *et al.*, 2014*b*). Redundancy can arise because of tallying and comparability. The tally of new fossils and new formations which yield the fossils both grow in tandem and both equally reflect the intensity of sampling (Benton, 2015). Comparability refers to the fact that greater collection effort might result in higher diversity, but higher genuine diversity might also result in more collecting (Benton, 2015). Each hypothesis can be summarised by the following drive-response relationship: sampling → diversity (RRB); environment → sampling and diversity (CC); and, diversity → sampling or diversity

←→ sampling (RED) (Benton, 2015).

Sampling metrics seek to capture the four aspects of bias or error in the fossil record: rock volume, rock accessibility, facies heterogeneity, and study effort (Raup, 1972; Benton *et al.*, 2011). The first three geological (including taphonomic) factors concern the extent to which the rock record preserves evolutionary events. Through the processes of burial and decay, the amount of exposed sediment, and range of depositional environments preserved, geological processes directly influence the preservation potential of taxa and number of opportunities to observe fossils. Geological processes such as uplifting, incision, surface erosion, and superficial deposition also affect the accessibility of fossiliferous sediments. Human or anthropogenic factors include the way in which palaeoanthropologists have sampled the fossil record. The use of different sampling protocols, highly variable collection effort (both geographically and stratigraphically), and other human factors such as economic stability and geopolitics, can all influence apparent species richness (Raup, 1972, 1976*b*). A positive correlation between palaeodiversity and sampling metrics is most frequently interpreted as evidence of a biased fossil record documenting an inaccurate depiction of diversification patterns (RRB hypothesis).

The most commonly used sampling metrics are estimates of sedimentary rock volume (Raup, 1972, 1976*b*) or rock outcrop (i.e., geological map) area (Smith, 2001; Crampton *et al.*, 2003; Uhen & Pyenson, 2007; Marx, 2009; Wall *et al.*, 2009), rock exposure area (Dunhill, 2011, 2012; Walker *et al.*, 2017), the number of hiatus- or gap-bound rock packages (Peters, 2005, 2006; Peters & Heim, 2010), counts of the number of fossiliferous formations (Barrett *et al.*, 2009; Butler *et al.*, 2009, 2011; Benson *et al.*, 2010; Marx & Uhen, 2010; Mannion *et al.*, 2011; Dunhill & Wills, 2015), counts of the number of collections (Upchurch *et al.*, 2011; Brocklehurst *et al.*, 2012; Cleary *et al.*, 2015, 2018; Tennant *et al.*, 2016), and counts of the number of fossiliferous localities or their total area (Fountaine *et al.*, 2005; Lloyd *et al.*, 2008; Lloyd & Friedman, 2013).

However, the pervasive nature of sampling has come under scrutiny in recent years because of the questionable validity of many of these putative sampling metrics and conceptual discrepancies in their formulation (Benton *et al.*, 2011, 2013; Dunhill *et al.*, 2014*b*; Benton, 2015). Of these sampling metrics, fossiliferous formation counts (FFC) have been used most frequently (see above), both to assess the influence of sampling on palaeodiversity patterns and to correct for it, and have consequently received the greatest criticism. The main criticisms against FFC are outlined below (Benton *et al.*, 2011), along with an explanation as to why some of these criticisms are inapplicable or of a lesser concern in the hominin fossil record:

1. Their definitions are arbitrary. Formations are defined on nuanced geological interpretations which are specific to each country or discipline (Benton *et al.*, 2011). However, in the East African Rift System (EARS) where 82% of hominin-bearing formations occur, formations are defined purely on lithological grounds and the majority have precise chronological controls (Feibel *et al.*, 1989; McDougall & Brown, 2006; McDougall *et al.*, 2012).

2. They do not consistently correlate with rock exposure (Dunhill, 2011), the simplest measure of rock accessibility. This finding is based on a small-scale study of England and Wales and has not been replicated elsewhere, making the results difficult to generalise. Formations have been shown to correlate with rock outcrop area in the terrestrial realm (Upchurch *et al.*, 2011; Dunhill *et al.*,

2014*b*), and the number of gap-bound packages in the marine realm (Peters, 2005). However, there is no consistent relationship between rock outcrop and exposure area: in four regional case studies, rock outcrop and exposure correlated in only half (Dunhill, 2012), and no correlation is found in the Dorset and East Devon Coast World Heritage Site (Dunhill *et al.*, 2012). On the other hand, rock outcrop and exposure area do correlate positively in the Chalk Group of Hampshire which the authors suggest may be due to the similarity of facies in this region (Walker *et al.*, 2017). The lack of a consistent link between FFC and rock exposure (and outcrop and exposure area) does not inherently render them useless but reiterates that the applicability of a sampling metric is specific to the clade of interest and should be considered in detail and at an appropriate scale before being used to assess fossil record quality. Given the narrow range of facies types hominins are found in (fluvio-lacustrine, palaeosol, and karst deposits), and the reduced vegetation cover and land use throughout the regions known to contain hominin fossils, rock outcrop and exposure area would be expected to correlate.

3. They do not consistently correlate with collection effort (Crampton *et al.*, 2003), but see Upchurch *et al.* (2011). It is assumed that the underlying driver of collection effort is the amount of fossiliferous sediment (i.e., rock exposure), however, there is no *a priori* reason why the rock record and collection effort should consistently correlate. Rock availability undoubtedly influences collection effort but sites of exceptional preservation or abundance (e.g., Lagerstätten) are likely to buck this trend: the pterosaur fossil record is a case in point, where pterosaur-bearing collections do not correlate with dinosaur-bearing formations ($p = 0.095$) (Butler *et al.*, 2013). While there are no hominin Lagerstätten *per se*, the very presence of hominin fossils leads to intense and sustained collection effort (e.g., Sterkfontein).

4. They may largely reflect facies heterogeneity. Formations that preserve a greater variety of facies types preserve a greater range of depositional environments (= habitats), leading to more finely-defined (= thinner) formations, and higher apparent diversity. Although this may be the case, this would mean that formation counts are capturing variability in depositional environment (Peters & Foote, 2001; Crampton *et al.*, 2006)—a fundamental aspect of fossil record bias (Raup, 1972, 1976*b*).

5. They depend on fossil abundance and diversity, and are therefore not independent of it (Benton, 2012). There are no known examples of formations being defined based on the apportioning of diversity (Benson & Upchurch, 2013).

6. They vary enormously in scale. Globally, formations vary in volume by eight orders of magnitude (Benton *et al.*, 2011; Dunhill *et al.*, 2012). However, this is not the case in the African Neogene where formations vary in thickness by one order of magnitude (from tens to hundreds of metres; see § 1.3 and Werdelin, 2010). Moreover, in the Upper Laetoli Beds, Tanzania, exposed rocks are limited to 1–5 km^2 despite the total area of these localities spanning approximately 100 km^2 (Su & Harrison, 2007; Harrison & Kweka in Harrison, 2011*a*); the three areas in the Koobi Fora Formation, Kenya, for which data are available have outcrop areas of 3.4, 6.4, and 8.4 km^2 (Thompson *et al.*, 2017); and in the late Miocene of Ethiopia, rock exposure area is small and isolated because

of dense faulting and superficial sediment deposition (WoldeGabriel *et al.*, 2009; DiMaggio *et al.*, 2015).

7. Only *fossiliferous* formations are counted. Total sampling effort should combine sampling failure (non-occurrence) and success (Upchurch & Barrett, 2005), however, early attempts at assessing the quality of the fossil record failed to do so (see Benton *et al.*, 2011). This problem can be mitigated using an FFC that includes ichnofossils and the formations of a closely related and taphonomically comparable group, in order to include any formations that have been searched but have not yielded fossils of the group of interest (Upchurch & Barrett, 2005).

Many of these criticisms relate to the notion that FFCs capture just one aspect of sampling (principally rock volume), however, this is incorrect: FFCs combine aspects of rock availability, the geographical extent of sampling, the heterogeneity of facies available for sampling, and the amount of study that has been undertaken (Benson & Upchurch, 2013:43).

Recently, correlations between hominin palaeodiversity (henceforth diversity) and sampling metrics (number of hominin-bearing formations and horizons) have been reported (Maxwell *et al.*, 2016; Hopley, 2018, *in review*), suggesting fluctuations in observed diversity may be more apparent than real. This interpretation, however, is easily questioned by their use of formation counts based solely on the clade of interest (i.e., hominins). In this chapter, the validity of numerous sampling metrics is tested. Climatic-forcing hypotheses of early hominin diversification are then tested alongside the RRB, CC, and RED hypotheses by comparing both a taxic diversity estimate (TDE) and four phylogenetic diversity estimates (PDE) to: (1) a *strict* FFC consisting of only those formations that have yielded a hominin fossil; (2) a *wider* FFC consisting of all formations that have yielded a primate fossil; and, (3) a *comprehensive* FFC consisting of all formations that have yielded a terrestrial macro-mammal fossil (geological sampling bias). In addition, diversity is compared to the amount and geographical spread of palaeoanthropological collection effort (anthropogenic sampling bias). Finally, this chapter addresses recent calls for a more objective interpretation of diversity dynamics in the hominin fossil record (Smith & Wood, 2017; Hopley, 2018, *in review*), and for the effect of sampling on diversity patterns to be assessed at finer spatial and temporal scales (Dunhill, 2011, 2012; Walker *et al.*, 2017).

3.2 Methodology

3.2.1 Diversity metrics

Protocols for estimating TDE and PDE follow those used in Chapter 2. The benefit of the PDE approach is that hominin systematics is the subject of considerable interest and so a wealth of competing phylogenetic hypotheses exist. SQS (Shareholder Quorum Subsampling; Alroy, 2010) and TRiPS (True Richness estimated using a Poisson Sampling model; Starrfelt & Liow, 2016) were not used to estimate early hominin diversity due to the large number of singleton and single-bin taxa. Further, the residual modelling method (Smith & McGowan, 2007; Lloyd, 2012) has recently come under intense scrutiny for statistical errors in its formulation and implementation (Sakamoto *et al.*, 2017; Dunhill *et al.*, 2018). For these reasons, modelled diversity estimates were not generated.

Table 3.1: List of abbreviations used in this chapter.

Abbreviation	Definition
D1GDE	Dembo *et al.* (2015) ghost lineage diversity estimate
D1PDE	Dembo *et al.* (2015) phylogenetic diversity estimate
D2GDE	Dembo *et al.* (2016) ghost lineage diversity estimate
D2PDE	Dembo *et al.* (2016) phylogenetic diversity estimate
EARS	East African Rift System
HBC	Hominin-bearing collection
HBF	Hominin-bearing formation
HBL	Hominin-bearing locality
HGDE	Haile-Selassie *et al.* (2015) ghost lineage diversity estimate
HPDE	Haile-Selassie *et al.* (2015) phylogenetic diversity estimate
LVI	Lake variability index
MBF	Mammal-bearing formation
PBF	Primate-bearing formation
SGDE	Strait & Grine (2004) ghost lineage diversity estimate
SPDE	Strait & Grine (2004) phylogenetic diversity estimate
TDE	Taxic diversity estimate

3.2.2 Gap-counting metrics

Gaps in fossil sampling can be inferred from ghost lineages and Lazarus taxa. Ghost lineages represent the minimum gap between the oldest fossil of a taxon or clade and the date of cladogenesis implied by a phylogeny (Norell, 1992, 1993), and Lazarus taxa are taxa that disappear from the fossil record only to re-appear at a later date (Flessa & Jablonski, 1983). For Lazarus taxa, the gap refers to the absence below the younger sampled horizon and above the older sampled horizon. Most gap analyses focus on Lazarus taxa (Paul, 1998; Smith, 2001, 2007*a*; Fara, 2002), however, in the current study gaps are counted as (1) ghost lineage diversity estimates (GDE), and (2) the relative proportion of implied to sampled lineages (GDE:PDE; Smith, 1994, 2007*b*; Benton *et al.*, 2011).

Ghost lineages can have major implications for evolutionary patterns: for example, the Cretaceous origin of modern euprimates (ca., 80–90 Ma) is supported by molecular evidence (e.g., Springer *et al.*, 2012; Pozzi *et al.*, 2014) and statistical analyses of the primate fossil record (e.g., Tavaré *et al.*, 2002; Wilkinson *et al.*, 2011). However, the earliest undisputed primate of modern aspect (= euprimate) is found in the late Eocene (ca., 56 Ma), implying a 20- to 30-Ma ghost lineage that crosses the Cretaceous-Paleogene (K-Pg) mass extinction event (Fleagle, 2013). Further, ghost lineages can also be substantial in duration: for example, the fossil record of Lemuriformes is less than 30 thousand years old (ka) (Crowley, 2010). However, fossils belonging to its sister group, the Lorisiformes, date to the middle to late Eocene (Seiffert *et al.*, 2003), implying a lemuriform ghost lineage of at least 35 Ma (Soligo *et al.*, 2007). Ghost lineages have also been used to disentangle genuine diversification from increased fossil sampling (Cavin & Forey, 2007). Genuine diversification results in many new taxa with short ghost lineages and therefore mean ghost lineage duration decreases when diversity increases. Sampling-driven increases in diversity can be distinguished from genuine diversification by an unchanging mean ghost lineage duration (Cavin & Forey, 2007).

Phylogenetic trees are constructed based solely on the distribution of morphological characters and are independent of the stratigraphic record. Therefore, once calibrated against time, the temporal distribution of ghost lineages represents an independent metric for sampling (Paul, 1998). The prediction being that (1) GDE should correlate negatively with sampling metrics such as collection, formation, and locality counts or rock outcrop area, and (2) that TDE should correlate better with sampling metrics than does PDE (Benton *et al.*, 2011). These predictions follow from their formulation: PDE = TDE (observed diversity) + GDE (implied diversity) and therefore GDE is calculated by subtracting TDE from PDE. GDE is presented for the same four phylogenies analysed in Chapter 2 (see § 2.2.1).

3.2.3 Preservational biases and sampling quality

In order to assess whether early hominin diversity is controlled by geological and anthropogenic sampling of their fossil record, we compared diversity to four metrics that each capture a different aspect of rock volume, rock accessibility, rock heterogeneity, and collecting effort (Raup, 1972; Benton *et al.*, 2011).

Rock availability. Information on sedimentary rock outcrop or exposure area is not available at the continental level for the African Neogene, so instead, fossiliferous formation counts (FFC) were used as a proxy for the amount of accessible rock that is available for sampling. FFC represents an estimate of the number of discrete depositional environments known to contain fossils, and—assuming the facies present are suitable for the clade of interest—is thus a proxy for geological controls on observed early hominin diversity. FFC has been shown to correlate significantly with the number of gap-bound packages (e.g., Peters, 2005) and rock outcrop area (e.g., Upchurch *et al.*, 2011).

Hominin-bearing formations (HBF) were taken from an exhaustive survey of the published literature. Primate-bearing formations (PBF) were taken from the chapters on cercopithecoids (Jablonski & Frost, 2010), hominoids (specifically hominins; MacLatchy *et al.*, 2010), and lorisoids (Harrison, 2010*b*) in *Cenozoic Mammals of Africa* (Werdelin & Sanders, 2010) and corroborated using the *Paleobiology Database* (PBDB). Mammal-bearing formations (MBF) were similarly gathered from *Cenozoic Mammals of Africa* (Werdelin & Sanders, 2010) and PBDB, and excluded small (i.e., Eulipotyphla, Hyracoidae, Lagomorpha, Macroscelidea, Rodentia) and non-terrestrial (i.e., Cetacea, Chiroptera, Sirenia) mammals.

Geological formations are rock units of defined areal and temporal extent, however, this is not the case in the Cradle of Humankind (CoH), South Africa, where each cave is informally assigned to its own geological formation (Partridge, 1978; Dirks *et al.*, 2010). Treating each CoH site as a distinct geological formation is more equivalent to a locality count (not a proxy for geological sampling). So to combat this the fossil-bearing deposits at Sterkfontein, Swartkrans, Kromdraai, Drimolen, Gondolin, Gladysvale, and Cooper's Cave were counted as one formation. Moreover, lumping the South African deposits into one formation is equivalent in geographical extent and geological time to a typical East African formation. Note this had minimal effect on the results as this sampling metric (more appropriately termed the number of hominin-bearing deposits) correlates strongly with HBF ($\rho = 0.941$, $p < 0.001$). Makapansgat and Taung are each counted as separate formations given their geographical separation from the CoH. The same treatment is applied to other karst deposits such as the Humpata Plateau (Angola; including the fossil-bearing localities of Cangalongue, Tchiua, and Malola), Koanaka (Botswana), and

the Otavi Mountains (Namibia; including the fossil-bearing localities of Jägersquelle and Uisib).

Collection effort. In-bin counts of the number of hominin-bearing collections (HBC) were compiled as a proxy for collecting effort. Benton (2015) defined a collection as an assemblage of fossils from one locality that were amassed in a single effort, or linked series of effort, and is roughly equivalent to a field season. Information on the duration and number of field seasons at a locality are not commonly provided so, instead, we used the number of years that have produced a hominin fossil per formation per bin. For example, *Sahelanthropus tchadensis* is known from the 7.0-Ma Anthracotheriid Unit (Chad) and the fossils that compose its hypodigm were collected in 2001 and 2002. The 7.0–6.8 Ma time bin therefore has a HBC count of 2. The number of HBC in a given time bin thus represents the number of discrete episodes of field study (= palaeoanthropological collection effort) that have yielded hominin fossils. HBC is probably better regarded as a Paleontologic Interest Unit (PIU), a measure of effort devoted to acquiring knowledge, counted in number of people, years, or publications on a particular time period or taxonomic group (Sheehan, 1977). HBC has been resolved as finely as possible given the published literature. Unfortunately, primate- (PBC) and mammal-bearing collections (MBC) could not be compiled in the same manner because discovery year is rarely reported in publications describing new primate and mammal fossils.

Palaeoarea of fossiliferous localities. The number of hominin-bearing localities (HBL) were taken from an exhaustive review of the published literature (see Compilation of the Hominin Fossil Database in Chapter 1). Latitude and longitude data for each HBL were taken from the Human Origins Locality Data Collective (http://www.fossilized.org) (see Appendix D), and the total sampled area (in km^2) of a spherical polygon calculated by drawing a convex-hull around all HBL in each time bin (R code from G.T. Lloyd: http://www.graemetlloyd.com/methspa.html). This method requires a minimum of three localities per time bin. However, four time bins contained only one locality (6.5–6.3 Ma, 6.3–6 Ma, 5.3–5 Ma, and 4.8–4.5 Ma) and two time bins (6.8–6.5 Ma and 5–4.8 Ma) contained no HBL (= 0 km^2).

These sampling metrics are up-to-date as of 1st December 2017.

3.2.4 Climate proxies

Many publications have proposed or found a relationship between early hominin evolution and aridity (deMenocal, 2004; Blumenthal *et al.*, 2017), temperature (Passey *et al.*, 2010), soil carbonate $\delta^{13}C$ (Levin *et al.*, 2015), and lake levels (Maslin & Trauth, 2009). However, most climate records lack the duration and/or temporal resolution to be comparable to the fossil record (Hopley, 2018, *in review*). Here, all diversity estimates and sampling metrics were compared to the 8.0-Ma East African terrigenous dust flux record from Ocean Drilling Program (ODP) Site 721/722 (Arabian Sea) (deMenocal, 1995). In addition, taxic diversity, sampling metrics, and climate data were compiled for the Plio-Pleistocene of Ethiopia, Kenya, and Tanzania, to test climatic-forcing hypotheses specific to the EARS. This included the 5.0-Ma West African terrigenous dust flux record from ODP Site 659 (East Atlantic) (Tiedemann *et al.*, 1994) and the lake variability index (LVI; Shultz & Maslin, 2013). Both dust flux records were interpolated to 0.05 Ma (50 thousand year [kyr]) intervals using the shape-preserving Piecewise Cubic Hermite Interpolating Polynomial and the mean calculated for each time bin. To convert the LVI into 0.25 Ma time bins (in the original publication LVI is given in 0.05 Ma time bins; see Fig. 1 in Shultz & Maslin 2013), the mean and maximum value in each time bin were taken. (LVI mean and maximum produce highly

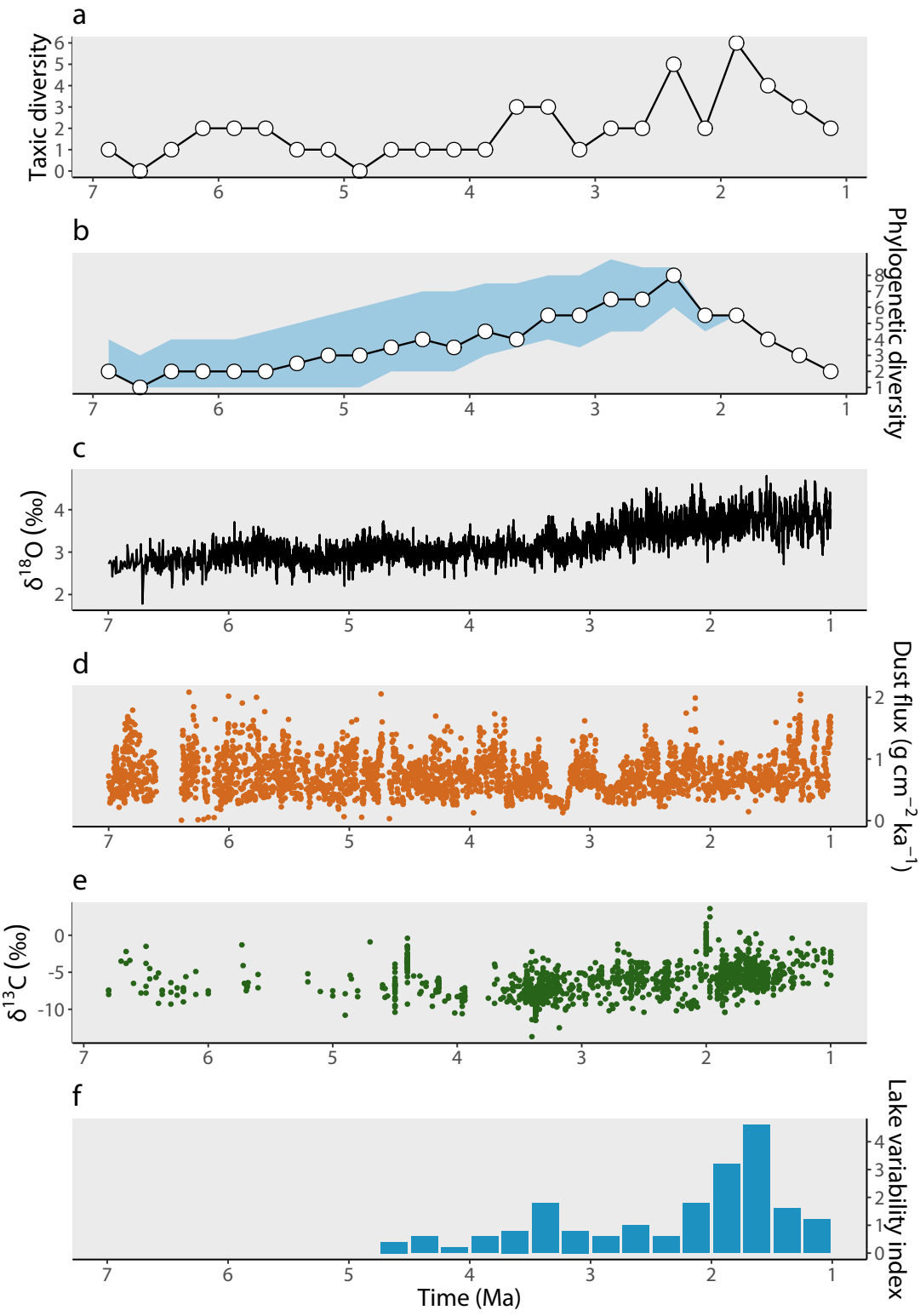

Figure 3.1: Early hominin diversity estimates and climate proxies through geological time. (**a**) Taxic diversity estimate (TDE). (**b**) Composite phylogenetic diversity estimate (PDE) based on the median of those in Fig. 2.2*b–e*. Data are plotted at the midpoint age of each time bin. (**c**) Global benthic δ^{18}O curve (Zachos *et al.*, 2001), where the higher the value the cooler the climate. (**d**) Terrigenous dust flux data (deMenocal, 1995), where the higher the value the drier the climate. (**e**) Eastern African soil carbonate δ^{13}C data (Levin, 2015), where the higher the value the more grassland (C_4) vegetation. (**f**) Lake variability index (Shultz & Maslin, 2013).

congruent results in all analyses, so only the results of the mean value are presented.)

3.2.5 Statistical tests

Spearman's rank (ρ) and the Kendall tau rank (τ) correlation coefficients were used to compare all diversity estimates, sampling metrics, and climate proxies (Hammer & Harper, 2006). Time series were de-trended and corrected for temporal autocorrelation by generalised differencing (GD) prior to regression (McKinney, 1990; R code from G.T. Lloyd: http://www.graemetlloyd.com/methgd.html). Long-term trends and autocorrelation tend to result in spurious detection of correlation between time series (McKinney, 1990) and must be removed prior to performing statistical tests (Benson & Butler, 2011). Given the large number of statistical tests performed, the significance of correlations were evaluated based on original p-values and p-values adjusted for the implementation of multiple tests using the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995).

In order to disentangle the underlying mechanism linking the rock record, fossil record, true diversity, and extrinsic abiotic factors, Generalised Least Squares (GLS) regression modelling was performed to explore the possibility of multiple explanatory variables controlling early hominin TDE (Benson & Mannion, 2012). GLS regression modelling has the benefit of assessing the fit of multiple dependent variables while simultaneously accounting for temporal autocorrelation using a first-order autoregressive model, which seeks autocorrelation up to a lag of 1 in either direction. The second-order Akaike Information Criterion (AICc), corrected for finite sample sizes, and the relative likelihood of each model based on Akaike weights (w_i) were used to assess model fit (Johnson & Omland, 2004). Models were created for all possible combination of parameters plus an intercept-only null model, representing statistically random variation around a constant mean. We calculated R^2 manually using the generalised R^2 of Nagelkerke (1991), which represents an estimate of the amount of variance explained by each model (Equation 3.1):

$$R^2 = 1 - e^{-1\times(2/n)\times[l(0)-l(\beta)]} \tag{3.1}$$

Where $l(0)$ represents the log-likelihood of the fitted model, $l(\beta)$ the log-likelihood of the null model, and n the sample size (or length of the time series). The Jarque-Bera test and Breusch-Pagan test were used to assess the normality and heteroskedasticity of residuals. Heteroskedasticity may cause overestimation of model fit, however, no cases of heteroskedasticity were found. Wald-Wolfowitz runs tests were also performed to investigate the null hypothesis of randomness and data independence in a time series (Hammer & Harper, 2006). All analyses were performed in R 3.5.0 (R Development Core Team, 2018) and an R script for Chapter 3 can be found in Appendix B.

3.2.6 Collector curves

Collector curves (or species discovery curves) (Cain, 1938) provide a tool for assessing the rate at which our knowledge of a group changes with increasing study effort. Knowledge by definition can only improve: each new fossil discovery can either increase diversity or diversity can remain unchanged; diversity cannot decrease unless through taxonomic revision. Collector curves were constructed by plotting the cumulative number of named taxa and HBF in each decade starting in 1920 (10 time bins in total),

where research time acts as a proxy for study effort (Benton, 1998). Information on nomen announcement year and the year each formation yielded a hominin fossil were taken from the HFBD and can be found in supplementary appendix Table S6, S7, and S8 in Maxwell *et al.* (2018).

3.3 Results

3.3.1 Statistical comparisons between diversity and sampling metrics

Early hominin diversity estimates are shown in Fig. 3.1. Sampling metrics are shown in Fig. 3.2. Taxic diversity (TDE) correlates significantly with both HBC (Fig. 3.3*a*) and HBF (Fig. 3.3*b*). Both correlations, however, become non-significant after the application of the False Discovery Rate (FDR) procedure, and these relationships disappear entirely when HBC and HBF are compared to each PDE (Table 3.2). Unexpectedly, the correlation between TDE and PBF is substantially stronger, and significant both before and after FDR correction (Fig. 3.3*c*), while there is no significant correlation between TDE and MBF before FDR correction (Fig. 3.3*d*). TDE also correlates significantly with the palaeoarea surrounding all hominin-bearing localities (HBL) (Table 3.2). However, there is no correlation between PBF, MBF, or palaeoarea and any PDE (Table 3.2). Finally, HBC ($p = 0.003$), HBF ($p < 0.001$), PBF ($p = 0.012$), MBF ($p < 0.001$), and palaeoarea ($p < 0.001$) all deviate significantly from randomness.

3.3.2 Gap analysis

The ghost lineage diversity estimate (GDE) for each phylogeny is shown in Fig. 3.4. The only statistically significant GDE to correlate with HBC is the SGDE, and each GDE correlates significantly with HBF except for HGDE (Fig. 3.5 and Table 3.3). However, these correlations are rendered non-significant after FDR correction and, except for a highly significant correlation between SGDE and PBF, no other GDE correlates with PBF, MBF, or palaeoarea after FDR correction. The ratio of cladistically-implied to sampled lineages (GDE:PDE) for each phylogeny is shown in Fig. 3.6. The GDE:PDE correlates strongly with HBF, PBF, and palaeoarea across all phylogenies, however, there is no correlation between any GDE:PDE and MBF (Fig. 3.7 and Table 3.3). In all statistically significant tests the correlation is negative, indicating that periods of poor sampling (= low collection/formation counts) correspond to periods with a higher relative proportion of ghost lineages (= high GDE:PDE).

3.3.3 Statistical comparisons between diversity and climate

East African aridity and lake abundance are shown in Fig. 3.1. Global palaeotemperature ($\delta^{18}O$) and East African soil carbonate $\delta^{13}C$ (a proxy for vegetation type) are also included to show the climatic and environmental context of early hominin diversification (Fig. 3.1).

Pairwise correlations. There is no significant correlation between diversity (either taxic or phylogenetic) and the interpolated aridity curve of deMenocal (1995) (Table 3.2). Eastern African TDE (TDE_{EA}) and lake variability (LVI) do not correlate (whether LVI is expressed as the mean or maximum value per time bin). The only significant correlations in the EARS are those between TDE_{EA} and HBC_{EA} and PBF_{EA} (Fig. 3.8 and Table 3.4). However, these correlations are rendered non-significant after FDR correction.

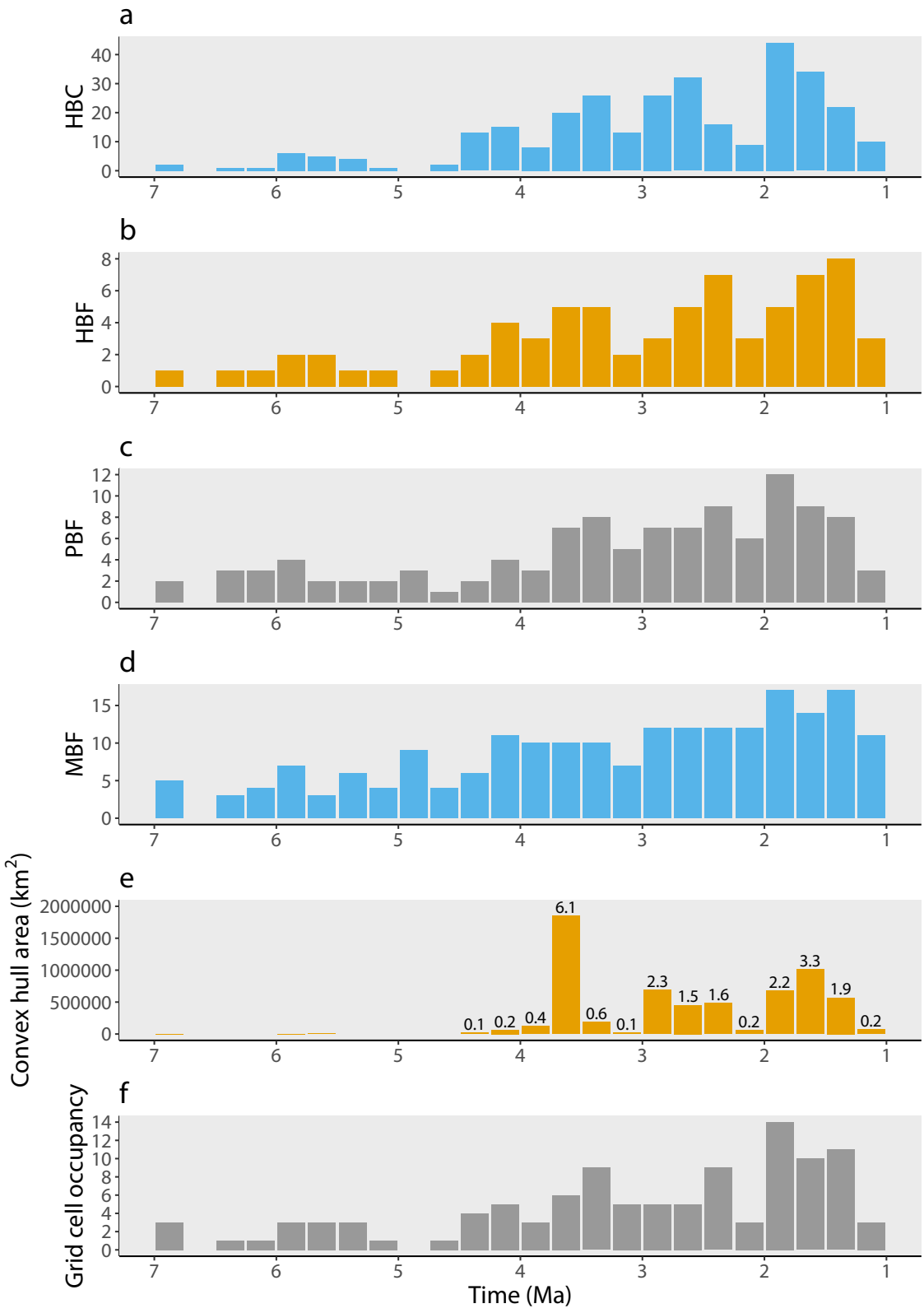

Figure 3.2: Sampling metrics through geological time. (**a**) HBC. (**b**) HBF. (**c**) PBF. (**d**) MBF. (**e**) Convex hull area (= palaeoarea). The number atop each bar refers to the percent of the African continent sampled in each time bin (percentage < 0.1% are not shown). The removal of the spatial outlier Koro Toro (Chad) from the 3.6 Ma time bin leads to a ten-fold reduction in the sampled area estimate from 6.1% to 0.6% (comparable to the adjacent time bins). (**f**) Grid cell occupancy. See Table 3.1 for an explanation of each abbreviation.

Table 3.2: Results of the statistical analyses comparing diversity estimates, sampling metrics, and terrigenous dust flux. See Table 3.1 for an explanation of each abbreviation. Statistically significant correlations are shown in bold. *Significant at $p \leq 0.05$. **Significant at $p \leq 0.05$ after false discovery rate (FDR) correction (Benjamini & Hochberg, 1995).

Comparison	Spearman's ρ	Kendall's τ
TDE *versus* HBC	**0.457***	**0.312***
SPDE *versus* HBC	0.065	0.028
D1PDE *versus* HBC	0.137	0.051
HPDE *versus* HBC	0.241	0.178
D2PDE *versus* HBC	0.073	0.020
TDE *versus* HBF	**0.618***	**0.439***
SPDE *versus* HBF	0.215	0.123
D1PDE *versus* HBF	0.171	0.099
HPDE *versus* HBF	0.262	0.178
D2PDE *versus* HBF	0.166	0.099
TDE *versus* PBF	**0.742****	**0.565****
SPDE *versus* PBF	0.235	0.154
D1PDE *versus* PBF	0.347	0.257
HPDE *versus* PBF	0.301	0.178
D2PDE *versus* PBF	0.354	0.241
TDE *versus* MBF	0.194	0.138
SPDE *versus* MBF	0.272	0.186
D1PDE *versus* MBF	0.294	0.209
HPDE *versus* MBF	0.067	0.067
D2PDE *versus* MBF	0.354	0.241
TDE *versus* palaeoarea	**0.495***	**0.383***
SPDE *versus* palaeoarea	0.184	0.099
D1PDE *versus* palaeoarea	0.121	0.075
HPDE *versus* palaeoarea	0.131	0.091
D2PDE *versus* palaeoarea	0.122	0.075
TDE *versus* aridity	0.123	0.075
SPDE *versus* aridity	0.126	0.107
D1PDE *versus* aridity	0.203	0.099
HPDE *versus* aridity	0.183	0.099
D2PDE *versus* aridity	0.243	0.146
HBC *versus* HBF	**0.490***	**0.368***
HBC *versus* PBF	**0.629***	**0.478****
HBC *versus* MBF	0.299	0.225
HBC *versus* palaeoarea	**0.580***	**0.439***
HBC *versus* aridity	0.070	0.051
HBF *versus* PBF	**0.638****	**0.447***
HBF *versus* MBF	0.246	0.162
HBF *versus* palaeoarea	**0.434***	**0.312***
HBF *versus* aridity	0.173	0.130
PBF *versus* MBF	**0.419***	**0.304***
PBF *versus* palaearea	**0.538***	**0.407***
PBF *versus* aridity	0.021	−0.028
MBF *versus* palaeoarea	0.399	0.281
MBF *versus* aridity	0.286	0.209

Table 3.3: Results of the statistical analyses comparing ghost lineage diversity estimates and sampling metrics. See Table 3.1 and text for an explanation of the abbreviations of each diversity estimate and sampling metric. Statistically significant correlations are shown in bold. *Significant at $p \leq 0.05$. **Significant at $p \leq 0.05$ after false discovery rate (FDR) correction (Benjamini & Hochberg, 1995).

Comparison	Spearman's ρ	Kendall's τ
SGDE *versus* HBC	**−0.455***	**−0.352***
D1GDE *versus* HBC	−0.325	−0.225
HGDE *versus* HBC	−0.181	−0.146
D2GDE *versus* HBC	−0.360	−0.257
SGDE *versus* HBF	**−0.644****	**−0.462****
D1GDE *versus* HBF	**−0.558***	**−0.399***
HGDE *versus* HBF	−0.294	−0.194
D2GDE *versus* HBF	**−0.540***	**−0.383***
SGDE *versus* PBF	**−0.627****	**−0.462****
D1GDE *versus* PBF	−0.403	**−0.304***
HGDE *versus* PBF	−0.312	−0.209
D2GDE *versus* PBF	−0.412	**−0.304***
SGDE *versus* MBF	−0.078	−0.036
D1GDE *versus* MBF	0.096	0.075
HGDE *versus* MBF	−0.166	−0.115
D2GDE *versus* MBF	0.095	0.075
SGDE *versus* palaeoarea	−0.32	−0.233
D1GDE *versus* palaeoarea	−0.166	−0.075
HGDE *versus* palaeoarea	−0.181	−0.154
D2GDE *versus* palaeoarea	−0.134	−0.059
SGDE:SPDE *versus* HBC	**−0.459***	**−0.328***
D1GDE:D1PDE *versus* HBC	−0.394	−0.273
HGDE:HPDE *versus* HBC	−0.320	−0.249
D2GDE:D2PDE *versus* HBC	**−0.436***	**−0.304***
SGDE:SPDE *versus* HBF	**−0.573***	**−0.360***
D1GDE:D1PDE *versus* HBF	**−0.603***	**−0.415***
HGDE:HPDE *versus* HBF	**−0.500***	**−0.328***
D2GDE:D2PDE *versus* HBF	**−0.599***	**−0.383***
SGDE:SPDE *versus* PBF	**−0.653****	**−0.486****
D1GDE:D1PDE *versus* PBF	**−0.604****	**−0.399***
HGDE:HPDE *versus* PBF	**−0.555***	**−0.360***
D2GDE:D2PDE *versus* PBF	**−0.617****	**−0.415***
SGDE:SPDE *versus* MBF	−0.213	−0.123
D1GDE:D1PDE *versus* MBF	−0.233	−0.146
HGDE:HPDE *versus* MBF	−0.297	−0.186
D2GDE:D2PDE *versus* MBF	−0.194	−0.115
SGDE:SPDE *versus* palaeoarea	**−0.573***	**−0.360***
D1GDE:D1PDE *versus* palaeoarea	**−0.603****	**−0.415***
HGDE:HPDE *versus* palaeoarea	**−0.500***	**−0.328***
D2GDE:D2PDE *versus* palaeoarea	**−0.599****	**−0.383***

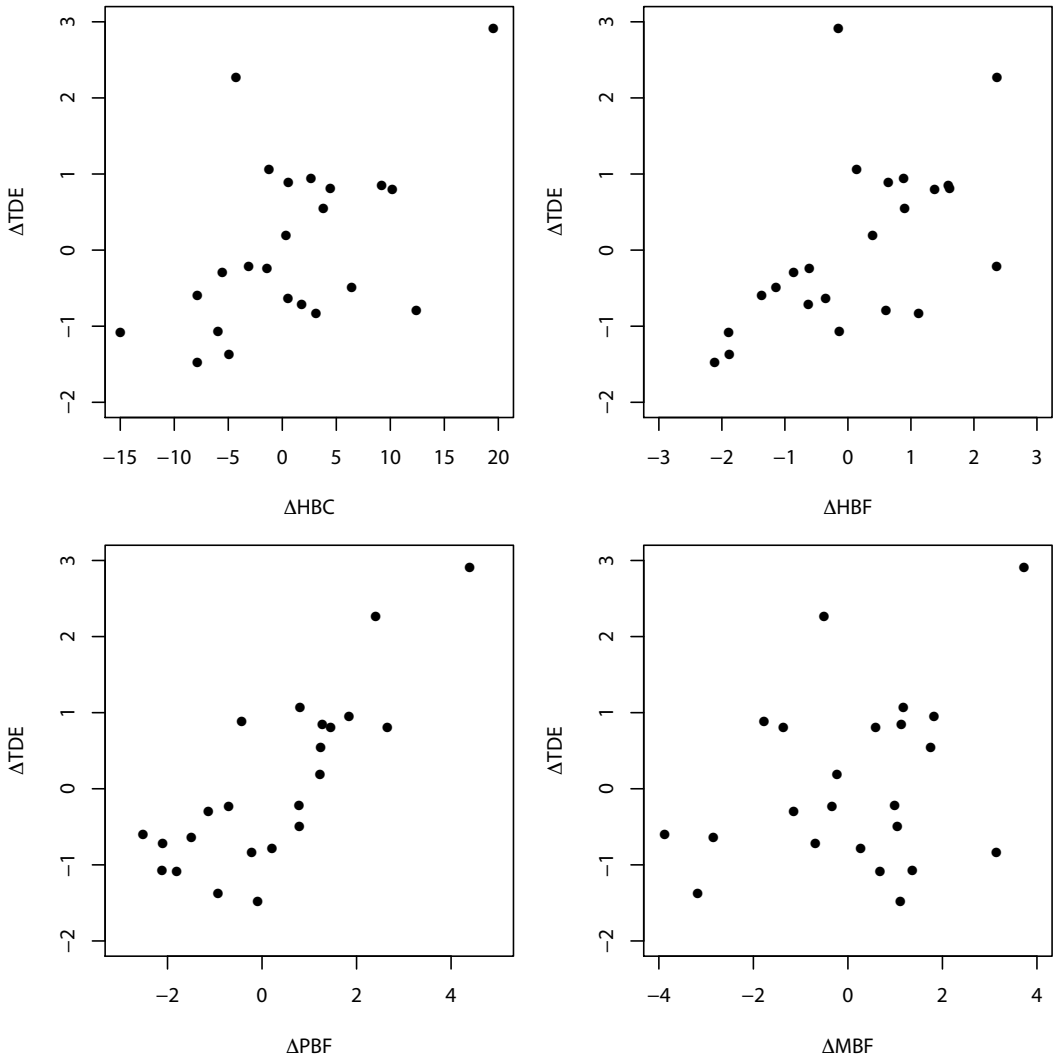

Figure 3.3: Scatter plots showing the relationships between the taxic diversity estimate (TDE) and sampling metrics. (**a**) TDE against HBC. (**b**) TDE against HBF. (**c**) TDE against PBF. (**d**) TDE against MBF. See Table 3.1 for an explanation of each abbreviation and Table 3.2 for correlation coefficients and p-values.

Generalised Least Squares. No model fits TDE better than PBF and East African aridity combined (R^2 = 0.75). However, the removal of aridity from the most supported model yields a model of equivalent explanatory power (R^2 = 0.72), and a model including only aridity is the least supported model overall (R^2 = 0.02) with a w_i less than the null. In every model with a non-negligible weight (w_i > 0.01), the only significant predictors of TDE are HBC and PBF (Table 3.5). The four models with the highest rank all contain PBF, while the lowest four contain collections and aridity.

In the EARS analysis, TDE_{EA} is best explained by PBF_{EA} (R^2 = 0.60) (Table 3.6). However, a combination of PBF_{EA} + East African aridity is the second-best model, with a difference in w_i of less than 0.001 (R^2 = 0.67), and a combination of PBF_{EA} + West African aridity the third-best model (R^2 = 0.67). In the four models with the highest rank, PBF_{EA} appears three times while each aridity proxy and HBC_{EA} appear only once (Table 3.6). However, within these models, the only significant predictors are HBC_{EA} and PBF_{EA} (shown in bold). The four models with the lowest rank contain each aridity proxy as single

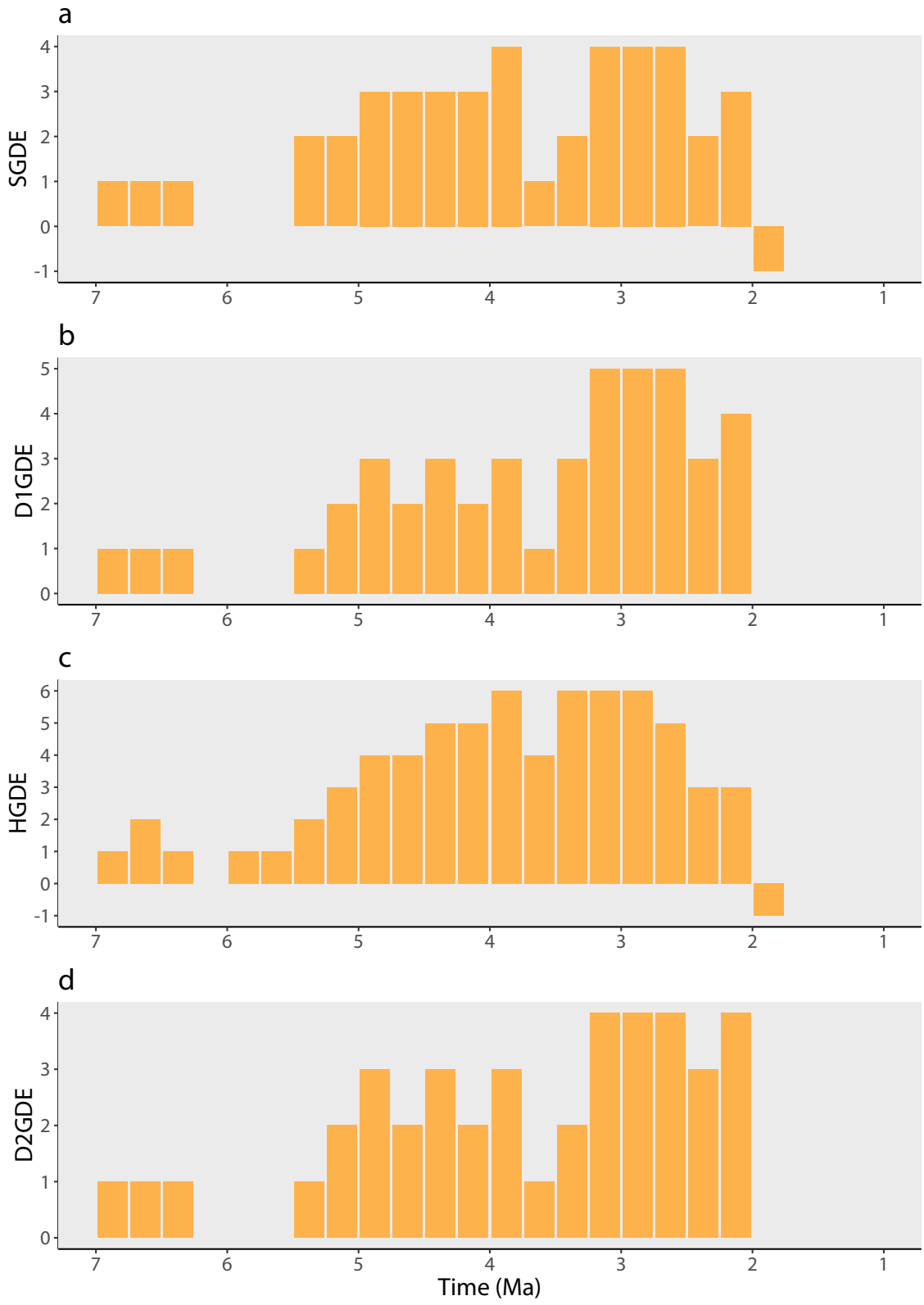

Figure 3.4: Ghost lineage diversity estimate (GDE) for each phylogeny through geological time. (**a**) SGDE. (**b**) D1GDE. (**c**) HGDE. (**d**) D2GDE. See Table 3.1 for an explanation of the abbreviations of each ghost lineage diversity estimate.

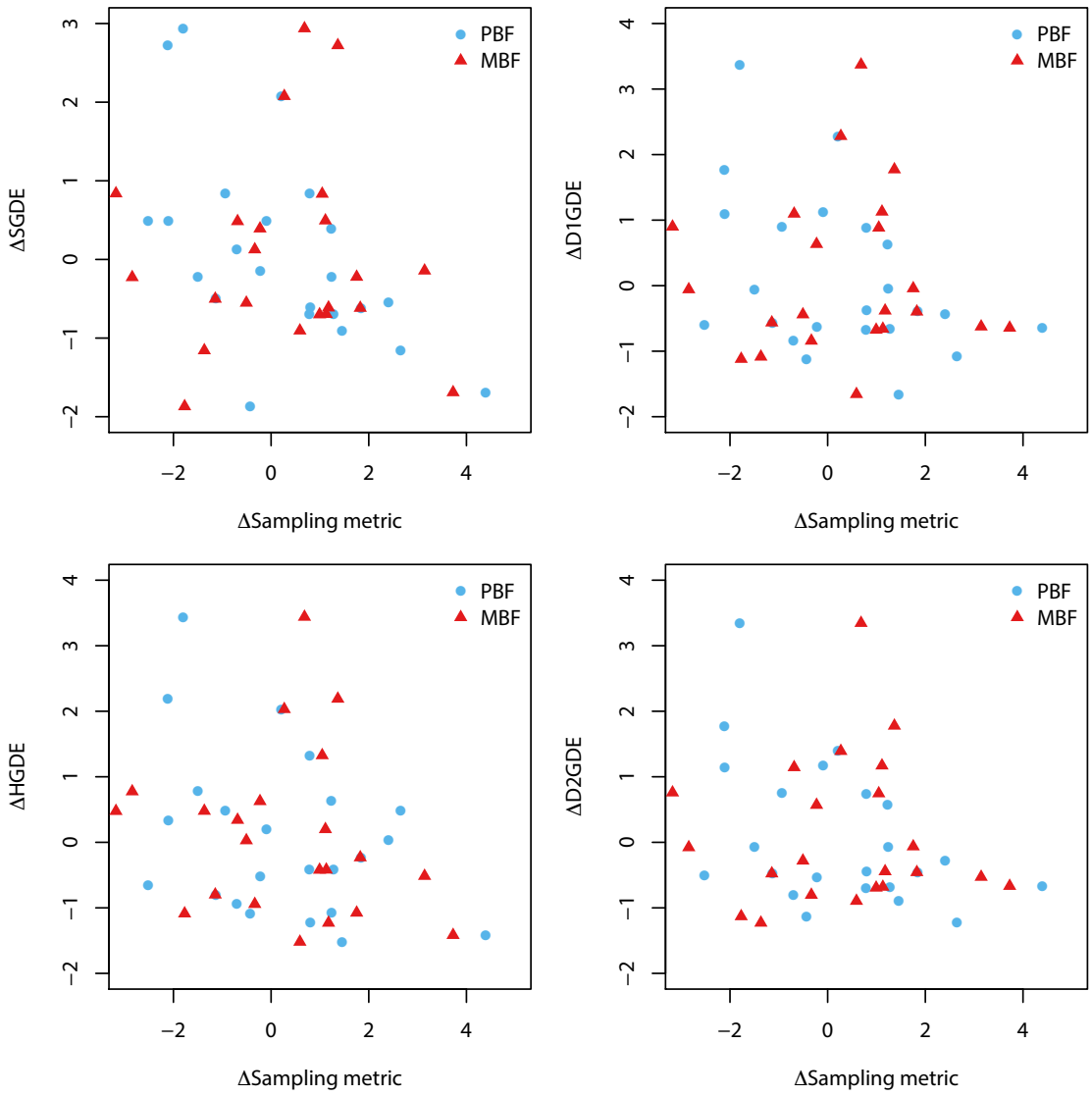

Figure 3.5: Scatter plots showing the relationships between each ghost lineage diversity estimate (GDE) and sampling metrics. (**a**) SGDE. (**b**) D1GDE. (**c**) HGDE. (**d**) D2GDE. See Table 3.1 for an explanation of each abbreviation and Table 3.3 for correlation coefficients and p-values.

predictors and in combination, plus a model combining LVI and both aridity proxies.

3.3.4 Collector curves

The cumulative number of early hominin species has increased exponentially over research time (R^2 = 0.98, $y = 2 \times 10^{-27} e^{0.032x}$), showing no sign of an asymptote (Fig. 3.9). The same pattern is found for the cumulative number of HBF (R^2 = 0.94, $y = 4 \times 10^{-30} e^{0.036x}$) and as a result both species discovery and formation discovery curves correlate strongly (R^2 = 0.97, $y = 0.57x + 0.17$). This implies that for the discovery of every two new HBF, one new hominin species is discovered (Fig. 3.9).

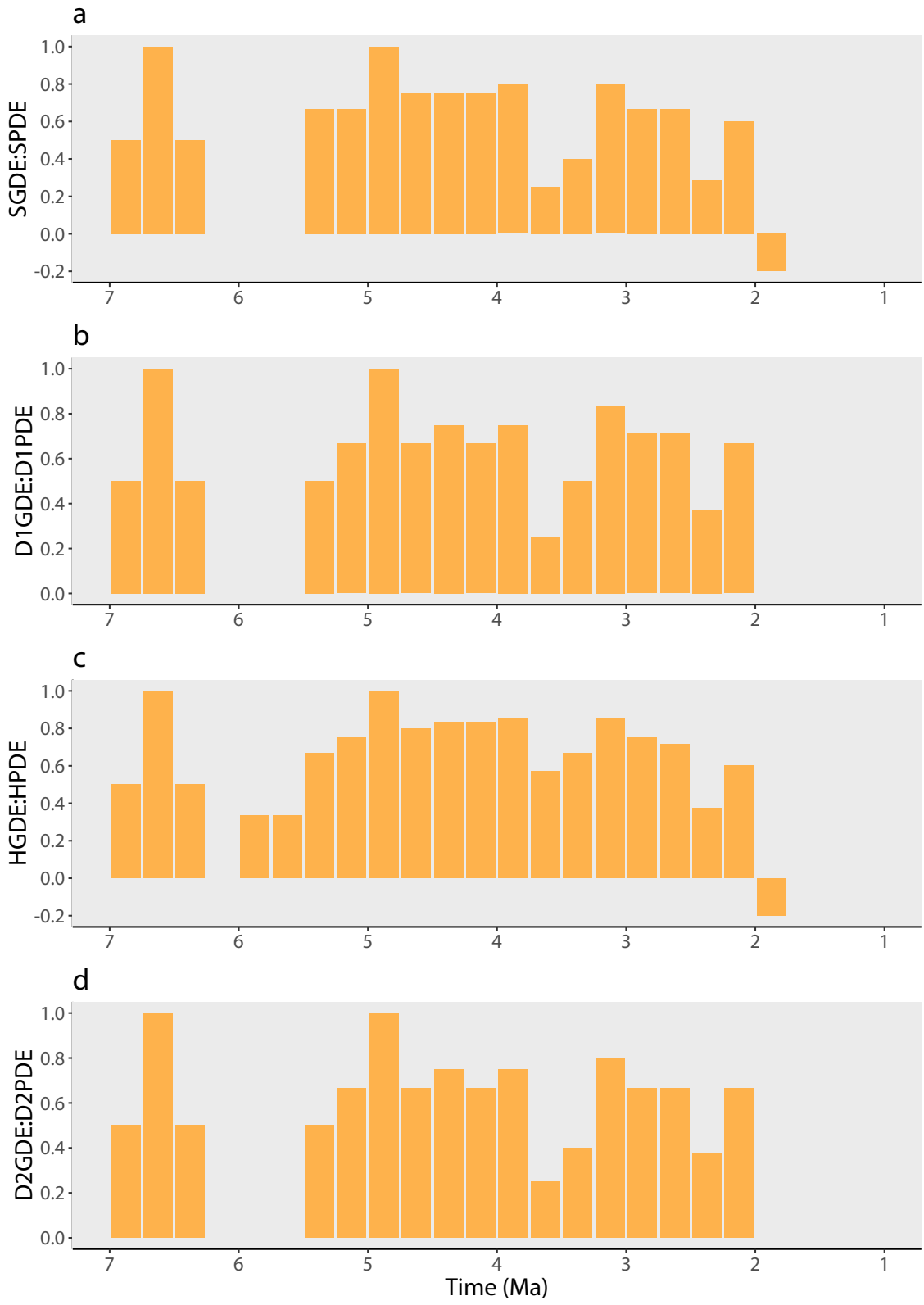

Figure 3.6: Ratio of cladistically-implied to sampled lineages (GDE:PDE) for each phylogeny through geological time. (**a**) SGDE:SPDE. (**b**) D1GDE:D1PDE. (**c**) HGDE:HPDE. (**d**) D2GDE:D2PDE. See Table 3.1 for an explanation of each abbreviation.

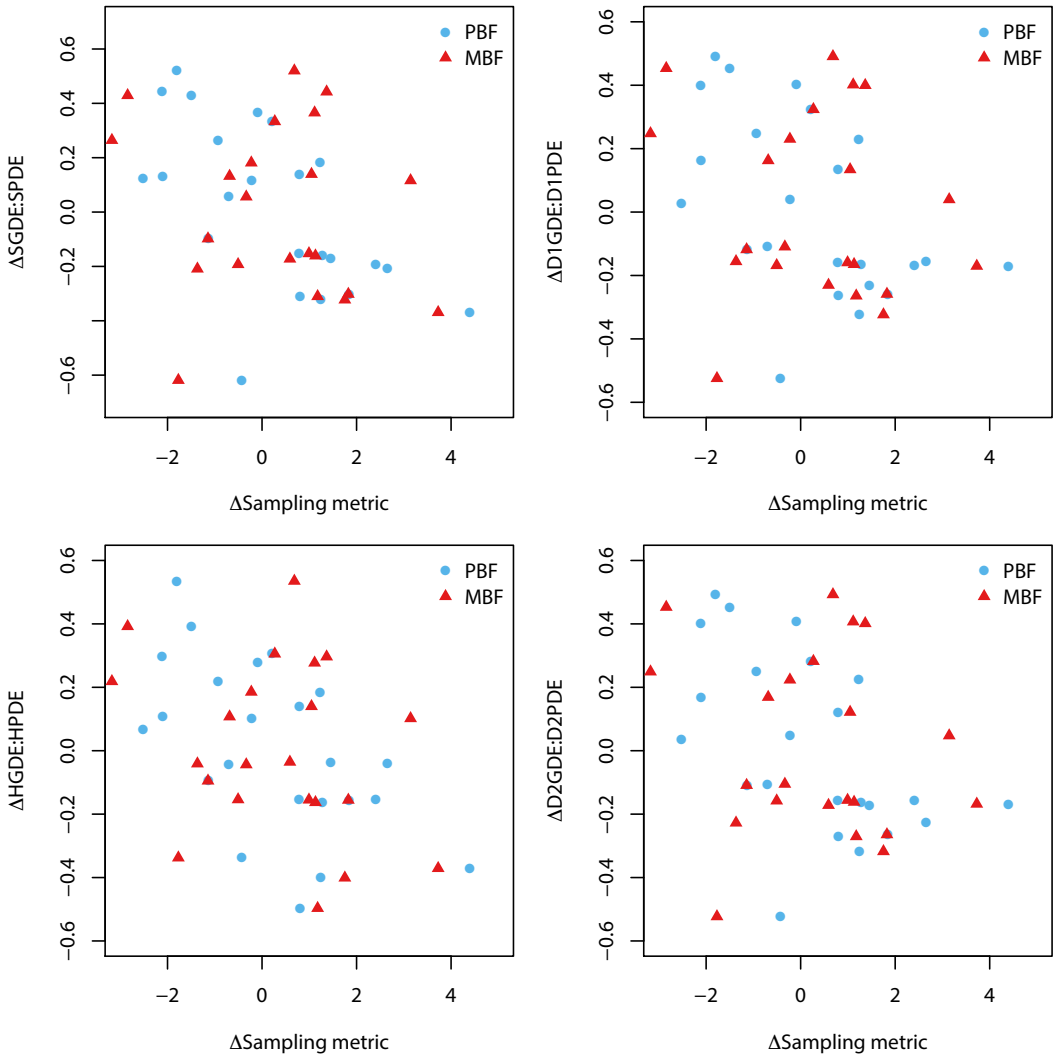

Figure 3.7: Scatter plots showing the relationships between each ratio of cladistically-impled to sampling lineages (GDE:PDE) and sampling metrics. (**a**) SGDE:SPDE. (**b**) D1GDE:D1PDE. (**c**) HGDE:HPDE. (**d**) D2GDE:D2PDE. See Table 3.1 for an explanation of each abbreviation and Table 3.3 for correlation coefficients and *p*-values.

3.4 Discussion

3.4.1 Is hominin diversity controlled by sampling?

The significant correlation between TDE and sampling metrics could indicate (1) major geological and anthropogenic controls on the sampling of the early hominin fossil record, or (2) redundancy between early hominin TDE and sampling metrics based solely on counts of early hominin fossils (Benton *et al.*, 2011; Dunhill *et al.*, 2014*b*). Hominins, like apes today, were probably a minor component of terrestrial ecosystems during their earliest evolution (Wood & Harrison, 2011) and are, therefore, expected to be found in a small number of collections/formations during periods of genuine relative low diversity. Conversely, during periods of genuine relative high diversity, hominin fossils are expected to make their way into a greater number of collections/formations. The drive-response relationship between TDE and

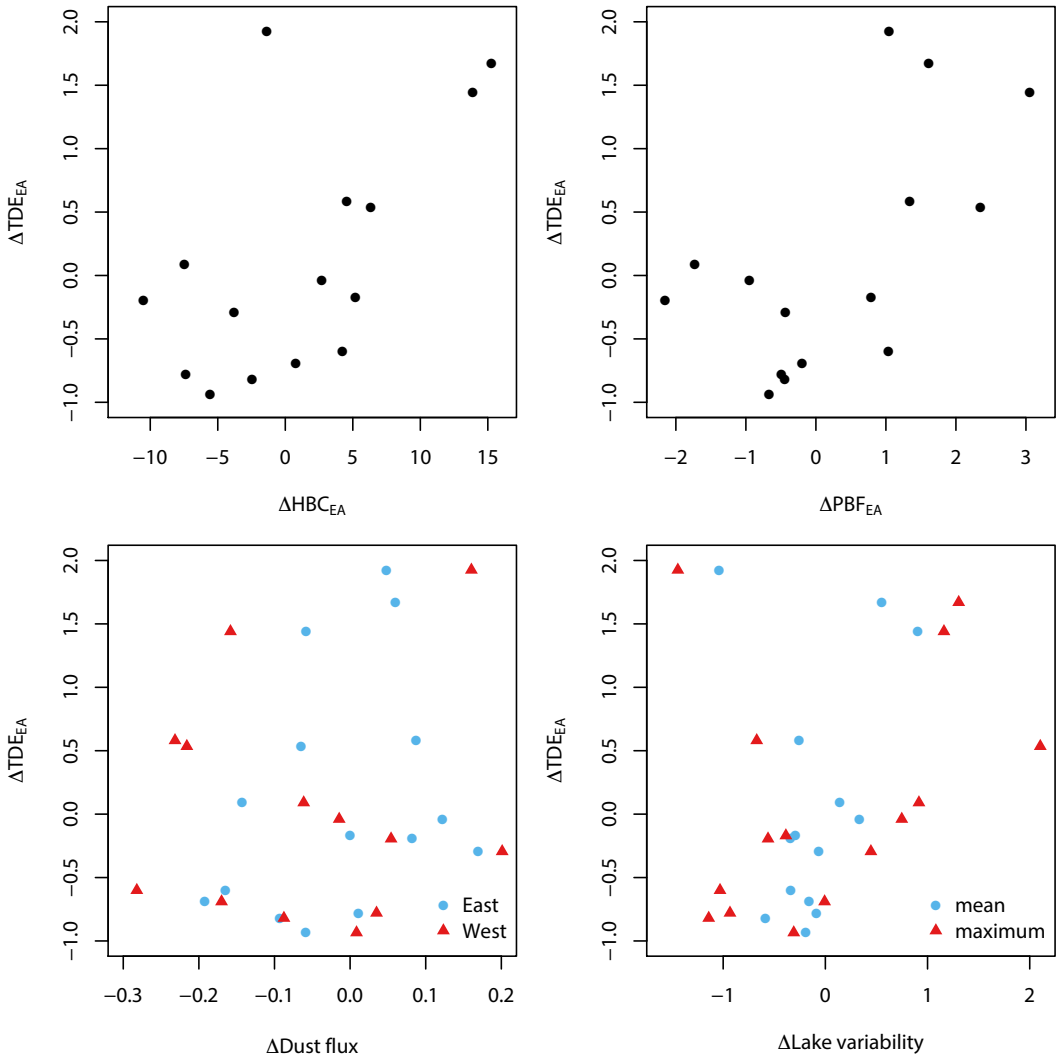

Figure 3.8: Scatter plots showing the relationships between Eastern African taxic diversity, sampling metrics, and climate proxies. (**a**). TDE_{EA} against HBC_{EA}. (**b**) TDE_{EA} against PBF_{EA}. (**c**) TDE_{EA} against East and West African aridity. (**d**) TDE_{EA} against the lake variability index (LVI), using both the mean and maximum value per time bin variants. See Table 3.1 for an explanation of each abbreviation and Table 3.4 for correlation coefficients and *p*-values.

HBC/HBF is, therefore, most likely bi-directional. This non-independence (HBF are as likely to drive TDE as TDE is HBF) calls into question their usefulness as a meaningful sampling metric (Benton *et al.*, 2011, 2013; Dunhill *et al.*, 2014*b*; Benton, 2015).

To test whether HBF is an independent sampling metric for the diversity of early hominins, a series of randomised trials were carried out (using the method described in Benton *et al.*, 2011) to determine whether the correlation between TDE and HBF is an inevitable result of using a strict FFC. For each HBF, we generated a random species diversity of 0, 1, or 2, where 0 is equivalent to no fossil finds in a formation and 2 the maximum in-bin taxon to formation ratio in the early hominin fossil record (specifically the 6.3–6.0 Ma time bin). These data were then summed for each time bin according to the number of formations present and forty simulations were performed to assess the statistical significance

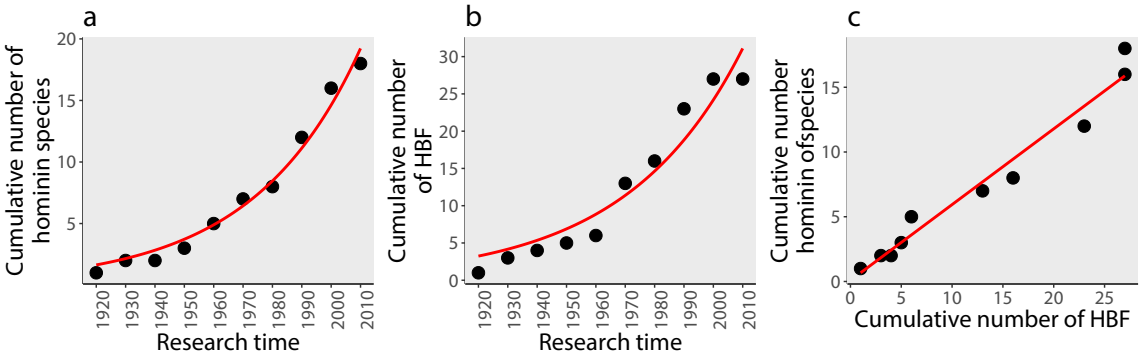

Figure 3.9: Collector curves. (**a**) Cumulative number of early African hominin species through research time (R^2 = 0.98, $y = 2 \times 10^{-27} e^{0.032x}$). (**b**) Cumulative number of hominin-bearing formations (HBF) through research time (R^2 = 0.94, $y = 4 \times 10^{-30} e^{0.036x}$). (**c**) Cumulative number of early African hominin species against cumulative number of HBF (R^2 = 0.97, $y = 0.57x + 0.17$). Data plotted at the first year of each decade in (a) and (b). The exponential rise in the cumulative number of early hominin species (a) has remained largely unchanged since Foley (2005:70). Note that the cumulative number of HBF through research time in (b) can be explained slightly better by a fourth-order polynomial (R^2 = 0.99) which reached an asymptote around the year 2000. The data used to create each collector curve can be found in supplementary appendix Table S6, S7, and S8 in Maxwell *et al.* (2018).

of the results at the 95% confidence level (Benton *et al.*, 2011). Not surprisingly, the simulated hominin diversity totals and HBF counts gave rank correlation coefficients equivalent to the actual data (mean = 0.666, median = 0.694). Moreover, the simulated hominin diversity totals showed a stronger correlation to HBF than the real data in 78% of cases. (Note we repeated the entire randomised trial 20 times and found this result to be highly robust.) Not only does this indicate that random data can produce a better fit than reality, but that this is by far the most likely result (*sensu* Benton *et al.*, 2011). These simulations indicate that the tight correlation observed between TDE and HBF is undoubtedly a mix of both redundancy—HBF are probably driven more by TDE than by sampling—and a genuine signal of episodic preservation and sporadic sampling.

Species and formation discovery (collector) curves reinforce this finding (Fig. 3.9*a–b*). The rate of species discovery has increased exponentially over the last century, showing no sign of an asymptote, and thus showing an identical trend to that reported by Foley (2005). That the rate of species discovery has continued to increase rapidly in the 13 years since the first collector curve (Foley, 2005) reinforces that it may be premature to directly link current patterns of taxic diversity to causal agents (Smith & Wood, 2017). The rate of HBF discovery mirrors the rate of species discovery almost identically (R^2 = 0.97, $p < 0.001$; Fig. 3.9*c*), and the half-life (the year in which each curve reached half their current number of species/formations; Bernard *et al.*, 2010) closely match (species = 1990, formations = 1981). This further indicates that, for early hominins, the discovery of new species and the discovery of new HBF are intimately linked, having tracked one another over research time.

To mitigate the issue of redundancy between TDE and HBC/HBF and more accurately quantify the extent to which sampling controls diversity, all statistical tests were repeated using a wider FFC based on the number of primate-bearing formations (PBF) and a comprehensive FFC based on the number of terrestrial (i.e., non-marine) macro-mammal-bearing formations (MBF). FFCs that include both HBF and those primate/macro-mammal-bearing formations that have not yielded a hominin are *a priori*

Table 3.4: Results of the statistical analyses comparing Eastern African taxic diversity, sampling metrics, and climate proxies. See Table 3.1 for an explanation of the abbreviations. East African aridity is based on terrigenous dust flux from deMenocal (1995). West African aridity is based on terrigenous dust flux from Tiedemann *et al.* (1994). Results are shown for mean LVI. *Significant at $p \leq 0.05$ (shown in bold). **Significant at $p \leq 0.05$ after false discovery rate (FDR) correction (Benjamini & Hochberg, 1995).

Comparison	Spearman's ρ	Kendall's τ
TDE_{EA} *versus* LVI	0.304	0.257
TDE_{EA} *versus* $aridity_{EA}$	0.339	0.200
TDE_{EA} *versus* $aridity_{WA}$	0.118	0.048
TDE_{EA} *versus* HBC_{EA}	**0.546***	**0.410***
TDE_{EA} *versus* PBF_{EA}	**0.575***	**0.390***
LVI *versus* $aridity_{EA}$	0.064	0.029
LVI *versus* $aridity_{WA}$	−0.068	−0.048
LVI *versus* HBC_{EA}	0.404	0.238
LVI *versus* PBF_{EA}	0.275	0.143
$Aridity_{EA}$ *versus* $aridity_{WA}$	**0.525***	**0.390***
$Aridity_{EA}$ *versus* HBC_{EA}	−0.004	−0.010
$Aridity_{EA}$ *versus* PBF_{EA}	−0.064	−0.029
$Aridity_{WA}$ *versus* HBC_{EA}	−0.164	−0.124
$Aridity_{WA}$ *versus* PBF_{EA}	−0.221	−0.105
HBC_{EA} *versus* PBF_{EA}	**0.864****	**0.714****

better sampling metrics than HBF alone, because they represent a closer approximation of supposed total sampling effort (i.e., collection effort and its underlying driver, the availability of sedimentary rock capable of preserving hominin fossils; Dunhill *et al.*, 2018). HBF alone, in contrast, ignores all sampling opportunities that failed to find a hominin (non-occurrence) and is, therefore, not an approximation of total sampling effort (Dunhill *et al.*, 2018).

The remarkably strong correlation between TDE and PBF supports the RRB hypothesis (Raup, 1972, 1976*b*; Smith, 2001), that observed diversity is largely controlled by the likelihood of sampling a primate fossil. That this correlation completely disappears for each PDE—despite a significant negative correlation between sampling (namely PBF and palaeoarea) and GDE:PDE—indicates (1) that the application of only a partial correction for sampling produces diversity estimates that show little relation to sampling metrics, and (2) that, for hominins at least, the relative proportion of cladistically-implied to sampled lineages is a good predictor of sampling (the same finding Benton *et al.* (2011) reported for dinosaurs). The fact that periods of reduced rock availability (= low PBF) also have reduced geographical spread of collection effort (= low palaeoarea), less observed diversity (= low TDE), and a greater number of ghost lineages (= high GDE and GDE:PDE) (Figs. 3.4 and 3.6) strengthens the validity of these independently-derived sampling metrics, and adds further support to the RRB hypothesis.

A highly significant correlation between TDE and PBF on the one hand, and lack of a correlation between TDE and MBF on the other could have three possible explanations: (1) PBF is information redundant with respect to TDE, and MBF (= sampling) does not control diversity; (2) PBF is information redundant with respect to TDE, and MBF is too broad a measure of the amount of sampling effort in rock suitable for the preservation of a hominin; or, (3) PBF captures a genuine signal of fossil sampling that MBF does not, and largely controls observed taxic diversity. To assess the redundancy

Table 3.5: Results of the Generalised Least Squares analysis comparing taxic diversity, sampling metrics, and terrigenous dust flux. Models 1–8 include all possible combinations of HBC (collections), PBF (formations), and aridity plus a null model. Significant predictors are shown in bold font ($p \leq 0.05$).

Model	Parameters	df	w_i	AICc	logLik	R^2
5	**Formations** + aridity	5	0.39	57.94	−23.37	0.75
7	**Formations**	4	0.34	58.23	−24.83	0.72
2	Collections + formations + aridity	6	0.16	59.67	−22.78	0.77
3	Collections + **formations**	5	0.11	60.48	−24.64	0.73
6	**Collections**	4	0.00	72.83	−32.13	0.49
4	**Collections** + aridity	5	0.00	74.56	−31.68	0.51
1	Null	3	0.00	86.62	−40.22	0.00
8	Aridity	4	0.00	88.56	−40.00	0.02

argument against PBF, we repeated the randomised trials above (Benton *et al.*, 2011) and found that, although the mean rank correlation coefficient is moderately high (ρ = 0.580), only 13% (10 – 20% across the 40 re-runs) of simulations produced correlation coefficients stronger than the actual data. A highly significant correlation between TDE and PBF is, therefore, not an inevitable consequence of the data, suggesting a reduced role for redundancy compared to HBF. This interpretation is reinforced by the fact that the positive correlation with TDE actually increases from HBF to PBF. If redundancy were the main cause of these correlations, we would expect the correlation to become weaker the more inclusive the FFC. Further, it is unlikely that TDE drives PBF to the same extent that PBF drives TDE: 39% of PBF are non-hominin-bearing and fossiliferous formations are defined purely on lithostratigraphic grounds. We know of no formations subdivided more finely based on the presence/absence of primate fossils, or fluctuations in primate taxic diversity. Nonetheless, PBF might be partially redundant with TDE and information transfer (a non-parametric method used to detect causation; Hannisdal & Peters, 2011; Dunhill *et al.*, 2014*b*; Hannisdal & Liow, 2018) could go some way to quantifying the nature of this interaction. However, information transfer is impossible for such a small, geologically-brief group.

For rare and sporadically sampled clades such as hominins, comprehensive FFC might not capture the idiosyncratic nature of fossil preservation and discovery that wider FFC can (see the case of pterosaurs; Butler *et al.*, 2009; Benton *et al.*, 2011). A lack of correlation between TDE and MBF may be a product of most macro-mammals living in, or being preserved in, habitats that lacked hominins or were unsuitable for them in some way. For example, periods with high MBF could have high TDE if the mammals suitable for preservation in those formations are taphonomically comparable to hominins; but equally, periods with high MBF could have low TDE if the majority of formations preserve habitats unsuitable for hominins, no matter the amount of palaeoanthropological interest a formation receives. This appears to be the case for MBF which, despite containing PBF, correlates weakly with it (Table 3.2), implying greater environmental (and therefore facies) heterogeneity in the mammal fossil record. While cercopithecoid and hominoid primates are taphonomically comparable to hominins in terms of body size, morphology, and habitat preference (Bobe & Leakey, 2009), macro-mammals differ markedly in body size (by several orders of magnitude) and morphology and, as a result, enter the fossil record via different taphonomic pathways. Consequently, the distribution of body sizes in terrestrial mammal assemblages differs markedly by habitat, agent of accumulation, and climate (Behrensmeyer *et al.*, 2000).

Table 3.6: Results of the Generalised Least Squares analysis in the East African Rift System. Models 1–31 comprise all possible combinations of collections, formations, East African (Arabian Sea) and West African (East Atlantic) aridity, and lake variability as explanatory predictors of taxic diversity. Models are ranked in order of their relative likelihood according to Akaike weights (w_i), where the larger the value the more likely the model. For those models with an w_i within one-eighth of the best model, significant predictors are shown in bold font ($p \leq 0.05$).

Model	Parameters	AICw	AICc	logLik	R^2
29	**F**	0.17	42.47	−16.77	0.60
22	**F**E	0.17	42.47	−15.24	0.67
23	**F**W	0.15	42.78	−15.39	0.67
28	**C**	0.06	44.52	−17.80	0.55
18	CF	0.06	44.72	−16.36	0.62
24	**F**L	0.05	45.13	−16.56	0.61
8	CFE	0.04	45.18	−14.77	0.69
20	**C**W	0.04	45.25	−16.63	0.61
9	CFW	0.04	45.26	−14.81	0.69
14	**F**EW	0.03	45.78	−15.07	0.68
19	**C**E	0.03	45.90	−16.95	0.59
15	**F**EL	0.03	45.98	−15.17	0.68
16	**F**WL	0.03	46.23	−15.30	0.67
21	CL	0.02	47.00	−17.50	0.57
10	CFL	0.01	48.13	−16.25	0.63
13	CWL	0.01	48.26	−16.31	0.63
11	CEW	0.01	48.79	−16.58	0.61
12	CEL	0.01	48.98	−16.67	0.61
3	CFEW	0.01	49.08	−14.54	0.70
4	CFEL	0.01	49.51	−14.75	0.69
5	CFWL	0.01	49.54	−14.77	0.69
32	L	0.01	49.56	−20.32	0.38
7	FEWL	0.00	50.03	−15.02	0.68
27	WL	0.00	51.87	−19.93	0.41
26	EL	0.00	51.97	−19.99	0.41
6	CEWL	0.00	52.56	−16.28	0.63
2	CFEWL	0.00	54.37	−14.52	0.70
1	Null	0.00	54.63	−24.17	0.00
17	EWL	0.00	55.46	−19.91	0.41
31	W	0.00	56.85	−23.96	0.03
30	E	0.00	57.23	−24.15	0.00
25	EW	0.00	59.80	−23.90	0.03

C: collections; F: formations; E: East African ardity; W: West African aridity; L: lake variability.

Mammals larger than 180 kg (e.g., Bovidae, Elephantidae, Rhinocerotidae) are over-represented relative to modern faunas, while the abundance of medium-sized taxa, including large-bodied primates, does not deviate significantly from modern analogues (Soligo, 2002; Soligo & Andrews, 2005). An FFC such as MBF, based on a clade that is preferentially preserved and found in a broader range of habitat types, is therefore less likely to depict a sampling signal relevant to a rarely preserved and poorly sampled clade.

Defining which formations might preserve a hominin is complex and, to a certain extent, subjec-

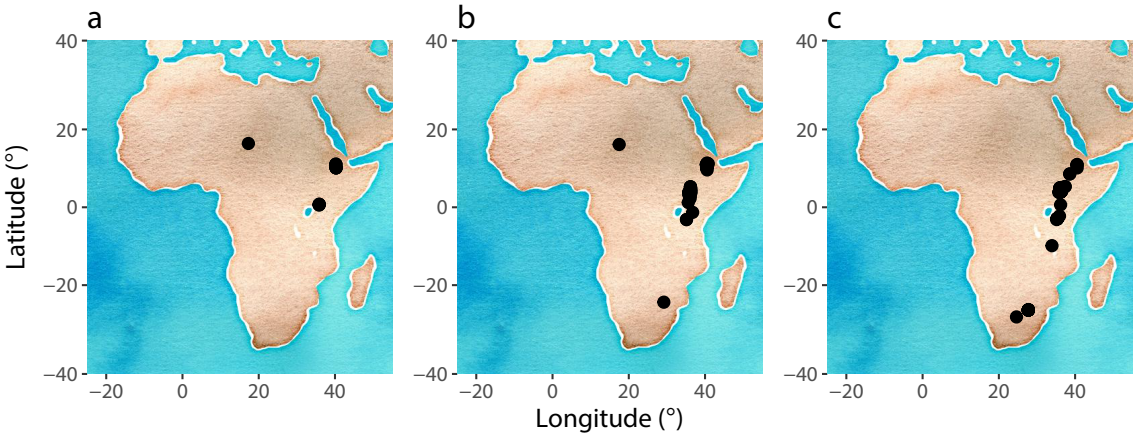

Figure 3.10: Map showing the location of all hominin-bearing localities (HBL) in the late Miocene, Pliocene, and early Pleistocene. (**a**): HBL in the late Miocene (7.00–5.33 Ma); (**b**): HBL in the Pliocene (5.33–2.58 Ma); and (**c**): HBL in the early Pleistocene (2.58–1.00 Ma). Sampled area estimates for each epoch are 1.38 million km² (n = 12), 4.76 million km² (n = 52), and 0.86 million km² (n = 76), corresponding to 5%, 16%, and 3% of the total area of the African continent (30.24 million km²).

tive. Although it is better to define a more inclusive clade of interest and compose an FFC based upon their occurrences, the question remains of how wide a clade is required to reach an optimum estimate of sampling intensity (Dunhill *et al.*, 2018). Recent model simulations have found that comprehensive FFC are the best predictor of true sampling, closely followed by formations that could potentially include the clade of interest and a FFC based on a wider clade of interest (Dunhill *et al.*, 2018). However, their simulations did not model habitat preference among taxa or different taphonomic pathways within the 10 notional geographical regions. Similarly, to simulate taphonomic processes, taxa were subject to random deletion from each fossil-bearing locality, which is an unrealistic scenario for fossilisation. The results of the current study indicate that a wider FFC based on primate fossils represents the most meaningful count of the number of preserved depositional environments suitable for the preservation of a hominin, precisely because they are a better approximation of those formations that could potentially include hominin fossils. FFCs have also been argued (e.g., Crampton *et al.*, 2003; Dunhill *et al.*, 2014*b*) to be poor predictors of sampling because they do not consistently correlate with collection effort (but see Upchurch *et al.*, 2011). However, we find a highly significant correlation between PBF and our proxy for palaeoanthropological sampling intensity (HBC) both at the continental (Table 3.2) and regional scale (Table 3.4). Generally speaking peaks in HBC are amplifications of the same peaks in PBF. This suggests that palaeoanthropologists tend to preferentially return to the same rich deposits and repeatedly amass new collections: a phenomenon known as the bonanza effect (Raup, 1977). The abundance of fossils and repeated field study at the Sagantole Fm. (4.4 Ma), Hadar Fm. (3.4–3.0 Ma), Sterkfontein (2.8–2.6 Ma), and Koobi Fora Fm./Olduvai Beds/Swartkrans (1.9 Ma) broadly coincide with the three major peaks in TDE at 3.6, 2.4, and 1.9 Ma (Fig. 3.1*a*).

The area of a convex-hull enclosing all HBL quantifies the known geographical spread of collection effort and may side-step the problems inherent to FFC (Barnosky *et al.*, 2005; Benton *et al.*, 2011). The moderate and significant positive correlation between TDE and palaeoarea supports the finding that the areal extent of collection effort directly influences taxic diversity patterns (either as a result

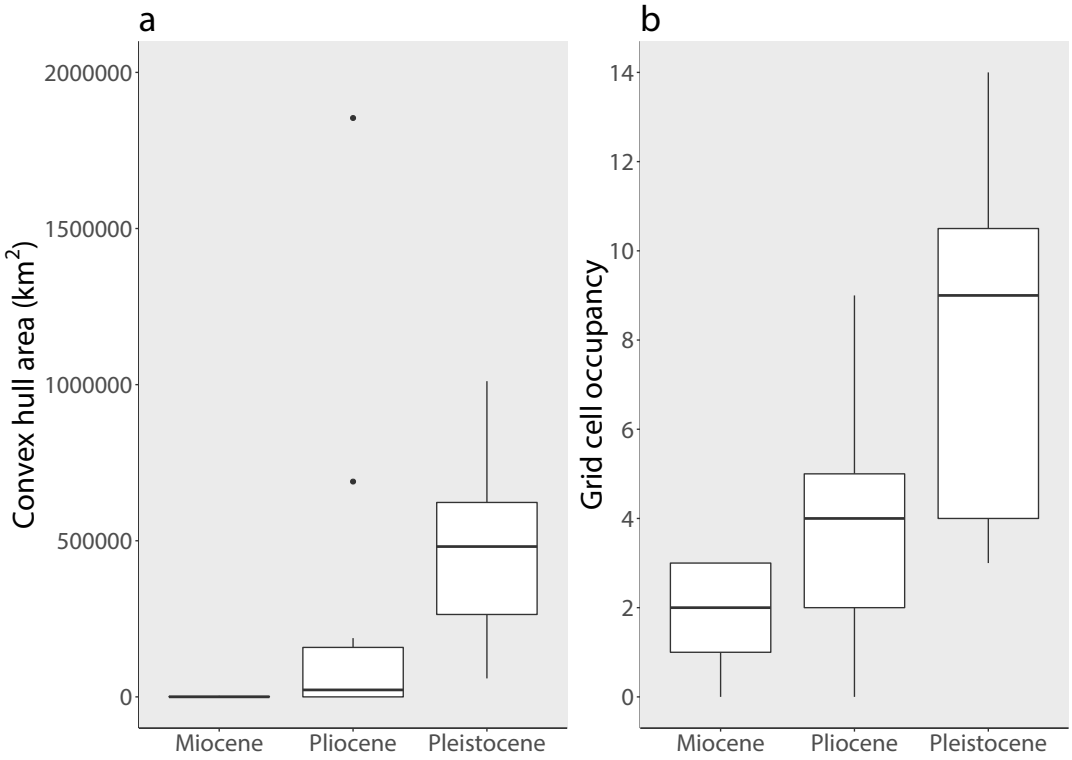

Figure 3.11: Boxplot showing per-time bin sampled area estimates for the late Miocene, Pliocene, and early Pleistocene. (**a**) Convex hull area (in km²). (**b**) Grid cell occupancy. The lower outlier (black dot) in (a) represents the 2.9 Ma time bin, which is the first to include HBL in the Cradle of Humankind (South Africa) and the EARS; the upper outlier represents the 3.6 Ma time bin, which is the only time bin to include HBL in Chad and the EARS. In both time bins, the anomalously high sampled area estimate is due to the inclusion of a single highly dispersed site (Makapansgat and Koro Toro), and not a realistic increase in sampled area. These inflated area estimates are absent when using grid cell occupancy (b), which specifically seeks to minimise the influence of spatial outliers.

of better sampling or a genuine species-area effect). In contrast, no PDE displays a correlation with the areal extent of collection effort. Further, significant correlations between palaeoarea, HBC, PBF, and GDE:PDE demonstrate that these sampling metrics describe the same sampling signal. However, convex-hull area is strongly influenced by spatial outliers and captures little-to-no information about the spatial completeness of sampling in a particular area (Close *et al.*, 2017). For example, the 3.8–3.5 Ma time bin contains, for the first time, fossils from Chad, in addition to fossils from Ethiopia, Kenya and Tanzania, and, as a result, records the largest sampled palaeoarea (~1.9 million km²) (Fig. 3.2*e*). The addition of Koro Toro (Chad) and other localities in the EARS, increases the number of HBL 340% (5 to 22) relative to the preceding time bin; yet the area enclosing these HBL increases by 1340% (0.13 million to 1.9 million km²). Moreover, the convex-hull enveloping these HBL spans—and therefore implies equivalent sampling effort in—Burundi, the Democratic Republic of Congo, Rwanda, South Sudan, and Uganda (Fig. 3.10). However, these countries (not including Uganda and the better-sampled Western Rift Valley; Pickford *et al.*, 1993) are poorly sampled and have a limited number of fossil mammal localities (Werdelin, 2010). In order to reduce any overestimation of sampled area, the number of (approximately equal-area) grid cells of 1° latitude/longitude (corresponding to ~111 by 111 km) occupied by HBL in each time bin was calculated using R code from Close *et al.* (2017) (Fig. 3.2*f*). Despite the

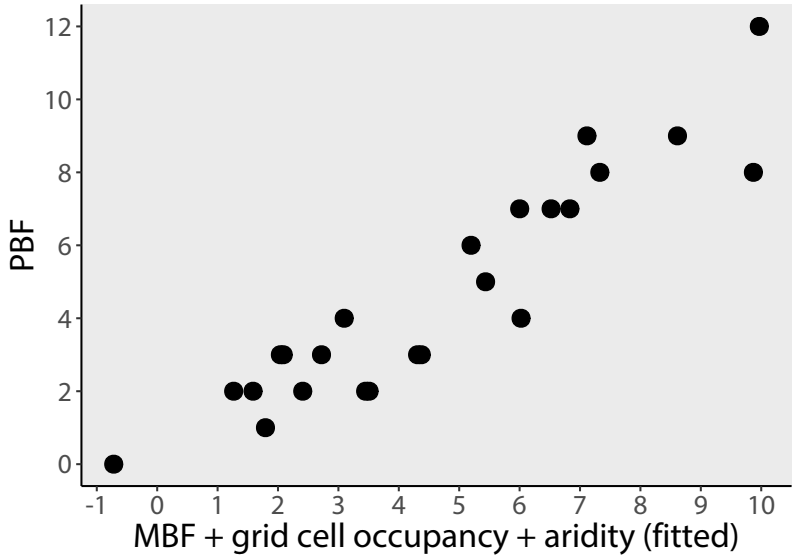

Figure 3.12: Number of primate-bearing formations against fitted data from the best-supported Generalised Least Squares model. The best-supported model for primate-bearing formations (PBF) includes terrestrial mammal-bearing formations (MBF), grid cell occupancy, and aridity (R^2 = 0.79). The relative contribution of each parameter to the coefficient of determination can be estimated by adding each parameter one by one and recording the increase in the model's R^2 value. The sum of each parameter's R^2 will not be equal to the total R^2 because MBF and grid cell occupancy are correlated or non-orthogonal (r = 0.804 for the raw data). Calculating the mean increase in R^2 for each parameter indicates approximately 60% of the variance can be explained by MBF, approximately 14% by grid cell occupancy, and approximately 5% by aridity.

lack of a correlation between convex-hull area and grid-cell occupancy (ρ = 0.389) and convex-hull area and HBL (ρ = 0.178), grid-cell occupancy correlates strongly with TDE (ρ = 0.727, p < 0.001), HBC (ρ = 0.673, p < 0.001), HBL (ρ = 0.558, p < 0.006), and PBF (ρ = 0.740, p < 0.001), suggesting, once again, that these distinct sampling metrics converge on the same signal. It is clear from the ubiquity and strength of these correlations that the influence of sampling is multiplex. It is also clear that (1) the areal extent of collection effort in each epoch differs considerably (Fig. 3.11), and (2) that a range of factors distort the perception of deep time patterns of species richness.

Repeating the GLS analysis with PBF as the dependent variable may go some may to further explore the underlying controls on the number of primate-bearing deposits. Regressing PBF against all possible combinations of sampling metrics (HBC, MBF, convex-hull area, grid cell occupancy, and East African aridity) shows that MBF + grid cell occupancy + aridity best explain PBF (R^2 = 0.79) (Fig. 3.12). Within this model MBF (R^2 ≈ 0.60) and grid cell occupancy (R^2 ≈ 0.14) have a statistically significant positive slope while aridity (R^2 ≈ 0.05) has a statistically significant negative slope (–3.63), suggesting that a higher number of PBF is a result of there being more suitable rock available in general (larger MBF), and that a drier climate results in less sediment deposition and a lower fossilisation potential for primates. The presence of grid cell occupancy (an estimate of the geographical spread of collection effort that is less biased by highly dispersed localities) suggests that PBF is also influenced by the distribution of early hominin fossils. The redundancy argument (Benton *et al.*, 2011) is therefore weak but cannot be ruled out entirely.

These findings are of critical importance for climatic forcing hypotheses of early hominin evolution

that interpret global and regional climate events, particularly in the EARS, as causal agents in hominin diversification, encephalisation, and dispersal (e.g., deMenocal, 1995, 2004; Vrba, 1995; Potts, 1996, 1998; Grove, 2011, 2012; Maslin & Trauth, 2009; Shultz & Maslin, 2013). Given the strong relationship between early hominin TDE and sampling found here, purported links between diversification and climate need to be re-assessed in a palaeobiological framework inclusive of this knowledge.

3.4.2 Did climate control hominin diversity?

Pairwise correlations indicate no relationship between diversity (either taxic or phylogenetic) and the interpolated aridity curve (Table 3.2). This relationship is confirmed by the weak and non-significant fit of aridity in the GLS multiple regression models (Table 3.5). Results from the GLS analysis indicate aridity explains roughly 2% of the variance in TDE (compared to 75% by PBF). Fluctuations in aridity probably played a key role in the emergent adaptive strategies taken by hominins during the Plio-Pleistocene. However, there is no evidence aridity had a sustained influence on diversification or that aridity had a significant taphonomic role. The same findings emerge in the EARS analysis (Fig. 3.8): East African aridity, West African aridity, and lake variability show no relation to diversity. Shultz & Maslin (2013) formed their hypothesis of lake-driven diversification based on a simple linear model including TDE, a suite of aridity proxies, and LVI with no consideration for sampling heterogeneity. Given the major sampling signal present in TDE, more sophisticated methods are necessary. In the regional GLS analysis, East and West African aridity and LVI are singly and in combination the poorest predictors of taxic diversity. The addition of either aridity record to PBF increases the explained variance by only 7% and once again demonstrates that (1) PBF are the dominant control of TDE, and (2) fluctuations in TDE caused by variation in aridity do not obscure the strong relationship between diversity and PBF.

The lack of a correlation between TDE and aridity may result from the treatment of the dust flux data: interpolating to smaller intervals may recover greater variation in the original time series and a potential link to diversity. However, means calculated using a dust flux curve interpolated to 25 ka intervals does not correlate with TDE (or any sampling metrics).

The lack of a correlation between TDE and aridity/LVI in the EARS could be a result of (1) different data sets used to estimate taxic diversity, (2) different first and last appearance dates, (3) temporal resolution (i.e., time bin size), or (4) the use of generalised differencing (GD), either singly or in combination. Of these explanations, the use of GD appears to be the key factor: TDE_{EA} ($r = -0.521$, $p = 0.039$) and LVI (mean: $r = -0.683$, $p = 0.007$; maximum: $r = -0.572$, $p = 0.021$) each display a significant linear trend (note the negative sign as time decreases towards the present), and TDE_{EA} and mean LVI correlate significantly before GD ($\rho = 0.687$, $p = 0.003$). This suggests that much of the support for a link between TDE_{EA} and LVI may relate to the comparison of two positive long-term trends which in reality show no tendency to increase or decrease in tandem over the short-term, as would be expected if they had a cause-and-effect relationship. In fact, peaks in TDE at 3.6, 2.4, and 1.9 Ma map directly onto peaks in HBC, PBF, sampled area estimates (either convex-hull area or grid-cell occupancy), and—in the latter case—MBF. Incidentally, peak MBF at 1.9 Ma also coincides with peak diversity of both EARS bovids and Turkana Basin large mammals (Bibi & Kiessling, 2015).

In summary, FFCs are a proxy for (1) the amount of rock available for sampling, (2) the geographical extent of sampled formations, (3) the heterogeneity of facies available for sampling, and (4) the

amount of study that has been undertaken (Benson & Upchurch, 2013:43). Thus, a strong relationship with PBF and weak relationship with aridity/LVI confirms that early hominin TDE is profoundly influenced by heterogeneous sampling of the fossil record. These data therefore offer no quantitative support for the *turnover pulse hypothesis* (Vrba, 1995), *aridity hypothesis* (deMenocal, 1995), *variability selection hypothesis* (Potts, 1996, 1998), or *pulsed climate variability hypothesis* (Maslin & Trauth, 2009) in the early hominin lineage. By failing to account for the temporal heterogeneity in fossil sampling, artefactual fluctuations in observed diversity have erroneously been linked to climatic events.

3.4.3 Testing the common-cause hypothesis

The CC hypothesis proposes that sampling metrics are driven by the same environmental factors that drove past diversity. Therefore any relationship between diversity and sampling is the result of a confounding variable (Peters, 2005, 2006; Hannisdal & Peters, 2011). In the marine realm, variation in continental flooding driven by sea level change offers the most plausible mechanism: high sea level increases the area of marine habitat promoting increases in diversity and accumulation and preservation of fossiliferous sediments. However, a CC mechanism in the terrestrial realm remains elusive (Butler *et al.*, 2011). In the case of hominins, a common-cause mechanism could be implied if aridity controlled both the likelihood of a hominin fossil becoming preserved (via changes in the rate of fluvio-lacustrine sediment deposition) and diversification rates (by habitat fragmentation and allopatric speciation). In the EARS, any such correlation could also result from the impact of lake levels on preservation and diversification rates. For example, during lake high stands, deposition of fluvio-lacustrine sediments will increase and the remains of terrestrial organisms will be more likely to reach aquatic environments and fossilise; conversely, during lake low stands or desiccation, sediment deposition will decrease and terrestrial remains will be less likely to reach aquatic environments and fossilise. Moreover, lake high stands could also promote population isolation and allopatric speciation in a spatially-constrained landscape, while lake low stands could increase competition and extinction given the limited resources (Shultz & Maslin, 2013).

The lack of a relationship between aridity/LVI and any sampling metric, both for the continental and EARS fossil record, cannot explain the strong correlation between TDE and PBF. Thus, rather than a CC explanation, these results support a simpler relationship in which observed diversity is controlled by sampling and climate does not appear to drive either of these parameters.

3.5 Conclusion

Long-term variation in aridity and climatic instability over the Plio-Pleistocene has long been thought to play a key role in human evolution. However, these results provide no evidence that short-term (bin-to-bin) fluctuations in climate relate to fluctuations in observed diversity. Instead, these data support a direct, causal relationship between observed early hominin TDE and temporal heterogeneity in fossil sampling—the rock record bias hypothesis. The near-linear increase in each PDE from the late Miocene to the mid-Pleistocene negates any explanation based on climate-driven pulsed turnover, or any explanation based on discrete change more generally, and gives further credence to the notion that the appearance of pulsed turnover in the early hominin fossil record is an artefact of uneven sam-

pling. This finding corroborates recent interpretations that events in hominin evolution once thought to be major transitions, when viewed in a phylogenetic (= lineage) context, actually represent gradual adaptive shifts (e.g., Foley, 2016; Kimbel & Villmoare, 2016). The identification of a major sampling component in the early hominin fossil record indicates that the pattern of diversification which many climatic forcing hypotheses purport to explain is more apparent than real.

This should come as no surprise: approximately one-quarter of early hominin species are point occurrences and the remainder have considerable uncertainties on their known stratigraphic ranges that have undoubtedly been artificially truncated by poor sampling (see Chapter 2 and Hopley, 2018, *in review*). Radiometric dating error associated with the earliest (latest) fossil is not equivalent to statistical uncertainty that such a date represents a speciation (extinction) event. Nor is the finding that radiometric dating error is random with respect to a climate event or period of climatic variability (e.g., Potts & Faith, 2015:13) an appropriate test of the quality of the fossil record. If error were randomly distributed in the early hominin fossil record any genuine evolutionary signal would be degraded not distorted (Raup, 1991). However, runs tests demonstrate that collection effort and rock availability are non-randomly distributed in the early hominin fossil record. The starting point for macroevolutionary analyses in palaeoanthropology ought to be that before any pattern in the fossil record is causally linked to climate, it is demonstrably shown that that pattern is not an artefact of sampling or poor fossil record quality. This requirement has been overlooked by palaeoanthropologists, archaeologists, and climatologists alike, and has severely impacted the interpretation of macroevolutionary pattern and process in the early hominin fossil record (Smith & Wood, 2017). This is especially surprising when one considers that, in palaeobiology, variation in fossil sampling is the preferred null hypothesis for short-term fluctuations in raw taxic diversity in the terrestrial realm (Butler *et al.*, 2011:1169), and therefore must be rejected, in full, before alternative explanations can be supported.

Summary

To summarise, the results of this chapter have shown:

1. Raw taxic diversity of early hominins is positively correlated with rock availability (number of primate-bearing formations) and the amount and areal extent of collection effort. Peaks and troughs in taxic diversity are more simply explained as artefacts of sporadic sampling.

2. Sampling metrics explain approximately three-quarters of the variance in taxic diversity, with roughly 25% unexplained by current sampling metrics. This indicates (1) that the rock record largely controls the sampling of the early hominin fossil record, and (2) that the early hominin fossil record probably reflects a mix of biological, geological, and anthropogenic signals.

3. Both continental and regional scale sampling metrics correlate with each other, implying that they converge on the same sampling signal.

4. The ratio of implied to sampled lineages offers a useful and independent proxy for sampling. Gaps in the fossil record (indicated by a higher proportion of implied lineages) correspond to reduced rock availability and geographical spread of collection effort by palaeoanthropologists.

5. Short-term fluctuations in taxic diversity do not relate to short-term variation in aridity. The role of climatic forcing in the diversification of early hominins is currently ambiguous and those hypotheses purporting a causal link are substantially weakened by points 1 and 2 (and Summary point 4 in Chapter 2).

Chapter 4

The completeness of the early hominin fossil record

This chapter is an extended version of the following publication: Maxwell SJ, Hopley PJ, Upchurch P, Soligo C, 2018. The completeness of the early hominin fossil record: implications for diversity patterns and the origin of Hominini (in preparation).

4.1 Introduction

Specimen completeness is a major issue in palaeoanthropology. It is frequently lamented that the hominin fossil record is "notoriously" incomplete (e.g., Villmoare *et al.*, 2015; Hopley, 2018, *in review*), particularly during the period molecular clock studies date the last common ancestor of hominins and panins, and the origin of genus *Homo* (Kimbel, 2009; Simpson, 2013). Yet, a comprehensive, in-depth assessment of the quality of the hominin fossil record has been neglected, despite the potential to yield significant insights into the timing of hominin origination and diversification. In light of this gap, the biological, geological, and anthropogenic factors controlling temporal variation in specimen and taxon completeness remain unclear.

Specimen completeness is of critical importance because it has a direct influence on all phylogenetic and macroevolutionary analyses. While deformation and incompleteness can be virtually corrected for a single specimen (e.g., Zollikofer *et al.*, 2005; Spoor *et al.*, 2015), analyses performed at higher systematic scales largely ignored specimen completeness until recently (Fountaine *et al.*, 2005; Mannion & Upchurch, 2010; Brocklehurst *et al.*, 2012). This is acutely true in the study of human evolution (Hopley, 2018, *in review*). Palaeoanthropology has primarily focussed on taphonomic processes affecting assemblage formation, bone quality, and compositional fidelity at single localities (e.g., Brain, 1981; Behrensmeyer *et al.*, 2003). However, large-scale taphonomic mega-biases (distortions caused by variation in the quality of the fossil record that affect palaeobiological analysis) remain poorly understood (Behrensmeyer *et al.*, 2000; Noto, 2010). In contrast, the effect of specimen and taxon completeness on macroevolutionary patterns is an active research agenda in palaeobiology (see Mannion & Upchurch, 2010 and references therein), and its systematic, non-random influence on phylogenetic analysis is now being fully realised (Sansom & Wills, 2013; Sansom, 2015; Sansom *et al.*, 2017). This interest stems from

the importance of a detailed description and accurate classification of specimens in studies of phylogeny and macroevolution: a detailed analysis of the completeness of the fossil record offers an essential check on data quality and all down-stream phylogenetic hypotheses and evolutionary inferences.

4.1.1 Past studies of specimen completeness

The first studies to assess specimen completeness employed a simple grade-based metric (e.g., Fountaine *et al.*, 2005; Smith, 2007*b*; Benton *et al.*, 2013), for example, with completeness ranked based on whether a taxon is represented by mostly fragmentary material (1), a partial skeleton (2), or one (3) or more complete skeletons (4). In recent years, specimen completeness has been scored based on the percentage of the skeleton or amount of phylogenetic information known. Once a scoring system has been determined (based on the relative volume or mass of a bone or the number of phylogenetic characters it contains), specimen completeness is relatively easy to quantify without subjective distinctions between completeness grades (Mannion & Upchurch, 2010). If, for example, a skull could be scored for 30% of characters in a phylogenetic analysis, a taxon known from only a complete skull would have a completeness score of 30% (Mannion & Upchurch, 2010). This approach has since become the default method in palaeobiology because it reduces subjectivity (the distinction between fragmentary remains and a partial skeleton might differ between palaeobiologists) and increases reproducibility compared to the former (Brocklehurst *et al.*, 2012; Walther & Fröbisch, 2013; Brocklehurst & Fröbisch, 2014; Cleary *et al.*, 2015; Dean *et al.*, 2016; Verrière *et al.*, 2016; Davies *et al.*, 2017; Tutin & Butler, 2017).

Recent studies that quantify variation in specimen and taxon completeness through geological time include early tetrapods (Benton *et al.*, 2004, 2013), Mesozoic birds (Fountaine *et al.*, 2005; Brocklehurst *et al.*, 2012), echinoids (Smith, 2007*b*), dinosaurs (Mannion & Upchurch, 2010; Bell *et al.*, 2013), anomodonts (Walther & Fröbisch, 2013), pelycosaur-grade synapsids (Brocklehurst & Fröbisch, 2014), ichthyosaurs (Cleary *et al.*, 2015), parareptiles (Verrière *et al.*, 2016), pterosaurs (Dean *et al.*, 2016), plesiosaurs (Tutin & Butler, 2017), eutherian mammals (Davies *et al.*, 2017), and mosasaurs (Driscoll *et al.*, 2018). In each case, temporal variation in specimen completeness has provided novel insights into fossil record quality and the evolutionary history of the clade of interest.

The relationship between palaeodiversity (henceforth diversity) and specimen completeness is complex and by no means universal across the Tree of Life. Intuitively, one might expect that raw taxic diversity would be highest when specimens are most complete and, indeed, a statistically significant positive correlation has been reported for dinosaurs, Mesozoic birds, and pterosaurs (Mannion & Upchurch, 2010; Brocklehurst *et al.*, 2012; Bell *et al.*, 2013; Dean *et al.*, 2016). Conversely, a negative correlation between specimen completeness and taxic diversity has been found in basal synapsids, suggesting a tendency to name species based on incomplete material (Brocklehurst & Fröbisch, 2014). The clade-specific nature of this relationship is further illustrated by the lack of a correlation between specimen completeness and taxic diversity in ichthyosaurs, parareptiles, plesiosaurs, eutherian mammals, and mosasaurs (Cleary *et al.*, 2015; Verrière *et al.*, 2016; Tutin & Butler, 2017; Davies *et al.*, 2017; Driscoll *et al.*, 2018).

The completeness of fossil specimens might also provide a useful sampling metric or proxy, capturing an aspect of sampling ignored by other metrics. One would expect, for example, that periods of good overall sampling would correspond with high specimen completeness, as more complete and

numerous skeletons should, in principle, mean more individual bones and regions of the skeleton are known. However, studies of sauropodomorph (Mannion & Upchurch, 2010), anomodont (Walther & Fröbisch, 2013), basal synapsid (Brocklehurst & Fröbisch, 2014), ichthyosaur (Cleary *et al.*, 2015), and pterosaur (Dean *et al.*, 2016) completeness found no correlation with sampling metrics such as rock outcrop area and fossiliferous formation counts. This implies that low specimen completeness cannot be explained by low overall sampling. Indeed, for some clades (e.g., birds and pterosaurs), times of poor sampling appear to correspond with high completeness, reflecting the influence of a small number of formations and localities with exceptional preservation (Lagerstätten). In the case of Mesozoic birds, completeness correlates significantly (albeit weakly) with both collection and fossiliferous formation counts, suggesting a complex mix of geological and anthropogenic controls on the Mesozoic bird fossil record (Brocklehurst *et al.*, 2012).

4.1.2 Rationale

Of particular relevance to this debate is the hominin branch of the Tree of Life. Palaeoanthropology abounds with claims that the early hominin fossil record is hopelessly incomplete, yet there has been very little consideration of (1) how specimen abundance and specimen completeness affects the recognition and interpretation of taxic diversity (Smith, 2005; Wood & Boyle, 2016); (2) how sampling heterogeneity can often distort the perception of large-scale macroevolutionary patterns (Maxwell *et al.*, 2018); and, (3) how incomplete Operational Taxonomic Units (OTUs) affects node support throughout hominin phylogeny. This lack of consideration might be because taphonomic processes at single sites are interpreted independently of time and thus the large-scale implications of taphonomic biases are not fully realised. Still, the unique aspect of the fossil record is its temporal dimension (Fleagle, 2002). Therefore, the architecture of the sedimentary rock record, and geographic and stratigraphic distribution of fossil data, should be thoroughly understood before any inferences are made about hominin origins and evolutionary history (Tarver *et al.*, 2011).

Here, I take a macroevolutionary perspective and ask how specimen completeness in the early African hominin fossil record changes through geological time; whether fluctuations in specimen completeness affect taxic diversity patterns; and whether temporal variation in specimen and taxon completeness is a useful guide to fossil sampling in the early African hominin fossil record. The main aims of this study are (1) to quantify the skeletal and character (phylogenetic) completeness of the early African hominin fossil record; (2) to assess the effect—if any—of completeness on taxic diversity patterns; (3) to assess the biological, geological, and anthropogenic factors controlling specimen and taxon completeness; and (4) to critically discuss these findings alongside our current understanding of hominin origins, and the issues inherent to identifying stem members of a clade.

4.2 Methodology

4.2.1 Completeness metrics

Character Completeness Metric. The Character Completeness Metric (CCM) assesses completeness based on the number of characters that can be scored for each skeletal element in phylogenetic analyses

Table 4.1: Holotype specimen, most complete skull, and most complete skeleton for early hominin taxa. Character Completeness Metric (CCM) scores are based solely on the skull because no phylogenetic analysis of hominins includes postcranial characters. Skeletal Completeness Metric (SCM) scores are based on the entire skeleton. For this reason, the most complete specimen may differ in the CCM1 and SCM1 analyses. See List of Hominin Fossil Abbreviations for an explanation of each fossil accession number.

Taxon	Holotype specimen	Most complete skull (CCM1)	Most complete skeleton (SCM1)
Sahelanthropus tchadensis	TM 266-01-060-1	TM 266-01-060-1	TM 266-01-060-1
Orrorin tugenensis	BAR 1000'00	BAR 1000'00	BAR 1002'00
Ardipithecus kadabba	ALA-VP-2/10	ALA-VP-2/10	ALA-VP-2/101
Ardipithecus ramidus	ARA-VP-6/1	ARA-VP-6/500	ARA-VP-6/500
Kenyanthropus platyops	KNM-WT 40000	KNM-WT 40000	KNM-WT 40000
Australopithecus anamensis	KNM-KP 29281	KNM-KP 29281	ASI-VP-5/154
Australopithecus afarensis	LH 4	AL 444-2	AL 288-1
Australopithecus bahrelghazali	KT 12/H1	KT 12/H1	KT 12/H1
Australopithecus deyiremeda	BRT-VP-3/1	BRT-VP-3/14	BRT-VP-3/14
Australopithecus africanus	Taung 1	Sts 5	Sts 14
Australopithecus garhi	BOU-VP-12/130	BOU-VP-12/130	BOU-VP-12/130
Australopithecus sediba	MH1	MH1	MH1
Paranthropus aethiopicus	Omo 18-1967-18	KNM-WT 17000	KNM-WT 17000
Paranthropus boisei	OH 5	OH 5	KNM-ER 406
Paranthropus robustus	TM 1517	DNH 7	DNH 7
Homo habilis	OH 7	KNM-ER 1813	OH 62
Homo rufolfensis	KNM-ER 1470	KNM-ER 1470	KNM-ER 1470
African *Homo erectus*/*Homo ergaster*	KNM-ER 992	KNM-WT 15000	KNM-WT 15000

(Mannion & Upchurch, 2010). Originally, skeletal regions were assigned a weighting based on their phylogenetic character richness such that phylogenetically informative regions contribute more to the CCM. Under this scoring scheme, if a skeletal element is present then it is assumed that all its characters can be scored. In the current analysis, however, the CCM is computed using the cladistic matrix method (Bell *et al.*, 2013) according to Equation (4.1).

$$CCM = \frac{Number\ of\ characters\ scored}{Total\ number\ of\ characters} \times 100 \tag{4.1}$$

Where the number of characters scored is calculated directly from each cladistic matrix according to Equation (4.2).

$$Number\ of\ characters\ scored = Total\ number\ of\ characters - Number\ of\ question\ marks \tag{4.2}$$

The number of question marks in a matrix reflects the number of inapplicable characters (skeletal regions not present) and the number of un-scored characters (skeletal regions not preserved). The CCM can be assessed according to the character completeness of the most complete specimen of each taxon (CCM1), the character completeness of a composite taxon (i.e., the maximum number of characters that can be scored for an OTU when based on the most complete specimen and any additional characters

Table 4.2: Percentages attributed to regions of the skull based on the Character Completeness Metric in the two early hominin phylogenetic analyses used in the current study. Percentages are rounded to the nearest whole number. Mean values for each skeletal region are shown in Table 4.3.

Skeletal region	Strait & Grine (2004)	Dembo *et al.* (2015)
Cranium	59	63
– Basicranium	11	17
– Facial skeleton	19	20
– Cranial vault	28	26
Mandible	13	16
Dentition	28	21

that can be used to fill in as many scoring gaps as possible) (CCM2), and the character completeness of the type specimen or holotype (CCM_H) (Mannion & Upchurch, 2010). The type specimen, most complete skull (CCM1), and most complete skeleton (SCM2) for each early hominin is shown in Table 4.1.

The 109 traditional characters in Strait & Grine (2004) and the 380 characters in Dembo *et al.* (2015) were compiled and the mean score across both matrices used in the CCM. Both of these cladistic matrices include characters of the skull only. These characters are assigned to 5 regions: (1) basicranium: includes characters inferior to the superior nuchal line, and characters inferior and posterior to the mastoid process of the temporal bone; (2) facial skeleton: includes characters of the face up to the supraorbital region extending laterally to the zygomaticotemporal suture; (3) cranial vault: includes characters superior to the frontonasal suture and glabella anteriorly, superior to the nuchal line posteriorly, and the zygomaticotemporal suture to the external auditory meatus laterally; (4) mandible; and (5) dentition (Dembo *et al.*, 2015). The percentages assigned to each region of the skull for each matrix can be found in Table 4.2 (means are shown in Table 4.3). It is immediately apparent that different elements of the skull are potentially phylogenetically more informative than others in hominin systematics: for example, on average a complete set of teeth can be scored for 25% of all characters, whereas a complete mandible can only be scored for 15% (Table 4.3). Caution is required, however, when assuming that more characters equals greater phylogenetic informativeness: some parts of the skeleton might

Table 4.3: Percentages attributed to each region of the skeleton based on the Character Completeness Metric (CCM) and Skeletal Completeness Metric (SCM). Percentages are rounded to one decimal place. Regions excluded from the CCM are represented by a hyphen (–). See § 4.2.1 and Appendix E for further details.

Skeletal region	CCM	SCM
Cranium	60.8	12.9
Mandible	14.6	1.7
Dentition	24.6	0.4
Pectoral girdle	–	4.2
Vertebrae and ribs	–	16.6
Forelimbs	–	14.6
Pelvic girdle	–	11.3
Hindlimbs	–	38.3

have been over- or under-atomized into characters by systematists; or a large suite of non-independent characters might occur when many different features co-evolve as a coherent module, and some of these modules might convey a strong, but nevertheless, homoplastic signal. Teeth, for example, have a high weighting in hominin systematics despite being especially poor at reconstructing phylogenetic relationships among mammals (Sansom *et al.*, 2017).

Orrorin tugenensis, *Ardipithecus kadabba*, and *Australopithecus bahrelghazali* have never been included in any analysis of hominin systematics because too few characters are preserved in each hypodigm (Strait *et al.*, 2007). However, though this may be the case and their inclusion would likely distort or weaken node support throughout the tree, there are a small number of mandibular and dental characters that can be scored for each taxon (no cranial characters can be scored). Those characters that can be scored for *Orrorin tugenensis* (Senut *et al.*, 2001:139–141), *Ardipithecus kadabba* (Haile-Selassie *et al.*, 2009:195–207), *Australopithecus bahrelghazali* (Brunet *et al.*, 1996:908), and *Australopithecus sediba* (Berger *et al.*, 2010) were taken from the original descriptions and subsequent monographs.

In order to score the CCM1 and CCM_H, characters were deleted from the original matrix by a three-step process. First, character states pertaining to those skeletal regions not preserved in the single most complete or holotype specimen were replaced with question marks. The *Homo ergaster* holotype specimen (the near-complete mandible KNM-ER 992) offers a simple illustration: scoring KNM-ER 992 required deleting all character state scores in the *Homo ergaster* OTU relating to the cranium and maxillary dentition. Second, the presence of any remaining characters were evaluated in high-quality casts (if available), or with reference to the original descriptions and *The Human Fossil Record Volumes 2 and 4* (Schwartz & Tattersall, 2003, 2005). In KNM-ER 992, this meant deleting all remaining characters of the incisors and mandibular condyle. Finally, the remaining characters were then vetted to ensure they were present. CCM1 and CCM_H were scored for both the Strait & Grine (2004) and Dembo *et al.* (2015) matrices, and the mean score for each variant used in all subsequent analyses.

Skeletal Completeness Metric. The Skeletal Completeness Metric (SCM) quantifies the relative bulk and amount of skeletal material known, by dividing the skeleton into different regions and assigning each a percentage based on how much of the skeleton it represents (Mannion & Upchurch, 2010). Like the CCM, the SCM can be assessed according to the skeletal completeness of the most complete specimen of each taxon (SCM1), the skeletal completeness of a composite including the most complete specimen and all other skeletal elements not represented in the most complete specimen (SCM2), and the skeletal completeness of the holotype (SCM_H) (Mannion & Upchurch, 2010). Within each region different bones are also weighted so that larger bones contribute more to the SCM than smaller bones. In the current study, this technique is modified by assigning percentages to each bone based on their proportional contribution to the total mass of an adult modern human (*Homo sapiens*) skeleton. For example, the mean mass of a skeleton is 4.98 kg and the mean mass of a complete femur is 0.46 kg (Ingalls, 1931). If a taxon is known only from a complete femur, its SCM score is (0.46/4.98) * 100 = 9%. One major implication of this is that the skull, the only skeletal region included in hominin phylogenetic analyses, contributes 100% in the CCM and 15% in the SCM (Fig. 4.1 and Table 4.3).

Bone and tooth masses were taken from Ingalls (1931) and can be found in Appendix E. Ingalls (1931) reported bone masses for nearly all individual bones in a modern human skeleton, but in some cases (e.g., vertebrae, metacarpals, metatarsals, and phalanges) only included the mass of the entire

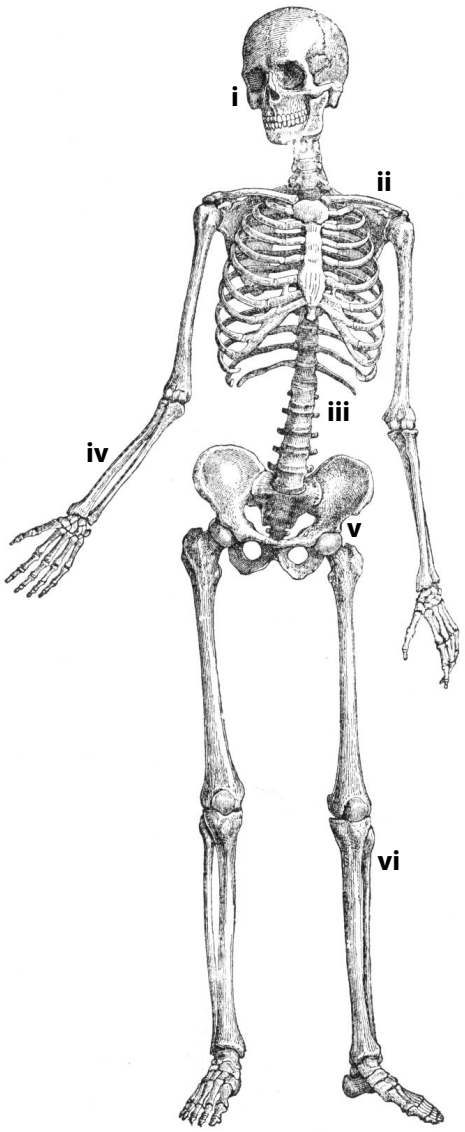

Figure 4.1: Skeleton of *Homo sapiens* (modified from the *Life of Animals* [Brehm, 1927]), showing the major regions of the skeleton based on the Skeletal Completeness Metric. Each bone is assigned a completeness score based on the percentage its weight contributes to the total weight of a *Homo sapiens* skeleton. Skeletal regions are indicated: i: skull (15%); ii: pectoral girdle (4%); iii: vertebrae and ribs (17%); iv: forelimbs: (15%); v: pelvic girdle (11%); vi: hindlimbs (38%). See Table 4.3 and Appendix E for details.

region. In such cases, bone masses were gathered from additional sources (Ingalls, 1932; Pyle, 1935; Cheyne & Oba, 1943; Lowrance & Latimer, 1967) and used to calculate the relative proportion of each bone within a region. The relative proportion of each bone could then be multiplied by the mass of the entire region reported in Ingalls (1931) in order to obtain an individual bone mass. This ensured that the entire skeleton could be assigned an SCM score despite the source data coming from multiple osteological collections. In order to assess more accurately the SCM for fragmentary cranial material, individual cranial bone masses were estimated using a Bone Clones Inc., 22-piece magnetic osteological teaching skull™ (Appendix F). The relative proportion of each bone in the teaching skull could then be multiplied by the total mass of the cranium reported in Ingalls (1931) to produce a mass for all 22 cranial

bones.

Rowbotham *et al.* (2017) recently proposed a standardised, bone volume-based approach for estimating skeletal completeness in modern humans using Computed Tomography (CT). Their percentages (assigned to each bone based on a single individual) agreed closely with those based on bone mass: skull: 12% (*versus* 15% in the current study), pectoral girdle: 4% (*versus* 4%), vertebrae and ribs: 23% (*versus* 17%), forelimbs: 12% (*versus* 15%), pelvic girdle: 13% (*versus* 11%), and hindlimbs: 36% (*versus* 38%) (Rowbotham *et al.*, 2017). By assigning percentages based on bone masses and not bone volume, this approach adds slightly more weight to denser (cortical > trabecular) bones, and is, therefore, a more taphonomically informative completeness metric.

Finally, we also created a variant of the SCM in which the skull comprises 100% of the completeness score (SCM_{skull}) in order to facilitate a more direct comparison with the purely cranio-dental CCM. This required summing all SCM scores relating to the skull and multiplying this value by 6.67 to scale each score from a maximum of 15% to 100% (Table 4.3).

4.2.2 Time bins and mean completeness scores

The completeness of the early hominin fossil record is represented by taking the mean completeness score for all taxa in each time bin based on their stratigraphic range. The mean scores can then be plotted through geological time to provide a representation of early hominin completeness from the late Miocene to the early Pleistocene. 0.25-million year time bins were used starting at 7 million years ago (Ma) and ending at 1 Ma. First and last appearance dates (FAD and LAD, respectively) were taken from Wood & Boyle (2016; see their Table 1 on page 40), and the mean completeness score calculated for each time bin based on whether a taxon occurs in that time bin or whether their stratigraphic range spans that time bin (range-through taxa). For example, 5 species are known from the 2–1.75 Ma time bin (*Australopithecus sediba*, *Homo habilis*, *Homo rudolfensis*, *Paranthropus boisei*, and *Paranthropus robustus*). By summing the CCM2 score for each of these species (29%, 84%, 60%, 81% and 70%, respectively) and dividing by the total number of species (i.e., 5) gives a mean CCM2 score of 65% for this time bin.

One major problem with assessing specimen completeness based on mean scores is that these values may be affected by the sample size and range of completeness scores in each time bin. For example, one can imagine a situation where we have two time bins. In the earlier time bin we have a 100% complete taxon and a 50% complete taxon so the mean completeness is 75%. In a later time bin we have a single taxon that is 80% complete. Just looking at the mean completeness scores, it would appear that completeness has risen from 75% to 80%. But this is not a good guide to an improvement of fossil record quality since the earlier time bin actually has two taxa, one of which is 100% complete. To combat this potential problem, 95% confidence intervals are plotted in orange to show the reader which periods are represented by single-bin species and the range of variation around the mean. Conclusions based on time bins with bin-to-bin changes with particularly large increases or decreases in sample size should be treated with caution. In addition, and for this reason, completeness metrics were also compiled based on the maximum completeness score in each time bin (denoted by the subscript max) and all statistical tests were performed on both the mean and maximum completeness scores.

The choice to assign completeness scores based on each taxon's stratigraphic range, and not those specimens actually present in a time bin, could be argued to distort trends in specimen completeness

by introducing autocorrelation into each completeness metric. However, any distortion is unlikely to be significant: 44% (8) of early hominins have a stratigraphic range that is less than a time bin duration of 250 kyr (Table 2.1) and 33% (6) occur in only one time bin. Inspection of correlograms shows no autocorrelation after generalised differencing, and Breusch-Godfrey tests indicate that there is no significant autocorrelation in each of the best-supported models.

4.2.3 Diversity metric

One of the main aims of this study is to assess whether specimen or taxon completeness affects the ability of palaeoanthropologists to recognise species in the fossil record, and to test whether apparent species richness in each time bin is correlated with specimen quality. While the issue of specimen abundance (or rather lack thereof) has been considered in relation to taxic diversity and the hominin status of many late Miocene fossils (Smith, 2005), the completeness of the fossils themselves is yet to be assessed empirically and not least within a temporal framework. We compiled the number of hominin taxa in each time bin (= the raw, empirical, uncorrected taxic diversity estimate [TDE]) using data from Maxwell *et al.* (2018).

Here, three hypotheses are tested:

H_0 There is no significant correlation between TDE and any completeness metric (null).

H_1 TDE is correlated significantly with CCM.

H_2 TDE is correlated significantly with SCM.

Note that H_1 and H_2 are not mutually exclusive; each hypothesis could support a significant correlation between TDE and specimen completeness, and this relationship might be positive or negative.

4.2.4 Sampling metrics

In order to assess whether specimen completeness is controlled by geological and anthropogenic sampling heterogeneity, all completeness metrics are compared to sampling metrics that each capture a different aspect of bias or error in the fossil record: non-uniform rock volume, rock accessibility, facies heterogeneity, and study effort (Raup, 1972, 1976*b*).

Geological sampling bias. Fossiliferous formation counts summarise the amount of rock available for sampling, facies diversity, the geographical and temporal spread of fossil-bearing rock, and collection effort (Benson & Upchurch, 2013:43). Fossiliferous formation counts have frequently been used in this context as a proxy for geological sampling in terrestrial depositional environments and are widely regarded as an effective sampling metric in the terrestrial realm (e.g., Upchurch *et al.*, 2011; Benson & Upchurch, 2013). However, fossiliferous formation counts have been criticised because they vary enormously in scale and are possibly redundant (non-independent) with respect to their fossil content (e.g., Benton *et al.*, 2011, 2013; Dunhill *et al.*, 2014*b*). In order to reduce redundancy, fossiliferous formations were compiled based on a more inclusive group: cercopithecoid, hominoid, and lorisoid primates. Counts of the number of primate-bearing formations (PBF) include formations that contain hominin fossils and formations that do not (non-occurrence). Cercopithecoid and hominoid primates are also

taphonomically comparable to hominins and, therefore, PBF represents a better proxy for supposed total sampling effort than hominin-bearing formations alone (Maxwell *et al.*, 2018). Moreover, PBF have been shown to correlate positively with TDE, supporting a sampling effect (Maxwell *et al.*, 2018). PBF were taken from Harrison (2010*b*) for lorisoids (specifically galagids), Jablonski & Frost (2010) for cercopithecoids, and MacLatchy *et al.* (2010) for hominoids (specifically hominins) and corroborated using the *Paleobiology Database* (PBDB) and an exhaustive survey of the published literature (Maxwell *et al.*, 2018).

Anthropogenic sampling bias. Counts of the number of hominin-bearing collections (HBC) were similarly gathered from the published literature (Maxwell *et al.*, 2018). A collection is defined as an assemblage of fossils that were amassed at one locality in a single effort, or linked series of efforts (Benton, 2015), with one collection equivelent to one field season. However, the duration and number of field seasons are not commonly reported so here the number of years that have produced a hominin fossil per formation per time bin were used as a proxy for the amount of collection effort and study interest in each time bin (Sheehan, 1977). For example, in the 5.5–5.3 Ma time bin, *Ardipithecus kadabba* is known from fossils collected in 1998, 2003, 2004, and 2013. Therefore this time bin has an HBC count of 4 (Maxwell *et al.*, 2018). Primate-bearing collections could not be compiled due to the inaccessibility and rarity of such data in the literature; the number of years that produced primate fossils is simply reported too infrequently.

While HBC describes the amount of collection effort in a particular time bin, it does not indicate whether that effort is expended at the same deposit or throughout a range of deposits of similar age. Palaeoanthropologists do not uniformly sample the entire area of exposed rock in each time bin. Further, the amount of exposed rock is not equally distributed through the rock record or the landscape as a direct consequence of its geological and tectonic history. Naturally, palaeoanthropologists preferentially return to rich deposits known to contain hominin fossils, resulting in over-sampling in some areas (= time periods) and under-sampling in others. This phenomenon is known as the bonanza effect (Raup, 1977; Dunhill *et al.*, 2012, 2013, 2014*b*), and is subtly distinct from the (rock record) bias effect (Benton, 2015). To quantify the influence of the bonanza effect, we calculated the ratio of hominin-bearing collections to formations (HBC:HBF). The HBC:HBF ratio differs from HBC by explicitly describing where many collections have been amassed from few formations. One can imagine a situation where a time bin has a high HBC count, but those collections were amassed in a similarly high number of formations, indicating an absence of any bonanza effect (Raup, 1977). The HBC:HBF ratio explicitly describes the extent to which collections were gathered in few or many formations and is, therefore, an additional proxy for palaeoanthropological field study.

4.2.5 Climate

Terrigenous dust flux (henceforth aridity) data were taken from deMenocal (1995), interpolated to 50-thousand-year intervals (using a Piecewise Cubic Hermite Interpolating Polynomial), and the mean value calculated for each 0.25-Ma time bin. If continental-scale aridity had a direct influence on specimen and taxon completeness one would expect a significant negative correlation between completeness metrics and aridity, as fossilisation potential is higher when aquatic environments (lakes, rivers) are more abundant. Similarly, the probability of hominin presence is likely to be higher when rainfall and

vegetation are more abundant due to increased carrying capacity.

4.2.6 Specimen completeness and phylogenetic informativeness

Specimen completeness is not necessarily equivalent to the diagnostic content or phylogenetic informativeness of a specimen. For example, one can imagine a situation where a specimen that is only 10% complete in terms of the number of phylogenetic characters scored could be rich in diagnostic characters (autapomorphies). Alternatively, one can imagine a specimen that is 90% complete in terms of the CCM but completely un-diagnosable if there are no autapomorphies. To identify whether the CCM relates to phylogenetic informativeness, the cladistic analysis in Strait & Grine (2004) was re-ran using the 109 traditional characters (note 2 characters in the original matrix are parsimony-uninformative) and identical model specifications and the number of autapomorphies that define each OTU were compiled. The number of autapomorphies were then directly compared with the CCM2 score for each OTU using log-transformed values and Pearson's r correlation coefficient.

4.2.7 Statistical tests

Two statistical tests were used to compare each completeness metric to one another, and each completeness metric to taxic diversity, sampling metrics, and aridity. Spearman's rank correlation coefficient (ρ) is a non-parametric measure of the correlation between two rank-ordered variables, and the Kendall tau rank correlation coefficient (τ) is a non-parametric measure of the extent to which two variables change synchronously (Hammer & Harper, 2006). Time series were de-trended and corrected for temporal autocorrelation by generalised differencing prior to regression (McKinney, 1990; R code from G.T. Lloyd: http://www.graemetlloyd.com/methgd.html). The significance of correlations was evaluated based on original p-values and p-values adjusted for the implementation of multiple tests using the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995). Non-time series (e.g., stratigraphic duration and number of autapomorphies) were compared to each completeness metric using Pearson's product-moment correlation coefficient (r). Pearson's r is a parametric measure of the strength of a linear correlation between two variables (Hammer & Harper, 2006), with the data log-transformed prior to analysis according to the function $[f(x) = \log(x+1)]$. Wald-Wolfowitz runs tests were performed to investigate the null hypothesis of randomness and data independence in a time series, and Kruskal-Wallis tests were performed to assess whether completeness scores are significantly different for the late Miocene (7.0–5.3 Ma), Pliocene (5.3–2.6 Ma), and early Pleistocene (2.6–1.0 Ma). Mann-Whitney U tests were performed to assess whether completeness scores in different geographical regions could be distinguished from one another statistically (Hammer & Harper, 2006).

In addition to pairwise correlations, Generalised Least Squares (GLS) modelling was used to explore the possibility of multiple explanatory variables controlling completeness. GLS models were constructed for all possible combinations of TDE, HBC, PBF, the bonanza effect, and aridity (n = 31) plus an intercept-only null model (representing statistically random variation around a constant mean) using the gls function in the R package nlme. GLS regression modelling has the benefit of assessing the fit of multiple dependent variables while simultaneously accounting for temporal autocorrelation using a first-order autoregressive model, which seeks autocorrelation up to a lag of 1 in either direction. The

Table 4.4: Mean completeness scores for each hominin genus. The number of species in each genus (n) is also shown. Percentages are rounded to one decimal place.

Genus	n	CCM1	CCM2	CCM_H	SCM1	SCM2	SCM_H
Sahelanthropus	1	28.3	28.3	28.3	12.7	13.8	12.7
Orrorin	1	1.9	5.1	1.8	6.2	8.7	0.7
Ardipithecus	2	6.3	11.8	6.3	19.7	27.8	0.4
Kenyanthropus	1	16.3	17.3	16.3	11.9	12.6	11.9
Australopithecus	7	27.1	34.3	18.0	15.2	33.3	6.7
Paranthropus	3	49.4	67.5	49.4	11.7	23.5	4.8
Homo	3	58.2	76.9	44.1	32.3	41.1	5.2

second-order Akaike Information Criterion (AICc) corrected for finite sample sizes and the relative likelihood of each model based on Akaike weights (w_i) were used to assess model fit (Johnson & Omland, 2004). The completeness metrics, taxic diversity, and sampling metrics were ln-transformed prior to analysis to ensure normality and homoskedasticity of residuals. The Jarque-Bera test and Breusch-Pagan test were used to assess the normality and heteroskedasticity of residuals. We also manually calculated generalised-R^2 using the likelihood-ratio of each model against the null (Nagelkerke, 1991; Equation 3.1). All analyses were performed in R 3.5.0 (R Development Core Team, 2018) and an R script can be found in Appendix B.

4.3 Results

4.3.1 Hominin fossil record completeness

Mean completeness scores for each hominin genus are shown in Fig. 4.2 and Table 4.4. CCM1-2 are generally higher than SCM1-2. The notable exceptions to this are *Orrorin* and *Ardipithecus* which include high-scoring skeletal elements (multiple partial femora and a partial skeleton, respectively) and few, fragmentary cranial elements. Holotype completeness is remarkably poor for the skeleton as a whole (mean SCM_H = 6%), demonstrating that new taxa are named based on fragmentary material that preserve few skeletal regions. Specimen completeness is also low when based on character completeness (mean CCM_H = 26%), indicating that holotype specimens cannot be scored for many of the characters used in phylogenetic analyses of hominins.

For the CCM1-2, mean completeness scores are low (<30%) from 7.0–3.4 Ma, with the greatest variability in character completeness at 3.6–3.4 Ma (Fig. 4.3). Both then increase rapidly (106%–120%) from 3.4–3.1 Ma due to the presence of *Australopithecus afarensis* in the Hadar Formation (Ethiopia) and the first appearance of *Australopithecus africanus* at Makapansgat (South Africa). These taxa are represented by the remarkably complete crania AL 444-2 and Sts 5, respectively (Broom *et al.*, 1950; Kimbel *et al.*, 2004). CCM1-2 then remain high (fluctuating around 56% and 74%, respectively) for the remainder of the early Pleistocene (2.6–1.0 Ma). These fluctuations reflect the sudden appearance and disappearance of single-bin taxa and taxa with short stratigraphic ranges. Such fluctuations do not occur in the $CCM1\text{-}2_{max}$ (Fig. 4.4). That CCM1-2 scores change little during the Pleistocene and display low uncertainty around the mean suggests these taxa are of comparable character completeness. This general

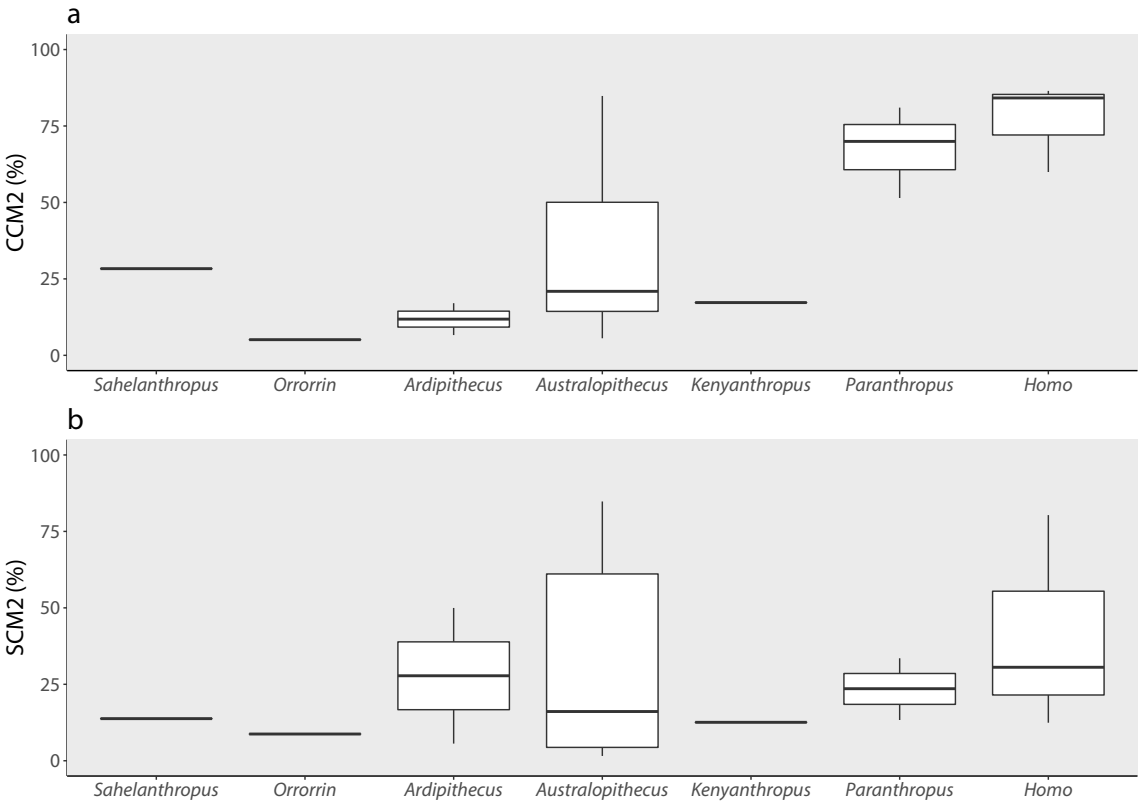

Figure 4.2: Boxplot showing mean completeness score for each hominin genus. (**a**) Character Completeness Metric 2 (CCM2). (**b**) Skeletal Completeness Metric 2 (SCM2). Mean scores for each completeness metric are shown in Table 4.4.

pattern is supported by the maximum CCM scores: both CCM1-2_{max} are low (<30%) from 7.0–3.6 Ma, increasing three-fold from 3.9 Ma to 3.6 Ma and remaining almost constant thereafter (Fig. 4.4). In short, there are clear, major differences in phylogenetic completeness between the first half and second half of hominin evolution.

For the SCM1-2, mean completeness scores are also low (<14%) from 7.0 to 4.6 Ma (Fig. 4.3). Thereafter, trends in character and skeletal completeness differ markedly, with the latter displaying three notable peaks through the Plio-Pleistocene. First, SCM1 and SCM2 increase fourteen- and nine-fold, respectively, at 4.4 Ma. This increase can be attributed to the appearance of *Ardipithecus ramidus* in the Sagantole Formation (Ethiopia) and the exceptionally complete associated skeleton ARA-VP-6/500 (White *et al.*, 2009). This peak is not recovered in the CCM1-2. Second, SCM1-2 reach their highest peak between 3.2–2.8 Ma, representing—as in the CCM1-2—the presence of *Australopithecus afarensis* (in a time bin with no other taxa) and the first appearance of *Australopithecus africanus*. Third, SCM1-2 display a gradual rise in skeletal completeness from 2.0 Ma to 1.0 Ma. This increase is caused by the disappearance of low-scoring taxa such as *Homo habilis* (SCM2 = 31%) and *Homo rudolfensis* (SCM2 = 12%), and the increasing contribution of *Homo erectus* (SCM2 = 80%). These peaks are separated by troughs from 4.1–3.4 Ma and ca., 2.0 Ma. The first trough can be attributed to the appearance of *Australopithecus anamensis* (SCM2 = 16%) in the Kanapoi Formation at 4.2 Ma and the highly incomplete fossils of *Australopithecus bahrelghazali* (a lower jaw fragment; SCM2 = 2%), *Australopithecus*

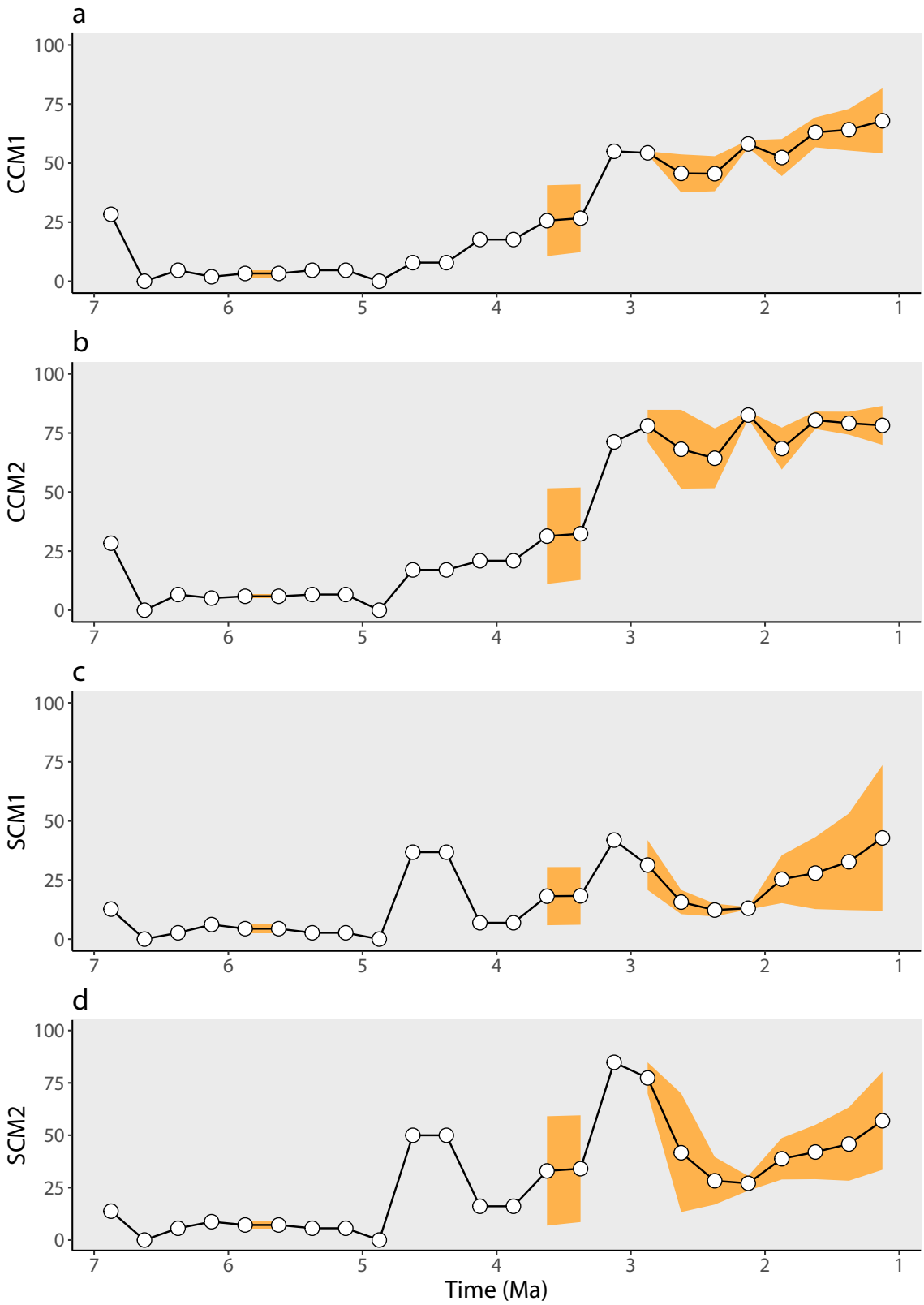

Figure 4.3: Mean early hominin completeness scores through geological time. (**a**) Character Completeness Metric 1 (CCM1). (**b**) Character Completeness Metric 2 (CCM2). (**c**) Skeletal Completeness Metric 1 (SCM1). (**d**) Skeletal Completeness Metric 2 (SCM2). Data are plotted at the midpoint age of each time bin. Orange areas surrounding completeness scores represent 95% confidence intervals.

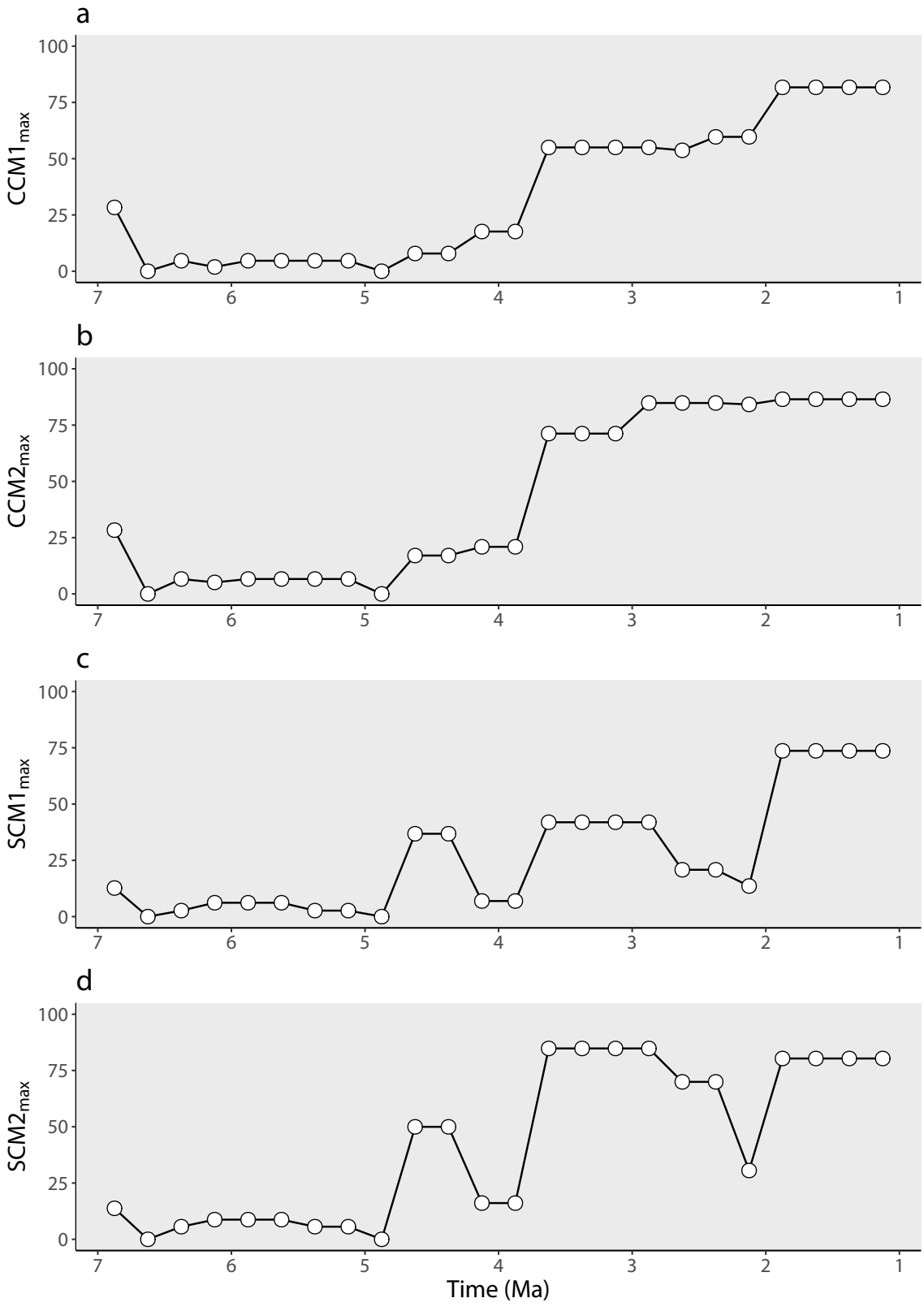

Figure 4.4: Maximum early hominin completeness scores through geological time. (**a**) Maximum Character Completeness Metric 1 ($CCM1_{max}$). (**b**) Maximum Character Completeness Metric 2 ($CCM2_{max}$). (**c**) Maximum Skeletal Completeness Metric 1 ($SCM1_{max}$). (**d**) Maximum Skeletal Completeness Metric 2 ($SCM2_{max}$). Data are plotted at the midpoint age of each time bin.

Table 4.5: Results of the statistical analyses comparing mean completeness scores through geological time. See text for an explanation of the abbreviations of each completeness score. Statistically significant correlations are shown in bold text. *Significant at $p \leq 0.05$. **Significant at $p \leq 0.05$ after false discovery rate (FDR) correction (Benjamini & Hochberg, 1995).

Comparison	Spearman's ρ	Kendall's τ
SCM1 *versus* SCM2	**0.861****	**0.715****
SCM1 *versus* CCM1	0.372	**0.304***
SCM1 *versus* CCM2	0.414	**0.328***
SCM2 *versus* CCM1	**0.431***	**0.352***
SCM2 *versus* CCM2	**0.502***	**0.423****
CCM1 *versus* CCM2	**0.980****	**0.913****

deyiremeda (an upper jaw fragment; SCM2 = 5%), and *Kenyanthropus platyops* (a relatively complete but plastically deformed cranium; SCM2 = 13%) in the middle Pliocene (3.6–3.4 Ma). The second trough represents *Australopithecus garhi* (SCM2 = 4%), *Paranthropus aethiopicus* (SCM2 = 13%), *Paranthropus boisei* (SCM2 = 24%), and early *Homo*. These taxa are mostly represented by fossil skulls; postcranial elements are either non-existent, cannot confidently be assigned to each taxon (Wood & Leakey, 2011), or are highly fragmentary, and therefore their SCM score is capped at 15% (Table 4.3). A similar pattern is found for the SCM1-2_{max} except for the middle Pleistocene (1.8–1.0 Ma) where completeness scores are high and invariant (Fig. 4.4).

4.3.2 Comparisons between completeness metrics

The only statistically significant correlations after false discovery rate (FDR) correction are those between mean SCM1 and SCM2, and CCM1 and CCM2 (Table 4.5). There are statistically significant correlations between SCM2 and CCM1, and SCM2 and CCM2. However, the latter is only significant after FDR correction for the Kendall τ, while the former is rendered non-significant after FDR correction. In addition, there are comparable correlations between SCM1 and CCM1, and SCM1 and CCM2. However, these correlations are only significant for the Kendall τ before FDR correction. When using maximum completeness scores per time bin, the only statistically significant correlations after FDR correction (for the Spearman's) are those between: SCM1 and SCM2; SCM2 and CCM2; and CCM1 and CCM2 (Table

Table 4.6: Results of the statistical analyses comparing maximum completeness scores through geological time. See text for an explanation of the abbreviations of each completeness score. Statistically significant correlations are shown in text. *Significant at $p \leq 0.05$. **Significant at $p \leq 0.05$ after false discovery rate (FDR) correction (Benjamini & Hochberg, 1995).

Comparison	Spearman's ρ	Kendall's τ
$SCM1_{max}$ *versus* $SCM2_{max}$	**0.650****	**0.542****
$SCM1_{max}$ *versus* $CCM1_{max}$	**0.564***	**0.502****
$SCM1_{max}$ *versus* $CCM2_{max}$	0.269	0.233
$SCM2_{max}$ *versus* $CCM1_{max}$	**0.582****	**0.470****
$SCM2_{max}$ *versus* $CCM2_{max}$	**0.607****	**0.534****
$CCM1_{max}$ *versus* $CCM2_{max}$	**0.752****	**0.605****

Table 4.7: Results of the statistical analyses comparing skull completeness scores through geological time. See text for an explanation of the abbreviations of each completeness score. Statistically significant correlations are shown in bold text. *Significant at $p \leq 0.05$. **Significant at $p \leq 0.05$ after false discovery rate (FDR) correction (Benjamini & Hochberg, 1995).

Comparison	Spearman's ρ	Kendall's τ
SCM_{skull} *versus* CCM1	**0.683****	**0.549****
SCM_{skull} *versus* CCM2	**0.703****	**0.557****
SCM_{skull} *versus* $CCM1_{max}$	**0.642****	**0.502****
SCM_{skull} *versus* $CCM2_{max}$	**0.807****	**0.628****
SCM_{skull} *versus* SCM1	0.382	**0.296***
SCM_{skull} *versus* SCM2	0.352	0.249
SCM_{skull} *versus* $SCM1_{max}$	0.384	**0.336***
SCM_{skull} *versus* $SCM2_{max}$	**0.576****	**0.462****

4.6). Finally, a statistically significant correlation is found when the SCM_{skull} is compared to CCM1-2 and CCM1-2_{max} after FDR correction (Table 4.7). Runs tests demonstrate a statistically significant deviation from randomness ($p < 0.003$) for all metrics (including the mean and maximum completeness scores).

4.3.3 Hominin diversity and fossil record completeness

Early hominin TDE is shown in Fig. 4.5. There is no statistically significant correlation between any mean completeness metrics (CCM1-2 and SCM1-2) and TDE (Figs. 4.6 and 4.7 and Table 4.8). However, TDE correlates significantly with CCM1-2_{max} before but not after FDR correction (Figs. 4.8 and 4.9 and Table 4.9), indicating that apparent diversity somewhat tracks maximal phylogenetic completeness.

In the GLS analysis, TDE is not included in any of the best-supported mean completeness models when ranked by Akaike weight (Tables 4.10, 4.11, 4.12, and 4.13). However, TDE is included in the best-supported model for the $CCM1_{max}$ (Table 4.14), the $SCM1_{max}$ (Table 4.16), the $SCM2_{max}$ (Table 4.17), and the SCM_{skull} (Table 4.18). Pairwise tests and GLS analyses both show that TDE consistently correlates with the maximum variant of each completeness metric, indicating that the occurrence of a single highly complete specimen potentially aids taxonomic identification of all other specimens in a particular time bin, raising apparent diversity.

A strong correlation between each taxon's CCM2 score and the number of autapomorphies recovered in the Strait & Grine (2004) cladistic analysis (Fig. 4.10) demonstrates that specimen completeness is proportional to the number of phylogenetically diagnostic characters in modern hominoids and fossil hominins. Therefore, specimen completeness can directly inform debates surrounding the fidelity of fluctuations in apparent diversity.

4.3.4 Controls on hominin fossil record completeness

Sampling metrics. Sampling metrics through geological time are shown in Fig. 4.5. The only statistically significant correlation is between SCM2 and the proxy for the bonanza effect, the ratio of hominin-bearing collections to formations (Table 4.8 and Fig. 4.7), suggesting a link between specimen completeness and the tendency of palaeoanthropologists to return to deposits rich in hominin fossils. However,

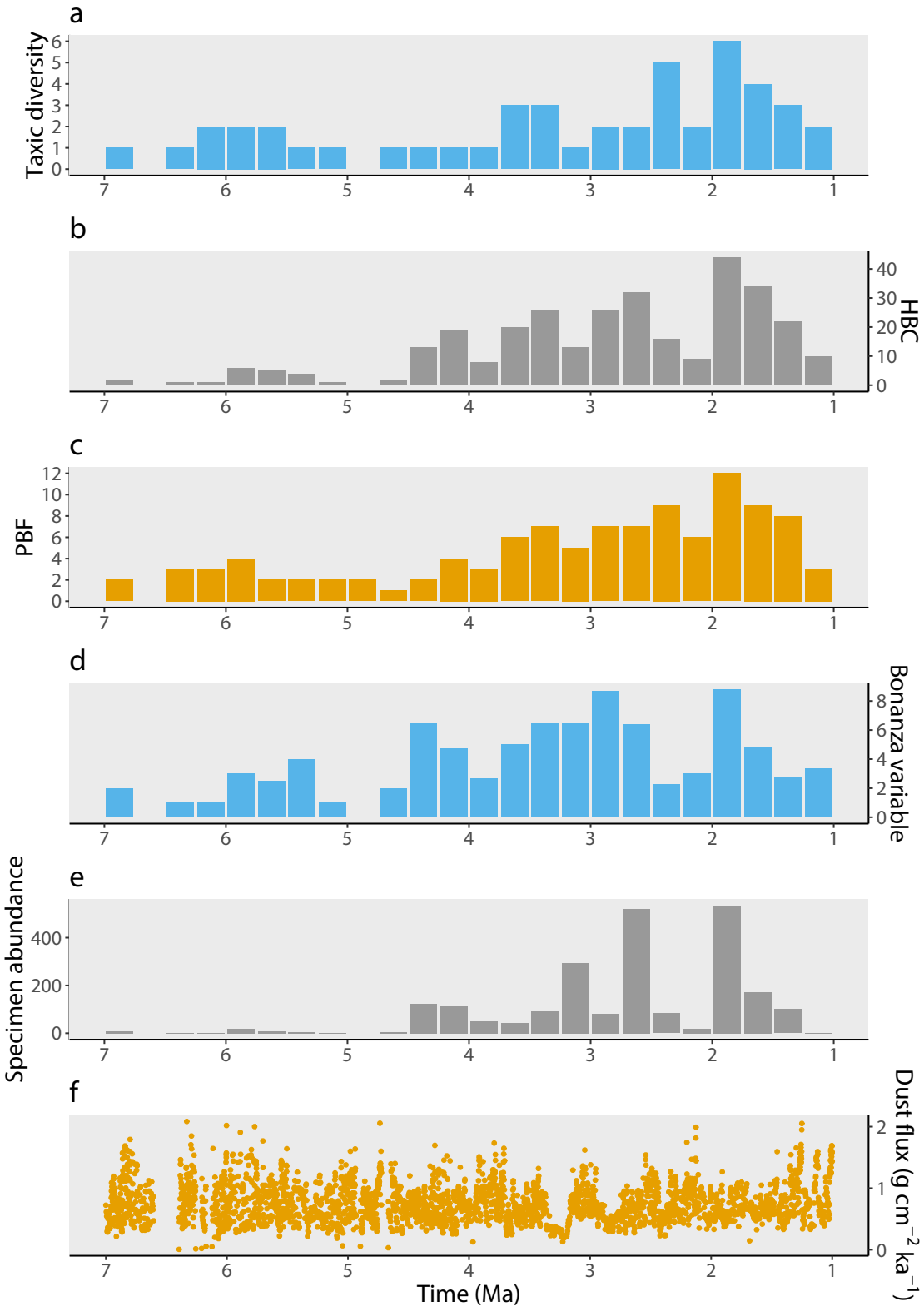

Figure 4.5: Early hominin raw taxic diversity, sampling metrics, and terrigenous dust flux through geological time. (**a**) Early hominin taxic diversity estimate (TDE). (**b**) Hominin-bearing collections (HBC). (**c**) Primate-bearing formations (PBF). (**d**) The bonanza variable: the ratio of hominin-bearing collections to formations (HBC:HBF). (**e**) Specimen abundance (in-bin unique accession numbers from the HFDB). Data are plotted at the midpoint age of each time bin. (**f**) Terrigenous dust flux (deMenocal, 1995).

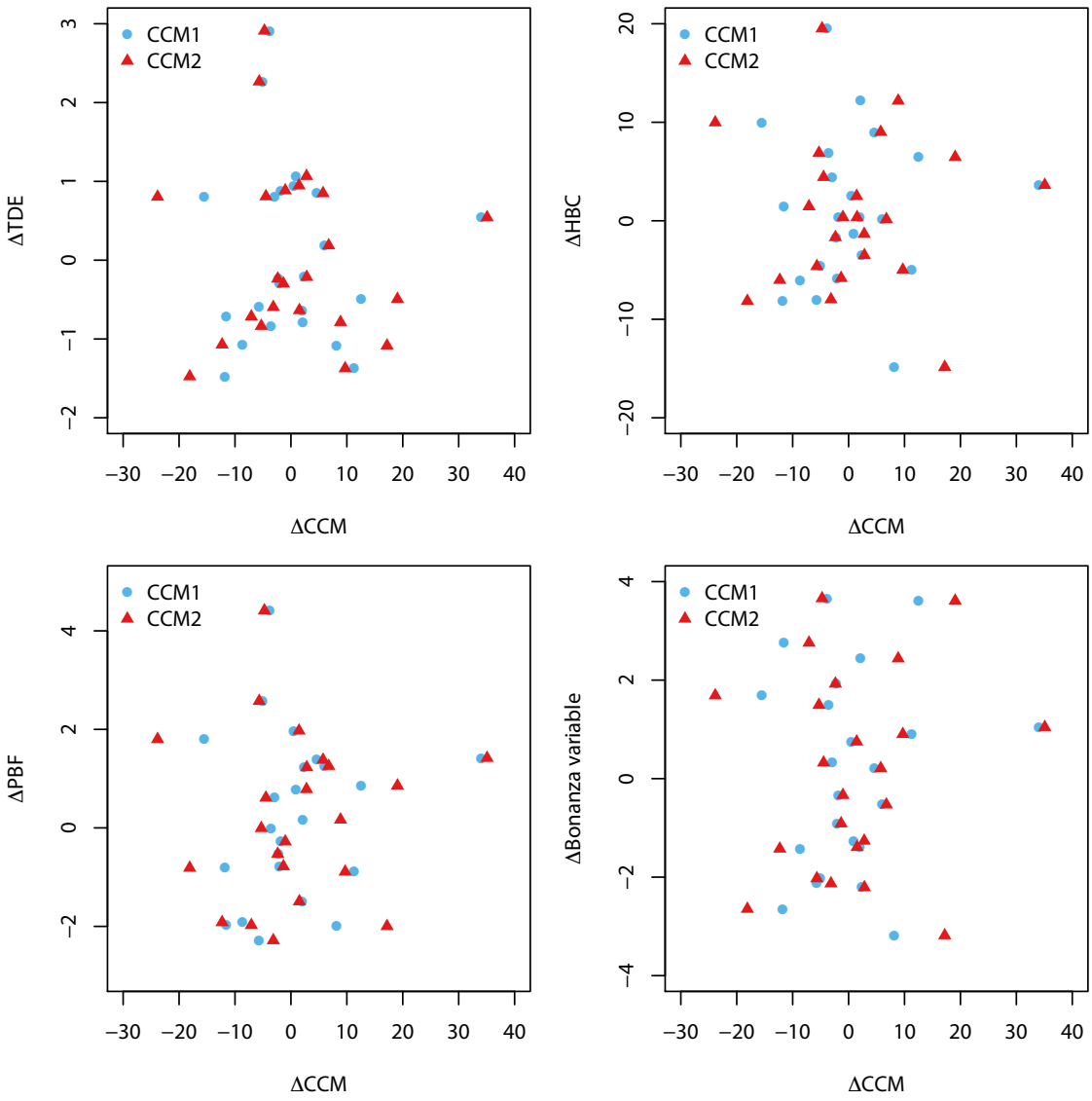

Figure 4.6: Scatter plots showing the relationships between mean character completeness (CCM1-2), raw taxic diversity, and sampling metrics. Δ indicates that the time series has undergone generalised differencing prior to statistical testing. (**a**) CCM1/2 against TDE. (**b**) CCM1/2 against HBC. (**c**) CCM1/2 against PBF. (**d**) CCM1/2 against the bonanza variable (HBC:HBF). Correlation coefficients and *p*-values are shown in Table 4.8.

this correlation becomes non-significant after FDR correction (Table 4.8). Using the highest completeness score per time bin (denoted by the subscript max) leads to slightly different correlations. In addition to the correlation between $SCM2_{max}$ and the bonanza effect (Fig. 4.9), each completeness metric correlates significantly with HBC before but not after FDR correction (Figs. 4.8 and 4.9). The only significant correlations after FDR correction are those between CCM1-2_{max} and PBF (Figs. 4.8 and Table 4.9).

In the GLS analysis, CCM1-2 are best explained by collection effort (model 30; Tables 4.10 and 4.11) when ranked by Akaike weight, while SCM1-2 are best explained by the bonanza variable (model 31; Tables 4.12 and 4.13). Regardless of whether specimen completeness is assessed according to character or skeletal completeness metrics, anthropogenic sampling bias is repeatedly found to be the biggest driver. However, the results are more complex when specimen completeness is assessed according to the max-

imum completeness score: the best-supported model for the $CCM1_{max}$ = diversity + bonanza variable (Table 4.14), for the $CCM2_{max}$ = collections (Table 4.15), for the $SCM1_{max}$ = diversity + formations + bonanza variable (Table 4.16), and for the $SCM2_{max}$ = diversity + bonanza variable (Table 4.17). These results support those from the pairwise correlations: the maximum completeness score correlates with TDE much more so than the mean. SCM_{skull} is also best explained by a model combining diversity + bonanza variable (Table 4.18).

Climate. There is no statistically significant correlation between any of the completeness metrics and aridity (Table 4.8 and 4.9). Interestingly, all correlation coefficients are negative, suggesting that, while aridity is not a major control of specimen completeness, completeness scores are generally lower

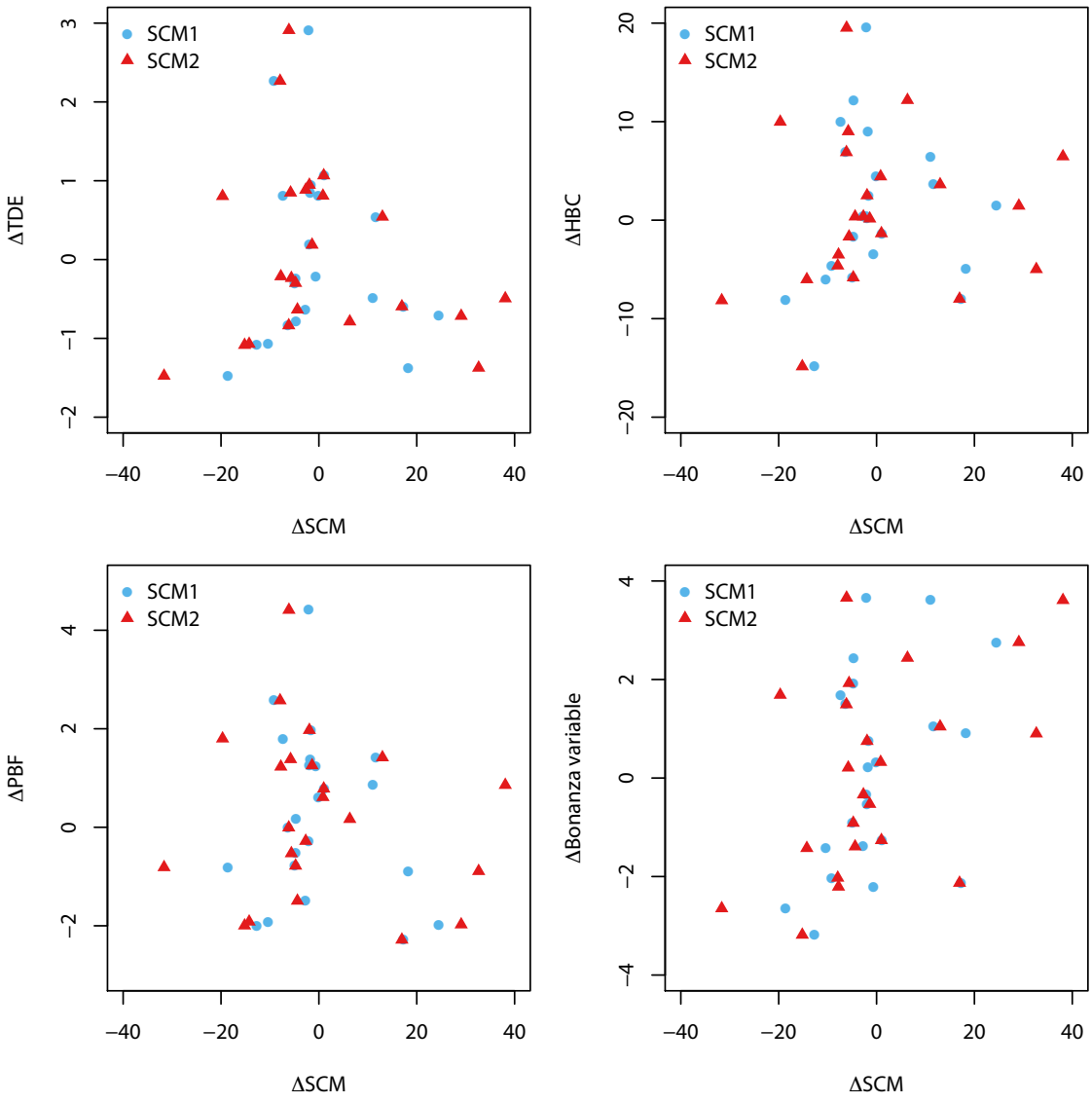

Figure 4.7: Scatter plots showing the relationships between mean skeletal completeness (SCM1-2), raw taxic diversity, and sampling metrics. Δ indicates that the time series has undergone generalised differencing prior to statistical testing. (**a**) SCM1/2 against TDE. (**b**) SCM1/2 against HBC. (**c**) SCM1/2 against PBF. (**d**) SCM1/2 against the bonanza variable (HBC:HBF). Correlation coefficients and *p*-values are shown in Table 4.8.

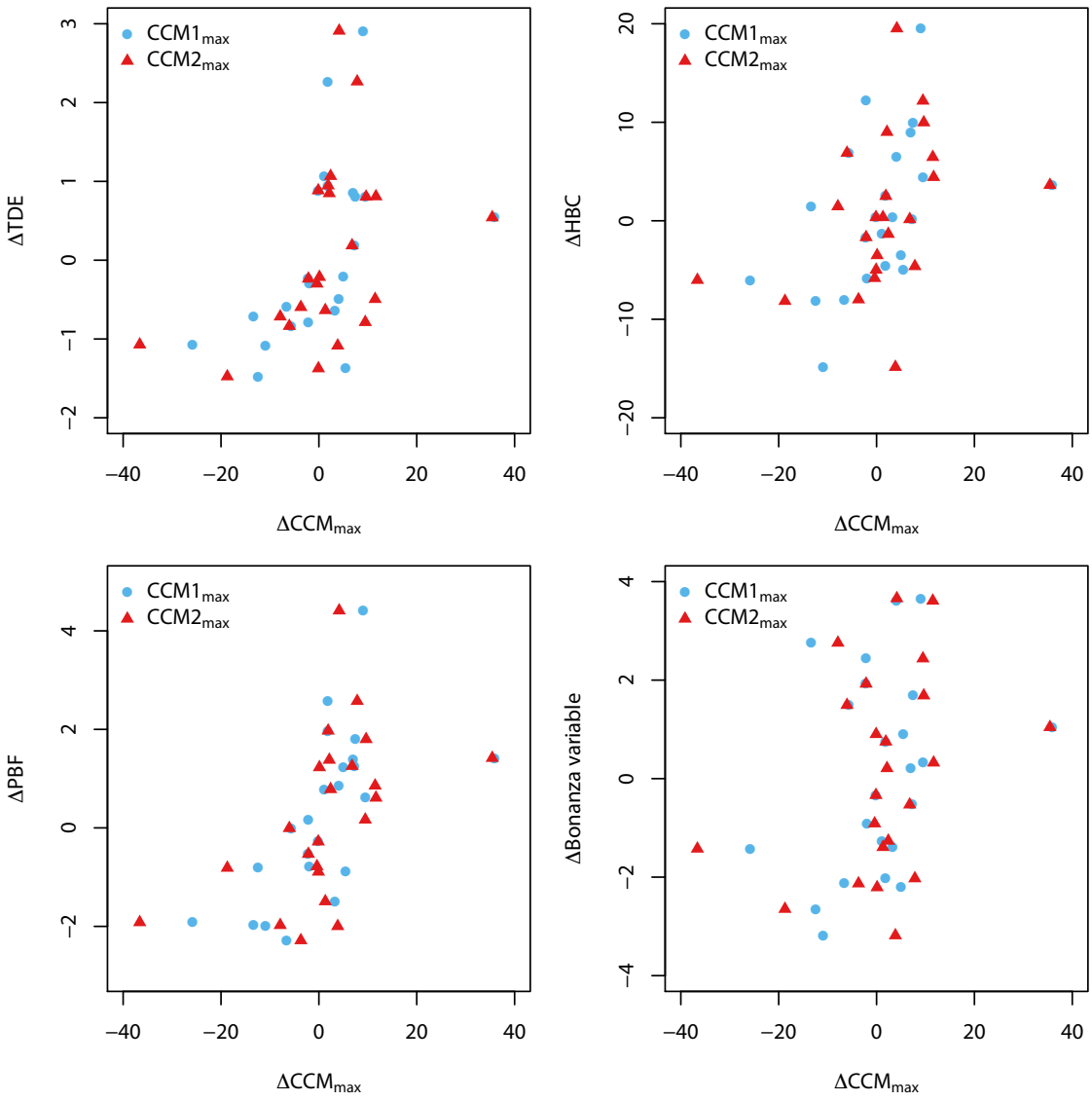

Figure 4.8: Scatter plots showing the relationships between maximum character completeness (CCM1-2$_{max}$), raw taxic diversity, and sampling metrics. Δ indicates that the time series has undergone generalised differencing prior to statistical testing. (**a**) CCM1-2$_{max}$ against TDE. (**b**) CCM1-2$_{max}$ against HBC. (**c**) CCM1-2$_{max}$ against PBF. (**d**) CCM1-2$_{max}$ against the bonanza variable (HBC:HBF). Correlation coefficients and *p*-values are shown in Table 4.9.

when aridity is higher (as predicted above). Moreover, aridity is consistently eliminated as a significant predictor in the GLS analyses (Table 4.10 to 4.18).

Geographical region and depositional environment. There is no statistically significant difference in per-taxon CCM2 ($U = 11, p = 0.296$) and SCM2 scores ($U = 7, p = 0.112$) in eastern and southern Africa, despite mean completeness scores being higher in southern Africa (CCM2 = 61%, SCM2 = 52%) than in eastern Africa (CCM2 = 41%, SCM2 = 27%). Despite the distinct depositional environments in the major geographical regions that yield hominin fossils (fluvio-lacustrine and palaeosol deposits in eastern Africa, karst deposits and cave infills in southern Africa), these results indicate that there is no statistical difference in early African hominin completeness scores (either phylogenetic or skeletal)

between fluvio-lacustrine sediments and cave infills.

Epoch-scale differences in specimen completeness. Mean completeness scores for each epoch are shown in Fig. 4.11. For both the CCM2 and SCM2, the late Miocene is the most incomplete (mean CCM2 = 9%, SCM2 = 7%), with completeness increasing through the Pliocene (mean CCM2 = 28%, SCM2 = 34%) and early Pleistocene (mean CCM2 = 74%, SCM2 = 40%). Kruskal-Wallis tests show that there are statistically significant differences in CCM2 (χ^2 = 14.23, p < 0.001), and SCM2 scores (χ^2 = 7.90, p = 0.019) for each epoch. Near-identical results are found using the maximum completeness scores per time bin. The trend of decreasing specimen and taxon completeness with increasing stratigraphic age is also present when completeness scores are regressed against the midpoint age of each time bin

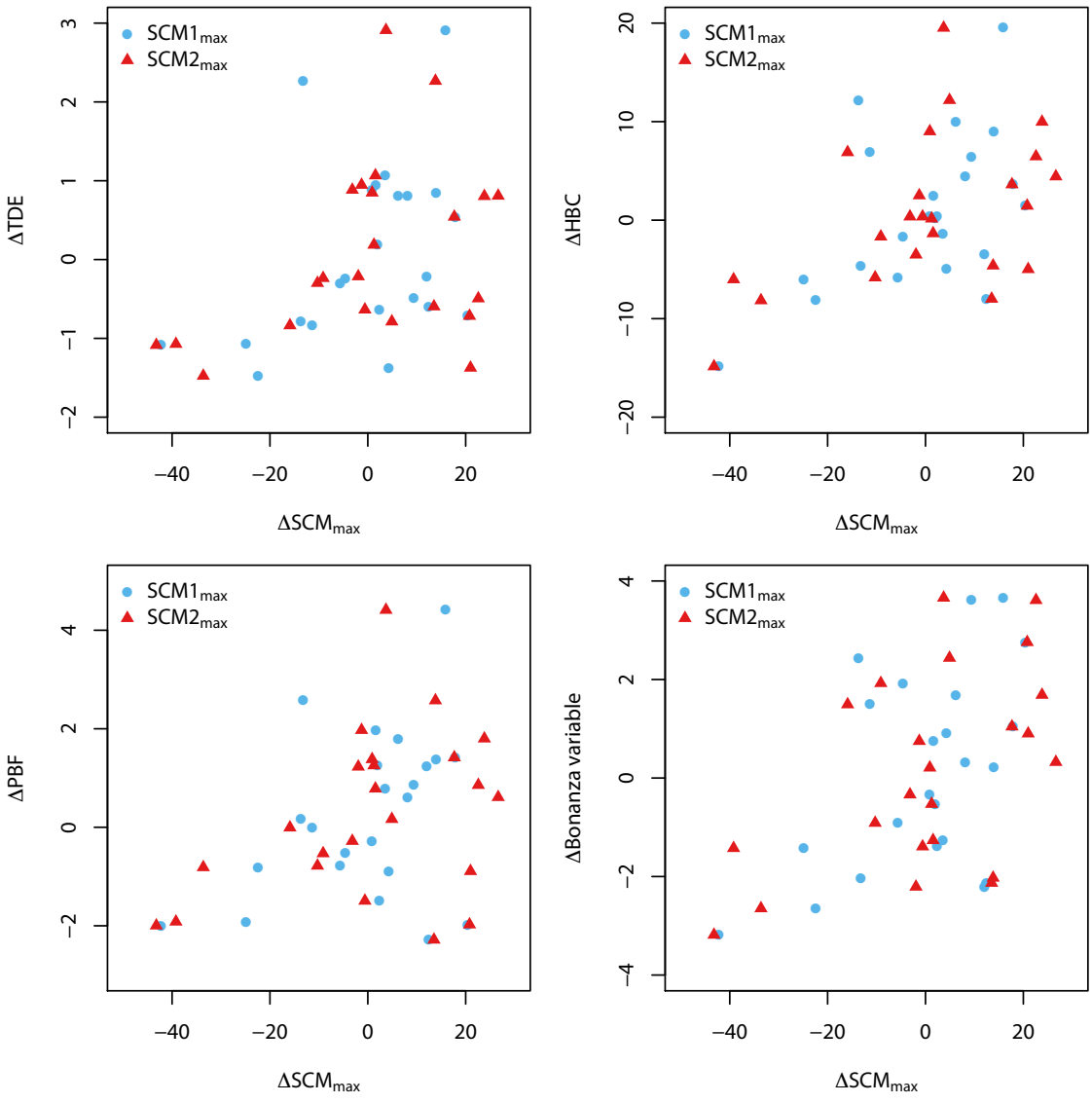

Figure 4.9: Scatter plots showing the relationships between maximum skeletal completeness (SCM1-2$_{max}$), raw taxic diversity, and sampling metrics. Δ indicates that the time series has undergone generalised differencing prior to statistical testing. (**a**) SCM1-2$_{max}$ against TDE. (**b**) SCM1-2$_{max}$ against HBC. (**c**) SCM1-2$_{max}$ against PBF. (**d**) SCM1-2$_{max}$ against the bonanza variable (HBC:HBF). Correlation coefficients and p-values are shown in Table 4.9.

Table 4.8: Results of the statistical analyses comparing mean completeness scores, sampling metrics, and aridity. See text for an explanation of the abbreviations of each completeness score and sampling metric. Significant correlations are shown in bold text. *Significant at $p \leq 0.05$. **Significant at $p \leq 0.05$ after false discovery rate (FDR) correction (Benjamini & Hochberg, 1995).

Comparison	Spearman's ρ	Kendall's τ
CCM1 *versus* TDE	−0.009	0.004
CCM2 *versus* TDE	−0.032	−0.004
SCM1 *versus* TDE	0.184	0.130
SCM2 *versus* TDE	0.012	0.004
CCM1 *versus* HBC	0.079	0.059
CCM2 *versus* HBC	0.058	0.051
SCM1 *versus* HBC	0.205	0.138
SCM2 *versus* HBC	0.203	0.170
CCM1 *versus* PBF	0.063	0.067
CCM2 *versus* PBF	0.016	0.028
SCM1 *versus* PBF	−0.088	−0.059
SCM2 *versus* PBF	−0.159	−0.091
CCM1 *versus* bonanza	0.044	0.036
CCM2 *versus* bonanza	0.045	0.043
SCM1 *versus* bonanza	0.324	0.257
SCM2 *versus* bonanza	**0.429***	**0.320***
CCM1 *versus* aridity	−0.222	−0.162
CCM2 *versus* aridity	−0.242	−0.170
SCM1 *versus* aridity	−0.250	−0.209
SCM2 *versus* aridity	−0.212	−0.146

(CCM2: $r = -0.789$, $p < 0.001$; SCM2: $r = -0.668$, $p < 0.001$; note the negative sign as time decreases towards the present).

Stratigraphic range. There is no statistically significant correlation between stratigraphic range and specimen completeness (CCM2: $r = 0.444$, $p = 0.059$; SCM2: $r = 0.438$, $p = 0.071$), suggesting that a longer stratigraphic range (= higher probability of more complete specimens) does not result in more complete taxa, or that more complete and easily identifiable specimens do not result in longer stratigraphic ranges (Table 2.1).

4.3.5 Completeness through research time

Holotype publication year display significant negative correlations with both per-taxon CCM2 ($r = -0.736$, $p < 0.001$) and SCM2 scores ($r = -0.516$, $p = 0.028$), indicating that taxa described more recently have typically been named based on more fragmentary fossils that contain less phylogenetic information (Fig. 4.12). This difference in specimen completeness is clearest in taxa named before (mean CCM2 = 74%, SCM2 = 44%) and after the year 1990 (mean CCM2 = 16%, SCM2 = 17%). Kruskal-Wallis tests show that this difference is statistically significant for both the CCM2 ($p < 0.001$) and SCM2 ($p = 0.026$).

Table 4.9: Results of the statistical analyses comparing maximum completeness scores, sampling metrics, and aridity. See text for an explanation of the abbreviations of each completeness score and sampling metric. Significant correlations are shown in bold text. *Significant at $p \leq 0.05$. **Significant at $p \leq 0.05$ after false discovery rate (FDR) correction (Benjamini & Hochberg, 1995).

Comparison	Spearman's	Kendall's
$CCM1_{max}$ *versus* TDE	**0.581***	**0.431***
$CCM2_{max}$ *versus* TDE	**0.489***	**0.368***
$SCM1_{max}$ *versus* TDE	0.377	0.265
$SCM2_{max}$ *versus* TDE	0.307	0.265
$CCM1_{max}$ *versus* HBC	**0.542***	**0.391***
$CCM2_{max}$ *versus* HBC	**0.519***	**0.360***
$SCM1_{max}$ *versus* HBC	**0.442***	**0.320***
$SCM2_{max}$ *versus* HBC	**0.471***	**0.352***
$CCM1_{max}$ *versus* PBF	**0.717****	**0.534****
$CCM2_{max}$ *versus* PBF	**0.646****	**0.423***
$SCM1_{max}$ *versus* PBF	0.281	0.241
$SCM2_{max}$ *versus* PBF	0.339	0.273
$CCM1_{max}$ *versus* bonanza	0.312	0.225
$CCM2_{max}$ *versus* bonanza	0.288	0.178
$SCM1_{max}$ *versus* bonanza	0.411	**0.296***
$SCM2_{max}$ *versus* bonanza	**0.496***	**0.344***
$CCM1_{max}$ *versus* aridity	−0.209	−0.146
$CCM2_{max}$ *versus* aridity	−0.064	−0.051
$SCM1_{max}$ *versus* aridity	−0.197	−0.123
$SCM2_{max}$ *versus* aridity	−0.233	−0.154

4.4 Discussion

4.4.1 Comparison of completeness metrics

Most past studies that have assessed both the CCM and SCM have reported a strong positive correlation between them and interpreted this as support for both metrics homing in on the same signal (e.g., Mannion & Upchurch, 2010; Brocklehurst & Fröbisch, 2014; Tutin & Butler, 2017). However, in these studies the CCM is based on the entire skeleton (skull + postcrania) whereas hominin phylogenetic analyses (and therefore CCM) only sample characters of the skull (Strait *et al.*, 2007). The absence of a strong correlation between CCM and SCM (whether using the mean or maximum variant) is a reflection of the former measuring skull completeness and the latter measuring whole skeleton completeness, and not necessarily evidence for an absence of signal. This is borne out by the skeletal completeness of the skull metric (SCM_{skull}) which closely tracks both the CCM1-2 and the CCM1-2_{max} but not the SCM1-2 or the $SCM1_{max}$ (SCM_{skull} and $SCM2_{max}$ correlate significantly after FDR correction; Table 4.7). Unlike past studies, unique insights into the distinct patterns of skull and skeleton completeness in the early African hominin fossil record are offered by those instances where CCM and SCM differ.

High amplitude fluctuations in both the SCM1-2 and SCM1-2_{max} indicate that skeletal completeness is primarily controlled by the sporadic occurrence of deposits with abundant and diverse fossils (Johanson *et al.*, 1982; Walker & Leakey, 1993; White *et al.*, 2009; Clarke, 2013). Palaeoanthropologists

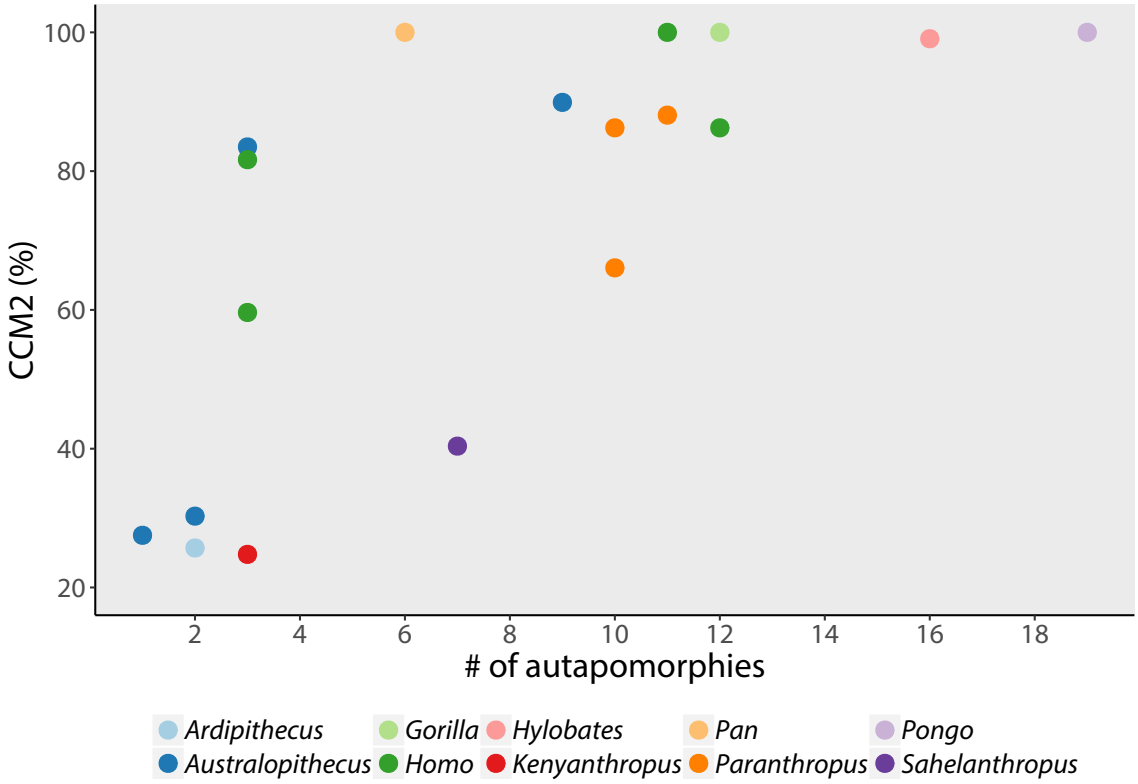

Figure 4.10: Scatter plot showing per-taxon completeness scores against the number (#) of uniquely derived characteristics (autapomorphies). Per-taxon CCM2 scores were calculated directly from the Strait & Grine (2004) character-taxon matrix of modern hominoids and fossil hominins (R^2 = 0.49, r = 0.720, p < 0.001). The four modern hominids (*Pongo*, *Gorilla*, *Pan*, and *Homo*) are 100% complete and the modern hylobatid (*Hylobates*) is 99.1% complete. In contrast, fossil hominins range from 24.7%–88.1% complete.

presumably, therefore, search according to the bonanza principle (Raup, 1977), preferentially sampling the richest rock units rather than working systematically over all exposed rock (Dunhill *et al.*, 2012, 2013, 2014*b*). In finding the richness rock units and preferentially re-sampling them, palaeoanthropologists directly relate the bonanza principle to specimen completeness: increased collection effort in a time bin will, statistically, lead to more complete specimens and more complete taxa. In contrast, high amplitude fluctuations are absent in both the CCM1-2 and CCM1-2_{max} indicating that character completeness is not controlled by the bonanza effect (Figs. 4.3 and 4.4).

The hominin fossil record is largely composed of crania and teeth which impose an upper limit of 15% on the SCM (Table 4.3). Consequently, the SCM is generally low, only increasing beyond this value with the addition of postcranial remains. In contrast, the skull constitutes 100% of characters in the CCM and, as a result, the early African hominin fossil record is best described as either incomplete (few, fragmentary skulls; 7.0–3.4 Ma) or complete (many, well-preserved skulls; 3.4–1.0 Ma). The CCM describes variation in specimen completeness for a single (albeit phylogenetically informative) skeletal region. However, it neglects a proportion of the skeleton that has changed dramatically during hominin evolution (Richmond & Hatala, 2013), and one that doubtless contains useful phylogenetic information (e.g., Begun, 2001; Finarelli & Clyde, 2004; Mounce *et al.*, 2016; Nengo *et al.*, 2017). The inclusion of postcrania in the SCM from deposits throughout the Afar Triangle (Ethiopia), Cradle of Humankind

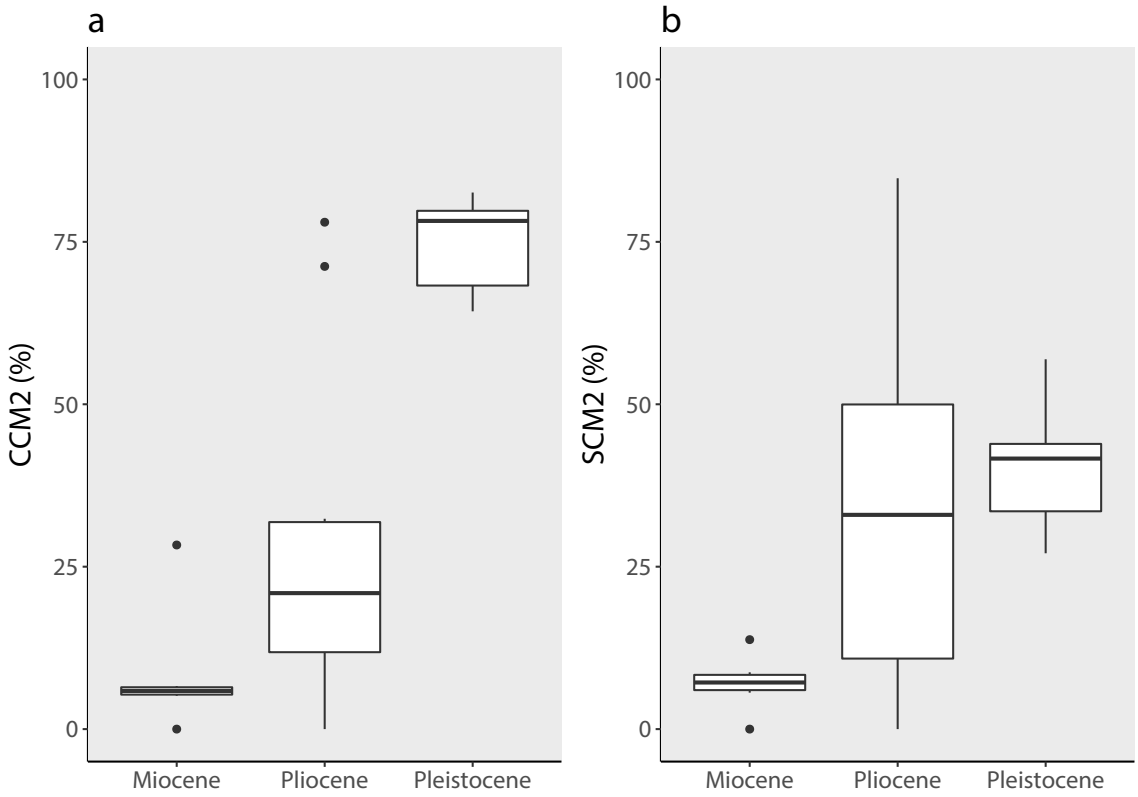

Figure 4.11: Boxplot showing per-time bin completeness scores for each epoch: late Miocene (7.0–5.3 Ma), Pliocene (5.33–2.6 Ma), and early Pleistocene (2.6–1.0 Ma). (**a**) Character Completeness Metric 2 (CCM2). (**b**) Skeletal Completeness Metric 2 (SCM2). Time bins are assigned to each epoch based on their start date. The increase in specimen completeness through geological time mirrors the increase in sampled area (Fig. 3.11), suggesting the larger the geographical spread of collection effort the higher the completeness score (= more of a skeleton known or characters scored in a cladistic matrix).

World Heritage Site (South Africa), and the Omo-Turkana Basin (Ethiopia-Kenya), for example, means the remaining 85% of the skeleton (by weight) can be used to produce a more detailed reconstruction of total specimen completeness. The SCM (specifically SCM2), therefore, offers a more nuanced depiction of specimen completeness and is the most meaningful metric for quantifying specimen completeness in hominins. Until the CCM reflects the entire hominin skeleton, CCM and SCM are unlikely to correlate as strongly as in earlier studies.

4.4.2 Completeness metrics and hominin diversity

Species are defined based on their unique combination of characters (autapomorphies) and the presence of shared derived characters (synapomorphies) identifies taxa as members of a clade (Smith, 1994). However, the finding that phylogenetically informative characters tend to be preferentially lost during fossilisation, and the fact that fossils are naturally fragmentary, means that key synapomorphies are commonly absent or unidentifiable (Sansom, 2015). The loss of key synapomorphies affects both phylogenetic reconstruction and all subsequent evolutionary inferences. For example, a recent meta-analysis of data from modern clades showed that incomplete taxa tend to resolve erroneously closer to the root of the tree—a phenomenon named stem-ward slippage (Sansom & Wills, 2013)—and incomplete taxa

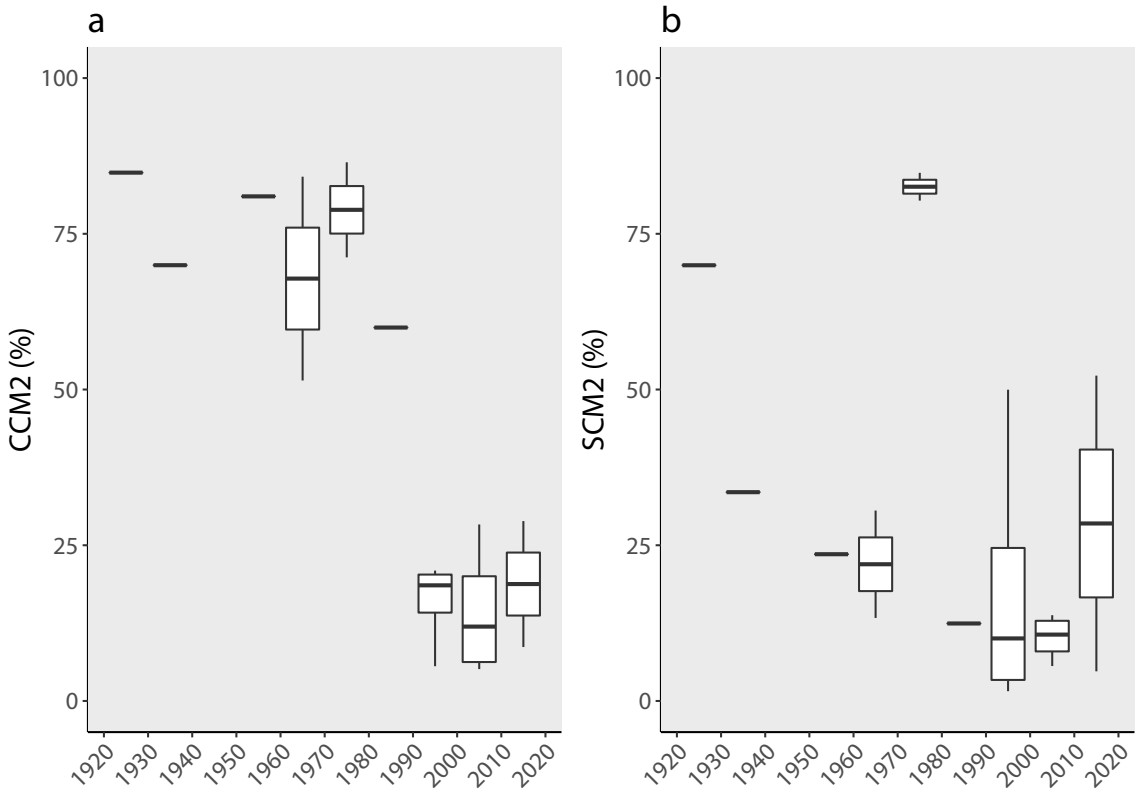

Figure 4.12: Boxplot showing per-taxon completeness scores through research time. (**a**) CCM2 through research time. (**b**) SCM2 through research time. Time bins were constructed from 1920–1929, 1930–1939, etc. Taxa were assigned to each decade based on the year the taxonomic nomen first formally appeared in the literature (Wood, 2011).

are known to reduce node support (e.g., Pattinson *et al.*, 2015). In relation to specimen completeness and apparent diversity, one possibility is that high specimen/taxon completeness promotes higher apparent diversity, because more skeletal material ought to mean more diagnostic characters can be assessed and taxonomic differences are easier to identify (Mannion & Upchurch, 2010). Alternatively, low specimen/taxon completeness could result in higher apparent diversity if taxonomists erroneously recognise more taxa than were actually present: a phenomenon named taxonomic over-splitting (Brocklehurst & Fröbisch, 2014). Indeed, the latter scenario has been raised as a possibility for the purported hominins of the late Miocene (*Sahelanthropus*, *Orrorin*, and *Ardipithecus*) whose hypodigms share very little anatomical overlap (Wood & Boyle, 2016:43).

In modern hominoids and fossil hominins, the number of autapomorphies that define each OTU (*sensu* Strait & Grine, 2004) is directly proportional to the number of characters that can be scored (Fig. 4.10), meaning that more complete taxa appear richer in uniquely derived characteristics and implying that temporal variation in specimen completeness can affect patterns of apparent species richness. Early hominin TDE is systematically biased by temporal variation in rock availability and collection effort (Maxwell *et al.*, 2018), so in order for specimen completeness and TDE to co-vary, completeness metrics ought to be controlled by the same—or similar—sampling artefacts. However, the absence of any consistent relationship between mean completeness scores (whether phylogenetic or skeletal) and

TDE across the different statistical tests leads to the conclusion that specimen completeness does not systematically affect the recognition of taxa in the same way that sampling metrics systematically affect apparent diversity. While TDE is largely controlled by PBF, completeness metrics are largely controlled by collection effort and the bonanza effect. The non-correlation between PBF and the bonanza variable, and TDE and the bonanza variable, supports the idea that taxic diversity patterns and specimen completeness are driven by different processes. On the other hand, the consistent finding that TDE and CCM1-2_{max} correlate with one another (albeit before FDR correction) indicates that TDE tracks the highest completeness score much more so than the mean (Figs. 4.8 and 4.9). This might be because the occurrence of a highly complete specimen/taxon in a time bin makes the taxonomic identification of other, contemporary specimens/taxa much easier. By including a large proportion of the skeleton, a single highly complete specimen/taxon potentially allows for other, more fragmentary fossils to be compared more easily and for taxonomic assignments to be made with greater certainty. The finding that all statistically significant correlations with taxic diversity are confined to the CCM is unsurprising considering the CCM quantifies phylogenetic information content while the SCM quantifies the number of skeletal elements and their physical bulk (Mannion & Upchurch, 2010).

Without consistent support for a relationship between each variant of the CCM and SCM and taxic diversity, one can only conclude that, if specimen completeness is distorting the diversity signal in the early African hominin fossil record, its effect is only minor, relating more so to the most complete specimen/taxon in each time bin. Nonetheless, these findings reiterate that hominin diversity patterns are no less immune from error in the fossil record than other groups and ought to be subject to greater scrutiny.

4.4.3 Controls on hominin fossil record completeness

Sampling. A relationship between completeness metrics and sampling metrics is repeatedly found across the different statistical tests, although with differing degrees of support. Pairwise correlations indicate that only the bonanza variable has any significant effect on mean completeness scores, and this is limited to the SCM2 (the variant based on the whole skeleton). Sites that have yielded postcrania but which cannot be regarded as bonanza sites are few, but include, for example, those in the Baringo County and Nariokotome III (Kenya) and Maka (Ethiopia). In each case, very few HBC have been amassed since the first discovery of postcranial remains. For most other sites that have yielded hominin postcrania exploration has continued largely uninterrupted to the present day. The bonanza variable shows no correlation with mean CCM1-2.

Maximum completeness scores show the same result but indicate a more widespread role for other sampling biases. In addition to the $SCM2_{max}$ correlation with the bonanza variable, collection effort also correlates with all maximum completeness scores. Most surprisingly, however, CCM1-2_{max} and PBF correlate very strongly and are the only parameters to do so after correcting for multiple comparisons. This suggests that, in early African hominins at least, PBF may drive TDE, in part, through a common-cause interaction with CCM1-2_{max} (higher rock availability results in a greater chance of finding more complete specimens which, in turn, inflates taxic diversity). It also suggests that more opportunities to sample fossils tends to result in a wider range of completeness scores. So, while greater sampling effort increases the maximum completeness score per time bin (hence the link with PBF), it also increases the

likelihood of finding fragmentary fossils, leading to a mean completeness score that contains very little useful information regarding short-term, bin-to-bin fluctuations in sampling effort. This reiterates that while completeness metrics offer a unique insight into sampling bias, they do not capture all aspects of sampling heterogeneity (Benton *et al.*, 2013).

Interestingly, the GLS model with the highest conditional probability (= Akaike weight) is collection effort for the CCM and the bonanza variable for the SCM (note that due to the large number of GLS models only the second variant of each completeness metric is discussed). However, sampling metrics explain only 53% (CCM2) to 59% (SCM2) of the variance in mean completeness scores meaning almost half remains unaccounted for. This is, in part, a result of no single sampling metric displaying the same pattern as the CCM and SCM_{skull}: completeness metrics based solely on the skull show a bimodal pattern (relatively incomplete fossil record from 7.0 to 3.6 Ma and relatively complete fossil record thereafter), while sampling metrics fluctuate considerably. This unusual bimodal trend may reflect a sudden increase in rock exposure area, fossilisation potential, or population density. Of these, a marked reduction in suitable rock exposure (and therefore collection effort) is most likely key to explaining the relative poverty of the early African hominin fossil record. With regards to the SCM2, moderate statistical support probably reflects the fact that skeletal (skull + postcrania) completeness is a product of many complex biological, geological, and anthropogenic factors which are not all described by current sampling metrics.

The best-supported GLS models using maximum completeness scores show the same results but differ in that they (1) also include taxic diversity, supporting earlier findings that taxic diversity patterns are influenced by the most complete specimen/taxon, and (2) consistently explain a larger proportion of model variance ($CCM2_{max}$ = 58% and $SCM2_{max}$ = 71%) compared to mean completeness scores. Collection counts (specifically, counts of the number of unique years that hominin fossils have been collected) and the bonanza variable aim to quantify study effort and the non-random behaviour of palaeoanthropologists, and so the presence of these parameters in the best-supported GLS models indicates that bias and bonanza factors have probably operated in concert in the early African hominin fossil record. The finding that there is no trend in the occurrence of bonanza sites (Wald-Wolfowitz runs test: p = 0.382) indicates that their temporal distribution is not linked to cyclical climatic or geological processes.

Similar results have also been reported in Mesozoic birds and pterosaurs (Brocklehurst *et al.*, 2012; Dean *et al.*, 2016), groups whose highly non-uniform fossil sampling (dominated by Lagerstätten deposits) closely resembles the sporadic sampling found in early African hominins. While there are no hominin Lagerstätten *per se*, the study of human origins is subject to intense scrutiny and considerable interest. Naturally, palaeoanthropologists preferentially return to rich hominin-bearing deposits in the Sagantole Fm. (4.4 Ma), Hadar Fm. (3.4–3.0 Ma), Sterkfontein (ca., 2.6 Ma), Koobi Fora Fm. (1.9–1.4 Ma), and Swartkrans (ca., 1.9 Ma). Repeated field seasons at these formations/localities (and others) have produced multiple collections and remarkably complete and abundant specimens, and hence the largest peaks in the SCM. In birds and pterosaurs, Lagerstätten deposits produce a similar pattern: specimen completeness is generally low throughout much of their evolutionary history with sudden, rapid peaks. However, in Mesozoic birds and pterosaurs these peaks in specimen completeness (in both cases the CCM2) are the result of deposits that contain beautifully complete specimens, whereas in hominins these peaks reflect sustained collection effort at rich sites. Like Dean *et al.* (2016), we find that, instead

of rock availability controlling fossil record completeness, the inclusion of a variable that quantifies the bonanza principle (or, in pterosaurs, Lagerstätten presence/absence) is the principal driver.

Fossil sites are rare occurrences. Within them, hominins are rare and complete skeletons are rarer still. Not surprisingly, the availability and subsequent study of hominin-bearing rock is highly non-random due to the rarity of sampling events, funding opportunities, and research logistics. While it is fair to say that mean completeness scores are a poor proxy for continental-scale sampling effort, maximum completeness scores offer new insights into the nature of sampling effort. Further work is needed, therefore, to assess whether these sampling metrics fully capture all aspects of sampling heterogeneity, and whether specimen completeness is influenced to a greater extent by systematic or stochastic processes. The generation and testing of alternative sampling metrics (e.g., rock exposure area, or formation counts that include evidence of footprints and stone tools—a better proxy for supposed total sampling effort) will bare directly on these issues and be a fruitful avenue of research into hominin fossil record quality.

Climate. Increased aridification and orbital-scale climatic instability are the causal agents that palaeoanthropologists most commonly invoke to explain patterns of speciation, extinction, dispersal, and adaptation in hominin evolution (e.g., Dart, 1925; Foley, 1994; Kimbel, 1995; Vrba, 1995; Potts, 1996, 1998; Grove, 2014). Increased and more variable dust deposition has been linked to habitat re-modelling and the appearance of arid-adapted mammals (including hominins) at 3.6, 2.7–2.5, and 1.9 Ma (Maslin *et al.*, 2014). Tectonic activity and orbitally-tuned variation in insolation are also known to have made the East African Rift System (EARS) sensitive to precessional or half-precessional pulses of wetter and more variable climate, resulting in the periodic appearance of amplifier lakes throughout the EARS (Maslin & Trauth, 2009; Trauth *et al.*, 2010). However, climatic and tectonic factors are known to effect temporal trends in specimen completeness. For example, large-scale fluctuations in erosion and, consequently, sediment deposition ($^{87}Sr/^{86}Sr$) best explain the completeness of the eutherian mammal fossil record (Davies *et al.*, 2017). Similarly, variation in the degree of continental flooding has been found to control, in some fashion, the completeness of the dinosaur fossil record (Mannion & Upchurch, 2010), however, the same pattern is not consistently found in other terrestrial groups (e.g., Brocklehurst *et al.*, 2012; Dean *et al.*, 2016).

In regard to hominins, it is unclear whether temporal trends in fossil record completeness are controlled, either wholly or in part, by large-scale climatic and/or tectonic factors. Maslin and co-workers (Maslin & Trauth, 2009; Trauth *et al.*, 2010; Shultz & Maslin, 2013) failed to consider the effect of humid conditions on the likelihood of terrestrial vertebrate remains entering the fossil record in their *pulse climate variability hypothesis*. Instead, fluctuations in Rift Valley climate are interpreted in a purely evolutionary framework, with any potential influence on the deposition of fossil-bearing formations unclear. For example, one might expect that higher aridity and the disappearance of lakes would reduce the rate of fluvio-lacustrine sediment deposition and terrestrial remains would be less likely to reach aquatic environments and fossilise (low completeness). In contrast, higher humidity and lake presence would be expected to increase sediment deposition and the likelihood of fossilisation (high completeness). However, aridity appears to have no discernible bearing on fossil record completeness. In each statistical test, aridity correlates negatively with completeness metrics (= increased aridity corresponds with low specimen completeness), however, it is both weak and non-significant. Likewise, there is no ap-

parent correspondence between fluctuations in any completeness metric and the lake variability index. This might be because orbital-scale variation in wet-dry cycles and lake levels are of insufficient resolution to identify fine-scale patterns when using quarter-million-year time bins, or that global climate records are not representative of regional or local climatic conditions in the EARS (Shultz & Maslin, 2013). However, the simplest explanation, at present, is that aridification played little role in shaping hominin fossil record completeness.

Habitat and depositional environment. Mean completeness scores in the late Miocene are startlingly low relative to the Plio-Pleistocene (Fig. 4.11). One possible explanation for the relative incompleteness of the late Miocene may be that the earliest hominins occupied a distinct habitat with a lower probability of fossilisation. Current interpretations of the palaeoenvironment of *Sahelanthropus tchadensis* suggests a mosaic of open grassland, extensive aquatic environments, gallery forest, and a near-by desert (Vignaud *et al.*, 2002)—akin to the modern Okavango Delta (Brunet *et al.*, 2005). Similarly, *Orrorin tugenensis* is associated with open country, with denser trees (based on the presence of Colobinae and Galagidae) and a lake margin near-by (Pickford & Senut, 2001). The presence of riverside forests and floodplain grasslands indicate that *Ardipithecus kadabba* is also associated with a wet and woody habitat (WoldeGabriel *et al.*, 2001; Su *et al.*, 2009).

Together, these interpretations suggest that the earliest hominins exploited a habitat mosaic, with a reliance on trees and permanent water sources much like later hominins. If the earliest hominins were primarily forest-dwelling and occupied a similar habitat to modern hominoids, one would expect them to be less complete than if they were savannah-dwelling. The moist acidic soils typical of tropical forest habitats are rich in humic acids and are, therefore, much more conducive to decomposition than fossilisation (Kingston, 2007). Compared with bone in dry, arid conditions, bone in humid, tropical forests weathers more slowly, but is soft and spongey from the action of bio-eroders such as fungi (Behrensmeyer *et al.*, 2000) and, therefore, rarely survives fossilisation. Moreover, if fossilised, erosion and rock exposure are minimal in tropical forests and so the likelihood of finding fossil-bearing rock is lower. The fact that the earliest purported hominins are found in habitats with a significant open habitat component, and not in the tropical forests typical of modern hominoids (Fleagle, 2013), suggests that a distinct habitat preference cannot, at present, explain their incompleteness.

4.4.4 Implications for the origin of Hominini

The striking poor quality of the Mio-Pliocene fossil record of early African hominin evolution has major implications for the origin of hominins, and our ability to distinguish the earliest hominin from a stem panin, a stem hominine, or an unrelated hominid for that matter (Maxwell *et al.*, 2018, *in preparation*).

Molecular estimates for the hominin-panin last common ancestor generally agree on a late Miocene to early Pliocene divergence date of 8.0–4.0 Ma (e.g., Glazko & Nei, 2003; Patterson *et al.*, 2006; Steiper & Young, 2006, 2009; Perelman *et al.*, 2011; Stone *et al.*, 2010; Langergraber *et al.*, 2012; Prüfer *et al.*, 2012; Scally *et al.*, 2012; Springer *et al.*, 2012; Pozzi *et al.*, 2014), an estimate also supported by statistical analyses of the primate fossil record that explicitly incorporate gaps in fossil sampling during divergence date estimation (e.g., Wilkinson *et al.*, 2011). Within this four-million-year range, the true speciation event is not yet clear, in part due to it being highly dependent on the precise mutation rate (which covaries with, e.g., metabolic rate, generation time, and body size) and choice of fossil calibrations (Scally

& Durbin, 2012).

In spite of the general agreement in molecular estimates, support for claims of hominin status in this time window are hindered by (1) uncertainty in whether the traditionally recognised evolutionary hallmarks of Hominini (canine reduction, loss of the C/P_3 honing complex, anteriorly positioned foramen magnum, and adaptations for bipedalism) are actually unique to hominins (Wood & Harrison, 2011; Wood & Boyle, 2017), and (2) the small amount of anatomical overlap in the earliest purported hominins: *Sahelanthropus*, *Orrorin*, and *Ardipithecus* (Wood & Boyle, 2016). This issue is complicated further by a historical tendency in palaeoanthropology to use extant hominoids (primarily *Gorilla* and *Pan*) as the only outgroups in phylogenetic analyses and not middle to late Miocene hominoids (Andrews & Harrison, 2005). *Gorilla* and *Pan* have undergone millions of years of independent evolutionary change along divergent paths (e.g., Young *et al.*, 2015; Fleagle *et al.*, 2016), and so "[r]ooting such an ingroup using an outgroup composed of modern taxa alone may mislead" phylogenetic inference (Smith, 1994:59). Smith suggests that "using taxa that lie closer to the basal node (fossils) reduces the chances of mistaking homoplasy for synapomorphy and improves the likelihood of arriving at the correct topology" (Smith, 1994:59). For example, hominin monophyly is consistently supported when the outgroup is entirely made up of modern hominoids (Strait *et al.*, 2007). However, a recent hierarchical Bayesian tip-dating analysis including 13 extinct hominoids and 100 non-coding genomic loci placed *Sahelanthropus* as the sister-taxon of a clade containing *Gorilla* + *Pan* + *Homo*, found only moderate support for *Ardipithecus ramidus* being a hominin, and produced a tighter divergence date of 4.4–5.5 Ma (Matzke *et al.*, 2013).

Claims that hominins genuinely occurred during the late Miocene-early Pliocene can only be supported if the early African hominin fossil record is thoroughly sampled—both stratigraphically and geographically—and in terms of specimen and taxon completeness. However, the results presented here show that the entire late Miocene has a CCM2 score less than 28% (mean = 9%), a SCM2 score less than 14% (mean = 7%), and there are multiple time bins (ca., 6.6 Ma and 4.9 Ma) which are zero percent complete because no identifiable hominin fossils are known.

The issue of specimen completeness is of critical importance to hominin origins because the number of phylogenetically informative characters that can be used to distinguish a hominin from a panin is inversely proportional to the proximity of an extinct taxon to modern humans (Andrews & Harrison, 2005). That is, the closer one approaches the hominin-panin last common ancestor, the more subtle the morphological differences, and the more difficult it becomes to recognise a hominin in the fossil record (Wood & Harrison, 2011). Therefore, "[w]ith fewer and fewer characters that are phylogenetically relevant to human origins, better and better material will need to become available" in order to confidently distinguish a stem hominin from a stem panin, or unrelated hominine (Andrews & Harrison, 2005:105). Irrespective of whether specimen completeness is assessed using metrics that specifically quantify phylogenetic information content or anatomical representation, the early African hominin fossil record is poorest during the time period most pertinent to hominin origins (7.0–4.6 Ma). Furthermore, these findings emerge whether one looks at mean or maximum completeness scores through geological time (Figs. 4.3 and 4.4) or mean time bin scores in each epoch (Fig. 4.11). The incompleteness of this time window, combined with (1) our poor understanding of what characteristics reliably define Hominini and (2) knowledge that the number of autapomorphies recovered is related to the completeness of our

character matrices, suggests that we should exercise caution when making claims of hominin status based on highly incomplete fossil material.

The earliest undisputed member of Hominini (*Australopithecus anamensis*) occurs at 4.2 Ma (Leakey *et al.*, 1998), while the earliest purported member (*Sahelanthropus tchadensis*) occurs at 7.0 Ma (Brunet *et al.*, 2002). Both fall within molecular divergence date estimates, however, *Sahelanthropus* is substantially older than most molecular estimates, including those based on complete genome sequence data (Prüfer *et al.*, 2012; Scally *et al.*, 2012). There are three possible reasons for a discrepancy in molecular and morphological (fossil) estimates of hominin origins (Benton, 2003):

1. ***Ancestors are cryptic or do not display all synapomorphies.*** If the earliest hominins could not be reliably distinguished from their hominine ancestors based purely on their morphology, this could indicate that molecular and morphological evolution were somehow de-coupled in early hominins. Under such a scenario, the molecular separation of hominins and panins could have occurred hundreds of thousands (potentially millions) of years before morphological differentiation. It has been suggested that primates may be particularly prone to such de-coupling if they accrue unique traits slowly or retain traits for long periods of time (Steiper & Young, 2008:181). Recently, Diogo *et al.* (2013) found that their rates of soft tissue (muscle) evolution differed from the molecular substitution rates obtained by Perelman *et al.* (2011) on multiple branches of the primate tree. However, within hominoids, molecular substitution rates and rates of soft tissue evolution are strongly coupled (Diogo *et al.*, 2013). Further work is required to quantify whether the same result is found in the hard tissue of primates.

2. ***Ancestors did not fossilise.*** Though the earliest hominins were probably a minor component of terrestrial ecosystems with low population densities and long life histories (*K*-selection), this explanation is unlikely in the extreme. Late Miocene hominoid-bearing deposits are known, for example, in Kenya (e.g., Namurungule Fm.), Ethiopia (e.g., Chorora Fm.), Niger (Pickford *et al.*, 2008), and Uganda (DeSilva *et al.*, 2006). Equally, taphonomically comparable cercopithecoid-bearing deposits are found at a wide range of sites, including the Nkondo Fm. (Uganda), Mpesida Beds and Lemudong'o (Kenya), Ongoliba Beds (Democratic Republic of Congo), and Wadi Natron (Egypt) (Werdelin, 2010). However, while hominoid-bearing deposits are found throughout the African Neogene, they represent a small proportion of the total number of mammal-bearing (= potentially primate-bearing) deposits available to sample (Werdelin, 2010). The incompleteness of the late Miocene fossil record is, therefore, unlikely to be a consequence of the earliest hominins being somehow more restricted in their ability to fossilise than other contemporary hominines. Instead, it is much more likely to relate to their rarity and the small number of opportunities to sample hominin fossils (see below).

3. ***Ancestors occurred in hitherto unexplored geographical regions.*** The early African hominin fossil record is very patchy in terms of spatial sampling, with the majority (more than 90%) of hominin-bearing localities found in the Cradle of Humankind (South Africa) and EARS. The earliest purported hominins are found in Chad, Ethiopia, and Kenya. The total area of a convex hull enclosing these late Miocene hominin-bearing localities is 1,380,400 km^2 (Fig. 3.10)—a meagre 5% of the total area of the African continent (30,244,000 km^2). However, this estimate

assumes comparable sampling effort in those countries between Chad and the EARS that are encompassed by the convex hull (namely parts of the Democratic Republic of Congo, Rwanda, South Sudan, and Uganda). Excluding the better-sampled deposits (e.g., Nkondo Fm., Warwire Fm., and Kaiso Village Fm.) of the Western Rift Valley, Uganda (Pickford *et al.*, 1993), homogeneous spatial sampling is highly unlikely. Therefore, this estimate is most certainly an upper limit. In reality, the actual value will be considerably lower. Hill (2007:xviii), for example, estimated that our current knowledge of ape and hominin evolution samples an area of approximately 0.1% of the African continent, an order of magnitude smaller than that estimated in the current study. It should be noted that the earliest purported hominin (*Sahelanthropus*) is found in Chad and not in the Cradle of Humankind or EARS, pointing to the need for greater exploration in hitherto un-sampled geographical regions (western and central Africa) and under-exploited geological formations.

In each of the three reasons outlined, a discrepancy in the estimated time of origin arises from molecular estimates being older than morphological estimates. However, in hominins, the opposite is true: morphological estimates are actually older than molecular estimates. Unlike green plants, angiosperms, and modern orders of birds and mammals (Benton, 2003), there is no gap in the hominin fossil record resulting from molecular estimates being significantly older than the oldest fossil; instead, there are purported hominins which are older than (or rather at the upper limits of) their estimated time of origin. This suggests that either the time of origin of hominins based on molecular sequence data is inaccurate, or that some non-hominin (ape) genera have been shoehorned into Hominini without due consideration (or both).

In spite of the discrepancy, current convention in palaeoanthropology is that *Sahelanthropus*, *Orrorin*, and *Ardipithecus* are hominins, and so the question remains of why specimen completeness in the late Miocene and earliest Pliocene is so poor. In this regard, the problem of non-uniform spatial sampling is not an isolated issue relating to the reliability of molecular and morphological divergence date estimates but a considerable source of bias in the hominin fossil record (Maxwell *et al.*, 2018). Brocklehurst *et al.* (2012) proposed three reasons for the differences in completeness of neornithine and non-neornithine birds, and these relate directly to the differences in specimen completeness of the purported hominins of the late Miocene and the undisputed hominins of the Plio-Pleistocene:

(i) ***Non-uniform fossil sampling.*** The early African hominin fossil record is poorly sampled resulting in low phylogenetic and skeletal completeness. Only 3 HBF are known from the late Miocene, compared to 14 in the Pliocene and 16 in the early Pleistocene (Maxwell *et al.*, 2018). Of those collections gathered in the Plio-Pleistocene, a substantial proportion were amassed at bonanza sites. Indeed, the modal number of HBC per formation per time bin in the late Miocene is one: in most time bins each formation yields only 1 collection. These data show that the tendency of palaeoanthropologists to preferentially return to abundant sites in the Plio-Pleistocene leads to a higher number of collections being gathered at each formation (the modal number of HBC per formation per time bin is 6 for the Pliocene and 7 for the early Pleistocene). Bonanza sites are the primary mechanism by which high-scoring postcranial elements enter our collections, and time bins without bonanza sites are largely composed of low-scoring cranio-dental remains. This

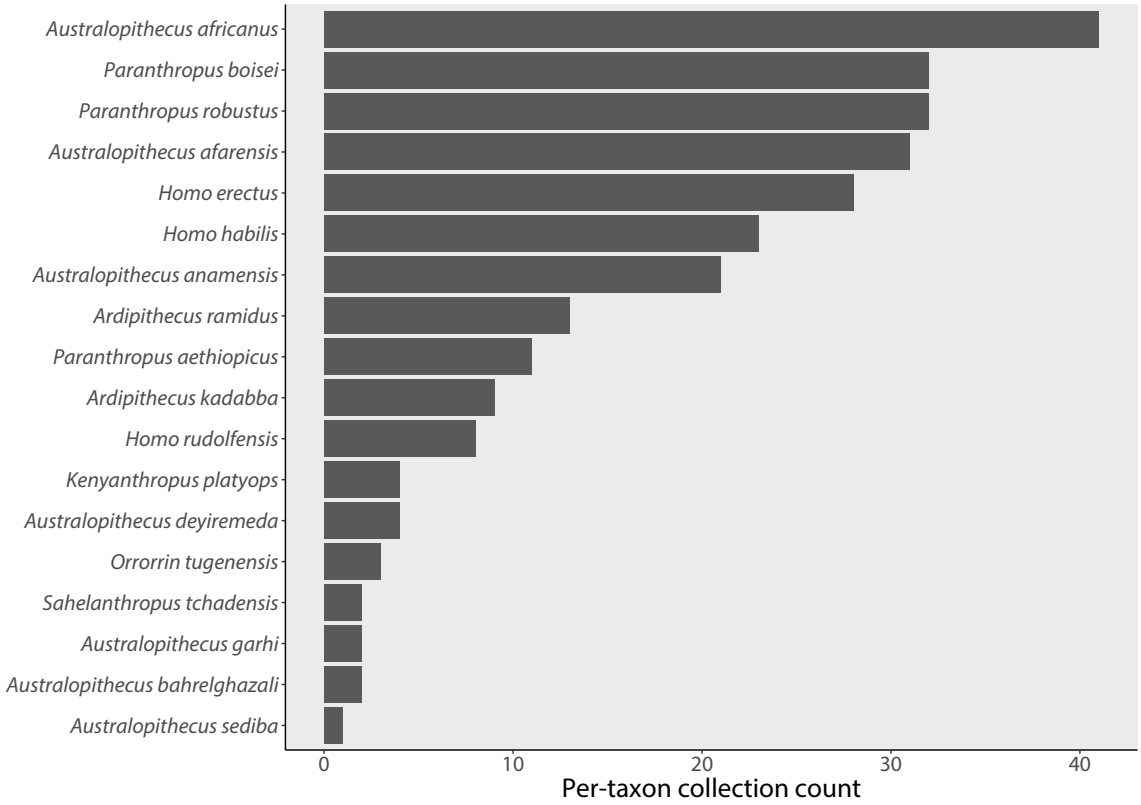

Figure 4.13: Per-taxon collection counts for each early African hominin. The number of collections is defined as the number of unique years that have produced a hominin fossil per formation and is an estimate of palaeoanthropological collection effort.

suggests that non-uniform fossil sampling and, in particular, a lack of productive deposits in the late Miocene and early Pliocene (7.0–4.6 Ma) do influence specimen completeness. Moreover, the amount of collection effort per taxon differs markedly (Figs. 4.13 and 4.14), with *Sahelanthropus* and *Orrorin* among the poorest sampled of all early hominins, in large part due to their having the shortest research history (Fig. 4.15).

(ii) ***Fragmentary fossils lead to ambiguity of identification.*** It is possible that high-quality hominoid fossils might include enough anatomical detail for them to be confidently assigned to non-hominin lineages. In contrast, low-quality hominoid fossils assigned to Hominini might actually belong to non-hominins, but their incompleteness leads to greater ambiguity in their identification. If, for example, some closely related non-hominins acquired character states that also occur in true Hominini (e.g., canine reduction in *Oreopithecus*, *Ouranopithecus*, and *Gigantopithecus*), then the homoplastic nature of these characters might only be detected when fossils are of sufficient quality to permit a thorough comparison.

(iii) ***The effect of population size on fossil abundance.*** It might be possible that the earliest hominins were genuinely less abundant than other mammal groups. A smaller population size would reduce the amount of skeletal material ultimately entering the fossil record and, therefore, the number of opportunities to sample hominin fossils. However, genetic sequence data indicate an effec-

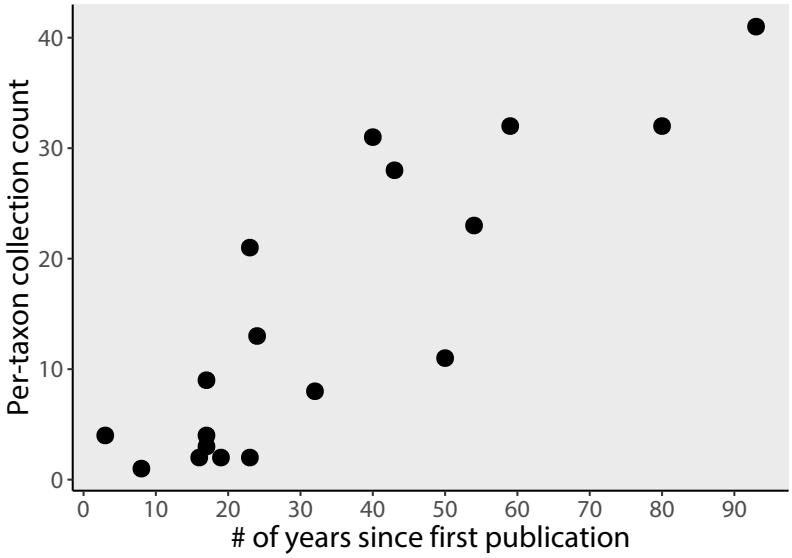

Figure 4.14: Per-taxon collection counts against the number of years since first publication. The significance of research history suggests that the number of collections amassed is directly proportional to the amount of research effort relating to fossil hominins: those hominins with the longest research history are also known from the highest number of collections ($R^2 = 0.73$, $r = 0.862$, $p < 0.001$). Such historiographic effects are discussed in further detail and in relation to cetacean diversity in Uhen & Pyenson (2007).

tive population size (N_e) of 50,000–100,000 for the *Pan-Homo* last common ancestor (Chen & Li, 2001; Satta *et al.*, 2004; Langergraber *et al.*, 2012; *contra* Yang, 2002), 5–10 larger than that estimated for modern humans. Population size is, therefore, unlikely to have had a major influence on early African hominin fossil record quality.

In short, fossil evidence for the occurrence of Hominini during the late Miocene and early Pliocene (7.0–4.2 Ma), as predicted by molecular clock studies, is limited by the fragmentary nature of the hominin fossil record and the incompleteness of the hominoid fossil record more generally. The early African hominin fossil record is particularly fragmentary because of a dearth of productive sites and pronounced spatial heterogeneity in sampling intensity. Large geographical regions remain un-sampled and as such produce no identifiable hominin fossils. Similarly, there are currently no known gorillin-bearing deposits, and no panin-bearing deposits older than 600 thousand years (McBrearty & Jablonski, 2005). The discovery of new sites in Mio-Pliocene age deposits should go some way to improve fossil record completeness. In addition, a wider study of the completeness of the hominoid, catarrhine, or primate fossil record would also place this result in its necessary context and determine whether these results are typical for a late Neogene primate group. Ultimately, a substantial increase in specimen quality is required before key phylogenetic and palaeobiogeographic hypotheses about the origin of hominins can be reliably tested (Maxwell *et al.*, 2018, *in preparation*).

4.5 Conclusion

The completeness of the early African hominin fossil record fluctuated substantially through geological time, with distinct patterns depending on whether specimen completeness is assessed using phyloge-

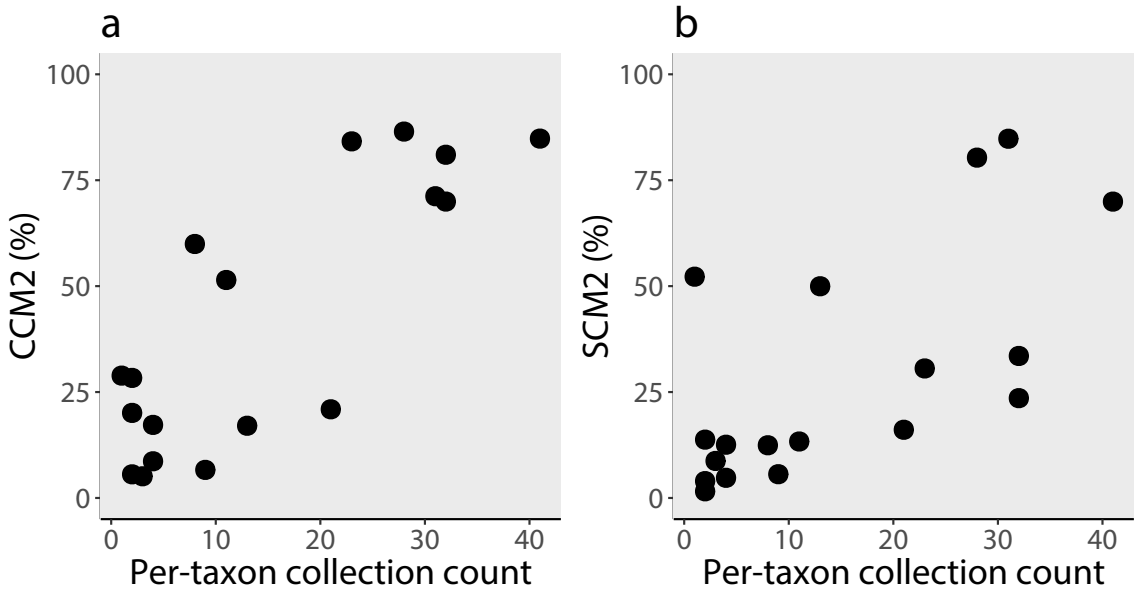

Figure 4.15: Per-taxon completeness scores against per-taxon collection counts. (**a**) CCM2 against per-taxon collection count (R^2 = 0.66, r = 0.654, p = 0.003). (**b**) SCM2 against per-taxon collection count (R^2 = 0.41, r = 0.598, p = 0.008). Those hominins for which the SCM2 is higher than one would expect given their collection count include *Ardipithecus ramidus*, *Australopithecus afarensis*, *Australopithecus sediba*, and *Homo erectus*, all of which are known from at least one rich rock unit (bonanza site).

netic character richness or the amount of the skeleton known. In general, CCM2 is low (<30%) during the first half of hominin evolution before a sudden rise ca., 3.6 Ma, after which completeness remains high (ca., 80%) for the rest of the Plio-Pleistocene. This bimodal pattern (the fossil record is known from either few, fragmentary skulls or many, well-preserved skulls) is not seen in the SCM2. The SCM2 is generally low (<20%) throughout much of hominin evolution, reflecting the over-representation of low-scoring cranio-dental material and under-representation of high-scoring postcranial material, with three major departures. First, at 4.4 Ma (Sagantole Fm.). Second, from 3.4–2.8 Ma (Hadar Fm. and Sterkfontein). Third, from 1.6–1.0 Ma (Nachukui Fm. and Swartkrans). These peaks coincide with bonanza sites (Raup, 1977), which are defined by rich formations that have produced a large number of collections as a result of palaeoanthropologists preferentially returning to conduct field study. Character and skeletal completeness generally capture a similar pattern of completeness but show clear differences because of the CCM being 100% cranio-dental. These findings reiterate that when calculating the CCM, one must consider the intended use of the original character matrix, and the specific question it sought to answer. Due to the long-standing view in palaeoanthropology that postcranial characters are riddled with homoplasy (despite evidence indicating that the prevalence of homoplasy in postcranial characters is no higher than in cranio-dental characters; Mounce *et al.*, 2016), they are currently universally excluded (Strait *et al.*, 2007; Pugh, 2016). Because of the historical tendency to exclude postcranial characters, a large amount of information relevant to hominin evolution is missing when cranio-dental-only matrices are used to quantify fossil record completeness. Whole-skeleton-based metrics are therefore more meaningful.

There is no link between mean completeness metrics and raw taxic diversity, suggesting speci-

men completeness does not control the ability of palaeoanthropologists to recognise species. However, CCM1-2_{max} do show significant positive correlations with raw taxic diversity, suggesting the presence of a single highly complete specimen may make the taxonomic identification of other, fragmentary specimens easier, increasing apparent species richness. Time bins with low data quality should be taken into account when examining apparent hominin diversity. Sampling metrics consistently provide the best explanation for variation in specimen completeness: CCM2 is most significantly controlled by collection effort, while SCM2 is controlled by the presence of bonanza sites. Together, CCM1-2 and SCM1-2 (mean and maximum variants) demonstrate that anthropogenic factors are the dominant control of fossil record completeness.

Finally, the hominin fossil record is most incomplete during the period most pertinent to their origin (7.0–4.4 Ma), which is problematic because (1) the number of autapomorphies in modern hominoids and fossil hominins (*sensu* Strait & Grine, 2004) is proportional to specimen completeness, and (2) the number of phylogenetically informative characters that can be used to distinguish a hominin from a panin (or closely related African ape) is inversely proportional to the proximity of an extinct taxon to modern humans. The current study suggests that the reasons for this incompleteness are (1) a lack or under-sampling of hominin-bearing deposits during the late Miocene and early Pliocene, and (2) highly non-uniform spatial sampling. Indeed, no more than 3% of Africa has been sampled in any one time bin, with the poorest coverage during this interval. If the fossil record is to increase in completeness by any substantial margin, palaeoanthropologists must radically alter their research strategy, increase funding for exploration, and search for new fossiliferous rock, in addition to returning to the well-known, rich deposits.

Summary

1. Character completeness is low (< 30%) during the first half of hominin evolution before a sudden rise in the middle Pliocene, after which completeness fluctuates around 80%. Skeletal completeness, on the other hand, is low (< 20%) throughout much of hominin evolution with three major departures from this pattern. These departures coincide with bonanza sites.

2. Character and skeletal completeness generally capture a similar pattern of completeness (R^2 = 53%–79%) but show clear differences because of the CCM being 100% cranial.

3. There is no link between mean completeness metrics and raw taxic diversity, suggesting specimen completeness does not control the ability of palaeoanthropologists to recognise species. However, maximum character completeness scores do show significant positive correlations with taxic diversity, suggesting that the presence of a single highly complete specimen may increase species richness by making the taxonomic identification of other, fragmentary specimens easier.

4. Sampling metrics consistently best explain variation in specimen completeness. Character completeness is most significantly controlled by collection effort, while skeletal completeness is controlled by the presence of bonanza sites (formations where palaeoanthropologists preferentially return each year and amass many collections).

5. Holotype specimen completeness is startlingly poor, and more recently described (post-1990) holotypes are more incomplete than holotypes described in between 1920 and 1980. One reason for this may be that there is now a better comparative framework of more complete specimens that makes it possible to identify new hominins on the basis of smaller and more fragmentary fossils. Trends in specimen completeness through geological time support this proposition.

6. The early African hominin fossil record is most incomplete during the period most pertinent to human origins (7.0–4.4 Ma). This is problematic because phylogenetic character richness co-varies with specimen completeness in modern hominoids and fossil hominins. The reasons for this incompleteness are a lack of rich bonanza sites during the Mio-Pliocene, and highly non-uniform spatial sampling.

Table 4.10: Results of the Generalised Least Squares analysis comparing the mean Character Completeness Metric 1 (CCM1) to diversity, sampling metrics, and aridity. Models 1–31 comprise all possible combinations of diversity, collections, formations, the bonanza effect, and terrigenous dust flux as explanatory predictors of CCM1. Models are ranked in order of their relative likelihood according to their Akaike weight (w_i), where the larger the value the higher the probability that the given model is the best model. R^2 is calculated using the log-likelihood (logLik) of the model in question and the logLik of the null (Nagelkerke, 1991).

Model	Parameters	df	w_i	AICc	logLik	R^2
30	C	4	0.15	61.33	−26.38	0.42
23	FB	5	0.11	62.01	−25.40	0.46
31	B	4	0.10	62.16	−26.79	0.40
22	FC	5	0.09	62.43	−25.62	0.45
20	DB	5	0.08	62.64	−25.72	0.45
19	DC	5	0.05	63.44	−26.12	0.43
26	CE	5	0.04	63.75	−26.28	0.42
16	FBE	6	0.04	63.84	−24.87	0.49
25	CB	5	0.04	63.90	−26.35	0.42
27	BE	5	0.04	64.11	−26.45	0.41
9	DFB	6	0.03	64.44	−25.17	0.47
14	FCB	6	0.03	64.59	−25.24	0.47
15	FCE	6	0.03	64.79	−25.34	0.47
13	DBE	6	0.03	64.82	−25.36	0.46
11	DCB	6	0.02	65.15	−25.52	0.46
8	DFC	6	0.02	65.34	−25.61	0.45
12	DCE	6	0.01	66.11	−26.00	0.44
29	F	4	0.01	66.44	−28.94	0.28
3	DFCB	7	0.01	66.49	−24.58	0.50
24	FE	5	0.01	66.60	−27.70	0.35
17	CBE	6	0.01	66.65	−26.27	0.42
7	FCBE	7	0.01	66.69	−24.68	0.49
5	DFBE	7	0.01	66.71	−24.69	0.49
28	D	4	0.01	67.42	−29.42	0.25
6	DCBE	7	0.01	67.49	−25.08	0.48
18	DF	5	0.01	67.71	−28.25	0.32
4	DFCE	7	0.01	68.00	−25.34	0.47
21	DE	5	0.00	68.33	−28.56	0.30
10	DFE	6	0.00	68.62	−27.26	0.37
2	DFCBE	8	0.00	68.87	−23.96	0.52
1	Null	3	0.00	71.89	−32.85	0.00
32	E	4	0.00	72.11	−31.77	0.09

D: diversity; C: collections; F: formations; B: bonanza; E: East African aridity.

Table 4.11: Results of the Generalised Least Squares analysis comparing the mean Character Completeness Metric 2 (CCM2) to diversity, sampling metrics, and aridity. Models 1–31 comprise all possible combinations of diversity, collections, formations, the bonanza effect, and terrigenous dust flux as explanatory predictors of CCM2. Models are ranked in order of their relative likelihood according to their Akaike weight (w_i), where the larger the value the higher the probability that the given model is the best model. R^2 is calculated using the log-likelihood (logLik) of the model in question and the logLik of the null (Nagelkerke, 1991).

Model	Parameters	df	w_i	AICc	logLik	R^2
30	C	4	0.21	59.16	−25.30	0.53
22	FC	5	0.15	59.83	−24.31	0.57
19	DC	5	0.10	60.69	−24.75	0.55
25	CB	5	0.06	61.67	−25.24	0.54
20	DB	5	0.06	61.76	−25.28	0.53
26	CE	5	0.06	61.79	−25.30	0.53
23	FB	5	0.06	61.79	−25.30	0.53
15	FCE	6	0.04	62.67	−24.28	0.57
8	DFC	6	0.04	62.71	−24.30	0.57
14	FCB	6	0.04	62.72	−24.31	0.57
31	B	4	0.03	63.18	−27.31	0.45
11	DCB	6	0.02	63.57	−24.73	0.56
12	DCE	6	0.02	63.59	−24.74	0.56
9	DFB	6	0.02	63.70	−24.80	0.55
16	FBE	6	0.02	63.83	−24.86	0.55
13	DBE	6	0.02	64.39	−25.14	0.54
17	CBE	6	0.01	64.54	−25.22	0.54
27	BE	5	0.01	65.52	−27.16	0.46
7	FCBE	7	0.01	65.84	−24.25	0.57
4	DFCE	7	0.01	65.90	−24.28	0.57
3	DFCB	7	0.01	65.91	−24.29	0.57
5	DFBE	7	0.01	66.43	−24.55	0.56
6	DCBE	7	0.00	66.80	−24.73	0.56
29	F	4	0.00	67.75	−29.59	0.33
28	D	4	0.00	68.16	−29.80	0.32
24	FE	5	0.00	68.16	−28.48	0.39
18	DF	5	0.00	68.73	−28.77	0.38
2	DFCBE	8	0.00	69.42	−24.24	0.57
21	DE	5	0.00	69.75	−29.28	0.35
10	DFE	6	0.00	70.24	−28.07	0.41
1	Null	3	0.00	75.10	−34.46	0.00
32	E	4	0.00	75.97	−33.70	0.06

D: diversity; C: collections; F: formations; B: bonanza; E: East African aridity.

Table 4.12: Results of the Generalised Least Squares analysis comparing the mean Skeletal Completeness Metric 1 (SCM1) to diversity, sampling metrics, and aridity. Models 1–31 comprise all possible combinations of diversity, collections, formations, the bonanza effect, and terrigenous dust flux as explanatory predictors of SCM1. Models are ranked in order of their relative likelihood according to their Akaike weight (w_i), where the larger the value the higher the probability that the given model is the best model. R^2 is calculated using the log-likelihood (logLik) of the model in question and the logLik of the null (Nagelkerke, 1991).

Model	Parameters	df	w_i	AICc	logLik	R^2
31	B	4	0.19	58.39	−24.91	0.39
20	DB	5	0.15	58.80	−23.80	0.44
11	DCB	6	0.14	59.03	−22.46	0.50
9	DFB	6	0.09	59.95	−22.92	0.48
25	CB	5	0.05	60.88	−24.84	0.39
23	FB	5	0.05	60.98	−24.89	0.39
27	BE	5	0.05	61.01	−24.91	0.39
3	DFCB	7	0.05	61.16	−21.91	0.52
30	C	4	0.04	61.57	−26.50	0.30
13	DBE	6	0.04	61.66	−23.78	0.44
6	DCBE	7	0.03	62.26	−22.46	0.50
5	DFBE	7	0.02	63.05	−22.86	0.48
22	FC	5	0.01	63.66	−26.23	0.32
17	CBE	6	0.01	63.76	−24.83	0.39
14	FCB	6	0.01	63.78	−24.84	0.39
16	FBE	6	0.01	63.88	−24.89	0.39
19	DC	5	0.01	64.00	−26.40	0.31
26	CE	5	0.01	64.07	−26.44	0.30
2	DFCBE	8	0.01	64.76	−21.91	0.52
8	DFC	6	0.01	65.31	−25.60	0.35
28	D	4	0.00	65.80	−28.61	0.17
15	FCE	6	0.00	66.19	−26.04	0.33
12	DCE	6	0.00	66.82	−26.36	0.31
7	FCBE	7	0.00	66.98	−24.82	0.39
4	DFCE	7	0.00	68.06	−25.37	0.36
21	DE	5	0.00	68.24	−28.52	0.17
18	DF	5	0.00	68.28	−28.54	0.17
29	F	4	0.00	70.77	−31.10	-0.03
10	DFE	6	0.00	71.00	−28.45	0.18
1	Null	3	0.00	72.51	−33.16	-0.22
24	FE	5	0.00	72.80	−30.80	0.00
32	E	4	0.00	74.17	−32.80	-0.18

D: diversity; C: collections; F: formations; B: bonanza; E: East African aridity.

Table 4.13: Results of the Generalised Least Squares analysis comparing the mean Skeletal Completeness Metric 2 (SCM2) to diversity, sampling metrics, and aridity. Models 1–31 comprise all possible combinations of diversity, collections, formations, the bonanza effect, and terrigenous dust flux as explanatory predictors of SCM2. Models are ranked in order of their relative likelihood according to their Akaike weight (w_i), where the larger the value the higher the probability that the given model is the best model. R^2 is calculated using the log-likelihood (logLik) of the model in question and the logLik of the null (Nagelkerke, 1991).

Model	Parameters	df	w_i	AICc	logLik	R^2
31	B	4	0.20	57.08	−24.25	0.59
20	DB	5	0.17	57.48	−23.14	0.62
25	CB	5	0.11	58.35	−23.57	0.61
23	FB	5	0.09	58.81	−23.81	0.60
30	C	4	0.06	59.51	−25.47	0.54
27	BE	5	0.06	59.60	−24.20	0.59
13	DBE	6	0.04	60.13	−23.01	0.63
9	DFB	6	0.04	60.20	−23.05	0.63
11	DCB	6	0.04	60.38	−23.14	0.62
17	CBE	6	0.03	60.71	−23.30	0.62
14	FCB	6	0.03	61.21	−23.55	0.61
26	CE	5	0.02	61.26	−25.03	0.56
16	FBE	6	0.02	61.62	−23.76	0.60
22	FC	5	0.02	61.84	−25.32	0.55
19	DC	5	0.02	62.11	−25.45	0.54
5	DFBE	7	0.01	63.09	−22.88	0.63
15	FCE	6	0.01	63.33	−24.61	0.57
6	DCBE	7	0.01	63.34	−23.00	0.63
3	DFCB	7	0.01	63.43	−23.05	0.63
7	FCBE	7	0.01	63.94	−23.30	0.62
12	DCE	6	0.01	64.07	−24.98	0.56
8	DFC	6	0.00	64.68	−25.29	0.55
4	DFCE	7	0.00	66.37	−24.52	0.58
2	DFCBE	8	0.00	66.63	−22.84	0.63
28	D	4	0.00	68.51	−29.97	0.34
21	DE	5	0.00	71.03	−29.91	0.34
18	DF	5	0.00	71.12	−29.96	0.34
29	F	4	0.00	72.14	−31.78	0.23
10	DFE	6	0.00	73.91	−29.90	0.34
24	FE	5	0.00	74.29	−31.54	0.24
1	Null	3	0.00	75.94	−34.88	0.00
32	E	4	0.00	77.74	−34.58	0.02

D: diversity; C: collections; F: formations; B: bonanza; E: East African aridity.

Table 4.14: Results of the Generalised Least Squares analysis comparing the maximum Character Completeness Metric 1 to diversity, sampling metrics, and aridity. Models 1–31 comprise all possible combinations of diversity, collections, formations, the bonanza effect, and terrigenous dust flux as explanatory predictors of $CCM1_{max}$. Models are ranked in order of their relative likelihood according to their Akaike weight (w_i), where the larger the value the higher the probability that the given model is the best model. R^2 is calculated using the log-likelihood (logLik) of the model in question and the logLik of the null (Nagelkerke, 1991).

Model	Parameters	df	w_i	AICc	logLik	R^2
20	DB	5	0.16	58.16	−23.48	0.56
30	C	4	0.09	59.29	−25.36	0.49
23	FB	5	0.08	59.57	−24.19	0.54
22	FC	5	0.07	59.72	−24.26	0.53
19	DC	5	0.07	59.86	−24.33	0.53
13	DBE	6	0.07	59.90	−22.90	0.58
9	DFB	6	0.06	60.00	−22.95	0.58
11	DCB	6	0.04	60.87	−23.38	0.57
16	FBE	6	0.04	60.99	−23.44	0.56
31	B	4	0.03	61.43	−26.43	0.44
26	CE	5	0.03	61.53	−25.16	0.50
15	FCE	6	0.03	61.74	−23.82	0.55
25	CB	5	0.03	61.80	−25.30	0.49
5	DFBE	7	0.02	61.87	−22.27	0.60
8	DFC	6	0.02	62.09	−23.99	0.54
12	DCE	6	0.02	62.18	−24.04	0.54
3	DFCB	7	0.02	62.39	−22.53	0.60
14	FCB	6	0.02	62.46	−24.18	0.54
6	DCBE	7	0.02	62.75	−22.71	0.59
27	BE	5	0.01	63.10	−25.95	0.46
28	D	4	0.01	63.51	−27.47	0.39
21	DE	5	0.01	63.78	−26.29	0.45
18	DF	5	0.01	63.88	−26.34	0.45
7	FCBE	7	0.01	64.22	−23.44	0.56
2	DFCBE	8	0.01	64.23	−21.64	0.63
10	DFE	6	0.01	64.23	−25.06	0.50
4	DFCE	7	0.01	64.41	−23.54	0.56
17	CBE	6	0.01	64.42	−25.16	0.50
24	FE	5	0.01	64.71	−26.75	0.43
29	F	4	0.00	65.14	−28.29	0.35
32	E	4	0.00	72.73	−32.08	0.11
1	Null	3	0.00	73.01	−33.41	0.00

D: diversity; C: collections; F: formations; B: bonanza; E: East African aridity.

Table 4.15: Results of the Generalised Least Squares analysis comparing the maximum Character Completeness Metric 2 to diversity, sampling metrics, and aridity. Models 1–31 comprise all possible combinations of diversity, collections, formations, the bonanza effect, and terrigenous dust flux as explanatory predictors of $CCM2_{max}$. Models are ranked in order of their relative likelihood according to their Akaike weight (w_i), where the larger the value the higher the probability that the given model is the best model. R^2 is calculated using the log-likelihood (logLik) of the model in question and the logLik of the null (Nagelkerke, 1991).

Model	Parameters	df	w_i	AICc	logLik	R^2
30	C	4	0.15	57.71	−24.57	0.58
22	FC	5	0.15	57.72	−23.26	0.62
20	DB	5	0.13	58.00	−23.40	0.62
19	DC	5	0.12	58.15	−23.47	0.61
25	CB	5	0.04	60.21	−24.50	0.58
23	FB	5	0.04	60.27	−24.54	0.58
26	CE	5	0.04	60.28	−24.54	0.58
8	DFC	6	0.04	60.36	−23.13	0.63
9	DFB	6	0.04	60.37	−23.13	0.63
14	FCB	6	0.03	60.60	−23.25	0.62
15	FCE	6	0.03	60.62	−23.26	0.62
13	DBE	6	0.03	60.67	−23.28	0.62
11	DCB	6	0.03	60.79	−23.34	0.62
12	DCE	6	0.03	60.99	−23.44	0.62
16	FBE	6	0.01	62.42	−24.16	0.59
31	B	4	0.01	62.83	−27.13	0.48
17	CBE	6	0.01	62.92	−24.41	0.58
5	DFBE	7	0.01	63.27	−22.97	0.63
3	DFCB	7	0.01	63.38	−23.03	0.63
6	DCBE	7	0.01	63.56	−23.11	0.63
4	DFCE	7	0.01	63.59	−23.13	0.63
7	FCBE	7	0.01	63.82	−23.24	0.62
28	D	4	0.00	64.74	−28.08	0.43
27	BE	5	0.00	65.22	−27.01	0.48
18	DF	5	0.00	65.96	−27.38	0.47
21	DE	5	0.00	66.30	−27.55	0.46
2	DFCBE	8	0.00	66.40	−22.73	0.64
29	F	4	0.00	67.35	−29.39	0.37
10	DFE	6	0.00	67.68	−26.79	0.49
24	FE	5	0.00	67.87	−28.33	0.42
1	Null	3	0.00	76.03	−34.92	0.00
32	E	4	0.00	76.86	−34.14	0.06

D: diversity; C: collections; F: formations; B: bonanza; E: East African aridity.

Table 4.16: Results of the Generalised Least Squares analysis comparing the maximum Skeletal Completeness Metric 1 to diversity, sampling metrics, and aridity. Models 1–31 comprise all possible combinations of diversity, collections, formations, the bonanza effect, and terrigenous dust flux as explanatory predictors of $SCM1_{max}$. Models are ranked in order of their relative likelihood according to their Akaike weight (w_i), where the larger the value the higher the probability that the given model is the best model. R^2 is calculated using the log-likelihood (logLik) of the model in question and the logLik of the null (Nagelkerke, 1991).

Model	Parameters	df	w_i	AICc	logLik	R^2
9	DFB	6	0.23	55.53	−20.71	0.71
11	DCB	6	0.22	55.57	−20.73	0.71
20	DB	5	0.21	55.72	−22.26	0.67
3	DFCB	7	0.12	56.79	−19.73	0.73
13	DBE	6	0.05	58.57	−22.23	0.67
5	DFBE	7	0.05	58.59	−20.63	0.71
6	DCBE	7	0.04	58.77	−20.72	0.71
2	DFCBE	8	0.02	60.32	−19.69	0.73
19	DC	5	0.01	61.58	−25.19	0.58
8	DFC	6	0.01	62.10	−24.00	0.62
28	D	4	0.01	62.31	−26.87	0.51
30	C	4	0.01	62.78	−27.10	0.50
31	B	4	0.01	62.96	−27.19	0.50
25	CB	5	0.00	64.08	−26.44	0.53
18	DF	5	0.00	64.48	−26.64	0.52
12	DCE	6	0.00	64.48	−25.19	0.58
21	DE	5	0.00	64.75	−26.77	0.52
23	FB	5	0.00	64.84	−26.82	0.52
4	DFCE	7	0.00	65.20	−23.94	0.62
26	CE	5	0.00	65.33	−27.06	0.51
22	FC	5	0.00	65.37	−27.09	0.51
27	BE	5	0.00	65.58	−27.19	0.50
14	FCB	6	0.00	66.95	−26.42	0.53
17	CBE	6	0.00	66.96	−26.43	0.53
10	DFE	6	0.00	67.20	−26.55	0.53
16	FBE	6	0.00	67.75	−26.82	0.52
15	FCE	6	0.00	68.17	−27.03	0.51
7	FCBE	7	0.00	70.16	−26.41	0.53
29	F	4	0.00	73.52	−32.47	0.23
24	FE	5	0.00	75.40	−32.10	0.25
1	Null	3	0.00	77.25	−35.54	0.00
32	E	4	0.00	78.88	−35.16	0.03

D: diversity; C: collections; F: formations; B: bonanza; E: East African aridity.

Table 4.17: Results of the Generalised Least Squares analysis comparing the maximum Skeletal Completeness Metric 2 to diversity, sampling metrics, and aridity. Models 1–31 comprise all possible combinations of diversity, collections, formations, the bonanza effect, and terrigenous dust flux as explanatory predictors of $SCM2_{max}$. Models are ranked in order of their relative likelihood according to their Akaike weight (w_i), where the larger the value the higher the probability that the given model is the best model. R^2 is calculated using the log-likelihood (logLik) of the model in question and the logLik of the null (Nagelkerke, 1991).

Model	Parameters	df	w_i	AICc	logLik	R^2
20	DB	5	0.32	54.71	−21.76	0.71
9	DFB	6	0.13	56.58	−21.24	0.72
11	DCB	6	0.12	56.71	−21.30	0.72
13	DBE	6	0.08	57.38	−21.64	0.71
30	C	4	0.06	57.92	−24.68	0.62
3	DFCB	7	0.04	59.02	−20.84	0.73
19	DC	5	0.04	59.08	−23.94	0.65
26	CE	5	0.03	59.43	−24.12	0.64
5	DFBE	7	0.03	59.47	−21.07	0.72
25	CB	5	0.03	59.70	−24.25	0.64
6	DCBE	7	0.02	59.83	−21.25	0.72
22	FC	5	0.02	60.51	−24.65	0.63
8	DFC	6	0.01	61.17	−23.53	0.66
12	DCE	6	0.01	61.19	−23.54	0.66
23	FB	5	0.01	61.61	−25.20	0.61
17	CBE	6	0.01	61.77	−23.83	0.65
14	FCB	6	0.01	62.26	−24.08	0.64
15	FCE	6	0.01	62.33	−24.11	0.64
31	B	4	0.01	62.37	−26.90	0.55
2	DFCBE	8	0.01	62.45	−20.75	0.73
28	D	4	0.01	63.02	−27.22	0.54
4	DFCE	7	0.00	63.15	−22.91	0.68
16	FBE	6	0.00	64.51	−25.20	0.61
7	FCBE	7	0.00	64.90	−23.78	0.65
27	BE	5	0.00	64.97	−26.88	0.55
21	DE	5	0.00	65.56	−27.18	0.54
18	DF	5	0.00	65.58	−27.19	0.54
10	DFE	6	0.00	68.40	−27.15	0.54
29	F	4	0.00	72.75	−32.09	0.30
24	FE	5	0.00	74.77	−31.78	0.32
1	Null	3	0.00	79.07	−36.45	0.00
32	E	4	0.00	80.87	−36.15	0.02

D: diversity; C: collections; F: formations; B: bonanza; E: East African aridity.

Table 4.18: Results of the Generalised Least Squares analysis comparing skull completeness to diversity, sampling metrics, and aridity. Models 1–31 comprise all possible combinations of diversity, collections, formations, the bonanza effect, and terrigenous dust flux as explanatory predictors of SCM_{skull}. Models are ranked in order of their relative likelihood according to their Akaike weight (w_i), where the larger the value the higher the probability that the given model is the best model. R^2 is calculated using the log-likelihood (logLik) of the model in question and the logLik of the null (Nagelkerke, 1991).

Model	Parameters	df	w_i	AICc	logLik	R^2
20	DB	5	0.16	67.91	−28.35	0.55
30	C	4	0.15	68.03	−29.73	0.50
23	FB	5	0.07	69.66	−29.23	0.52
31	B	4	0.07	69.72	−30.57	0.46
19	DC	5	0.06	69.81	−29.30	0.51
11	DCB	6	0.06	69.87	−27.88	0.57
22	FC	5	0.06	69.90	−29.35	0.51
13	DBE	6	0.05	70.34	−28.12	0.56
25	CB	5	0.04	70.64	−29.72	0.50
9	DFB	6	0.04	70.64	−28.27	0.55
26	CE	5	0.04	70.66	−29.73	0.50
16	FBE	6	0.02	71.76	−28.83	0.53
27	BE	5	0.02	71.82	−30.31	0.47
6	DCBE	7	0.02	72.32	−27.49	0.58
14	FCB	6	0.02	72.41	−29.15	0.52
3	DFCB	7	0.02	72.44	−27.56	0.58
8	DFC	6	0.02	72.59	−29.24	0.52
12	DCE	6	0.01	72.70	−29.30	0.51
15	FCE	6	0.01	72.76	−29.33	0.51
5	DFBE	7	0.01	73.34	−28.00	0.56
17	CBE	6	0.01	73.52	−29.71	0.50
28	D	4	0.01	73.79	−32.61	0.36
7	FCBE	7	0.00	74.99	−28.83	0.53
21	DE	5	0.00	75.04	−31.92	0.40
2	DFCBE	8	0.00	75.09	−27.07	0.60
18	DF	5	0.00	75.31	−32.06	0.39
29	F	4	0.00	75.44	−33.44	0.31
4	DFCE	7	0.00	75.78	−29.22	0.52
24	FE	5	0.00	75.96	−32.38	0.37
10	DFE	6	0.00	76.74	−31.31	0.42
1	Null	3	0.00	82.09	−37.95	0.00
32	E	4	0.00	82.46	−36.95	0.08

D: diversity; C: collections; F: formations; B: bonanza; E: East African aridity.

Chapter 5

Conclusion

5.1 Summary

Chapter 2

In Chapter 2 I showed that early hominin diversity patterns generated using the taxic diversity estimate (TDE) and phylogenetic diversity estimate (PDE) approach bear little similarity to one another in terms of the timing and magnitude of their peaks and troughs. The TDE is near-identical to that produced by Foley (2016), displaying low-standing diversity during the late Miocene and early Pliocene (7.0–3.9 Ma), followed by a pulse-like pattern in the Plio-Pleistocene with major turnover events at 3.6 Ma, 2.4 Ma, and 1.9 Ma. It is widely accepted in human evolutionary theory that these pulsed turnover events were driven by large-scale patterns of climatic change and variability, without the fidelity of the TDE or alternative (non-climatic) hypotheses ever being tested (Smith & Wood, 2017). In contrast, each PDE—irrespective of tree topology, tree size, and phylogenetic and node age uncertainty—supports a near-linear increase in diversity from 7.0–2.4 Ma (positive net diversification), followed by a near-linear decrease from 2.4–1.0 Ma (negative net diversification). This suggests that many of the events in hominin evolution once thought to be major transitions actually form part of a continuum of steady diversification (Kimbel & Villmoare, 2016). Comparison of the TDE and PDE suggest that the application of a partial correction for sampling has a fundamental impact on the resultant diversity pattern, and that the fossil record is at present inadequate for studying diversity patterns without correction for fossil sampling. The gradual rise and fall in phylogenetically corrected diversity renders any explanation based on climate-driven pulsed turnover, or any explanation based on discrete change more generally, at best substantially weakened and at worst unsupported (Maxwell *et al.*, 2018).

The question of what drove the monotonic increase in PDE for the majority of early hominin evolution remains unclear. Fitting different mathematical models, representing a straight line and an exponential curve, to the first three-quarters of each PDE (7.0–3.4/2.4 Ma) shows that the increase is best explained by an exponential curve. A constant-rate birth-death model will produce an exponential (or log-linear) increase in diversity (Nee, 2006) identical to that recovered in each PDE. This supports an "expansionist" mode of diversification, under which unbounded diversification led to an ever-increasing diversity of early hominins. Climate-driven bursts of diversification therefore cannot be reconciled with a pattern of constant-rate speciation and extinction. The finding that early hominin di-

versification proceeded with little to no variation in speciation and extinction rate should be integrated with data showing that within-lineage morphological diversification associated with hominin speciation is infrequently adaptive and more often can be explained by neutral genetic drift (e.g., Schroeder *et al.*, 2014; Schroeder & Ackermann, 2017). There have also been few attempts to link phenotypic variation to meaningful evidence of niche differentiation, and to demonstrate a trait's adaptive advantage (Ackermann & Smith, 2007). Each of these aspects of macroevolution must be fully understood to facilitate an accurate understanding of early hominin diversity dynamics.

Chapter 3

In Chapter 3 I compared the TDE and PDE produced in Chapter 2 to a range of sampling metrics that quantify temporal variation in the availability of fossil-bearing rock and the amount and geographical spread of palaeoanthropological collection effort. I also compared these diversity patterns and sampling metrics to a continental-scale aridity proxy and a regional-scale lake variability proxy. The finding that raw taxic diversity is positively correlated with the number of primate-bearing formations (a proxy for rock availability), the number of unique years that have produced a hominin fossil per formation (a proxy for collection effort), and the geographic spread of collection effort (area of the land surface sampled), supports a RRB scenario. Raw taxic diversity therefore cannot be considered a reliable estimate of early hominin diversity because it largely reflects geological and anthropogenic factors. These findings support those from Chapter 2 that climate-driven hypotheses of hominin diversification have therefore sought to explain a pattern that is more apparent than real (Maxwell *et al.*, 2018). In contrast, none of the PDEs correlate with any sampling metric, suggesting that phylogenetically corrected diversity is unaffected by the sampling metrics used and that the PDE represents a closer approximation of true diversity than the TDE (Maxwell *et al.*, 2018). The key argument against fossil sampling driving TDE would be if sampling metrics and taxic diversity were independently driven by the same environmental factor, or if the sampling metrics used were non-independent with respect to taxic diversity. The available data show that there is no support for the former, and the evidence for the latter is weakened by the use of sampling metrics based on the occurrence of primate and mammal fossils. The application of information theoretic methods could go some way to disentangle RRB from redundancy, if developments are made to allow for small data sets (Hannisdal, 2011).

By comparing different sampling metrics I also sought to explore the spatial and temporal architecture of the African Neogene rock record more generally. Sampling metrics consistently show that the availability of terrestrial mammal-bearing rock is highly non-random, and that the amount and areal extent of collection effort closely tracks variation in the rock record (Fig. 3.2). Consequently, some time periods are more intensely studied than others, and it is these time periods with the greatest study interest that record the highest early hominin raw taxic diversity. Multiple sampling metrics (with the exception of MBF) converge on the same signal and correlate strongly with traditional geological sampling metrics. These include the relative proportion of ghost lineage to sampled diversity and counts of the number of grid cells occupied by hominin fossils. Further study is now necessary in order to explore whether these sampling metrics correlate with the raw taxic diversity of other African mammals. Bovids and cercopithecids have long been used as a proxy for hominins thanks to their more abundant fossil records (e.g., Vrba, 1995; Elton, 2006; Patterson *et al.*, 2014). However, cercopithecids display peak raw

taxic diversity at 2.4 Ma (Frost, 2007) and bovids at 1.9 Ma (Bibi & Kiessling, 2015), when the number of primate- and mammal-bearing formations is highest.

Chapter 4

In Chapter 4 I showed that the completeness of the early African hominin fossil record has fluctuated throughout geological time, with distinct patterns depending on whether specimen completeness is assessed using phylogenetic character richness or the amount of anatomical representation, and differing levels of support for a relationship with raw taxic diversity and sampling metrics. The CCM is low throughout the first half of hominin evolution before a sudden tripling in the middle Pliocene (3.6 Ma), after which completeness remains high for the rest of the Plio-Pleistocene. This bimodal pattern (the fossil record is known from either few, fragmentary skulls or many, well-preserved skulls) is unusual in that CCM1-2_{max} and TDE both correlate with PBF. The finding that TDE and PBF strongly correlate (Chapter 3) may, therefore, be the result of a common-cause interaction (increased rock availability may increase the likelihood of highly complete specimens and highly complete specimens may make taxonomic identification easier, raising apparent diversity). The SCM is also generally low throughout much of hominin evolution, particularly during the late Miocene, reflecting an abundance of low-scoring cranio-dental material. However, there are three irregular peaks during the Plio-Pleistocene, each of which coincide with rich formations that are preferentially re-sampled, yielding many collections and high-scoring postcranial material (Raup, 1977). These peaks in the SCM show no relation to raw taxic diversity. Sampling metrics appear to be the strongest control on specimen quality, however, for some completeness metrics they explain around only half of the variance. Further study should therefore concentrate on whether the completeness and sampling metrics used are the most appropriate for early hominins. For example, these results reiterate that when calculating the CCM one must consider the intended use of the original character matrix, and the specific question it sought to answer. A large amount of information relevant to hominin evolution is being missed when cranio-dental-only matrices are used to quantify specimen completeness.

These results demonstrate quantitatively that the early African hominin fossil record is notoriously poor during the period that molecular clock studies date the most recent common ancestor of hominins and panins (8.0–4.0 Ma), and the origin of genus *Homo* (3.0–2.0 Ma). The incompleteness of the former time window is problematic for reconstructions of the ancestral morphotype of panins and hominins. With the number of phylogenetically relevant characters diminishing with increasing proximity to the separation of hominins and panins, and so few phylogenetically relevant characters present in the earliest purported hominins, uncertainties around character state polarity and the confounding influence of homoplasy mire the phylogenetic placement of hominins (Wood & Harrison, 2011). These problems are compounded by the highly non-uniform spatial sampling of the fossil record and absence of rich (bonanza) deposits during this time window. There are very few hominin-bearing formations from 7.0–4.6 Ma and those formations that are known sample no more than 4% of Africa's land surface. Moreover, the African Neogene rock record is heavily biased toward the depositional basins of the East African Rift System where rifting, uplift, and erosion have created the conditions perfect for the recovery of terrestrial vertebrate fossils. If the fossil record is to increase in quality by any substantial margin, palaeoanthropologists must radically alter their research strategy and search for new fossil-bearing de-

posits, in addition to returning to the well-known, rich deposits in the Cradle of Humankind and East African Rift Valley.

5.2 Limitations

Phylogenetic diversity estimates are only as robust as the source cladogram. Phylogenetic *trees* represent phylogenetic *hypotheses* which are subject to evaluation against alternative topologies and branch lengths. A phylogenetic diversity estimate (PDE) based on an incorrect cladogram will produce an incorrect diversity pattern, and will lead to incorrect evolutionary inferences about the causal mechanism behind such a pattern (Lane *et al.*, 2005). In Chapter 2 I adopted a pluralistic PDE approach using four cladograms and a probabilistic time-scaling method, in addition to the TDE (e.g., Wagner, 1995; Mannion *et al.*, 2011, 2014; Smith *et al.*, 2012). Comparison of the different PDEs demonstrated they are surprisingly robust to differences in tree topology, tree size, and branch lengths for hominins (Maxwell *et al.*, 2018). However, limitations in the PDE approach apply equally to the four different cladograms so this congruence is unsurprising. For example, all excluded the purported hominins from the late Miocene to early Pliocene (they are also the most incomplete), all used living hominoids (principally *Gorilla* and *Pan*) as the outgroup and excluded Miocene fossil hominoids (which is problematic for ancestral morphotype reconstruction), and all were based on a character matrix comprised entirely of cranio-dental characters. Singly and in combination, each of these factors have the potential to profoundly influence the reliability of the resultant tree topology, and therefore any PDE based upon them.

Chronological uncertainty in the South African cave deposits. One key limitation of using hominin fossil occurrence data is the poor dating of the hominin-bearing palaeocave deposits (e.g., Makapansgat, Taung, and the Cradle of Humankind) compared to the well-constrained deposits found throughout the EARS. The often substantial dating error adds much ambiguity to any diversity curve (be it a TDE or PDE) and may obscure genuine large-scale peaks and troughs. One possible way to attempt to ameliorate this problem in the future would be to implement sensitivity tests, whereby the effect of stratigraphic uncertainty in the age of a deposit is incorporated into the diversity estimate by varying the age of the oldest (FAD) and youngest horizon (LAD) within the limits of the uncertainty of the dating technique. This could be repeated numerous times, with an overall randomised diversity curve produced (Mannion *et al.*, 2011). Similar methods are used for *a posteriori* phylogenetic time-scaling (Bapst, 2012, 2013).

Sampling metrics. Unfortunately, the vast majority of papers describing new primate and mammal fossils (from both hominin- and non-hominin-bearing deposits) rarely include discovery year. Moreover, counts of the number of primate- (PBC) and mammal-bearing collections (MBC) could not be compiled in the same manner as HBC because the African Neogene is too poorly represented in the *Paleobiology Database* (PBDB), and updating the PBDB occurrence list for all Neogene African terrestrial mammals became, on assessment, too large a task. This is especially unfortunate because PBC and MBC could provide key insights into temporal variation in sampling effort, and improve our ability to reliably determine whether time bins with low diversity are a result of poor sampling or a genuine feature of a clade (good sampling). Compiling the number of PBC and MBC would also directly inform the redundancy argument (Benton *et al.*, 2011) applied to hominins (Maxwell *et al.*, 2018).

A preliminary assessment of the number of late Miocene PBC (from a small number of papers which included the relevant information), suggests that this time period may be better sampled than originally thought. For example, while most formations in the 6.3–6.0 Ma time bin include only one PBC (e.g., Mpesida Beds and Nkondo Fm., Kenya; Ongoliba Beds, Democratic Republic of Congo; Wadi Natrun, Egypt), the site of Lemudong'o (Kenya) is well-sampled, yielding 7 PBC that can be identified (Hlusko, 2007) and, more recently, 16 fossils of unknown year (Gilbert *et al.*, 2010). The presence of only one HBC in this time bin suggests that the absence of hominins at Lemudong'o may be a genuine biological signal.

Rock exposure data. Dunhill (2011, 2012) contends that the most appropriate measure of rock availability is rock exposure area. However, there is probably no *a priori* reason why rock exposure area should accurately predict the number of opportunities to sample fossils. Either way, data on rock exposure area is not currently available at the continental scale. Moreover, the tight chronological controls required to quantify rock exposure area in time bins less than a million years in duration simply do not exist, and so an epoch-scale assessment of rock outcrop (= geological map) area is probably the only feasible alternative (at present). Preliminary work looking at the relative abundance of cercopithecid and hominin fossils per square kilometre of outcropping rock in the Koobi Fora Formation, East Turkana (Kenya), did so for only three collection areas and did not consider the age of the rock (Thompson *et al.*, 2017). Data coverage is better at Laetoli, Tanzania, where outcropping rocks of the Upper Laetolil and Upper Ndolanya Beds have been mapped (Harrison & Kweka in Harrison, 2011*a*), but which, to the best of my knowledge, have only been published graphically. Obtaining continental-scale information on rock exposure area will be a difficult task, but should nonetheless be encouraged using remote sensing and a Geographic Information System (GIS) (Dunhill, 2012). This would provide a crucial test of whether PBF is an appropriate proxy for rock availability, and therefore of whether geological factors do drive early hominin TDE (Maxwell *et al.*, 2018). Because primate-bearing rock is distributed widely in any one time bin (Werdelin, 2010), and because outcrop area and exposure area tend to correlate in arid and semi-arid environments (Dunhill, 2012), I would expect rock exposure area to correlate with rock outcrop area and for both to show the same peaks and troughs seen in PBF. I would also expect rock exposure area and MBF to correlate more weakly (compared to PBF) because, while many time bins have the same MBF count due to the presence of several long-ranging formations, the exposure area of those formations per time bin is unlikely to follow the same pattern.

Specimen completeness. It could be argued that the peaks and troughs in CCM and SCM reported in Chapter 4 are inaccurate because of the choice to bin each taxon's completeness score based on their total stratigraphic range, and not on those specimens actually present in a particular time bin. While this issue may be of little concern when assessing specimen completeness per geological stage (e.g., Mannion & Upchurch, 2010; Brocklehurst *et al.*, 2012; Dean *et al.*, 2016), it could be problematic when using time bins with an average duration of less than one million years. For example, the SCM2 score for *Australopithecus afarensis* is 84% and so this value is added to each of the time bins spanning its stratigraphic range (3.7–3.0 Ma). However, half of this taxon's completeness score is the result of a single partial skeleton, the 3.2-Ma AL 288-1 (SCM1 = 42%), which occurs in only one time bin. This has the result of artificially inflating specimen completeness in adjacent time bins by adding the entire *Australopithecus afarensis* hypodigm to each time bin that spans its first and last appearance date. So, while

the time bin containing the 3.2-Ma partial skeleton has a genuinely high completeness score given the presence of a partial skeleton, adjacent time bins may not necessarily be equally complete. The choice of methodology could hide the fact an adjacent time bin contains highly fragmentary and incomplete specimens of *Australopithecus afarensis* as a result of there being less available rock to sample or poorer overall collection effort.

One solution would be to quantify specimen completeness based on the top 5 or top 10 most complete specimens per time bin (e.g., Cleary *et al.*, 2015; Tutin & Butler, 2017). This would avoid artificially inflating a time bin's mean completeness score based on the presence of a highly complete specimen in a neighbouring time bin, although it would have the drawback of being time consuming when assessing character completeness.

5.3 Future directions

Generating more comprehensive phylogenetic hypotheses. Postcranial characters are often judged *a priori* to be more homoplastic than cranio-dental characters. Nowhere is this assumption more apparent than in the study of ape, and in particular, hominin phylogeny (Ward, 1997; Young, 2005; Pugh, 2016). The idea that there is little useful phylogenetic signal in postcranial characters compared to cranio-dental characters appears to be firmly entrenched in the minds of palaeoanthropologists (Pilbeam, 1996) considering no study of hominin phylogeny has included postcrania to date (Strait *et al.*, 2007). However, there is little solid evidence for such an opinion (Lockwood & Fleagle, 1999), and any support is piecemeal and largely anecdotal. In fact, there is mounting evidence to suggest the opposite is true (postcranial characters are less homoplastic than cranio-dental characters), or at least that cranio-dental and postcranial characters display similar levels of homoplasy. For example, in a recent meta-analysis of over 80 vertebrate character matrices, Mounce *et al.* (2016) reported a non-significant difference in the level of homoplasy for cranio-dental and postcranial characters. However, they also found that 32% of the time cranio-dental and postcranial characters produce an incongruent tree topology, much more often than expected (Mounce *et al.*, 2016). The assumption that postcranial characters are particularly prone to homoplasy among modern hominoids is prevalent because of evidence suggesting that orthograde posture and suspensory locomotion are convergent (Fleagle & Lieberman, 2015). However, Finarelli & Clyde (2004) reported similar levels of homoplasy in their study of early to late Miocene fossil hominoids. In addition, Collard, Gibbs, & Wood (2001) and, more recently, Pugh (2016) have reported (in conference proceedings) that postcranial characters do recover a well-supported cladogram compatible with the molecular phylogeny, including a *Pan-Homo* clade. In summary, there are three major directions in which future phylogenetic research on hominoids and hominins can increase resolution and phylogenetic accuracy:

1. Incorporate postcranial characters.
2. Incorporate early to late Miocene fossil hominoids.
3. Combine hard tissue, soft tissue, non-coding genomic loci, and stratigraphic data into a single total-evidence approach, in order to simultaneously estimate tree topology, branch lengths, and the probability of sampled ancestors.

Pugh & Gilbert (2018) recently published a phylogeny of papionins (Old World monkeys) using a combination of molecular and morphological data and cranio-dental and postcranial characters, demonstrating there is similar scope for hominoids (Pugh, 2016).

Generating better sampling metrics. Palaeoanthropology has one major advantage over palaeobiology with respect to constructing sampling metrics and that is the archaeological record, which may also be used, in combination with the fossil record, to obtain a more accurate estimate of sampling intensity. Sampling metrics that include, for example, formations that are known to contain ichnofossils (e.g., footprints) and lithic artefacts (stone tools) would provide a better estimate of supposed total sampling effort compared to sampling metrics based purely on primate or mammal fossils. Formations that include palaeobiological (fossil), behavioural (footprints), and cultural evidence (stone tools) represent a closer approximation of the total number of possible opportunities to sample hominin fossils (= possible hominin-bearing formations), and therefore a closer approximation of supposed total sampling effort. Lithic- and ichnofossil-bearing formations, in combination with PBF or MBF, would represent the most comprehensive sampling metric currently available.

The bonanza variable used throughout Chapter 4 could also be modified to reduce the influence of any non-bonanza deposits. The bonanza variable is a measure of the tendency of palaeoanthropologists to preferentially return to rich deposits and amass a large number of collections. Here, it is calculated as the ratio of hominin-bearing collections to formations (HBC:HBF), which is essentially the average number of collections in any one formation per time bin. One could also use the maximum HBC count in a single formation per time bin. For example, the 4.6–4.4 Ma time bin has one HBC from the Chemeron Fm. and 12 from the Sagantole Fm. (Maxwell *et al.*, 2018). Using the HBC:HBF ratio would give a value of 6.5 for this time bin, however, using the maximum HBC would give a value of 12. Alternatively, one could ignore those formations which have yielded a HBC count below an arbitrary cut-off (e.g., 4), and calculate the HBC:HBF ratio based on this smaller data set. Both would give a lower or no weight to non-bonanza sites (e.g., formations that have yielded 1–3 collections in any one time bin) and may provide better sampling metrics for specifically quantifying the bonanza effect (Raup, 1977).

Further work is needed to assess whether the sampling metrics used herein fully capture all aspects of geological and anthropogenic sampling bias and whether these sampling metrics (and also PBC and MBC if they are compiled in the future) are indeed controlled by the thing they seek to quantify (the rock record), or by their fossil content. The generation and testing of alternative sampling metrics (especially those that are independent of the precise definition of what is actually an opportunity to sample hominin fossils, e.g., sampling area estimates) will bare directly on these issues and be a fruitful avenue of research into hominin fossil record quality.

Quantitative palaeobiology. The rate at which new methods and statistical tests are being introduced to the study of macroevolution is startling, driven in part by the proliferation of R (R Development Core Team, 2018) in evolutionary biology, ecology, and palaeobiology. However, palaeoanthropologists, archaeologists, climatologists, and physical geographers tend to approach human evolution with the view that modern humans are special and somehow exceptional (Gee, 2013). One unfortunate and rather significant consequence of such a view is that sound evolutionary hypotheses are usually second to just-so stories (Smith, 2016). Palaeoanthropologists tackle the same questions as palaeobiologists, the only difference being that the former are concerned with primate and human evolution while

the latter with the evolution of non-primates. Other than the clade of interest, the methods and statistical tests used are directly applicable to many questions in primate and human evolution. Becoming cognisant of the rapidly advancing field of quantitative paleobiology should be at the center of future research in paleoanthropology.

> *"We're responding to what we find, rather than trying to find what we need."*
>
> – Richard Leakey (2010).*

Field study. In Chapter 3 I showed that rock availability and palaeoanthropological collection effort are both highly non-random (Figs. 3.2) and geographically limited (Figs. 3.10 and 3.11). In Chapter 4 I showed that the tendency of palaeoanthropologists to preferentially re-sample known rich deposits is the dominant control (among the sampling metrics used) on temporal trends in specimen completeness. Those time bins that contain few, well-sampled formations are more complete in terms of specimen quality than those time bins which contain many, poorly sampled formations. However, if the fossil record is to improve by any substantial margin then palaeoanthropologists must shift from a *re-sampling regime* to an *exploratory regime*, and undertake exploration in un-sampled and under-sampled geographical regions and geological formations.

The initial drive to survey a new area is usually the result of either commercial exploration or land-use change leading to the discovery of new fossils (fossils → formations; Raup, 1977) or intense field study in a sedimentary basin leading to the chance discovery of un-sampled geological formations (formations → fossils; Sheehan, 1977). Once the fossil content of a new formation has been thoroughly described then either (1) attention shifts back to the discovery of new formations, (2) old formations are repeatedly re-sampled (with a formation or deposit potentially becoming a bonanza site if collection effort is concentrated there for a prolonged period of time), or (3) new fossils are found through a combination of exploration and re-sampling. In palaeobiology, the tendency to undertake both exploration and re-sampling is readily apparent, with the rate of taxon discovery in both new and old formations on the rise (Benton, 2015). In contrast, palaeoanthropologists predominantly re-sample known formations, and have done so for more than a decade (note the curve in Fig. 3.9*b* is best explained by a fourth-order polynomial (R^2 = 0.99) which reached an asymptote around the year 2000).

The underlying reason why palaeoanthropologists subscribe to a model in which fossils drive collection effort is unclear. It may be that exploration is unlikely to be funded in new formations due to the rarity of—and therefore improbability of finding—hominin fossils, or perhaps the hostile working environment leads palaeoanthropologists to conduct their research at field sites where the infrastructure (accommodation, transportation, etc.) is already in place to do so. Nonetheless, non-occurrence is valuable data that represents a singular, unique unit of collection effort that did not result in a hominin fossil: non-occurrence at a locality or in a formation in spite of intense collection effort lends support to hominins being genuinely absent.

If the early African hominin fossil record is to increase in completeness by any substantial margin, palaeoanthropologists must radically alter their research strategy and search for new fossiliferous rock, in addition to returning to the well-known, rich deposits. The adoption of an exploratory regime, combined with an increased appreciation of the quality and fidelity of the African Neogene hominoid fossil

*The 9th Human Evolution Workshop of the Turkana Basin Institute (Kenya).

record more generally, is the best way to increase our understanding of the evolutionary history of our fossil ancestors.

Bibliography

Ackermann, R.R. and Smith, R.J., 2007. The macroevolution of our ancient lineage: what we know (or think we know) about early hominin diversity. *Evolutionary Biology*, **34**(1–2), 72–85.

Alexeev, V.P., 1986. *The Origin of the Human Race*. Progress Publishers, Moscow.

Almécija, S., Tallman, M., Alba, D.M., Pina, M., Moyà-Solà, S., and Jungers, W.L., 2013. The femur of *Orrorin tugenensis* exhibits morphometric affinities with both Miocene apes and later hominins. *Nature Communications*, **4**, 2888.

Alroy, J., 2000. Successive approximations of diversity curves: ten more years in the library. *Geology*, **28**(11), 1023–1026.

Alroy, J., 2010. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. In: Alroy, J. and Hunt, G., (Eds.), *Quantification Methods in Palaeobiology*. *Paleontological Society Papers*, **16**, pp. 55–80.

Andrews, P., 1995. Ecological apes and ancestors. *Nature*, **376**, 555–556.

Andrews, P. and Harrison, T., 2005. The last common ancestor of apes and humans. In: Lieberman, D.E., Smith, R.J., and Kelley, J., (Eds.), *Interpreting the Past: Essays on Human, Primate, and Mammal Evolution*. Brill Academic Publishers, Boston, pp. 103–121.

Antón, S.C., 2003. Natural history of *Homo erectus*. *American Journal of Physical Anthropology*, **S37**, 126–170.

Antón, S.C. and Swisher, C.C., 2004. Early dispersals of *Homo* from Africa. *Annual Review of Anthropology*, **33**, 271–296.

Arambourg, C. and Coppens, Y., 1968. Découverte d'un australopithecien nouveau dans les Gisements de L'Omo (Ethiopie). *South African Journal of Science*, **64**(2), 58–59.

Asfaw, B., 1987. The Belohdelie frontal: new evidence of early hominid cranial morphology from the Afar of Ethiopia. *Journal of Human Evolution*, **16**(7–8), 611–624.

Asfaw, B., Gilbert, W.H., Beyene, Y., Hart, W.K., Renne, P.R., WoldeGabriel, G., Vrba, E.S., and White, T.D., 2002. Remains of *Homo erectus* from Bouri, Middle Awash, Ethiopia. *Nature*, **416**(6878), 317–320.

Asfaw, B., White, T.D., Lovejoy, C.O., Latimer, H.B., Simpson, S.W., and Suwa, G., 1999. *Australopithecus garhi*: a new species of early hominid from Ethiopia. *Science*, **284**(5414), 629–635.

Bapst, D.W., 2012. paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, **3**(5), 803–807.

Bapst, D.W., 2013. A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. *Methods in Ecology and Evolution*, **4**(8), 724–733.

Bapst, D.W. and Hopkins, M.J., 2017. Comparing *cal3* and other *a posteriori* time-scaling approaches in a case study with the pterocephaliid trilobites. *Paleobiology*, **43**(1), 49–67.

Bapst, D.W., Wright, A.M., Matzke, N.J., and Lloyd, G.T., 2016. Topology, divergence dates, and macroevolutionary inferences vary between tip-dating approaches applied to fossil theropods (Dinosauria). *Biology Letters*, **12**(7), 1–4.

Barnosky, A.D., 2001. Distinguishing the effects of the Red queen and Court Jester on Miocene mammal evolution in the northern Rocky Mountains. *Journal of Vertebrate Paleontology*, **21**(1), 172–185.

Barnosky, A.D., Carrasco, M.A., and Davis, E.B., 2005. The impact of the species-area relationship on

estimates of paleodiversity. *PLOS Biology*, **3**(8), e266.
Barr, W.A., 2017. Signal or noise? A null model method for evaluating the significance of turnover pulses. *Paleobiology*, **43**(4), 656–666.
Barrett, P.M., McGowan, A.J., and Page, V., 2009. Dinosaur diversity and the rock record. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**(1667), 2667–2674.
Begun, D.R., 2001. African and Eurasian Miocene hominoids and the origins of the Hominidae. In: De Bonis, L., Koufos, G.D., and Andrews, P., (Eds.), *Hominoid Evolution and Climatic Change in Europe: Volume 2, Phylogeny of the Neogene Hominoid Primates of Eurasia*. Cambridge University Press, Cambridge, pp. 231–253.
Begun, D.R., 2004. The earliest hominins—is less more? *Science*, **303**(5663), 1478–1480.
Begun, D.R., 2007. Fossil record of Miocene hominoids. In: Henke, W., Hardt, T., and Tattersall, I., (Eds.), *Handbook of Paleoanthropology*. Springer, New York, pp. 921–977.
Begun, D.R., 2010. Miocene hominids and the origins of the African apes and humans. *Annual Review of Anthropology*, **39**, 67–84.
Behrensmeyer, A.K., Harmon, E.H., and Kimbel, W.H., 2003. Environmental context and taphonomy of the First Family hominid locality, Hadar, Ethiopia. *Journal of Vertebrate Paleontology*, **23**(3), 33.
Behrensmeyer, A.K., Kidwell, S.M., and Gastaldo, R.A., 2000. Taphonomy and paleobiology. *Paleobiology*, **26**(4), 103–147.
Behrensmeyer, A.K. and Reed, K.E., 2013. Reconstructing the habitats of *Australopithecus*: paleoenvironments, site taphonomy, and faunas. In: Reed, K.E., Fleagle, J.G., and Leakey, R.E.F., (Eds.), *The Paleobiology of Australopithecus*. Springer, Dordrecht, pp. 41–60.
Behrensmeyer, A.K., Todd, N.E., Potts, R., and McBrinn, G.E., 1997. Late Pliocene faunal turnover in the Turkana Basin, Kenya. *Science*, **278**(5343), 1589–1594.
Bell, M.A. and Lloyd, G.T., 2015. strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. *Palaeontology*, **58**(2), 379–389.
Bell, M.A., Upchurch, P., Mannion, P.D., and Lloyd, G.T. Using the character completeness metric to examine completeness of Mesozoic dinosaurs: a Maastrichtian high and a paleoequatorial low. Paper presented at the Society of Vertebrate Paleontology annual meeting, Los Angeles, 2013.
Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B: Methodological*, **57**(1), 289–300.
Bennett, C.V., Upchurch, P., Goin, F.J., and Goswami, A., 2018. Deep time diversity of metatherian mammals: implications for evolutionary history and fossil-record quality. *Paleobiology*, **44**(2), 171–198.
Benson, R.B.J. and Butler, R.J., 2011. Uncovering the diversification history of marine tetrapods: ecology influences the effect of geological sampling biases. In: McGowan, A.J. and Smith, A.B., (Eds.), *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society of London Special Publications, **358**, pp. 191–207.
Benson, R.B.J., Butler, R.J., Lindgren, J., and Smith, A.S., 2010. Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**(1683), 829–834.
Benson, R.B.J. and Mannion, P.D., 2012. Multi-variate models are essential for understanding vertebrate diversification in deep time. *Biology Letters*, **8**(1), 127–130.
Benson, R.B.J. and Upchurch, P., 2013. Diversity trends in the establishment of terrestrial vertebrate ecosystems: interactions between spatial and temporal sampling biases. *Geology*, **41**(1), 43–46.
Benton, M.J., 1998. The quality of the fossil record of vertebrates. In: Donovan, S.K. and Paul, C.R.C, (Eds.), *The Adequacy of the Fossil Record*. Wiley, New York, pp. 269–303.
Benton, M.J., 2003. The quality of the fossil record. In: Donoghue, P.C.J. and Smith, M.P., (Eds.), *Telling the Evolutionary Time: Molecular Clocks and the Fossil Record*. Taylor and Francis, London, pp. 66–90.
Benton, M.J., 2007. Biodiversity through time. In: Briggs, D.E.G. and Crowther, P.R., (Eds.), *Palaeobi-*

ology II. Blackwell Science, Oxford, pp. 211–220.
Benton, M.J., 2009. The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science*, **323**(5915), 728–732.
Benton, M.J., 2012. No gap in the Middle Permian record of terrestrial vertebrates. *Geology*, **41**(9), e293.
Benton, M.J., 2015. Palaeodiversity and formation counts: redundancy or bias? *Palaeontology*, **58**(6), 1003–1029.
Benton, M.J., Dunhill, A.M., Lloyd, G.T., and Marx, F.G., 2011. Assessing the quality of the fossil record: insights from vertebrates. In: McGowan, A.J. and Smith, A.B., (Eds.), *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society of London Special Publications, **358**, pp. 63–94.
Benton, M.J. and Emerson, B.C., 2007. How did life become to diverse? The dynamics of diversification according to the fossil record and molecular phylogenies. *Palaeontology*, **50**(1), 23–40.
Benton, M.J., Ruta, M., Dunhill, A.M., and Sakamoto, M., 2013. The first half of tetrapod evolution, sampling proxies, and fossil record quality. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 18–41.
Benton, M.J., Tverdokhlebov, V.P., and Surkov, M.V., 2004. Ecosystem remodelling among vertebrates at the Permian–Triassic boundary in Russia. *Nature*, **432**(7013), 97–100.
Berger, L.R., de Ruiter, D.J., Churchill, S.E., Schmid, P., Carlson, K.J., Dirks, P.H.G.M., and Kibii, J.M., 2010. *Australopithecus sediba*: a new species of *Homo*-like australopith from South Africa. *Science*, **328**(5975), 195–204.
Bernard, E.L., Ruta, M., Tarver, J.E., and Benton, M.J., 2010. The fossil record of early tetrapods: worker effort and the end-Permian mass extinction. *Acta Palaeontologica Polonica*, **55**(2), 229–239.
Bibi, F. and Kiessling, W., 2015. Continuous evolutionary change in Plio-Pleistocene mammals of eastern Africa. *Proceedings of the National Academy of Sciences*, **112**(34), 10623–10628.
Bigazzi, G., Balestrieri, M.L., Norelli, P., Oddone, M., and Tewolde, M.T., 2004. Fission-track dating of a tephra layer in the Alat Formation of the Dandiero Group (Danakil Depression, Eritrea). *Rivista Italiana di Paleontologia e Stratigrafia*, **110**, 45–49.
Blumenschine, R.J., Peters, C.R., Masao, F.T., Clarke, R.J., Deino, A.L., Hay, R.L., Swisher, C.C., Stanistreet, I.G., Ashley, G.M., McHenry, L.J., Sikes, N.E., Van Der Merwe, N.J., Tactikos, J.C., Cushing, A.E., Deocampo, D.M., Njau, J.K., and Ebert, J.I., 2003. Late Pliocene *Homo* and hominid land use from Western Olduvai Gorge, Tanzania. *Science*, **299**(5610), 1217–1221.
Blumenthal, S.A., Levin, N.E., Brown, F.H., J.-P., Brugal., Chritz, K.L., Harris, J.M., Jehle, G.E., and Cerling, T.E., 2017. Aridity and hominin environments. *Proceedings of the National Academy of Sciences*, **114**(28), 7331–7336.
Bobe, R. and Behrensmeyer, A.K., 2004. The expansion of grassland ecosystems in Africa in relation to mammalian evolution and the origin of the genus *Homo*. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **207**, 399–420.
Bobe, R., Behrensmeyer, A.K., and Chapman, R.E., 2002. Faunal change, environmental variability and late Pliocene hominin evolution. *Journal of Human Evolution*, **42**(4), 475–497.
Bobe, R., Behrensmeyer, A.K., Leakey, M.G., and Mbua, E., 2011. The Turkana Database: an archive of vertebrate evolution in East Africa. *Evolutionary Anthropology*, **20**(6), 256.
Bobe, R. and Leakey, M.G., 2009. Ecology of Plio-Pleistocene mammals in the Omo-Turkana Basin and the emergence of *Homo*. In: Grine, F.E., Fleagle, J.G., and Leakey, R.E.F., (Eds.), *The First Humans: Origins of the Genus Homo*. Springer, Dordrecht, pp. 173–184.
Bonnefille, R., Potts, R., Chalié, F., Jolly, D., and Peyron, O., 2004. High-resolution vegetation and climate change associated with Pliocene *Australopithecus afarensis*. *Proceedings of the National Academy of Sciences*, **101**(33), 12125–12129.
Bowen, B.E., 1974. *The geology of the upper Cenozoic sediments in the East Rudolf embayment of the Lake Rudolf basin, Kenya*. PhD thesis, Iowa State University.
Bradley, B.J., 2008. Reconstructing phylogenies and phenotypes: a molecular view of human evolution.

Journal of Anatomy, **212**(4), 337–353.

Brain, C.K., 1981. *The Hunters or the Hunted? An introduction to African Cave Taphonomy*. University of Chicago Press, Chicago.

Brocklehurst, N., 2015. A simulation-based examination of residual diversity estimates as a method of correcting for sampling bias. *Paleontologica Electronica*, **18**(3), 1–15.

Brocklehurst, N. and Fröbisch, J., 2014. Current and historical perspectives on the completeness of the fossil record of pelycosaurian-grade synapsids. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **399**, 114–126.

Brocklehurst, N., Kammerer, C.F., and Fröbisch, J., 2013. The early evolution of synapsids, and the influence of sampling on their fossil record. *Paleobiology*, **39**(3), 470–490.

Brocklehurst, N., Upchurch, P., Mannion, P.D., and O'Connor, J., 2012. The completeness of the fossil record of Mesozoic birds: implications for early avian evolution. *PLOS ONE*, **7**(6), e39056.

Broom, R., 1938. The Pleistocene anthropoid apes of South Africa. *Nature*, **142**(3591), 377–379.

Broom, R., Roberton, J.T., and Schepers, G.W.H, 1950. Sterkfontein ape-man, *Plesianthropus*. *Transvaal Museum Memoir*, **4**(1), 1–117.

Brown, C.M., Evans, D.C., Campione, N.E., O'Brien, L.J., and Eberth, D.A., 2013. Evidence for taphonomic size bias in the Dinosaur Park Formation (Campanian, Alberta), a model Mesozoic terrestrial alluvial-paralic system. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 108–122.

Brown, F.H. and Feibel, C.S., 1985. Stratigraphical notes on the Okote Tuff Complex at Koobi Fora, Kenya. *Nature*, **316**, 794–797.

Brunet, M., Beauvilain, A., Coppens, Y., Heintz, E., Moutaye, A.H.E., and Pilbeam, D.R., 1995. The first australopithecine 2,500 kilometres west of the Rift Valley (Chad). *Nature*, **378**(6554), 273–275.

Brunet, M., Beauvilain, A., Coppens, Y., Heintz, E., Moutaye, A.H.E., and Pilbeam, D.R., 1996. *Australopithecus bahrelghazali*, une nouvelle espèce d'Hominidé ancien de la région de Koro Toro (Tchad). *Comptes Rendus de l'Académie des Sciences*, **322**(10), 907–913.

Brunet, M., Frank, G., Pilbeam, D.R., Lieberman, D.E., Likius, A., Mackaye, H.T., Ponce De Leon, M.S., Zollikofer, C.P.E., and Vignaud, P., 2005. New material of the earliest hominid from the Upper Miocene of Chad. *Nature*, **434**(7034), 752–755.

Brunet, M., Guy, F., Pilbeam, D.R., Mackaye, H.T., Likius, A., Ahounta, D., Beauvilain, A., Blondel, C., Bocherens, H., Boisserie, J.-R., De Bonis, L., Coppens, Y., Dejax, L., Denys, C., Duringer, P., Eisenmann, V., Fanone, G., Fronty, P., Geraads, D., Lehmann, T., Lihoreau, F., Louchart, A., Mahamat, A., Merceron, G., Mouchelin, G., Otero, O., Campomanes, P.P., Ponce De Leon, M.S., Rage, J.-C., Sapanet, M., Schuster, M., Sudre, J., Tassy, P., Valentin, X., Vignaud, P., Viriot, L., Zazzo, A., and Zollikofer, C.P.E., 2002. A new hominid from the Upper Miocene of Chad, Central Africa. *Nature*, **418** (6894), 145–151.

Brusatte, S.L., Benton, M.J., Ruta, M., and Lloyd, G.T., 2008. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science*, **321**(5895), 1485–1488.

Butler, R.J., Barrett, P.M., Nowbath, S., and Upchurch, P., 2009. Estimating the effects of sampling biases on pterosaur diversity patterns: implications for hypotheses of bird/pterosaur competitive replacement. *Paleobiology*, **35**(3), 432–446.

Butler, R.J., Benson, R.B.J., and Barrett, P.M., 2013. Pterosaur diversity: untangling the influence of sampling biases, Lagerstätten, and genuine biodiversity signals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 78–87.

Butler, R.J., Benson, R.B.J., Carrano, M.T., Mannion, P.D., and Upchurch, P., 2011. Sea level, dinosaur diversity and sampling biases: investigating the "common cause" hypothesis in the terrestrial realm. *Proceedings of the Royal Society of London B: Biological Sciences*, **278**(1709), 1165–1170.

Cain, S.A., 1938. The species-area curve. *American Midland Naturalist*, **19**, 573–581.

Cameron, D.W., 2003. Early hominin speciation at the Plio/Pleistocene transition. *HOMO – Journal of Comparative Human Biology*, **54**(1), 1–28.

Cavin, L. and Forey, P.L., 2007. Using ghost lineages to identify diversification events in the fossil record.

Biology Letters, **3**(2), 201–204.

Cerling, T.E., Manthi, F.K., Mbua, E.N., Leakey, L.N., Leakey, M.G., Leakey, R.E.F., Brown, F.H., Grine, F.E., Hart, J.A., Kaleme, P., Roche, H., Uno, K.T., and Wood, B.A., 2013. Stable isotope-based diet reconstructions of Turkana Basin hominins. *Proceedings of the National Academy of Sciences*, **110** (26), 10501–10506.

Chamberlain, A.T. and Wood, B.A., 1987. Early hominid phylogeny. *Journal of Human Evolution*, **16**, 118–133.

Chavaillon, J., Brahimi, C., and Coppens, Y., 1974. Première découverte d'Hominidé dans l'un des sites acheuléens de Melka-Kunturé (Ethiopie). *Comptes Rendus de l'Académie des Sciences*, **278**, 3299–3302.

Chavaillon, J., Chavaillon, N., Coppens, Y., and Senut, B., 1977. Présence d'Hominidé dans le site oldowayen de Gomboré I à Melka-Kunturé. *Comptes Rendus de l'Académie des Sciences*, **285**, 961–963.

Chen, F.C. and Li, W.-H., 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *American Journal of Human Genetics*, **68**(2), 444–456.

Cheyne, V.D. and Oba, J.T., 1943. Average weights of the permanent teeth, including the relative amounts of enamel to dentin and cementum. *Journal of Dental Research*, **22**, 181–184.

Clark, J.D., Asfaw, B., Assefa, G., Harris, J.W.K., Kurashina, H., Walter, R.C., White, T.D., and Williams, M.A.J., 1984. Palaeoanthropological discoveries in the Middle Awash Valley, Ethiopia. *Nature*, **307**, 423–428.

Clarke, R., 2008. Latest information on Sterkfontein's *Australopithecus* skeleton and a new look at *Australopithecus*. *South African Journal of Science*, **104**(11–12), 443–449.

Clarke, R., 2013. *Australopithecus* from Sterkfontein Caves, South Africa. In: Reed, K.E., Fleagle, J.G., and Leakey, R.E.F., (Eds.), *The Paleobiology of Australopithecus*. Springer, Dordrecht, pp. 105–123.

Clarke, R.J., Howell, F.C., and Brain, C.K., 1970. More evidence of an advanced hominid at Swartkrans. *Nature*, **225**(5239), 1219–1222.

Cleary, T.J., Benson, R.B.J., Evans, S.E., and Barrett, P.M., 2018. Lepidosaurian diversity in the Mesozoic–Palaeogene: the potential roles of sampling biases and environmental drivers. *Royal Society Open Science*, **5**, 171830.

Cleary, T.J., Moon, B.C., Dunhill, A.M., and Benton, M.J., 2015. The fossil record of ichthyosaurs, completeness metrics and sampling biases. *Palaeontology*, **58**(3), 521–536.

Close, R.A., Benson, R.B.J., Upchurch, P., and Butler, R.J., 2017. Controlling for the species-area effect supports constrained long-term Mesozoic terrestrial vertebrate diversification. *Nature Communications*, **8**, 15381.

Collard, M., Gibbs, S., and Wood, B.A. Phylogenetic utility of higher primatepostcranial morphology. Paper presented at the American Association of Physical Anthropogists annual meeting, Kansas City, 2001.

Condemi, S., 2004. The Garba IV E mandible. In: Chavaillon, J. and Piperno, M., (Eds.), *Studies on the Early Paleolithic Site of Melka Kunture*. Origines, Florence, pp. 687–701.

Crampton, J.S., Beu, A.G., Cooper, R.A., Jones, C.A., Marshall, B., and Maxwell, P.A., 2003. Estimating the rock volume bias in paleobiodiversity studies. *Science*, **301**(5631), 358–360.

Crampton, J.S., Foote, M., Beu, A.G., Cooper, R.A., Matcham, I., Jones, C.M., Maxwell, P.A., and Marshall, B., 2006. all. 2006. Second-order sequence stratigraphie controls on the quality of the fossil record at an active margin: New Zealand Eocene to Recent shelf molluscs. *Palaios*, **21**(1), 86–105.

Crompton, R.H., Vereecke, E.E., and Thorpe, S.K.S., 2008. Locomotion and posture from the common hominoid ancestor to fully modern hominins, with special reference to the last common panin/hominin ancestor. *Journal of Anatomy*, **212**(4), 501–543.

Cronin, J.E., Boaz, N.T., Stringer, C.B., and Rak, Y., 1981. Tempo and mode in hominid evolution. *Nature*, **292**(5819), 113–122.

Crowley, B.E., 2010. A refined chronology of prehistoric Madagascar and the demise of the megafauna. *Quaternary Science Reviews*, **29**(19–20), 2591–2603.

Dart, R.A., 1925. *Australopithecus africanus*: the man-ape of South Africa. *Nature*, **115**(2884), 195–199.

Darwin, C.R., 1859. *On the Origin of Species by Means of Natural Selection*. John Murray, London.

Davies, T.W., Bell, M.A., Goswami, A., and Halliday, T.J.D., 2017. Completeness of the eutherian mammal fossil record and implications for reconstructing mammal evolution through the Cretaceous/-Paleogene mass extinction. *Paleobiology*, **43**(4), 521–536.

de Heinzelin, J., Clark, J.D., White, T.D., Hart, W.K., Renne, P.R., WoldeGabriel, G., Beyene, Y., and Vrba, E.S., 1999. Environment and behavior of 2.5-million-year-old Bouri hominids. *Science*, **284** (5414), 625–629.

Dean, C.D., Mannion, P.D., and Butler, R.J., 2016. Preservation bias controls the fossil record of pterosaurs. *Palaeontology*, **59**(2), 225–247.

Deino, A.L., 2012. $^{40}Ar/^{39}Ar$ dating of Bed I, Olduvai Gorge, Tanzania, and the chronology of early Pleistocene climate change. *Journal of Human Evolution*, **63**(2), 251–273.

Delezene, L.K., 2015. Modularity of the anthropoid dentition: implications for the evolution of the hominin canine honing complex. *Journal of Human Evolution*, **86**, 1–12.

Dembo, M., Matzke, N.J., Mooers, A.Ø., and Collard, M., 2015. Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. *Proceedings of the Royal Society of London B: Biological Sciences*, **282**(1812), 1–9.

Dembo, M., Radovčić, D., Garvin, H.M., Laird, M.F., Schroeder, L., Scott, J.E., Brophy, J., Ackermann, R.R., Musiba, C.M., de Ruiter, D.J., Mooers, A.Ø., and Collard, M., 2016. The evolutionary relationships and age of *Homo naledi*: an assessment using dated Bayesian phylogenetic methods. *Journal of Human Evolution*, **97**, 17–26.

deMenocal, P.D., 1995. Plio-Pleistocene African climate. *Science*, **270**(5233), 53–59.

deMenocal, P.D., 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, **220**, 3–24.

DeSilva, J.M., Shoreman, E., and MacLatchy, L.M., 2006. A fossil hominoid proximal femur from Kikorongo Crater, Southwestern Uganda. *Journal of Human Evolution*, **50**(6), 687–695.

DiMaggio, E.N., Campisano, C.J., Rowan, J., Dupont-Nivet, G., Deino, A.L., Bibi, F., Lewis, M.E., Souron, A., Garello, D., Werdelin, L., Reed, K.E., and Arrowsmith, J.R., 2015. Late Pliocene fossiliferous sedimentary record and the environmental context of early *Homo* from Afar, Ethiopia. *Science*, **347**(6228), 1355–1359.

Diogo, R., Peng, Z., and Wood, B.A., 2013. First comparative study of primate morphological and molecular evolutionary rates including muscle data: implications for the tempo and mode of primate and human evolution. *Journal of Anatomy*, **222**(4), 410–418.

Dirks, P.H.G.M., Kibii, J.M., Steininger, C., Churchill, S.E., Kramers, J.D., Pickering, R., Farber, D.L., Mériaux, A.-S., Herries, A.I.R., King, G.C.P., and Berger, L.R., 2010. Geological setting and age of *Australopithecus sediba* from southern Africa. *Science*, **328**(5975), 205–208.

Domínguez-Rodrigo, M., Mabulla, A., Bunn, H.T., Barba, R., Diez-Martin, F., Egeland, C.P., Espilez, E., Egeland, A., Yravedra, J., and Sánchez, P., 2009. Unraveling hominin behavior at another anthropogenic site from Olduvai Gorge (Tanzania): new archaeological and taphonomic research at BK, Upper Bed II. *Journal of Human Evolution*, **57**(3), 260–283.

Domínguez-Rodrigo, M., Pickering, T.R., Baquedano, E., Mabulla, A., Mark, D.F., Musiba, C., Bunn, H.T., Uribelarrea, D., Smith, V., Diez-Martin, F., Pérez-González, A., Sánchez, P., Santonja, M., Barboni, D., Gidna, A., Ashley, G., Yravedra, J., Heaton, J. L., , and Arriaza, M.C., 2013. First partial skeleton of a 1.34-million-year-old *Paranthropus boisei* from Bed II, Olduvai Gorge, Tanzania. *PLOS ONE*, **8**(12), e80347.

Driscoll, D.A., Dunhill, A.M., Stubbs, T.L., and Benton, M.J., 2018. The mosasaur fossil record through the lens of fossil completeness. *Palaeontology,* early view.

Dunhill, A.M., 2011. Using remote sensing and a geographic information system to quantify rock exposure area in England and Wales: implications for paleodiversity studies. *Geology*, **39**(2), 111–114.

Dunhill, A.M., 2012. Problems with using rock outcrop area as a paleontological sampling proxy: rock

outcrop and exposure area compared with coastal proximity, topography, land use, and lithology. *Paleobiology*, **38**(1), 126–143.

Dunhill, A.M., Benton, M.J., Newell, A.J., and Twitchett, R.J., 2013. Completeness of the fossil record and the validity of sampling proxies: a case study from the Triassic of England and Wales. *Journal of the Geological Society of London*, **170**(2), 291–300.

Dunhill, A.M., Benton, M.J., Twitchett, R.J., and Newell, A.J., 2012. Completeness of the fossil record and the validity of sampling proxies at outcrop level. *Palaeontology*, **55**(6), 1155–1175.

Dunhill, A.M., Hannisdal, B., and Benton, M.J., 2014*b*. Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nature Communications*, **5**, 4818.

Dunhill, A.M., Hannisdal, B., Brocklehurst, N., and Benton, M.J., 2018. On formation-based sampling proxies and why they should not be used to correct the fossil record. *Palaeontology*, **61**(1), 119–132.

Dunhill, A.M. and Wills, M.A., 2015. Geographic range did not confer resilience to extinction in terrestrial vertebrates at the end-Triassic crisis. *Nature Communications*, **7980**, 6.

Eldredge, N. and Gould, S.J., 1972. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf, T.J.M., (Ed.), *Models in Paleobiology*. Freeman, Cooper and Company, San Francisco, pp. 82–115.

Eldredge, N. and Tattersall, I., 1975. Evolutionary models, phylogenetic reconstruction and another look at hominid phylogeny. In: Szalay, F.S., (Ed.), *Contributions to Primatology 5: Approaches to Primate Paleobiology*. Karger, Basel, pp. 218–242.

Elton, S., 2006. Forty years on and still going strong: the use of hominin-cercopithecid comparisons in palaeoanthropology. *Journal of the Royal Anthropological Institute*, **12**(1), 19–38.

Emonet, E.-G., Andossa, L., Mackaye, H.T., and Brunet, M., 2014. Subocclusal dental morphology of *Sahelanthropus tchadensis* and the evolution of teeth in hominins. *American Journal of Physical Anthropology*, **153**(1), 116–123.

Fara, E., 2002. Sea-level variation and quality of the continental fossil record. *Journal of the Geological Society of London*, **159**(5), 489–491.

Feibel, C.S., 2003. Stratigraphy and depositional setting of the Pliocene Kanapoi Formation, Lower Kerio Valley, Kenya. *Contributions in Science*, **498**, 9–20.

Feibel, C.S., Brown, F.H., and McDougall, I., 1989. Stratigraphic context of fossil hominids from the Omo Group deposits: Nortern Turkana Basin, Kenya and Ethiopia. *American Journal of Physical Anthropology*, **78**(4), 595–622.

Finarelli, J.A. and Clyde, W.C., 2004. Reassessing hominoid phylogeny: evaluating congruence in the morphological and temporal data. *Paleobiology*, **30**(4), 614–651.

Fleagle, J.G., 1995. Too many species? *Evolutionary Anthropology*, **4**(2), 37–38.

Fleagle, J.G., 2002. The primate fossil record. *Evolutionary Anthropology*, **11**(S1), 20–23.

Fleagle, J.G., 2013. *Primate Adaptation and Evolution*. Academic Press, Boston.

Fleagle, J.G., Gilbert, C.C., and Badan, A.L., 2016. Comparing primate crania: the importance of fossils. *American Journal of Physical Anthropology*, **161**(2), 259–275.

Fleagle, J.G. and Lieberman, D.E., 2015. Major transformations in the evolution of primate locomotion. In: Dial, K.P, Shubin, N., and Brainerd, E.L., (Eds.), *Great Transformations in Vertebrate Evolution*. University of Chicago Press, Chicago, pp. 257–278.

Flessa, K.W. and Jablonski, D., 1983. Extinction is here to stay. *Paleobiology*, **9**(4), 315–321.

Foley, R.A., 1991. How many hominid species should there be? *Journal of Human Evolution*, **20**, 413–427.

Foley, R.A., 1994. Speciation, extinction and climatic change in hominid evolution. *Journal of Human Evolution*, **26**, 275–289.

Foley, R.A., 2002. Adaptive radiations and dispersals in hominin evolutionary ecology. *Evolutionary Anthropology*, **11**(S1), 32–37.

Foley, R.A., 2005. Species diversity in human evolution: challenges and opportunities. *Transactions of the Royal Society of South Africa*, **60**(2), 67–72.

Foley, R.A., 2016. Mosaic evolution and the pattern of transitions in the hominin lineage. *Proceedings*

of the Royal Society of London B: Biological Sciences, **371**(1698), 1–14.
Foote, M., 1994. Temporal variation in extinction risk and temporal scaling of extinction metrics. *Paleobiology*, **20**(4), 424–444.
Foote, M., 1996*a*. Perspective: evolutionary patterns in the fossil record. *Evolution*, **50**(1), 1–11.
Foote, M., 1996*b*. On the probability of ancestors in the fossil record. *Paleobiology*, **22**(2), 141–151.
Foote, M., 1997. Estimating taxonomic durations and preservation probability. *Paleobiology*, **23**(3), 278–300.
Foote, M. and Raup, D.M., 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*, **22**(2), 121–140.
Fountaine, T., Benton, M.J., Dyke, G.J., and Nudds, R., 2005. The quality of the fossil record of Mesozoic birds. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**(1560), 289–294.
Friedman, M. and Sallan, L.C., 2012. Five hundred million years of extinction and recovery: a Phanerozoic survey of large-scale diversity patterns in fishes. *Palaeontology*, **55**(4), 707–742.
Fröbisch, J., 2008. Global taxonomic diversity of anomodonts (Tetrapoda, Therapsida) and the terrestrial rock record across the Permian-Triassic boundary. *PLOS ONE*, **3**(11), e3733.
Fröbisch, J., 2013. Vertebrate diversity across the end-Permian mass extinction – Separating biological and geological signals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 50–61.
Frost, S.R., 2007. African Pliocene and Pleistocene cercopithecid evolution and global climatic change. In: Bobe, R., Alemseged, Z., and Behrensmeyer, A.K., (Eds.), *Hominin Environments in the East African Pliocene: An Assessment of the Faunal Evidence*. Springer, Dordrecht, pp. 51–76.
Fuss, J., Spassov, N., Begun, D.R., and Böhme, M., 2017. Potential hominin affinities of *Graecopithecus* from the Late Miocene of Europe. *PLOS ONE*, **12**(5), e0177127.
Galik, K., Senut, B., Pickford, M., Gommery, D., Treil, J., Kuperavage, A.J., and Eckhardt, R.B., 2004. External and internal morphology of the BAR 1002'00 *Orrorin tugenensis* femur. *Science*, **305**(5689), 1450–1453.
Gathogo, P.N. and Brown, F.H., 2006. Stratigraphy of the Koobi Fora Formation (Pliocene and Pleistocene) in the Ileret region of northern Kenya. *Journal of African Earth Sciences*, **45**(4–5), 369–390.
Gavryushkina, A., Heath, T.A., Ksepka, D.T., Stadler, T., Welch, D., and Drummond, A.J., 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic Biology*, **66**(1), 57–73.
Gee, H., 2013. *The Accidental Species: Misunderstandings of Human Evolution*. University of Chicago Press, Chicago.
Gilbert, C.C., Goble, E.D., and Hill, A., 2010. Miocene Cercopithecoidea from the Tugen Hills, Kenya. *Journal of Human Evolution*, **59**(5), 465–483.
Glazko, G.V. and Nei, M., 2003. Estimation of divergence times for major lineages of primate species. *Molecular Biology and Evolution*, **20**(3), 424–434.
Gould, S.J. and Eldredge, N., 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*, **3**(2), 115–151.
Grine, F.E., Ungar, P.S., Teaford, M.F., and El-Zaatari, S., 2013. Molar microwear, diet and adaptation in a purported hominin species lineage from the Pliocene of East Africa. In: Reed, K.E., Fleagle, J.G., and Leakey, M.G., (Eds.), *The Paleobiology of Australopithecus*. Springer, Dordrecht, pp. 213–224.
Grove, M., 2011. Speciation, diversity, and Mode 1 technologies: the impact of variability selection. *Journal of Human Evolution*, **61**(3), 306–319.
Grove, M., 2012. Amplitudes of orbitally induced climatic cycles and patterns of hominin speciation. *Journal of Archaeological Science*, **39**(10), 3085–3094.
Grove, M., 2014. Evolution and dispersal under climatic instability: a simple evolutionary algorithm. *Adaptive Behavior*, **22**(4), 235–254.
Groves, C.P. and Mazák, V., 1975. An approach to the taxonomy of Homindiae: gracile Villafranchian hominids from Africa. *Casopis pro mineralogii a geologii*, **20**(3), 225–247.
Guy, F., Lieberman, D.E., Pilbeam, D.R., Ponce De Leon, M.S., Likius, A., Mackaye, H.T., Vignaud, P., Zollikofer, C.P.E., and Brunet, M., 2005. Morphological affinities of the *Sahelanthropus tchadensis*

(Late Miocene hominid from Chad) cranium. *Proceedings of the National Academy of Sciences*, **102** (52), 18836–18841.

Haile-Selassie, Y., 2001. Late Miocene hominids from the Middle Awash, Ethiopia. *Science*, **412**(6843), 178–181.

Haile-Selassie, Y., Gibert, L., Melillo, S.M., Ryan, T.M., Alene, M., Deino, A.L., Levin, N.E., Scott, G., and Saylor, B.Z., 2015. New species from Ethiopia further expands Middle Pliocene hominin diversity. *Nature*, **521**(7553), 483–488.

Haile-Selassie, Y., Latimer, B., Alene, M., Deino, A.L., Gibert, L., Melillo, S.M., Saylor, B.Z., Scott, G.R., and Lovejoy, C.O., 2010. An early *Australopithecus afarensis* postcranium from Woranso-Mille, Ethiopia. *Proceedings of the National Academy of Sciences*, **107**(27), 12121–12126.

Haile-Selassie, Y., Melillo, S.M., and Su, D.F., 2016. The Pliocene hominin diversity conundrum: do more fossils mean less clarity? *Proceedings of the National Academy of Sciences*, **113**(23), 6364–6371.

Haile-Selassie, Y., Suwa, G., and White, T.D., 2004. Late Miocene teeth from Middle Awash, Ethiopia, and early hominid dental evolution. *Science*, **303**(5663), 1503 –1505.

Haile-Selassie, Y., Suwa, G., and White, T.D., 2009. Hominidae. In: Haile-Selassie, Y. and WoldeGabriel, G., (Eds.), *Ardipithecus kadabba: Late Miocene Evidence from the Middle Awash, Ethiopia*. University of California Press, Berkeley, pp. 159–236.

Hammer, Ø. and Harper, D.A.T., 2006. *Palaeontological Data Analysis*. Blackwell Publishing, Oxford.

Hammer, Ø., Harper, D.A.T., and Ryan, P.D., 2001. PAST: Paleontological Statistics software package for education and data analysis. *Palaeontologia Electronica*, **4**(1), 1–9.

Hannisdal, B., 2011. Detecting common-cause relationships with directional information transfer. In: McGowan, A.J. and Smith, A.B., (Eds.), *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society of London Special Publications, **358**, pp. 19–30.

Hannisdal, B. and Liow, L.H., 2018. Causality from palaeontological time series. *Palaeontology*, **61**(4), 495–509.

Hannisdal, B. and Peters, S.E., 2011. Phanerozoic Earth system evolution and marine biodiversity. *Science*, **334**(6059), 1121–1124.

Harcourt-Smith, W.E.H., 2007. The origins of bipedal locomotion. In: Henke, W., Hardt, T., and Tattersall, I., (Eds.), *Handbook of Paleoanthropology*. Springer, New York, pp. 1483–1518.

Harcourt-Smith, W.E.H. and Aiello, L.C., 2004. Fossils, feet and the evolution of human bipedal locomotion. *Journal of Anatomy*, **204**(5), 403–416.

Harris, J.M., Brown, F.H., and Leakey, M.G., 1998a. Stratigraphy and paleontology of the Nachukui Formation, Lake Turkana Region, Kenya. *Contributions in Science*, **399**, 1–128.

Harris, J.M., Brown, F.H., Leakey, M.G., Walker, A.C., and Leakey, R.E.F., 1988b. Plio-Pleistocene hominid bearing sites of west of Lake Turkana, Kenya. *Science*, **139**(4835), 27–33.

Harrison, T., 2010*a*. Apes among the tangled branches of human origins. *Science*, **327**(5965), 532–534.

Harrison, T., 2010*b*. Late Tertiary lorisiformes. In: Werdelin, L. and Sanders, W.J., (Eds.), *Cenozoic Mammals of Africa*. University of California Press, Berkeley, pp. 333–350.

Harrison, T., (Ed.), 2011*a*. *Paleontology and Geology of Laetoli: Human Evolution in Context. Volume 1: Geology, Geochronology, Paleoecology and Paleoenvironment*. Springer, Dordrecht.

Harrison, T., (Ed.), 2011*b*. *Paleontology and Geology of Laetoli: Human Evolution in Context. Volume 2: Fossil Hominins and the Associated Fauna*. Springer, Dordrecht.

Hedman, M.M., 2010. Constraints on clade ages from fossil outgroups. *Paleobiology*, **36**(1), 16–31.

Herries, A.I.R. and Adams, J.W., 2013. Clarifying the context, dating and age range of the Gondolin hominins and *Paranthropus* in South Africa. *Journal of Human Evolution*, **65**(5), 676–681.

Herries, A.I.R., Adams, J.W., Kuykendall, K.L., and Shaw, J., 2006. Speleology and magnetobiostratigraphic chronology of the GD2 locality of the Gondolin hominin-bearing paleocave deposits, North West Province, South Africa. *Journal of Human Evolution*, **51**(6), 617–631.

Herries, A.I.R., Curnoe, D., and Adams, J.W., 2009. A multi-disciplinary seriation of early *Homo* and *Paranthropus* bearing palaeocaves in southern Africa. *Quaternary International*, **202**(1–2), 14–28.

Herries, A.I.R. and Shaw, J., 2011. Palaeomagnetic analysis of the Sterkfontein palaeocave deposits: implications for the age of the hominin fossils and stone tool industries. *Journal of Human Evolution*, **60**(5), 523–539.

Hill, A., 2007. Preface. In: Bobe, R., Alemseged, Z., and Behrensmeyer, A.K., (Eds.), *Hominin Environments in the East African Pliocene: An Assessment of the Faunal Evidence*. Springer, Dordrecht, pp. xvii–xx.

Hlusko, L.J., 2007. A new Late Miocene species of *Paracolobus* and other Cercopithecoidea (Mammalia: Primates) fossils from Lemudong'o, Kenya. *Kirtlandia*, **56**, 72–85.

Hooker, P.J. and Miller, J.A., 1979. K-Ar dating of the Pleistocene fossil hominid site at Chesowanja, North Kenya. *Nature*, **282**(5740), 710–712.

Hopley, P.J., 2018, *in review*. Environmental, stratigraphic and taxonomic bias in the hominin fossil record: implications for theories of the climatic forcing of human evolution. In: Reynolds, S.C. and Bobe, R., (Eds.), *African Paleoenvironments*. Cambridge University Press, Cambridge.

Hopley, P.J., Herries, A.I.R., Baker, S.E., Kuhn, B.F., and Menter, C.G., 2013. Beyond the South African cave paradigm—*Australopithecus africanus* from Plio-Pleistocene paleosol deposits at Taung. *American Journal of Physical Anthropology*, **151**(2), 316–324.

Hopley, P.J. and Maslin, M.A., 2010. Climate-averaging of terrestrial faunas: an example from the Plio-Pleistocene of South Africa. *Paleobiology*, **36**(1), 32–50.

Hublin, J.-J., Neubauer, S., and Gunz, P., 2015. Brain ontogeny and life history in Pleistocene hominins. *Proceedings of the Royal Society of London B: Biological Sciences*, **370**(1663), 20140062.

Hughes, A.R. and Tobias, P.V., 1977. A fossil skull probably of the genus *Homo* from Sterkfontein, Transvaal. *Nature*, **265**(5592), 310–312.

Hunt, A.P., Lockley, M.G., Lucas, S.G., and Meyer, C.A., 1994. The global sauropod fossil record. *Gaia*, **10**, 261–279.

Hunt, G., 2010. Evolution in fossil lineages: paleontology and *The Origin of Species*. *The American Naturalist*, **176**, S61–S76.

Hunt, K. and Vitzthum, V.J., 1986. Dental metric assessment of the Omo fossils: implications for the phylogenetic position of *Australopithecus africanus*. *American Journal of Physical Anthropology*, **71**(2), 141–155.

Huxley, J.S., 1958. Evolutionary processes and taxonomy with special reference to grades. *Uppsala University Arsskrift*, **6**, 21–38.

Huxley, J.S., 1959. Clades and grades. In: *Function and Taxonomic Importance*. Systematics Association, London, pp. 21–22.

Ingalls, N.W., 1931. Observations on bone weights. *American Journal of Anatomy*, **48**(1), 45–98.

Ingalls, N.W., 1932. Observations on bone weights. II. The bones of the foot. *American Journal of Anatomy*, **50**(3), 435–450.

Irish, J.D., Guatelli-Steinberg, D., Legge, S.S., de Ruiter, D.J., and Berger, L.R., 2013. Dental morphology and the phylogenetic "place" of *Australopithecus sediba*. *Science*, **340**(6129), 1233062.

Jablonski, N.G. and Frost, S.R., 2010. Cercopithecoidea. In: Werdelin, L. and Sanders, W.J., (Eds.), *Cenozoic Mammals of Africa*. University of California Press, Berkeley, pp. 393–428.

Johanson, D.C., Lovejoy, C.O., Kimbel, W.H., White, T.D., Ward, S.C., Bush, M.E., Latimer, B.M., and Coppens, Y., 1982. Morphology of the Pliocene partial hominid skeleton (A.L. 288-1) from the Hadar Formation, Ethiopia. *American Journal of Physical Anthropology*, **57**(4), 403–451.

Johanson, D.C., Masao, F.T., Eck, G.G., White, T.D., Walter, R.C., Kimbel, W.H., Asfaw, B., Manega, P., Ndessokia, P., and Suwa, G., 1987. New partial skeleton of *Homo habilis* from Olduvai Gorge, Tanzania. *Nature*, **327**(6119), 205–209.

Johanson, D.C., White, T.D., and Coppens, Y., 1978. A new species of the genus *Australopithecus* (Primates: Hominidae) from the Pliocene of Eastern Africa. *Kirtlandia*, **28**, 1–14.

Johnson, J.B. and Omland, K.S., 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution*, **19**(2), 101–108.

Jolly, C.J., 1970. The seed-eaters: a new model of hominid differentiation based on a baboon analogy. *Man*, **5**(1), 5–26.

Kaiser, T.M., Seiffert, C., Hertler, C., Fiedler, L., Schwartz, H.L., Frost, S.R., Giemsch, L., Bernor, R.L., Wolf, D., Semprebon, G., Nelson, S.V., Schrenk, F., Harvati, K., Bromage, T.G., and Saanane, C., 2010. Makuyuni, a new Lower Palaeolithic hominid site in Tanzania. *Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut*, **106**, 69–110.

Kappelman, J.W., Swisher, C.C., Fleagle, J.G., Yirga, S., Brown, T.M., and Feseha, M., 1996. Age of *Australopithecus afarensis* from Fejej, Ethiopia. *Journal of Human Evolution*, **30**(2), 139–146.

Katoh, S., Nagaoka, S., WoldeGabriel, G., Renne, P.R., Snow, M.G., Beyene, Y., and Suwa, G., 2000. Chronostratigraphy and correlation of the Plio-Pleistocene tephra layers of the Konso Formation, southern Main Ethiopian Rift, Ethiopia. *Quaternary Science Reviews*, **19**(13), 1305–1317.

Kelley, J., 2001. Hominoid evolution and climate change in Europe: phylogeny of the Neogene hominoid primates of Eurasia. In: De Bonis, L., Koufos, G.D., and Andrews, P., (Eds.), *Hominoid Evolution and Climatic Change in Europe, Phylogeny of the Neogene Hominoid Primates of Eurasia*. Cambridge University Press, Cambridge, pp. 269–271.

Kimbel, W.H., 1988. Identification of a partial cranium of *Australopithecus afarensis* from the Koobi Fora Formation, Kenya. *Journal of Human Evolution*, **17**(7), 649–656.

Kimbel, W.H., 1995. Hominid speciation and Pliocene climatic change. In: Vrba, E.S., Denton, G.H., Partridge, T.C., and Burckle, L.H., (Eds.), *Paleoclimate and Evolution, with Emphasis on Human Origins*. Yale University Press, New Haven, pp. 425–437.

Kimbel, W.H., 2007. The species and diversity of australopiths. In: Henke, W., Hardt, T., and Tattersall, I., (Eds.), *Handbook of Paleoanthropology*. Springer, New York, pp. 1539–1574.

Kimbel, W.H., 2009. The origin of *Homo*. In: Grine, F.E., Fleagle, J.G., and Leakey, R.E.F., (Eds.), *The First Humans: Origin and Early Evolution of the Genus Homo*. Springer, Dordrecht, pp. 31–38.

Kimbel, W.H. and Delezene, L.K., 2009. "Lucy" redux: a review of research on *Australopithecus afarensis*. *American Journal of Physical Anthropology*, **140**(S49), 2–48.

Kimbel, W.H., Johanson, D.C., and Rak, Y., 1997. Systematic assessment of a maxilla of *Homo* from Hadar, Ethiopia. *American Journal of Physical Anthropology*, **103**(2), 235–262.

Kimbel, W.H., Lockwood, C.A., Ward, C.V., Leakey, M.G., Rak, Y., and Johanson, D.C., 2006. Was *Australopithecus anamensis* ancestral to *A. afarensis*? A case of anagenesis in the hominin fossil record. *Journal of Human Evolution*, **51**(2), 134–152.

Kimbel, W.H. and Rak, Y., 2010. The cranial base of *Australopithecus afarensis*: new insights from the female skull. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365** (1556), 3365–3376.

Kimbel, W.H. and Rak, Y., 2017. *Australopithecus sediba* and the emergence of *Homo*: questionable evidence from the cranium of the juvenile holotype MH1. *Journal of Human Evolution*, **107**, 94–106.

Kimbel, W.H., Rak, Y., and Johanson, D.C., 2004. *The Skull of Australopithecus afarensis*. Oxford University Press, Oxford.

Kimbel, W.H., Suwa, G., Asfaw, B., Rak, Y., and White, T.D., 2014. *Ardipithecus ramidus* and the evolution of the human cranial base. *Proceedings of the National Academy of Sciences*, **111**(3), 948–953.

Kimbel, W.H. and Villmoare, B.A., 2016. From *Australopithecus* to *Homo*: the transition that wasn't. *Proceedings of the Royal Society of London B: Biological Sciences*, **37**, 20150248.

Kingston, J.D., 2007. Shifting adaptive landscapes: progress and challenges in reconstructing early hominid environments. *Yearbook of Physical Anthropology*, **134**, 20–58.

Kissel, M. and Hawks, J., 2015. What are the Lothagam and Tabarin Mandibles? *PaleoAnthropology*, pp. 37–43.

Kleinsasser, L.L., Quade, J., McIntosh, W.C., Levin, N.E., Simpson, S.W., and Semaw, S., 2008. Stratigraphy and geochronology of the late Miocene Adu-Asa Formation at Gona, Ethiopia. In: Quade, J. and Wynn, J.G., (Eds.), *The Geology of Early Humans in the Horn of Africa*. Geological Society of America Special Paper, Boulder, **446**, pp. 33–5.

Kowalewski, M., 1996. Time-averaging, overcompleteness, and the geological record. *Journal of Geology*, **104**(3), 317–326.

Kullmer, O., 2008. The Fossil Suidae from the Plio-Pleistocene Chiwondo Beds of Northern Malawi, Africa. *Journal of Vertebrate Paleontology*, **28**(1), 208–216.

Lane, A., Janis, C.M., and Sepkoski, J.J., 2005. Estimating paleodiversities: a test of the taxic and phylogenetic methods. *Paleobiology*, **31**(1), 21–34.

Langergraber, K.E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyamag, M., Mitani, J.C., Muller, M.N., Robbins, M.M., Schubert, G., Stoinski, T.S., Viola, B., Watts, D., Wittig, R.M., Wrangham, R.W., Zuberbühler, K., Pääbo, S., and Vigilant, L., 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proceedings of the National Academy of Sciences*, **109**(39), 15716–15721.

Laurin, M., 2004. The evolution of body size, Cope's Rule and the origin of amniotes. *Systematic Biology*, **53**(4), 594–622.

Leakey, L.S.B., 1959. A new fossil skull from Olduvai. *Nature*, **184**, 491–493.

Leakey, L.S.B., Tobias, P.V., and Napier, J.R., 1964. A new species of the genus *Homo* from Olduvai Gorge. *Nature*, **202**, 7–9.

Leakey, M.D., Clarke, R.J., and Leakey, L.S.B., 1971. New hominid skull from Bed I, Olduvai Gorge, Tanzania. *Nature*, **232**(5309), 308–312.

Leakey, M.G., Feibel, C.S., McDougall, I., and Walker, A.C., 1995. New four-million-year-old hominid species from Kanapoi and Allia Bay, Kenya. *Nature*, **376**(6541), 565–571.

Leakey, M.G., Feibel, C.S., McDougall, I., Ward, C.V., and Walker, A.C., 1998. New specimens and confirmation of an early age for *Australopithecus anamensis*. *Nature*, **393**(6680), 62–66.

Leakey, M.G., Spoor, F., Brown, F.H., Gathogo, P.N., Kiarie, C., Leakey, L.N., and McDougall, I., 2001. New hominin genus from eastern Africa shows diverse middle Pliocene lineages. *Nature*, **410**(6827), 433–440.

Leakey, M.G., Spoor, F., Dean, M.C., Feibel, C.S., Antón, S.C., Kiarie, C., and Leakey, L.N., 2012. New fossils from Koobi Fora in northern Kenya confirm taxonomic diversity in early *Homo*. *Nature*, **488** (7410), 201–204.

Leakey, M.G. and Walker, A.C., 2003. The Lothagam hominds. In: Leakey, M.G. and Harris, J.M., (Eds.), *Lothagam: The Dawn of Humanity in Eastern Africa*. Columbia University Press, New York, pp. 249–260.

Lebatard, A.-E., Bourlès, D.L., Duringer, P., Jolivet, M., Braucher, R., Carcaillet, J., Schuster, M., Arnaud, N., Monié, P., Lihoreau, F., Likius, A., Mackaye, H.T., Vignaud, P., and Brunet, M., 2008. Cosmogenic nuclide dating of *Sahelanthropus tchadensis* and *Australopithecus bahrelghazali*: Mio-Pliocene hominids from Chad. *Proceedings of the National Academy of Sciences*, **105**(9), 3226–3231.

Levin, N.E., Haile-Selassie, Y., Frost, S.R., and Saylor, B.Z., 2015. Dietary change among hominins and cercopithecids in Ethiopia during the early Pliocene. *Proceedings of the National Academy of Sciences*, **112**(40), 12304–12309.

Levinton, J., 1988. *Genetics, Paleontology, and Macroevolution*. Cambridge University Press, Cambridge.

Lieberman, D.E., Ross, C.F., and Ravosa, M.J., 2000. The primate cranial base: ontogeny, function, and integration. *American Journal of Physical Anthropology*, **31**, 117–169.

Lieberman, D.E., Wood, B.A., and Pilbeam, D.R., 1996. Homoplasy and early *Homo*: an analysis of the evolutionary relationships of *H. habilis sensu stricto* and *H. rudolfensis*. *Journal of Human Evolution*, **30**, 97–120.

Lloyd, G.T., 2012. A refined modelling approach to assess the influence of sampling on palaeobiodiversity curves: new support for declining Cretaceous dinosaur richness. *Biology Letters*, **8**(1), 123–126.

Lloyd, G.T., Bapst, D.W., Friedman, M., and Davis, K.E., 2016. Probabilistic divergence time estimation without branch lengths: dating the origins of dinosaurs, avian flight and crown birds. *Biology Letters*, **12**(11), 20160609.

Lloyd, G.T., Davis, K.E., Pisani, D., Tarver, J.E., Ruta, M., Sakamoto, M., Hone, D.W.E., Jennings, R., and

Benton, M.J., 2008. Dinosaurs and the Cretaceous terrestrial revolution. *Proceedings of the Royal Society of London B: Biological Sciences*, **275**(1650), 2483–2490.

Lloyd, G.T. and Friedman, M., 2013. A survey of palaeontological sampling biases in fishes based on the Phanerozoic record of Great Britain. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 5–17.

Lloyd, G.T., Young, J.R., and Smith, A.B., 2012*a*. Taxonomic structure of the fossil record is shaped by sampling bias. *Systematic Biology*, **61**(1), 80–89.

Lockwood, C.A. and Fleagle, J.G., 1999. The recognition and evaluation of homoplasy in primate and human evolution. *Yearbook of Physical Anthropology*, **42**, 189–232.

Lovejoy, C.O., Latimer, B., Suwa, G., Asfaw, B., and White, T.D., 2009*c*. Combining prehension and propulsion: the foot of *Ardipithecus ramidus*. *Science*, **326**(5949), 72e1–72e8.

Lovejoy, C.O., Meindl, R.S., Ohman, J.C., Heiple, K.G., and White, T.D., 2002. The Maka femur and its bearing on the antiquity of human walking: applying contemporary concepts of morphogenesis to the human fossil record. *American Journal of Physical Anthropology*, **119**(2), 97–133.

Lovejoy, C.O., Simpson, S.W., White, T.D., Asfaw, B., and Suwa, G., 2009*a*. Careful climbing in the Miocene: the forelimbs of *Ardipithecus ramidus* and humans are primitive. *Science*, **326**(5949), 70e1–70e8.

Lovejoy, C.O., Suwa, G., Simpson, S.W., Matternes, J.H., and White, T.D., 2009*d*. The great divides: *Ardipithecus ramidus* reveals the postcrania of our last common ancestors with African apes. *Science*, **326**(5949), 100–106.

Lovejoy, C.O., Suwa, G., Spurlock, L., Asfaw, B., and White, T.D., 2009*b*. The pelvis and femur of *Ardipithecus ramidus*: the emergence of upright walking. *Science*, **326**(5949), 71e1–71e6.

Lowrance, E.W. and Latimer, H.B., 1967. Weights and variability of components of the human vertebral column. *The Anatomical Record*, **159**(1), 83–88.

MacLatchy, L.M., DeSilva, J.M., Sanders, W.J., and Wood, B.A., 2010. Hominini. In: Werdelin, L. and Sanders, W.J., (Eds.), *Cenozoic Mammals of Africa*. University of California Press, Berkeley, pp. 471–540.

Mannion, P.D., Benson, R.B.J., Carrano, M.T., Tennant, J.P., Judd, J., and Butler, R.J., 2015. Climate constrains the evolutionary history and biodiversity of crocodylians. *Nature Communications*, **6**, 8438.

Mannion, P.D. and Upchurch, P., 2010. Completeness metrics and the quality of the sauropodomorph fossil record through geological and historical time. *Paleobiology*, **36**(2), 283–302.

Mannion, P.D., Upchurch, P., Benson, R.B.J., and Goswami, A., 2014. The latitudinal biodiversity gradient through deep time. *Trends in Ecology and Evolution*, **29**(1), 42–50.

Mannion, P.D., Upchurch, P., and Carrano, M.T., 2011. Testing the effect of the rock record on diversity: a multidisciplinary approach to elucidating the generic richness of sauropodomorph dinosaurs through time. *Biological Reviews*, **86**(1), 157–181.

Marshall, C.R., 1990. Confidence intervals on stratigraphic ranges. *Paleobiology*, **16**(1), 1–10.

Marshall, C.R., 1994. Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. *Paleobiology*, **20**(4), 459–469.

Marshall, C.R., 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology*, **23**(2), 165–173.

Marshall, C.R., 2017. Five palaeobiological laws needed to understand the evolution of the living biota. *Nature Ecology and Evolution*, **1**, 0165.

Martin, R.D., 1993. Primate origins: plugging the gaps. *Nature*, **363**(6426), 223–234.

Marx, F.G., 2009. Marine mammals through time: when less is more in studying palaeodiversity. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**(1658), 887–892.

Marx, F.G. and Uhen, M.D., 2010. Climate, critters, and cetaceans: Cenozoic drivers of the evolution of modern whales. *Science*, **327**(5968), 993–996.

Maslin, M.A., Brierley, C.M., Milner, A.M., Shultz, S., Trauth, M.H., and Wilson, K.E., 2014. East African climate pulses and early human evolution. *Quaternary Science Reviews*, **101**, 1–17.

Maslin, M.A. and Trauth, M.H., 2009. Plio-Pleistocene East African pulsed climate variability and its influence on early human evolution. In: Grine, F.E., Leakey, R.E.F., and Fleagle, J.G., (Eds.), *The First Humans: Origins of the Genus Homo*. Springer, Dordrecht, pp. 151–158.

Matzke, N.J., Schraiber, J.G., Collard, M., Dembo, M., and Yang, M.A. Tigher estimation of hominoid divergence times by hierarchical Bayesian analysis of dated fossil morphology and incompletely sorted genes. Paper presented at the Society for Molecular Biology and Evolution annual meeting, Chicago, 2013.

Maxwell, S.J., Hopley, P.J., Upchurch, P., and Soligo, C., 2016. The completeness of the early hominin fossil record. *Proceedings of the European Society for the study of Human Evolution*, **6**, 160.

Maxwell, S.J., Hopley, P.J., Upchurch, P., and Soligo, C., 2018. Sporadic sampling, not climate forcing, drives observed early hominin diversity. *Proceedings of the National Academy of Sciences*, **115**(19), 4891–4896.

Maxwell, S.J., Hopley, P.J., Upchurch, P., and Soligo, C., 2018, *in preparation*. The completeness of the early hominin fossil record: implcations for diversity patterns and the origin of Hominini.

Mbua, E., Kusaka, S., Kunimatsu, Y., Geraads, D., Sawada, Y., Brown, F.H., Sakai, T., Boisserie, J.-R., Saneyoshi, M., Omuombo, C., Muteti, S., Hirata, T., Hayashida, A., Iwano, H., Danhara, T., Bobe, R., Jicha, B., and Nakatsukasa, M., 2016. Kantis: a new *Australopithecus* site on the shoulders of the Rift Valley near Nairobi, Kenya. *Journal of Human Evolution*, **94**, 28–44.

McBrearty, S. and Jablonski, N.G., 2005. First fossil chimpanzee. *Nature*, **437**(7055), 105–108.

McDougall, I. and Brown, F.H., 2006. Precise $^{40}Ar/^{39}Ar$ geochronology for the upper Koobi Fora Formation, Turkana Basin, northern Kenya. *Journal of the Geological Society of London*, **163**(1), 205–220.

McDougall, I., Brown, F.H., Vasconcelos, P.M., Cohen, B.E., Thiese, D.S., and Buchanan, M.J., 2012. New single crystal $^{40}Ar/^{39}Ar$ ages improve time scale for deposition of the Omo Group, Omo-Turkana Basin, East Africa. *Journal of the Geological Society of London*, **169**(2), 213–226.

McDougall, I. and Feibel, C.S., 1999. Numerical age control for the Miocene-Pliocene succession at Lothagam, a hominoid-bearing sequence in the northern Kenya Rift. *Journal of the Geological Society of London*, **156**(4), 731–745.

McGowan, A.J. and Smith, A.B., 2008. Are global Phanerozoic marine diversity curves truly global? A study of the relationship between regional rock records and global Phanerozoic marine diversity. *Paleobiology*, **34**(1), 80–103.

McGowan, A.J. and Smith, A.B., (Eds.), 2011. *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society of London Special Publications, London.

McHenry, H.M. and Corrucchini, R.S., 1980. Late Tertiary hominoids and human origins. *Nature*, **285**, 397–398.

McKee, J.K., 1996. Faunal turnover patterns in the Pliocene and Pleistocene of southern Africa. *South African Journal of Science*, **92**, 111–113.

McKee, J.K., 2001. Faunal turnover rates and mammalian biodiversity of the late Pliocene and Pleistocene of eastern Africa. *Paleobiology*, **27**(3), 500–511.

McKinney, M.L., 1990. Classifying and analyzing evolutionary trends. In: McNamara, K.J., (Ed.), *Evolutionary Trends*. Belhaven, London, pp. 28–58.

Miller, A.I. and Foote, M., 1996. Calibrating the Ordovician Radiation of marine life: implications for Phanerozoic diversity trends. *Paleobiology*, **22**(2), 304–309.

Mongle, C.S., Strait, D.S., and Grine, F.E., 2018, *in review*. Phylogenetic implications of new craniodental character data for *Ardipithecus ramidus*. *Journal of Human Evolution*.

Mounce, R.C.P., Sansom, R.S., and Wills, M.A., 2016. Sampling diverse characters improves phylogenies: craniodental and postcranial characters of vertebrates often imply different trees. *Evolution*, **70** (3), 666–686.

Nagelkerke, N.J.D., 1991. A note on a general de nition of the coe cient of determination. *Biometrika*, **78**(3), 691–692.

Nee, S., 2006. Birth-death models in macroevolution. *Annual Review of Ecology, Evolution, and Sys-*

tematics, **37**, 1–17.

Nengo, I., Tafforeau, P., Gilbert, C.C., Fleagle, J.G., Miller, E.R., Feibel, C.S., Fox, D.L., Feinberg, J., Pugh, K.D., Berruyer, C., Mana, S., Engle, Z., and Spoor, F., 2017. New infant cranium from the African Miocene sheds light on ape evolution. *Nature*, **548**(7666), 169–174.

Newham, E., Benson, R.B.J., Upchurch, P., and Goswami, A., 2014. Mesozoic mammaliaform diversity: the effect of sampling corrections on reconstructions of evolutionary dynamics. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **412**, 32–44.

Noble, S.N., 2009. How does multiple testing correction work? *Nature Biotechnology*, **27**(12), 1135–1137.

Norell, M.A., 1992. Taxic origin and temporal diversity: the effect of phylogeny. In: Novacek, M.J. and Wheeler, Q.D., (Eds.), *Extinction and Phylogeny*. Columbia University Press, New York, pp. 89–118.

Norell, M.A., 1993. Tree-based approaches to understanding history: comments on ranks, rules, and the quality of the fossil record. *American Journal of Science*, **293**(A), 407–417.

Norell, M.A. and Novacek, M.J., 1992. The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. *Science*, **255**(5052), 1690–1693.

Noto, C.R., 2010. Hierarchical control of terrestrial vertebrate taphonomy over space and time: discussion of mechanisms and implications for vertebrate paleobiology. In: Allison, P.J. and Bottjer, D.J., (Eds.), *Taphonomy: Process and Bias Through Time. Topics in Geobiology*. Springer, Dordrecht, **32**, pp. 287–336.

Oakley, K.P., Campbell, B.G., and Molleson, T.I., 1977. *Catalogue of Fossil Hominids. Part I: Africa*. Trustees of the British Museum (Natural History), London.

Ohman, J.C., Lovejoy, C.O., White, T.D., Eckhardt, R.B., Galik, K., and Kuperavage, A.J., 2005. Questions about *Orrorin* femur. *Science*, **307**(5711), 845.

Owen, R.B., Potts, R., Behrensmeyer, A.K., and Ditchfield, P., 2008. Diatomaceous sediments and environmental change in the Pleistocene Olorgesailie Formation, southern Kenya Rift Valley. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **269**, 17–37.

Partridge, T.C., 1978. Re-appraisal of lithostratigraphy of Sterkfontein hominid site. *Nature*, **275**(5678), 282–287.

Partridge, T.C., Granger, D.E., Caffee, M.W., and Clarke, R., 2003. Lower Pliocene hominid remains from Sterkfontein. *Science*, **300**(5619), 607–612.

Passey, B.H., Levin, N.E., Cerling, T.E., Brown, F.H., and Eiler, J.M., 2010. High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates. *Proceedings of the National Academy of Sciences*, **107**(25), 11245–11249.

Patterson, D.B., Faith, J.T., Bobe, R., and Wood, B.A., 2014. Regional diversity patterns in African bovids, hyaenids, and felids during the past 3 million years: the role of taphonomic bias and implications for the evolution of *Paranthropus*. *Quaternary Science Reviews*, **96**, 9–22.

Patterson, N., Richter, D.J., Gnerre, S., Lander, E.S., and Reich, D., 2006. Genetic evidence for complex speciation of humans and chimpanzees. *Nature*, **441**(7097), 1103–1108.

Pattinson, D.J., Thompson, R.S., Piotrowski, A.K., and Asher, R.J., 2015. Phylogeny, paleontology, and primates: do incomplete fossils bias the tree of life? *Systematic Biology*, **64**(2), 169–186.

Paul, C.R.C, 1991. Completeness of the fossil record. In: Briggs, D.E.G. and Crowther, P.R., (Eds.), *Palaeobiology: a Synthesis*. Blackwell Scientific, Oxford, pp. 298–303.

Paul, C.R.C, 1998. Adequacy, completeness and the fossil record. In: Donovan, S.K. and Paul, C.R.C, (Eds.), *The Adequacy of the Fossil Record*. Wiley, New York, pp. 1–22.

Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M.A. M., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M.P.C., Silva, A., O'Brien, S.J., and Pecon-Slattery, J., 2011. A molecular phylogeny of living primates. *PLOS Genetics*, **7**(3), e1001342.

Peters, S.E., 2005. Geologic constraints on the macroevolutionary history of marine animals. *Proceedings of the National Academy of Sciences*, **102**(35), 12326–12331.

Peters, S.E., 2006. Genus extinction, origination, and the durations ofsedimentary hiatuses. *Paleobiology*, **32**(3), 387–407.

Peters, S.E. and Foote, M., 2001. Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology*, **27** (4), 583–601.

Peters, S.E. and Foote, M., 2002. Determinants of extinction in the fossil record. *Nature*, **416**(6879), 420–424.

Peters, S.E. and Heim, N.A., 2010. The geological completeness ofpaleontological sampling in North America. *Paleobiology*, **36**(1), 61–79.

Pickering, R., Kramers, J.D., Hancox, P.J., de Ruiter, D.J., and Woodhead, J.D., 2011. Contemporary flowstone development links early hominin bearing cave deposits in South Africa. *Earth and Planetary Science Letters*, **306**(1), 23–32.

Pickford, M., 1975. Late Miocene sediments and fossils from the Northern Kenya Rift Valley. *Nature*, **256**, 279–284.

Pickford, M., 2012. *Orrorin* and the African ape/hominid dichotomy. In: Reynolds, S.C. and Gallagher, A., (Eds.), *African Genesis: Perspectives on Hominin Evolution*. Cambridge University Press, Cambridge, pp. 99–119.

Pickford, M. and Senut, B., 2001. The geological and faunal context of Late Miocene hominid remains from Lukeino, Kenya. *Comptes Rendus de l'Académie des Sciences*, **332**(2), 145–152.

Pickford, M. and Senut, B., 2004. Hominoid teeth with chimpanzee- and gorilla-like features from the Miocene of Kenya: implications for the chronology of ape-human divergence and biogeography of Miocene hominoids. *Anthropological Science*, **113**(1), 95–102.

Pickford, M., Senut, B., Gommery, D., and Treil, J., 2002. Bipedalism in *Orrorin tugenensis* revealed by its femora. *Comptes Rendus Palevol*, **1**(4), 191–203.

Pickford, M., Senut, B., and Hadoto, D., 1993. *Geology and Palaeobiology of the Albertine Rift Valley Uganda-Zaire*. CIFEG Occasional Publications, Orléans.

Pickford, M., Senut, B., Morales, J., and Braga, J., 2008. First hominoid from the Late Miocene of Niger. *South African Journal of Science*, **104**(9–10), 337–339.

Pilbeam, D.R., 1996. Genetic and morphological records of the Hominoidea and hominid origins: a synthesis. *Molecular Phylogenetics and Evolution*, **5**(1), 155–168.

Potts, R., 1996. Evolution and climate variability. *Science*, **273**(5277), 922–923.

Potts, R., 1998. Variability selection in hominid evolution. *Evolutionary Anthropology*, **7**(3), 81–96.

Potts, R., 2013. Hominin evolution in settings of strong environmental variability. *Quaternary Science Reviews*, **73**, 1–13.

Potts, R., Behrensmeyer, A.K., Deino, A.L., Ditchfield, P., and Clark, J., 2004. Small mid-Pleistocene hominin associated with East African Acheulean technology. *Science*, **305**(5680), 75–78.

Potts, R. and Faith, J.T., 2015. Alternating high and low climate variability: the context of natural selection and speciation in Plio-Pleistocene hominin evolution. *Journal of Human Evolution*, **87**, 5–20.

Pozzi, L., Hodgson, J.A., Burrell, A.S., Sterner, K.N., Raaum, R.L., and Disotell, T.R., 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **75**, 165–183.

Prang, T.C., 2016. The subtalar joint complex of *Australopithecus sediba*. *Journal of Human Evolution*, **90**, 105–119.

Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J.R., Walenz, B., Koren, S., Sutton, G., Kodira, C., Winer, R., Knight, J.R., Mullikin, J.C., Meader, S.J., Ponting, C.P., Lunter, G., Higashino, S., Hobolth, A., Dutheil, J., Karakoç, E., Alkan, C., Sajjadian, S., Catacchio, C.R., Ventura, M., Marques-Bonet, T., Eichler, E.E., André, C., Atencia, R., Mugisha, L., Junhold, J., Patterson, N., Siebauer, M., Good, J.M., Fischer, A., Ptak, S.E., Lachmann, M., Symer, D.E., Mailund, T., Schierup, M.H., Andrés, A.M., Kelso, J., and Pääbo, S., 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature*, **486**(7404), 527–531.

Pugh, K.D. Phylogenetic analysis of extant hominoid postcranial characters recovers molecular clades. Paper presented at the American Association of Physical Anthropologists annual meeting, Atlanta, 2016.

Pugh, K.D. and Gilbert, C.C., 2018. Phylogenetic relationships of living and fossil African papionins: combined evidence from morphology and molecules. *Journal of Human Evolution*.

Pyle, S.I., 1935. Bone weight in the human carpus. *Human Biology*, **7**(1), 108–118.

R Development Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Rabosky, D.L., 2013. Diversity-dependence, ecological speciation, and the role of competition in macroevolution. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 481–502.

Raup, D.M., 1972. Taxonomic diversity during the Phanerozoic. *Science*, **177**(4054), 1065–1071.

Raup, D.M., 1975. Taxonomic diversity estimation using rarefaction. *Paleobiology*, **1**(4), 333–342.

Raup, D.M., 1976*a*. Species diversity in the Phanerozoic: a tabulation. *Paleobiology*, **2**(4), 279–288.

Raup, D.M., 1976*b*. Species diversity in the Phanerozoic: an interpretation. *Paleobiology*, **2**(4), 289–297.

Raup, D.M., 1977. Systematists follow the fossils. *Paleobiology*, **3**(3), 328–329.

Raup, D.M., 1979. Biases in the fossil record of species and genera. *Bulletin of the Carnegie Museum of Natural History*, **13**, 85–91.

Raup, D.M., 1991. The future of analytical paleobiology. In: Gilinksy, N.L. and Signor, P.W., (Eds.), *Analytical Paleobiology*. Short Courses in Paleontology, **4**, pp. 207–216.

Richmond, B.G., Aiello, L.C., and Wood, B.A., 2002. Early hominin limb proportions. *Journal of Human Evolution*, **43**(4), 529–548.

Richmond, B.G., Begun, D.R., and Strait, D.S., 2001. Origin of human bipedalism: the knuckle-walking hypothesis revisited. *American Journal of Physical Anthropology*, **116**(33), 70–105.

Richmond, B.G. and Hatala, K.G., 2013. Origin and evolution of human postcranial anatomy. In: Begun, D.R., (Ed.), *A Companion to Paleoanthropology*. Blackwell Publishing, Chichester, pp. 183–202.

Richmond, B.G. and Jungers, W.L., 2008. *Orrorin tugenensis* femoral morphology and the evolution of hominin bipedalism. *Science*, **319**(5870), 1662–1665.

Robinson, J.T., 1963. Adaptive radiation in the australopithecines and the origin of man. In: Howell, F.C. and Bourliere, F., (Eds.), *African Ecology and Human Evolution*. Aldine, Chicago, pp. 385–416.

Rowbotham, S.K., Blau, S., and Hislop-Jambrich, J., 2017. Recording skeletal completeness: a standardised approach. *Forensic Science International*, **275**, 117–123.

Russo, G.A. and Kirk, C., 2013. Foramen magnum position in bipedal mammals. *Journal of Human Evolution*, **65**(5), 656–670.

Russo, G.A. and Kirk, E.C., 2017. Another look at the foramen magnum in bipedal mammals. *Journal of Human Evolution*, **105**, 24–40.

Ruta, M., Wagner, P.J., and Coates, M.I., 2006. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proceedings of the Royal Society of London B: Biological Sciences*, **273**(1598), 2107–2111.

Ruth, A.A., Raghanti, M.A., Meind, R.S., and Lovejoy, C.O., 2016. Locomotor pattern fails to predict foramen magnum angle in rodents, strepsirrhine primates, and marsupials. *Journal of Human Evolution*, **94**, 45–52.

Sahney, S. and Benton, M.J., 2017. The impact of the Pull of the Recent on the fossil record of tetrapods. *Evolutionary Ecology Research*, **18**(2), 7–23.

Sakamoto, M., Venditti, C., and Benton, M.J., 2017. "Residual diversity estimates" do not correct for sampling bias in palaeodiversity data. *Methods in Ecology and Evolution*, **8**, 453–459.

Sansom, R.S., 2015. Bias and sensitivity in the placement of fossil taxa resulting from interpretations of missing data. *Systematic Biology*, **64**(2), 256–266.

Sansom, R.S. and Wills, M.A., 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Scientific Reports*, **3**, 2545.

Sansom, R.S., Wills, M.A., and Williams, T., 2017. Dental data perform relatively poorly in reconstructing mammal phylogenies: morphological partitions evaluated with molecular benchmarks. *Systematic Biology*, **66**(5), 813–522.

Satta, Y., Hickerson, M.J., Watanabe, H., O'hUigin, C., and Klein, J., 2004. Ancestral population sizes

and species divergence times in the primate lineage on the basis of intron and BAC end sequences. *Journal of Molecular Evolution*, **59**(4), 478–487.

Scally, A. and Durbin, R., 2012. Revising the human mutation rate: implications for understanding human evolution. *Nature Reviews Genetics*, **13**(10), 745–753.

Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., McCarthy, S., Montgomery, S.H., Schwalie, P.C., Tang, Y.A., Ward, M.C., Xue, Y., Yngvadottir, B., Alkan, C., Andersen, L.N., Ayub, Q., Ball, E.V., Beal, K., Bradley, B.J., Chen, Y., Clee, C.M., Fitzgerald, S., Graves, T.A., Gu, Y., Heath, P., Heger, A., Karakoc, E., Kolb-Kokocinski, A., Laird, G.K., Lunter, G., Meader, S., Mort, M., Mullikin, J.C., Munch, K., O'Connor, T.D., Phillips, A.D., Prado-Martinez, J., Rogers, A.S., Sajjadian, S., Schmidt, D., Shaw, K., Simpson, J.T., Stenson, P.D., Turner, D.J., Vigilant, L., Vilella, A.J., Whitener, W., Zhu, B., Cooper, D.N., de Jong, P., Dermitzakis, E.T., Eichler, E.E., Flicek, P., Goldman, N., Mundy, N.I., Ning, Z., Odom, D.T., Ponting, C.P., Quail, M.A., Ryder, O.A., Searle, S.M., Warren, W.C., Wilson, R.K., Schierup, M.H., Rogers, J., Tyler-Smith, C., and Durbin, R., 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature*, **483**(7388), 169–175.

Schrenk, F., Bromage, T.G., Betzler, C.G., Ring, U., and Juwayeyi, Y.M., 1993. Oldest *Homo* and Pliocene biogeography of the Malawi Rift. *Nature*, **365**(6449), 833–836.

Schroeder, L. and Ackermann, R.R., 2017. Evolutionary processes shaping diversity across the *Homo* lineage. *Journal of Human Evolution*, **111**, 1–17.

Schroeder, L., Roseman, C.C., Cheverud, J.M., and Ackermann, R.R., 2014. Characterizing the evolutionary path(s) to early *Homo*. *PLOS ONE*, **9**(12), e114307.

Schwartz, J.H. and Tattersall, I., 2003. *The Human Fossil Record: Craniodental Morphology of Genus Homo (Africa and Asia)*. Wiley-Liss, New York.

Schwartz, J.H. and Tattersall, I., 2005. *The Human Fossil Record: Craniodental Morphology of Early Hominids (Genera Australopithecus, Paranthropus, Orrorin), and Overview*. Wiley-Liss, New York.

Seiffert, E.R., Simons, E.L., and Attia, Y., 2003. Fossil evidence for an ancient divergence of lorises and galagos. *Nature*, **422**(6930), 421–424.

Semaw, S., Simpson, S.W., Quade, J., Renne, P.R., Butler, R.F., McIntosh, W.C., Levin, N.E., Domínguez-Rodrigo, M., and Rogers, M.J., 2005. Early Pliocene hominids from Gona, Ethiopia. *Nature*, **433** (7023), 301–305.

Senut, B., 2007. The earliest putative hominids. In: Henke, W., Hardt, T., and Tattersall, I., (Eds.), *Handbook of Paleoanthropology*. Springer, New York, pp. 1519–1538.

Senut, B., Pickford, M., Gommery, D., Meind, P., Cheboie, K., and Coppens, Y., 2001. First hominid from the Miocene (Lukeino Formation, Kenya). *Comptes Rendus de l'Académie des Sciences*, **332**(2), 137–144.

Sepkoski, J.J., Bambach, R.K., Raup, D.M., and Valentine, J.W., 1981. Phanerozoic marine diversity and the fossil record. *Nature*, **293**, 435–437.

Sheehan, P.M., 1977. A reflection of labor by systematics? *Paleobiology*, **3**(3), 325–328.

Shultz, S. and Maslin, M.A., 2013. Early human speciation, brain expansion and dispersal influenced by African climate pulses. *PLOS ONE*, **8**(10), 1–7.

Signor, P.W. and Lipps, J.H., 1982. Sampling bias, gradual extinction patterns and catastrophes in the fossil record. In: Silver, L.T. and Schultz, P.H., (Eds.), *Geological Implications of Impacts of Large Asteroids and Comets on the Earth*. Geological Society of America Special Paper, **190**, pp. 291–296.

Simpson, S.W., 2013. Before *Australopithecus*: the earliest hominins. In: Begun, D.R., (Ed.), *A Companion to Paleoanthropology*. Blackwell Publishing, Chichester, pp. 417–433.

Simpson, S.W., Kleinsasser, L., Quade, J., Levin, N.E., McIntosh, W.C., Dunbar, N., Semaw, S., and Rogers, M.J., 2015. Late Miocene hominin teeth from the Gona Paleoanthropological Research Project area, Afar, Ethiopia. *Journal of Human Evolution*, **81**, 68–82.

Simpson, S.W., Quade, J., Levin, N.E., Butler, R., Dupont-Nivet, G., Everett, M., and Semaw, S., 2008. A female *Homo erectus* pelvis from Gona, Ethiopia. *Science*, **322**(5904), 1089–1092.

Singleton, M., 2003. Functional and phylogenetic implications of molar flare variation in Miocene hominoids. *Journal of Human Evolution*, **45**(1), 57–79.

Skelton, R.R. and McHenry, H.M., 1992. Evolutionary relationships among early hominids. *Journal of Human Evolution*, **23**(4), 309–349.

Smith, A.B., 1994. *Systematics and the Fossil Record*. Blackwell Scientific, Oxford.

Smith, A.B., 2001. Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **356** (1407), 351–367.

Smith, A.B., 2007*a*. Marine diversity through the Phanerozoic: problems and prospects. *Journal of the Geological Society of London*, **164**(4), 731–745.

Smith, A.B., 2007*b*. Intrinsic *versus* extrinsic biases in the fossil record: contrasting the fossil record of echinoids in the Triassic and early Jurassic using sampling data, phylogenetic analysis and molecular clocks. *Paleobiology*, **33**(2), 311–324.

Smith, A.B., Lloyd, G.T., and McGowan, A.J., 2012. Phanerozoic marine diversity: rock record modelling provides an independent test of large-scale trends. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**(1746), 4489–4495.

Smith, A.B. and McGowan, A.J., 2007. The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of Western Europe? *Palaeontology*, **50**(4), 765–774.

Smith, A.B. and McGowan, A.J., 2011. The ties linking rock and fossil records and why they are important for palaeobiodiversity studies. In: McGowan, A.J. and Smith, A.B., (Eds.), *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society of London Special Publications, **358**, pp. 1–8.

Smith, R.J., 2005. Species recognition in paleoanthropology: implications of small sample sizes. In: Lieberman, D.E., Smith, R.J., and Kelley, J., (Eds.), *Interpreting the Past: Essays on Human, Primate, and Mammal Evolution in Honor of David Pilbeam*. Brill Academic Publishers, Boston, pp. 207–219.

Smith, R.J., 2016. Explanations for adaptations, just-so stories, and limitations on evidence in evolutionary biology. *Evolutionary Anthropology*, **25**(6), 276–287.

Smith, R.J. and Wood, B.A., 2017. The principles and practice of human evolution research: are we asking questions that can be answered? *Comptes Rendus Palevol*, **16**(5–6), 670–679.

Soligo, C., 2002. Primatology, paleoecology, and a new method for assessing taphonomic bias in fossil assemblages. *Evolutionary Anthropology*, **11**(S1), 24–27.

Soligo, C. and Andrews, P., 2005. Taphonomic bias, taxonomic bias and historical non-equivalence of faunal structure in early hominin localities. *Journal of Human Evolution*, **49**(2), 206–229.

Soligo, C., Will, O., Tavaré, S., Marshall, C.R., and Martin, R.D., 2007. New light on the dates of primate origins and divergence. In: Ravosa, M.J. and Dagosto, M., (Eds.), *Primate Origins: Adaptations and Evolution*. Springer, New York, pp. 29–49.

Sponheimer, M., Alemseged, Z., Cerling, T.E., Grine, F.E., Kimbel, W.H., Leakey, M.G., Lee-Thorp, J.A., Manthi, F.K., Reed, K.E., Wood, B.A., and Wynn, J.G., 2013. Isotopic evidence for early hominin diets. *Proceedings of the National Academy of Sciences*, **110**(26), 10513–10518.

Spoor, F., 2015. The middle Pliocene gets crowded. *Nature*, **521**(7553), 432–433.

Spoor, F., Gunz, P., Neubauer, S., Stelzer, S., Scott, N., Kwekason, A., and Dean, M.C., 2015. Reconstructed *Homo habilis* type OH 7 suggests deep-rooted species diversity in early *Homo*. *Nature*, **519** (7541), 83–86.

Spoor, F., Leakey, M.G., Gathogo, P.N., Brown, F.H., Antón, S.C., McDougall, I., Kiarie, C., Manthi, F.K., and Leakey, L.N., 2007. Implications of new early *Homo* fossils from Ileret, east of Lake Turkana, Kenya. *Nature*, **448**(7154), 688–691.

Spoor, F., Leakey, M.G., and Leakey, L.N., 2010. Hominin diversity in the Middle Pliocene of eastern Africa: the maxilla of KNM-WT 40000. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365**(1556), 3377–3388.

Spoor, F., Leakey, M.G., and O'Higgins, P., 2016. Middle Pliocene hominin diversity: *Australopithecus deyiremeda* and *Kenyanthropus platyops*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **371**(1698), 20150231.

Springer, M.S., Meredith, R.W., Gatesy, J., Emerling, C.A., Park, J., Rabosky, R., Stadler, T., Steiner, C., Ryder, O.A., Janečka, J.E., Fisher, C.A., and Murphy, W.J., 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLOS ONE*, **7**(11), e49521.

Stanley, S.M., 1979. *Macroevolution: Patterns and Process*. W. H. Freeman & Company, San Francisco.

Starrfelt, J. and Liow, L.H., 2016. How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **371**(1691), 20150219.

Steiper, M.E. and Young, N.M., 2006. Primate molecular divergence dates. *Molecular Phylogenetics and Evolution*, **41**(2), 384–394.

Steiper, M.E. and Young, N.M., 2008. Timing primate evolution: lessons from the discordance between molecular and paleontological estimates. *Evolutionary Anthropology*, **17**(4), 179–188.

Steiper, M.E. and Young, N.M., 2009. Primates. In: Hedges, S.B. and Kumar, S., (Eds.), *The Timetree of Life*. Oxford University Press, Oxford, pp. 482–486.

Stern, J.T. and Susman, R.L., 1983. The locomotor anatomy of *Australopithecus afarensis*. *American Journal of Physical Anthropology*, **60**(3), 279–317.

Stone, A.C., Battistuzzi, F.U., Kubatko, L.S., Perry, G.H., Trudeau, E., Lin, H., and Kumar, S., 2010. More reliable estimates of divergence times in *Pan* using complete mtDNA sequences and accounting for population structure. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365**(1556), 3277–3288.

Strait, D.S., 2001. Integration, phylogeny, and the hominid cranial base. *American Journal of Physical Anthropology*, **114**(4), 273–297.

Strait, D.S., 2010. The evolutionary history of the australopiths. *Evolution: Education and Outreach*, **3** (3), 341–352.

Strait, D.S. and Grine, F.E., 2001. The systematics of *Australopithecus garhi*. *Ludus Vitalis*, **9**(15), 109–135.

Strait, D.S. and Grine, F.E., 2004. Inferring hominoid and early hominid phylogeny using craniodental characters: the role of fossil taxa. *Journal of Human Evolution*, **47**(6), 399–452.

Strait, D.S., Grine, F.E., and Fleagle, J.G., 2007. Analyzing hominid phylogeny. In: Henke, W., Hardt, T., and Tattersall, I., (Eds.), *Handbook of Paleoanthropology*. Springer, New York, pp. 1781–1806.

Strait, D.S., Grine, F.E., and Moniz, M.A., 1997. A reappraisal of early hominid phylogeny. *Journal of Human Evolution*, **32**(1), 17–82.

Strauss, D. and Sadler, P.M., 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Mathematical geology*, **21**(4), 411–427.

Su, D.F., Ambrose, S.H., DeGusta, D., and Haile-Selassie, Y., 2009. Paleoenvironment. In: Haile-Selassie, Y. and WoldeGabriel, G., (Eds.), *Ardipithecus kadabba: Late Miocene Evidence from the Middle Awash, Ethiopia*. University of California Press, Berkeley, pp. 521–548.

Su, D.F. and Harrison, T., 2007. The paleoecology of the Upper Laetolil Beds at Laetoli: a reconsideration of the large mammal evidence. In: Bobe, R., Alemseged, Z., and Behrensmeyer, A.K., (Eds.), *Hominin Environments in the East African Pliocene: An Assessment of the Faunal Evidence*. Springer, Dordrecht, pp. 279–314.

Suwa, G., Asfaw, B., Kono, R.T., Kubo, D., Lovejoy, C.O., and White, T.D., 2009*a*. The *Ardipithecus ramidus* skull and its implications for hominid origins. *Science*, **326**(5949), 68e1–68e7.

Suwa, G., Asfaw, B., Kono, R.T., Kubo, D., Lovejoy, C.O., and White, T.D., 2009*b*. Paleobiological implications of the *Ardipithecus ramidus* dentition. *Science*, **326**(5949), 94–99.

Suwa, G., Kono, R.T., Katosh, S., Asfaw, B., and Beyene, Y., 2007. A new species of great ape from the late Miocene epoch in Ethiopia. *Nature*, **448**(7156), 921–924.

Tarver, J.E., Donoghue, P.C.J., and Benton, M.J., 2011. Is evolutionary history repeatedly rewritten in

light of new fossil discoveries? *Proceedings of the Royal Society of London B: Biological Sciences*, **278** (1705), 599–604.

Tavaré, S., Marshall, C.R., Will, O., Soligo, C., and Martin, R.D., 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature*, **416**(6882), 726–729.

Teaford, M.F. and Ungar, P.S., 2000. Diet and the evolution of the earliest human ancestors. *Proceedings of the National Academy of Sciences*, **97**(25), 13506–13511.

Tennant, J.P., Mannion, P.D., and Upchurch, P., 2016. Sea level regulated tetrapod diversity dynamics through the Jurassic/Cretaceous interval. *Nature Communications*, **2**(7), 12737.

Thompson, B., Arenson, J., Biernat, M., Barr, W.A., Reeves, J., Braun, D.R., and Hammond, A. A preliminary study of primate abundance in East Turkana collection areas relative to outcrop size. Paper presented at the American Association of Physical Anthropogists 86th annual meeting, New Orleans, 2017.

Tiedemann, R., Sarnthein, M., and Shackleton, N.J., 1994. Astronomic timescale for the Pliocene Atlantic δ^{18}O and dust flux records of Ocean Drilling Program Site 659. *Paleoceanography*, **9**(4), 619–638.

Toussaint, M., Macho, G.A., Tobias, P.V., Partridge, T.C., and Hughes, A.R., 2003. The third partial skeleton of a late Pliocene hominin (Stw 431) from Sterkfontein, South Africa. *South African Journal of Science*, **99**(5), 215–223.

Trauth, M.H., Maslin, M.A., Deino, A.L., Junginger, A., Lesoloyiae, M., Odada, E.O., Olago, D.O., Olaka, L.A., Strecker, M.R., and Tiedemann, R., 2010. Human evolution in a variable environment: the amplifier lakes of Eastern Africa. *Quaternary Science Reviews*, **29**(23–24), 2981–2988.

Tutin, S.L. and Butler, R.J., 2017. The completeness of the fossil record of plesiosaurs, marine reptiles from the Mesozoic. *Acta Palaeontologica Polonica*, **62**(3), 563–573.

Uhen, M.D. and Pyenson, N.D., 2007. Diversity estimates, biases, and historiographic effects: resolving cetacean diversity in the Tertiary. *Palaeontologia Electronica*, **10**(2), 1–22.

Ungar, P.S., Walker, A.C., and Coffing, K., 1994. Reanalysis of the Lukeino Molar (KNM-LU 335). *American Journal of Physical Anthropology*, **94**(2), 165–173.

Upchurch, P. and Barrett, P.M., 2005. Phylogenetic and taxic perspectives on sauropod diversity. In: Curry Rogers, K.A. and Wilson, J.A., (Eds.), *The Sauropods: Evolution and Paleobiology*. University of California Press, Berkeley, pp. 104–124.

Upchurch, P., Mannion, P.D., Benson, R.B.J., Butler, R.J., and Carrano, M.T., 2011. Geological and anthropogenic controls on the sampling of the terrestrial fossil record: a case study from the Dinosauria. In: McGowan, A.J. and Smith, A.B., (Eds.), *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society of London Special Publications, **358**, pp. 209–240.

Valentine, J.W., 1969. Patterns of taxonomic and ecological structure of the shelf benthos during Phanerozoic time. *Palaeontology*, **12**(4), 684–709.

Verrière, A., Brocklehurst, N., and Fröbisch, J., 2016. Assessing the completeness of the fossil record: comparison of different methods applied to parareptilian tetrapods (Vertebrata: Sauropsida). *Paleobiology*, **42**(4), 680–695.

Vignaud, P., Duringer, P., Mackaye, H.T., Likius, A., Blondel, C., Boisserie, J.-R., De Bonis, L., Eisenmann, V., Etienne, M.E., Geraads, D., Guy, F., Lehmann, T., Lihoreau, F., Lopez-Martinez, N., Mourer-Chauviré, Otero, O., Rage, J.-C., Schuster, M., Viriot, L., Zazzo, A., and Brunet, M., 2002. Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Chad. *Nature*, **418**(6894), 152–155.

Villmoare, B.A. and Kimbel, W.H., 2011. CT-based study of internal structure of the anterior pillar in extinct hominins and its implications for the phylogeny of robust *Australopithecus*. *Proceedings of the National Academy of Sciences*, **108**(39), 16200–16205.

Villmoare, B.A., Kimbel, W.H., Seyoum, C., Campisano, C.J., DiMaggio, E.N., Rowan, J., Braun, D.R., Arrowsmith, J.R., and Reed, K.E., 2015. Early *Homo* at 2.8 Ma from Ledi-Geraru, Afar, Ethiopia. *Science*, **347**(6228), 1352–1355.

Vrba, E.S., 1984. What is species selection? *Systematic Zoology*, **33**(3), 318–328.
Vrba, E.S., 1985*a*. Environment and evolution: alternative causes for the temporal distribution of evolutionary events. *South African Journal of Science*, **81**, 229–236.
Vrba, E.S., 1985*b*. African Bovidae: evolutionary events since the Miocene. *South African Journal of Science*, **81**, 263–266.
Vrba, E.S., 1988. Late Pliocene climatic events and hominid evolution. In: Grine, F.E., (Ed.), *Evolutionary History of the "Robust" Australopithecines*. Aldine de Gruyter, New York, pp. 405–426.
Vrba, E.S., 1993. Turnover-pulses, the Red Queen, and related topics. *American Journal of Science*, **293**, 418–452.
Vrba, E.S., 1995. The fossil record of African antelopes (Mammalia, Bovidae) in relation to human evolution and paleoclimate. In: Vrba, E.S., Denton, G.H., Partridge, T.C., and Burckle, L.H., (Eds.), *Paleoclimate and Evolution, with Emphasis on Human Origins*. Yale University Press, New Haven, pp. 385–424.
Vrba, E.S., 2000. Major features of Neogene mammalian evolution in Africa. In: Partridge, T.C. and Maud, R.R., (Eds.), *The Cenozoic of Southern Africa*. Oxford University Press, Oxford, pp. 277–304.
Wagner, P.J., 1995. Diversity patterns among early gastropods: contrasting taxonomic and phylogenetic descriptions. *Paleobiology*, **21**(4), 410–439.
Wagner, P.J., 2000*a*. The quality of the fossil record and the accuracy of phylogenetic inferences about sampling and diversity. *Systematic Biology*, **49**(1), 65–86.
Wagner, P.J., 2000*b*. Phylogenetic analyses and the fossil record: tests and inferences, hypotheses and models. *Paleobiology*, **26**(4), 341–371.
Wagner, P.J. and Erwin, D.H., 1995. Phylogenetic tests of speciation hypotheses. In: Erwin, D.H. and Anstey, R.L., (Eds.), *New Approaches to Studying Speciation in the Fossil Record*. Columbia University Press, New York, pp. 87–122.
Wagner, P.J. and Sidor, C.A., 2000. Age rank/clade rank metrics—sampling, taxonomy, and the meaning of stratigraphic consistency. *Systematic Biology*, **49**(3), 463–479.
Walker, A.C. and Leakey, R.E.F., 1993. *The Nariokotome Homo erectus Skeleton*. Harvard University Press, Cambridge.
Walker, F.M., Dunhill, A.M., Woods, M.A., Newell, A.J., and Benton, M.J., 2017. Assessing sampling of the fossil record in a geographically and stratigraphically constrained dataset: the Chalk Group of Hampshire, southern UK. *Journal of the Geological Society of London*, **174**, 509–521.
Wall, P.D., Ivany, L.C., and Wilkinson, B.H., 2009. Revisiting Raup: exploring the influence of outcrop area on diversity in light of modern sample standardization techniques. *Paleobiology*, **35**(1), 146–167.
Wall, P.D., Ivany, L.C., and Wilkinson, B.H., 2011. Impact of outcrop area on estimates of Phanerozoic terrestrial biodiversity trends. In: McGowan, A.J. and Smith, A.B., (Eds.), *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*. Geological Society of London Special Publications, **358**, pp. 53–62.
Walther, M. and Fröbisch, J., 2013. The quality of the fossil record of anomodonts (Synapsida, Therapsida). *Comptes Rendus Palevol*, **12**, 495–504.
Ward, C.V., 1997. Functional anatomy and phyletic implications of the hominoid trunk and hindlimb. In: Begun, D.R., Ward, C.V., and Rose, M., (Eds.), *Function, Phylogeny, and Fossils: Moicene Hominoid Evolution and Adaptations*. Plenum Press, New York, pp. 101–130.
Ward, C.V., Leakey, M.G., Brown, F.H., Harris, J.M., and Walker, A.C., 1999*a*. South Turkwel: a new Pliocene hominid site in Kenya. *Journal of Human Evolution*, **36**(1), 69–95.
Ward, C.V., Leakey, M.G., and Walker, A.C., 1999*b*. The new hominid species *Australopithecus anamensis*. *Evolutionary Anthropology*, **7**(6), 197–205.
Ward, C.V., Leakey, M.G., and Walker, A.C., 2001. Morphology of *Australopithecus anamensis* from Kanapoi and Allia Bay, Kenya. *Journal of Human Evolution*, **41**(4), 255–368.
Ward, C.V., Manthi, F.K., and Plavcan, J.M., 2013. New fossils of *Australopithecus anamensis* from Kanapoi, West Turkana, Kenya (2003–2008). *Journal of Human Evolution*, **65**(5), 501–524.

Ward, C.V., Plavcan, J.M., and Manthi, F.K., 2017, *in press*. New fossils of *Australopithecus anamensis* from Kanapoi, West Turkana, Kenya (2012–2015). *Journal of Human Evolution*.

Werdelin, L., 2010. Chronology of Neogene mammal localities. In: Werdelin, L. and Sanders, W.J., (Eds.), *Cenozoic Mammals of Africa*. University of California Press, Berkeley, pp. 27–43.

Werdelin, L. and Lewis, M.E., 2005. Plio-Pleistocene Carnivora of eastern Africa: species richness and turnover patterns. *Zoological Journal of the Linnean Society*, **144**(2), 121–144.

Werdelin, L. and Sanders, W.J., 2010. *Cenozoic Mammals of Africa*. University of California Press, Berkeley.

White, T.D., 1995. African omnivores: global climatic change and Plio-Pleistocene hominids and suids. In: Vrba, E.S., Denton, G.H., Partridge, T.C., and Burckle, L.H., (Eds.), *Paleoclimate and Evolution, with Emphasis on Human Origins*. Yale University Press, New Haven, pp. 369–384.

White, T.D., 2002. Earliest hominids. In: Hartwig, W., (Ed.), *The Primate Fossil Record*. Cambridge University Press, Cambridge, pp. 407–417.

White, T.D., 2003. Early hominids—diversity or distortion? *Science*, **299**(5615), 1994–1997.

White, T.D., Asfaw, B., Beyene, Y., Haile-Selassie, Y., Lovejoy, C.O., Suwa, G., and WoldeGabriel, G., 2009. *Ardipithecus ramidus* and the paleobiology of early hominids. *Science*, **326**(5949), 75–86.

White, T.D., Lovejoy, C.O., Asfaw, B., Carlson, J.P., and Suwa, G., 2015. Neither chimpanzee nor human, *Ardipithecus* reveals the surprising ancestry of both. *Proceedings of the National Academy of Sciences*, **112**(16), 4877–4884.

White, T.D., Suwa, G., and Asfaw, B., 1994. *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Nature*, **371**(6526), 306–312.

White, T.D., Suwa, G., and Asfaw, B., 1995. Erratum: *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Science*, **375**(6526), 88.

White, T.D., Suwa, G., Simpson, S.W., and Asfaw, B., 2000. Jaws and teeth of *Australopithecus afarensis* from Maka, Middle Awash, Ethiopia. *American Journal of Physical Anthropology*, **111**(1), 45–68.

White, T.D., WoldeGabriel, G., Asfaw, B., Ambrose, S.H., Beyene, Y., Bernor, R.L., Boisserie, J.-R., Currie, B., Gilbert, H., Haile-Selassie, Y., Hart, W.K., Hlusko, L.J., Howell, F.C., Kono, R.T., Lehmann, T., Louchart, A., Lovejoy, C.O., Renne, P.R., Saegusa, H., Vrba, E.S., Wesselman, H., and Suwa, G., 2006. Asa Issie, Aramis and the origin of *Australopithecus*. *Nature*, **440**(7086), 883–889.

Wilkinson, R.D., Steiper, M.E., Soligo, C., Martin, R.D., Yang, Z., and Tavaré, S., 2011. Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology*, **60**(1), 16–31.

WoldeGabriel, G., Haile-Selassie, Y., Renne, P.R., Hart, W.K., Ambrose, S.H., Asfaw, B., Heiken, G., and White, T.D., 2001. Geology and palaeontology of the Late Miocene Middle Awash valley, Afar rift, Ethiopia. *Nature*, **412**(6843), 175–178.

WoldeGabriel, G., Hart, W.K., Renne, P.R., Haile-Selassie, Y., and White, T.D., 2009. Stratigraphy of the Adu-Asa Formation. In: Haile-Selassie, Y. and WoldeGabriel, G., (Eds.), *Ardipithecus kadabba: Late Miocene Evidence from the Middle Awash, Ethiopia*. University of California Press, Berkeley, pp. 27–62.

Wolpoff, M.H., Hawks, J., Senut, B., Pickford, M., and Ahern, J., 2006. An ape or *the* ape: is the Toumaï cranium TM 266 a hominid? *PaleoAnthropology*, **2006**, 36–50.

Wolpoff, M.H., Senut, B., Pickford, M., and Hawks, J., 2002. *Sahelanthropus* or '*Sahelpithecus*'? *Nature*, **419**(6907), 581–582.

Wood, B.A., 1988. Are "robust" australopithecines a monophyletic group? In: Grine, F.E., (Ed.), *Evolutionary History of the 'Robust' Australopithecines*. Aldine de Gruyter, New York, pp. 269–284.

Wood, B.A., 1991. *Koobi Fora Research Project. Volume 4: Hominid Cranial Remains*. Clarendon Press, Oxford.

Wood, B.A., 1992*a*. Origin and evolution of the genus *Homo*. *Nature*, **355**(6363), 783–790.

Wood, B.A., 1992*b*. Early hominid species and speciation. *Journal of Human Evolution*, **22**(4–5), 351–365.

Wood, B.A., (Ed.), 2011. *Wiley-Blackwell Encyclopedia of Human Evolution*. Wiley-Blackwell, Hoboken.

Wood, B.A. and Boyle, E.K., 2016. Hominin taxic diversity: fact or fantasy? *Yearbook of Physical Anthropology*, **159**, 37–78.

Wood, B.A. and Boyle, E.K., 2017. Hominins: context, origins, and taxic diversity. In: Tibayrenc, M. and Ayala, F.J., (Eds.), *On Human Nature: Biology, Psychology, Ethics, Politics, and Religion*. Academic Press, Amsterdam, pp. 17–44.

Wood, B.A. and Grabowski, M.W., 2015. Macroevolution in and around the hominin clade. In: Serrelli, E. and Gontier, N., (Eds.), *Macroevolution: Explanation, Interpretation and Evidence*. Springer, Cham, pp. 345–376.

Wood, B.A. and Harrison, T., 2011. The evolutionary context of the first hominins. *Nature*, **470**(7334), 347–352.

Wood, B.A. and Leakey, M.G., 2011. The Omo-Turkana Basin fossil hominins and their contribution to our understanding of human evolution in Africa. *Evolutionary Anthropology*, **20**(6), 264–292.

Wood, B.A. and Lonergan, N., 2008. The hominin fossil record: taxa, grades and clades. *Journal of Anatomy*, **212**(4), 354–376.

Wood, B.A. and Richmond, B.G., 2000. Human evolution: taxonomy and paleobiology. *Journal of Anatomy*, **197**(1), 19–60.

Wood, B.A. and Schroer, K., 2013. *Paranthropus*. In: Begun, D.R., (Ed.), *A Companion to Paleoanthropology*. Wiley-Blackwell, Chichester, pp. 457–478.

Wynn, J.G., Alemseged, Z., Bobe, R., Geraads, D., Reed, D., and Roman, D.C., 2006. Geological and palaeontological context of a Pliocene juvenile hominin at Dikika, Ethiopia. *Nature*, **443**(7109), 332–336.

Yang, Z., 2002. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics*, **162**(4), 1811–1823.

Young, N.M., 2005. Estimating hominoid phylogeny from morphological data: character choice, phylogenetic signal and postcranial data. In: Lieberman, D.E., Smith, R.J., and Kelley, J., (Eds.), *Interpreting the Past: Essays on Human, Primate, and Mammal Evolution*. Brill Academic Publishers, Boston, pp. 19–31.

Young, N.M., Capellini, T.D., Roach, N.T., and Alemseged, Z., 2015. Fossil hominin shoulders support an African ape-like last common ancestor of humans and chimpanzees. *Proceedings of the National Academy of Sciences*, **112**(38), 11829–11834.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., and Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**(5517), 686–693.

Zollikofer, C.P.E., Ponce De Leon, M.S., Lieberman, D.E., Guy, F., Pilbeam, D.R., Likius, A., Vignaud, P., and Brunet, M., 2005. Virtual cranial reconstruction of *Sahelanthropus tchadensis*. *Nature*, **434** (7034), 755–759.

Appendix A

Taxonomy

The hominoid taxonomy used in this thesis is that of Wood & Harrison (2011). Vernacular names are given in parentheses.

Superfamily Hominoidea (hominoids)
- **Family** Hylobatidae (hylobatids)
- **Family** Hominidae (hominids)
  - **Subfamily** Ponginae (pongines)
    - **Genus** *Pongo*
  - **Subfamily** Homininae (hominines)
    - **Tribe** Gorillini (gorillins)
      - **Genus** *Gorilla*
    - **Tribe** Panini (panins)
      - **Genus** *Pan*
    - **Tribe** Hominini (hominins)
      - **Subtribe** Australopithecina (australopiths)
        - **Genus** *Australopithecus**
        - **Genus** *Kenyanthropus**
        - **Genus** *Paranthropus**
      - **Subtribe** Hominina (hominans)
        - **Genus** *Homo*
    - **Tribe** *incertae sedis*
      - **Genus** *Sahelanthropus*†
      - **Genus** *Orrorin*†
      - **Genus** *Ardipithecus*†

*These are almost certainly early hominins.
†These are purported early hominins.

Appendix B

R scripts

Chapter 2

```
# R code for Chapter 2: Phylogenetic and taxic perspectives on early hominin
    diversity

# Set working directory:
setwd("~/git-repo/thesis-scripts/Chapter2")

# Load packages:
library(ape) # if (!require("ape")) install.packages("ape")
library(paleotree) # if (!require("paleotree")) install.packages("paleotree")
library(strap) # if (!require("strap")) install.packages("strap")
library(picante) # if (!require("picante")) install.packages("picante")
library(matrixStats) # if (!require("matrixStats")) install.packages("
    matrixStats")

# Load packages pre-installed from GitHub (requires devtools):
library(Claddis) # devtools::install_github("graemetlloyd/Claddis")
library(egg) # devtools::install_github("baptiste/egg")

# Load packages for plotting:
library(ggplot2) # if (!require("ggplot2")) install.packages("ggplot2")
library(RColorBrewer) # if (!require("RColorBrewer")) install.packages("
    RColorBrewer")

# Load time-scaled trees:
load("Chapter2datedtrees.rda")

# Load cal3 time scaled trees:
# load("Chapter2datedtrees.rda")

# Load time bins for diversity analysis:
time.bins <- read.table(file = "time.bins.txt", header = TRUE)

# Load time bins that span the full period of hominin evolution (for rates
    calculation):
# time.bins.all <- read.table(file = "~/R/time.bins.txt", header = TRUE)

# Load range data from Wood and Boyle (2016):
ages <- read.table("ages.txt", header = TRUE)

# TDE:
tde <- taxicDivCont(timeData = ages, int.times = time.bins, plot = FALSE)[,3]
```

```
# Time-sclaing cladograms using cal3 requires a sampling, branching, and
    extinction rate. These are determined empirically using the frequency of
    taxon durations.
# Load appearance dates in discrete time:
# homininRanges <- [object composed of list of time bins and list of first and
     last appearances for each taxon in discrete time bins]

# Construct maximum likelihood function of the observed frequency of taxon
    durations:
likFun <- make_durationFreqDisc(homininRanges)

# Find the best fitting sampling probability and extinction rate:
Rates <- optim(parInit(likFun), likFun, lower = parLower(likFun), upper =
    parUpper(likFun),
               method = "L-BFGS-B", control = list(maxit = 1000000))

# Calculate duration of each time bin:
intLength <- apply(homininRanges[[1]], 1, diff)

# Calculate mean time bin duration:
meanInt <- mean(abs(intLength))

# Calculate instantaneous samping rate (R) (per L Ma):
sRate <- sProb2sRate(Rates$par[[2]], int.length = meanInt) # data does not fit
     model expectation that log of taxon durations are exponentially
    distributed: model rejected.

# Set sRate to that of Tavare et al. (2002):
sRate <- 0.023

# Calculate extinction rate:
extRate <- Rates$par[[1]]/meanInt

# Set number of trees to produce:
n <- 1000

# Read Dembo et al.'s (2016) tree:
d16.tree <- read.tree("trees/dembo_etal_2016.tre")

# Prune Gorilla and Pan from tree:
d16.tree <- drop.tip(phy = d16.tree, tip = c("Gorilla_gorilla","Pan_
    troglodytes"))

# Load ages data for this tree:
d16.ages <- read.table("ages/d16.ages.txt", header = TRUE)

# Create empty list for output:
d16.cal3.trees <- list()

# For each tree:
for (i in 1:n) {

  # Create random tree as condition:
  d16.cal3.trees[[i]] <- rtree(10)

  # Set all edge lengths in random tree to zero:
  d16.cal3.trees[[i]]$edge.length <- rep(0, 11)

  # Time scale tree until condition is satistfied (i.e., all edge lengths are
      non-zero):
  while (min(d16.cal3.trees[[i]]$edge.length) == 0) {
```

```
    # cal3 time scale:
    d16.cal3.trees[[i]] <- cal3TimePaleoPhy(tree = d16.tree, timeData = d16.
        ages, brRate = 1.775051, extRate = 1.775051, sampRate = sRate, ntrees
        = 1, root.max = 1, FAD.only = TRUE)

  }
}

# For each tree:
for (i in 1:length(d16.cal3.trees)) {

  # Prune Euraian and FAD < 1 Ma taxa:
  d16.cal3.trees[[i]] <- dropPaleoTip(tree = d16.cal3.trees[[i]], tip = c("
      Homo_floresiensis","Georgian_Homo_erectus","Asian_Homo_erectus","Homo_
      naledi","Homo_antecessor","Homo_heidelbergensis","Homo_neanderthalensis"
      ,"Homo_sapiens"))

}

# For each tree:
for (i in 1:length(d16.cal3.trees)) {

  # Rename Homo erectus:
  d16.cal3.trees[[i]]$tip.label[1] <- c("Homo_ergaster")

}

# D2PDE:
d16.pde <- multiDiv(d16.cal3.trees, int.times = time.bins)

# Extract node ages from each tree:
d16.nodes <- lapply(d16.cal3.trees, function(x){GetNodeAges(x)})

# Extract root age for each tree:
d16.roots <- unlist(lapply(d16.cal3.trees, function(x){x$root.time}))

d16.table <- matrix(unlist(d16.nodes), ncol = 27, byrow = TRUE)

d16.node.ages <- d16.table[,c((length(d16.cal3.trees[[1]]$tip.label) + 1):ncol
    (d16.table))]

# Calculate median node age of cal3 trees:
d16.cal3.nodes <- colMedians(d16.node.ages)

# Calculate variation node age:
sd.d16 <- apply(d16.node.ages, 2, sd)

# Test if older nodes have higher uncertainty in their estimate:
cor.test(log10(d16.cal3.nodes), log10(sd.d16), method = "pearson")

# Correlation without root:
cor.test(log10(d16.cal3.nodes[2:13]), log10(sd.d16[2:13]), method = "pearson")

# Node ages in Dembo et al.'s (2016) Bayesian tip-dating:
d16.nodes <- c
    (7.52,7.08,6.57,6.20,5.71,5.46,4.46,3.72,3.49,4.83,3.69,2.90,4.54)

# Test cal3 and Bayesian tip dated node ages:
cor.test(log(d16.nodes), log(d16.cal3.nodes), method = "pearson")
```

```
# # Read Haile-Selassie et al.'s (2015) majority-rule tree:
h.tree <- read.tree("trees/haileselassie_etal_2015.tre")

# Prune Pan from tree:
h.tree <- drop.tip(phy = h.tree, tip = c("Pan"))

# Remove range data for pruned taxa:
h.ages <- read.table("ages/h.ages.txt", header = TRUE)

# Resolve polytomies by first appearance:
h.tree <- timeLadderTree(tree = h.tree, timeData = h.ages)

# Ladderise tree:
h.tree <- ladderize(h.tree)

# Create empty list for output:
h.cal3.trees <- list()

# For each tree:
for (i in 1:n) {

  # Create random tree as condition:
  h.cal3.trees[[i]] <- rtree(10)

  # Set all edge lengths in random tree to zero:
  h.cal3.trees[[i]]$edge.length <- rep(0, 11)

  # Time scale tree until condition is satistfied (i.e., all edge lengths are
      non-zero):
  while (min(h.cal3.trees[[i]]$edge.length) == 0) {

    # cal3 time scale:
    h.cal3.trees[[i]] <- cal3TimePaleoPhy(tree = h.tree, timeData = h.ages,
        brRate = 0.8174231, extRate = 0.8174231, sampRate = sRate, ntrees = 1,
         root.max = 1, FAD.only = TRUE)

  }
}

# For each tree:
for (i in 1:length(h.cal3.trees)) {

  # Prune Homo sapiens:
  h.cal3.trees[[i]] <- dropPaleoTip(tree = h.cal3.trees[[i]], tip = c("Homo_
      sapiens"))

}

# For each tree:
for (i in 1:length(h.cal3.trees)) {

  # Rename Homo erectus:
  h.cal3.trees[[i]]$tip.label[9] <- c("Homo_ergaster")

}

# HPDE:
h.pde <- multiDiv(data = h.cal3.trees, int.times = time.bins)

# Extract node ages from each tree:
h.nodes <- lapply(h.cal3.trees, function(x){GetNodeAges(x)})
```

```
# Extract root age for each tree:
h.roots <- unlist(lapply(h.cal3.trees, function(x){x$root.time}))

h.table <- matrix(unlist(h.nodes), ncol = 27, byrow = TRUE)

h.node.ages <- h.table[,c((length(h.cal3.trees[[1]]$tip.label) + 1):ncol(h.
    table))]

# Calculate median node age of cal3 trees:
h.cal3.nodes <- colMedians(h.node.ages)

# Calculate variation node age:
sd.h <- apply(h.node.ages, 2, sd)

# Test if older nodes have higher uncertainty in their estimate:
cor.test(log10(h.cal3.nodes), log10(sd.h), method = "pearson")

# Correlation without root:
cor.test(log10(h.cal3.nodes[2:13]), log10(sd.h[2:13]), method = "pearson")

# Drop chronospecies:
h.tree.ana <- drop.tip(phy = h.tree, tip = c("Australopithecus_afarensis", "
    Paranthropus_boisei"))

# Rename chronospecies:
h.tree.ana$tip.label[4] <- c("Australopithecus_afarensis")
h.tree.ana$tip.label[12] <- c("Paranthropus_boisei")

# Remove (later) chronospecies from range data:
h.ages.ana <- h.ages[-c(5, 10),]

# Rename taxa from range data:
rownames(h.ages.ana)[4] <- c("Australopithecus_afarensis")
rownames(h.ages.ana)[8] <- c("Paranthropus_boisei")

# Update range data to match first and last appearance of lineage:
h.ages.ana$LAD[4] <- 3.0
h.ages.ana$LAD[8] <- 1.3

# Create empty list for output:
h.cal3.trees.ana <- list()

# For each tree:
for (i in 1:n) {

  # Create random tree as condition:
  h.cal3.trees.ana[[i]] <- rtree(10)

  # Set all edge lengths in random tree to zero:
  h.cal3.trees.ana[[i]]$edge.length <- rep(0, 11)

  # Time scale tree until condition is satistfied (i.e., all edge lengths are
      non-zero):
  while (min(h.cal3.trees.ana[[i]]$edge.length) == 0) {

    # cal3 time scale:
    h.cal3.trees.ana[[i]] <- cal3TimePaleoPhy(tree = h.tree.ana, timeData = h.
        ages.ana, brRate = 0.5525464, extRate = 0.5525464, sampRate = sRate,
        ntrees = 1, root.max = 1, FAD.only = TRUE)
```

```
  }
}

# For each tree:
for (i in 1:length(h.cal3.trees.ana)) {

  # Prune Homo sapiens:
  h.cal3.trees.ana[[i]] <- dropPaleoTip(tree = h.cal3.trees.ana[[i]], tip = c(
      "Homo_sapiens"))

}

# aHPDE:
h.pde.ana <- multiDiv(h.cal3.trees.ana, int.times = time.bins)

# Read Dembo et al.'s (2015) tree:
d15.tree <- read.tree("trees/dembo_etal_2015.tre")

# Prune Gorilla and Pan from tree:
d15.tree <- drop.tip(phy = d15.tree, tip = c("Gorilla_gorilla","Pan_
    troglodytes"))

# Load consensus ages file for this tree:
d15.ages <- read.table("ages/d15.ages.txt", header = TRUE)

# Create empty list for output:
d15.cal3.trees <- list()

# For each tree:
for (i in 1:n) {

  # Create random tree as condition:
  d15.cal3.trees[[i]] <- rtree(10)

  # Set all edge lengths in random tree to zero:
  d15.cal3.trees[[i]]$edge.length <- rep(0, 11)

  # Time scale tree until condition is satistfied (i.e., all edge lengths are
      non-zero):
  while (min(d15.cal3.trees[[i]]$edge.length) == 0) {

    # cal3 time scale:
    d15.cal3.trees[[i]] <- cal3TimePaleoPhy(tree = d15.tree, timeData = d15.
        ages, brRate = 1.469807, extRate = 1.469807, sampRate = sRate, ntrees
        = 1, root.max = 1, FAD.only = TRUE)

  }
}

# For each tree:
for(i in 1:length(d15.cal3.trees)){

  # Prune Euraian and FAD < 1 Ma taxa:
  d15.cal3.trees[[i]] <- dropPaleoTip(tree = d15.cal3.trees[[i]], tip = c("
      Homo_erectus", "Homo_floresiensis", "Homo_antecessor","Homo_
      heidelbergensis","Homo_neanderthalensis","Homo_sapiens"))

}

# D1PDE:
d15.pde <- multiDiv(d15.cal3.trees, int.times = time.bins)
```

```
# Extract node ages from each tree:
d15.nodes <- lapply(d15.cal3.trees, function(x){GetNodeAges(x)})

# Extract root age for each tree:
d15.roots <- unlist(lapply(d15.cal3.trees, function(x){x$root.time}))

d15.table <- matrix(unlist(d15.nodes), ncol = 27, byrow = TRUE)

d15.node.ages <- d15.table[,c((length(d15.cal3.trees[[1]]$tip.label) + 1):ncol
    (d15.table))]

# Calculate median node age of cal3 trees:
d15.cal3.nodes <- colMedians(d15.node.ages)

# Calculate variation node age:
sd.d15 <- apply(d15.node.ages, 2, sd)

# Test if older nodes have higher uncertainty in their estimate:
cor.test(log10(d15.cal3.nodes), log10(sd.d15), method = "pearson")

# Correlation without root:
cor.test(log10(d15.cal3.nodes[2:13]), log10(sd.d15[2:13]), method = "pearson")

# Node ages in Dembo et al.'s (2015) Bayesian tip-dating:
d15.nodes <- c
    (7.45,6.52,5.67,5.16,4.87,4.3,3.86,3.45,2.90,2.81,3.24,2.64,4.16)

# Test cal3 and Bayesian tip dated node ages:
cor.test(log(d15.nodes), log(d15.cal3.nodes), method = "pearson")

# Read Strait and Grine's (2004) strict consensus tree:
sg.tree <- read.tree("trees/strait_grine_2004.tre")

# Prune non-hominin apes from tree:
sg.tree <- drop.tip(phy = sg.tree, tip = c("Hylobates","Pongo_pygmaeus","
    Gorilla_gorilla","Pan_troglodytes"))

# Load ages data for this tree:
sg.ages <- read.table("ages/sg.ages.txt", header = TRUE)

# Resolve polytomies by first appearance:
sg.tree <- timeLadderTree(tree = sg.tree, timeData = sg.ages)

# Ladderise tree:
sg.tree <- ladderize(sg.tree)

# Create empty list for output:
sg.cal3.trees <- list()

# For each tree:
for (i in 1:n) {

  # Create random tree as condition:
  sg.cal3.trees[[i]] <- rtree(10)

  # Set all edge lengths in random tree to zero:
  sg.cal3.trees[[i]]$edge.length <- rep(0, 11)

  # Time scale tree until condition is satistfied (i.e., all edge lengths are
      non-zero):
```

```
  while (min(sg.cal3.trees[[i]]$edge.length) == 0) {

    # cal3 time scale:
    sg.cal3.trees[[i]] <- cal3TimePaleoPhy(tree = sg.tree, timeData = sg.ages,
         brRate = 0.5787885, extRate = 0.5787885, sampRate = sRate, ntrees =
        1, root.max = 1, FAD.only = TRUE)

  }
}

# For each tree:
for (i in 1:length(sg.cal3.trees)) {

  # Prune Homo sapiens:
  sg.cal3.trees[[i]] <- dropPaleoTip(tree = sg.cal3.trees[[i]], tip = c("Homo_
      sapiens"))

}

# SPDE:
sg.pde <- multiDiv(data = sg.cal3.trees, int.times = time.bins)

# Extract node ages from each tree:
sg.nodes <- lapply(sg.cal3.trees, function(x){GetNodeAges(x)})

# Extract root age for each tree:
sg.roots <- unlist(lapply(sg.cal3.trees, function(x){x$root.time}))

sg.table <- matrix(unlist(sg.nodes), ncol = 25, byrow = TRUE)

sg.node.ages <- sg.table[,c((length(sg.cal3.trees[[1]]$tip.label) + 1):ncol(sg
    .table))]

# Calculate median node age of cal3 trees:
sg.cal3.nodes <- colMedians(sg.node.ages)

# Calculate variation node age:
sd.sg <- apply(sg.node.ages, 2, sd)

# Test if older nodes have higher uncertainty in their estimate:
cor.test(log10(sg.cal3.nodes), log10(sd.sg), method = "pearson")

# Correlation without root:
cor.test(log10(sg.cal3.nodes[2:13]), log10(sd.sg[2:13]), method = "pearson")

# Drop chronospecies:
sg.tree.ana <- drop.tip(phy = sg.tree, tip = c("Australopithecus_afarensis", "
    Paranthropus_boisei"))

# Rename chronospecies:
sg.tree.ana$tip.label[3] <- c("Australopithecus_afarensis")
sg.tree.ana$tip.label[7] <- c("Paranthropus_boisei")

# Remove chronospecies from range data:
sg.ages.ana <- sg.ages[-c(5, 9),]

# Re-name taxa:
rownames(sg.ages.ana)[4] <- c("Australopithecus_afarensis")
rownames(sg.ages.ana)[7] <- c("Paranthropus_boisei")

# Update range data to match first and last appearance of lineage:
```

```
sg.ages.ana$LAD[4] <- 3.0 # Australopithecus afarensis LAD
sg.ages.ana$LAD[7] <- 1.3 # Paranthropus boisei LAD

# Create empty list for output:
sg.cal3.trees.ana <- list()

# For each tree:
for (i in 1:n) {

  # Create random tree as condition:
  sg.cal3.trees.ana[[i]] <- rtree(10)

  # Set all edge lengths in random tree to zero:
  sg.cal3.trees.ana[[i]]$edge.length <- rep(0, 11)

  # Time scale tree until condition is satistfied (i.e., all edge lengths are
      non-zero):
  while (min(sg.cal3.trees.ana[[i]]$edge.length) == 0) {

    # cal3 time scale:
    sg.cal3.trees.ana[[i]] <- cal3TimePaleoPhy(tree = sg.tree.ana, timeData =
        sg.ages.ana, brRate = 0.5525464, extRate = 0.5525464, sampRate = sRate
        , ntrees = 1, root.max = 1, FAD.only = TRUE)

  }
}

# For each tree:
for (i in 1:length(sg.cal3.trees.ana)) {

  # Prune Homo sapiens:
  sg.cal3.trees.ana[[i]] <- dropPaleoTip(tree = sg.cal3.trees.ana[[i]], tip =
      c("Homo_sapiens"))

}

# aSPDE:
sg.pde.ana <- multiDiv(sg.cal3.trees.ana, int.times = time.bins)

# Get mid-point age of each interval:
time <- apply(time.bins, 1, median)

# Call Graeme T. Lloyd's generalised differencing function:
gen.diff <- function(x, time) {
  # Suppress warning message:
  # if(cor.test(time, x)$p.value > 0.05) print("Warning: variables not
      significantly correlated, generalised differencing not recommended.")
  dt <- x - ((lsfit(time, x)$coefficients[2] * time) + lsfit(time, x)$
      coefficients[1])
  m <- lsfit(dt[1:(length(dt)-1)], dt[2:length(dt)])$coefficients[2]
  gendiffs <- dt[1:(length(dt) - 1)]- (dt[2:length(dt)] * m)
  gendiffs
}

# Perform generalised differencing function on each variable:
gd.tde <- gen.diff(tde, time) # TDE
gd.sg.pde <- gen.diff(sg.pde$median.curve[,1], time) # SPDE
gd.d15.pde <- gen.diff(d15.pde$median.curve[,1], time) # D1PDE
gd.h.pde <- gen.diff(h.pde$median.curve[,1], time) # HPDE
gd.d16.pde <- gen.diff(d16.pde$median.curve[,1], time) # D2PDE
gd.h.pde.ana <- gen.diff(h.pde.ana$median.curve[,1], time) # HPDEa
```

```
gd.sg.pde.ana <- gen.diff(sg.pde.ana$median.curve[,1], time) # SPDEa

# Call cor function (from Cleary et al., 2015):
cor = function(x, y, method = c("pearson", "spearman", "kendall"), N) {
  test = cor.test(x, y, method = method)
  adj = p.adjust(test$p.value, "BH", n = N)
  display = c(test$est, test$p.value, adj)
  names(display)[1] = "Coefficient"
  names(display)[2] = "p-value"
  names(display)[3] = "p.adjusted"
  return(display)
}

# TDE vs. SPDE:
cor(gd.sg.pde, gd.tde, method = "spearman", N = 12)
cor(gd.sg.pde, gd.tde, method = "kendall", N = 12)

# TDE vs. D1PDE:
cor(gd.d15.pde, gd.tde, method = "spearman", N = 12)
cor(gd.d15.pde, gd.tde, method = "kendall", N = 12)

# TDE vs. HPDE:
cor(gd.h.pde, gd.tde, method = "spearman", N = 12)
cor(gd.h.pde, gd.tde, method = "kendall", N = 12)

# TDE vs. D2PDE:
cor(gd.d16.pde, gd.tde, method = "spearman", N = 12)
cor(gd.d16.pde, gd.tde, method = "kendall", N = 12)

# SPDE vs. D1PDE:
cor(gd.sg.pde, gd.d15.pde, method = "spearman", N = 12)
cor(gd.sg.pde, gd.d15.pde, method = "kendall", N = 12)

# SPDE vs. HPDE:
cor(gd.sg.pde, gd.h.pde, method = "spearman", N = 12)
cor(gd.sg.pde, gd.h.pde, method = "kendall", N = 12)

# SPDE vs. D2PDE:
cor(gd.sg.pde, gd.d16.pde, method = "spearman", N = 12)
cor(gd.sg.pde, gd.d16.pde, method = "kendall", N = 12)

# D1PDE vs. HPDE:
cor(gd.d15.pde, gd.h.pde, method = "spearman", N = 12)
cor(gd.d15.pde, gd.h.pde, method = "kendall", N = 12)

# D1PDE vs. D2PDE:
cor(gd.d15.pde, gd.d16.pde, method = "spearman", N = 12)
cor(gd.d15.pde, gd.d16.pde, method = "kendall", N = 12)

# HPDE vs. D2PDE:
cor(gd.h.pde, gd.d16.pde, method = "spearman", N = 22)
cor(gd.h.pde, gd.d16.pde, method = "kendall", N = 22)

# SPDE vs. SPDE.ana:
cor(gd.sg.pde, gd.sg.pde.ana, method = "spearman", N = 12)
cor(gd.sg.pde, gd.sg.pde.ana, method = "kendall", N = 12)

# HPDE vs. HPDE.ana:
cor(gd.h.pde, gd.h.pde.ana, method = "spearman", N = 12)
cor(gd.h.pde, gd.h.pde.ana, method = "kendall", N = 12)
```

```
# Creat mid-point time vector for plotting:
midpoint <- seq(6.875, 1.125, length.out = 24)

# FIGURE 2:
# Put diversity estimates in a data frame:
dframe1 <- data.frame(midpoint, tde, sg.pde$median.curve, d15.pde$median.curve
    , h.pde$median.curve, d16.pde$median.curve)

# TDE:
a <- ggplot(dframe1, aes(x = midpoint, y = tde)) +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab("TDE") +
  scale_y_continuous(breaks = seq(0,6,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 18))

# SPDE:
b <- ggplot(dframe1, aes(x = midpoint, y = median)) +
  geom_ribbon(aes(ymin = low.95.quantile, ymax = high.95.quantile), fill = "#9
      ecae1") +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  xlab("Time (Ma)") +
  ylab("SPDE") +
  scale_y_continuous(breaks = seq(1,10,1), position = "right") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 18))

# D1PDE:
c <- ggplot(dframe1, aes(x = midpoint, y = median)) +
  geom_ribbon(aes(ymin = low.95.quantile, ymax = high.95.quantile), fill = "#9
      ecae1") +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab("D1PDE") +
  scale_y_continuous(breaks = seq(1,10,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('c') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 18))

# HPDE:
d <- ggplot(dframe1, aes(x = midpoint, y = median)) +
  geom_ribbon(aes(ymin = low.95.quantile, ymax = high.95.quantile), fill = "#9
      ecae1") +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
```

```
  ylab("HPDE") +
  scale_y_continuous(breaks = seq(1,10,1), position = "right") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('d') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 18))

# D2PDE:
e <- ggplot(dframe1, aes(x = midpoint, y = median)) +
  geom_ribbon(aes(ymin = low.95.quantile, ymax = high.95.quantile), fill = "#9
      ecae1") +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  xlab("Time (Ma)") +
  ylab("D2PDE") +
  scale_y_continuous(breaks = seq(1,10,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('e') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 18))

# Create multiplot:
ggarrange(a, b, c, d, e, ncol = 1)

# FIGURE 3:
# TDE against each PDE:
pdf("xyTDE.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))
# A: TDE against SPDE:
plot(gd.sg.pde, gd.tde, pch = 16, xlim = c(-1.5,2.5), ylim = c(-2,3), family =
     "Myriad Pro", xlab = expression(paste(Delta,"SPDE")), ylab = expression(
    paste(Delta,"TDE")))
# B: TDE against D1PDE:
plot(gd.d15.pde, gd.tde, pch = 16, xlim = c(-2,2.5), ylim = c(-2,3), family =
    "Myriad Pro", xlab = expression(paste(Delta,"D1PDE")), ylab = expression(
    paste(Delta,"TDE")))
# C: TDE against HPDE:
plot(gd.h.pde, gd.tde, pch = 16, xlim = c(-2,3.5), ylim = c(-2,3), family = "
    Myriad Pro", xlab = expression(paste(Delta,"HPDE")), ylab = expression(
    paste(Delta,"TDE")))
# D: TDE against D1PDE:
plot(gd.d16.pde, gd.tde, pch = 16, xlim = c(-1.5,2.5), ylim = c(-2,3), family
    = "Myriad Pro", xlab = expression(paste(Delta,"D2PDE")), ylab = expression
    (paste(Delta,"TDE")))
par(mfrow = c(1,1))
dev.off()

# FIGURE 4:
# PDE against PDE:
pdf("xyPDE.pdf", height = 5, width = 7)
par(mfrow = c(2,3))
par(mar = c(4,4.1,1,1))
# A: SPDE against D1PDE:
plot(gd.d15.pde, gd.sg.pde, pch = 16, cex.lab = 1.4, xlim = c(-2,2.5), ylim =
    c(-2,3), family = "Myriad Pro", xlab = expression(paste(Delta,"D1PDE")),
```

```
    ylab = expression(paste(Delta,"SPDE")))
# B: SPDE against HPDE:
plot(gd.h.pde, gd.sg.pde, pch = 16, cex.lab = 1.4, xlim = c(-2,3.1), ylim = c
    (-2,3), family = "Myriad Pro", xlab = expression(paste(Delta,"HPDE")),
    ylab = expression(paste(Delta,"SPDE")))
# C: SPDE against D2PDE:
plot(gd.d16.pde, gd.sg.pde, pch = 16, cex.lab = 1.4, xlim = c(-2,3), ylim = c
    (-2,3), family = "Myriad Pro", xlab = expression(paste(Delta,"D2PDE")),
    ylab = expression(paste(Delta,"SPDE")))
# D: D1PDE against HPDE:
plot(gd.h.pde, gd.d15.pde, pch = 16, cex.lab = 1.4, xlim = c(-2,3.1), ylim = c
    (-2,2.5), family = "Myriad Pro", xlab = expression(paste(Delta,"HPDE")),
    ylab = expression(paste(Delta,"D1PDE")))
# E: D1PDE against D2PDE:
plot(gd.d16.pde, gd.d15.pde, pch = 16, cex.lab = 1.4, xlim = c(-1.5,2.5), ylim
     = c(-2,2.5), family = "Myriad Pro", xlab = expression(paste(Delta,"D2PDE"
    )), ylab = expression(paste(Delta,"D1PDE")))
# F: HPDE against D2PDE:
plot(gd.d16.pde, gd.h.pde, pch = 16, cex.lab = 1.4, xlim = c(-1.5,2.5), ylim =
     c(-2,3.1), family = "Myriad Pro", xlab = expression(paste(Delta,"D2PDE"))
    , ylab = expression(paste(Delta,"HPDE")))
par(mfrow = c(1,1))
dev.off()

# FIGURE 5:
# Enter # per-taxon horizon counts:
pt.hbh <- c(1,4,7,3,4,8,22,1,2,4,1,1,13,31,8,13,4,26)

# Enter taxonomic names:
taxon <- c("Sahelanthropus tchadensis",
           "Orrorrin tugenensis",
           "Ardipithecus kadabba",
           "Ardipithecus ramidus",
           "Kenyanthropus platyops",
           "Australopithecus anamensis",
           "Australopithecus afarensis",
           "Australopithecus bahrelghazali",
           "Australopithecus deyiremeda",
           "Australopithecus africanus",
           "Australopithecus garhi",
           "Australopithecus sediba",
           "Paranthropus aethiopicus",
           "Paranthropus boisei",
           "Paranthropus robustus",
           "Homo habilis",
           "Homo rudolfensis",
           "Homo erectus")

data <- data.frame(pt.hbh, taxon)

# FIGURE 5:
ggplot(data, aes(x = reorder(taxon, pt.hbh), y = pt.hbh)) +
  geom_bar(stat = "identity") +
  ylab("Per-taxon horizon count") +
  coord_flip() +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.y = element_blank(),
        axis.text.y = element_text(face = "italic"),
        text = element_text(family = "Myriad Pro", size = 20))
```

```
# FIGURE 6:
time <- c(6.875,6.625,6.375,6.125,5.875,5.625,5.375,5.125,
  4.875,4.625,4.375,4.125,3.875,3.625,3.375,3.125,2.875,
  2.625,2.375,2.125,1.875,1.625,1.375,1.125)
FAD <- c(1,0,1,1,0,0,0,0,0,1,0,1,0,3,1,0,1,1,3,0,4,0,0,0)
LAD <- c(1,0,0,0,0,1,0,1,0,0,1,0,1,1,2,0,1,0,3,0,2,1,1,0)

horizons <- c(1,1,3,5,5,6,5,4,2,1,2,2,3,2,1,1,2,0,1,0,2,2,1,0,0,1)
horizons <- rev(horizons)

data <- data.frame(time, FAD, LAD, horizons)

# FIGURE y:
m <- ggplot(data, aes(x = time, y = FAD)) +
  geom_bar(stat = "identity") +
  ylab("FAD") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

n <- ggplot(data, aes(x = time, y = LAD)) +
  geom_bar(stat = "identity") +
  ylab("LAD") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

o <- ggplot(data, aes(x = time, y = horizons)) +
  geom_bar(stat = "identity") +
  xlab("Time (Ma)") +
  ylab("HBH") +
  scale_y_continuous(breaks = seq(0,24,4)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('c') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# Create multiplot:
ggarrange(m, n, o, ncol = 1)

# FIGURE 7:
# Median and SD node ages:
pdf("xyNodes.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))
# A: TDE against SPDE:
plot(dframe3$sg.cal3.nodes, dframe3$sd.sg, pch = 16, xlim = c(2,7), ylim = c
    (0.2,1), family = "Myriad Pro", xlab = expression(paste("Median ", italic(
    cal3), " node age (Ma)")), ylab = expression(paste("Std of ", italic(cal3)
    , " node ages")))
```

```
# B: TDE against D1PDE:
plot(dframe4$d15.cal3.nodes, dframe4$sd.d15, pch = 16, xlim = c(2,7), ylim = c
    (0.2,1), family = "Myriad Pro", xlab = expression(paste("Median ", italic(
    cal3), " node age (Ma)")), ylab = expression(paste("Std of ", italic(cal3)
    , " node ages")))
# C: TDE against HPDE:
plot(dframe4$h.cal3.nodes, dframe4$sd.h, pch = 16, xlim = c(2,7), ylim = c
    (0.2,0.8), family = "Myriad Pro", xlab = expression(paste("Median ",
    italic(cal3), " node age (Ma)")), ylab = expression(paste("Std of ",
    italic(cal3), " node ages")))
# D: TDE against D1PDE:
plot(dframe4$d16.cal3.nodes, dframe4$sd.d16, pch = 16, xlim = c(2,7), ylim = c
    (0.2,1), family = "Myriad Pro", xlab = expression(paste("Median ", italic(
    cal3), " node age (Ma)")), ylab = expression(paste("Std of ", italic(cal3)
    , " node ages")))
par(mfrow = c(1,1))
dev.off()

# End of Chapter 2 script
```

Chapter 3

```
# R code for Chapter 3: Geological and anthropogenic controls on the sampling
    of the early hominin fossil record

# Set working directory:
setwd("~/git-repo/thesis-scripts/Chapter3")

# Load packages:
require(openxlsx)
require(gdata)
require(Claddis)
require(pracma) # for interpolating
require(nlme) # for GLS
require(qpcR) # for AICc
require(egg) # for multi-plots
require(lmtest) # for Breusch-Pagan test
require(tseries) # for Jarque-Bera test
require(extrafont)
require(extrafontdb)

# Load data:
data <- read.xlsx("Chapter3_data.xlsx", sheet = 1)

# Load data:
ea.data <- read.xlsx("Chapter3_data.xlsx", sheet = 2)

# Load time interval data:
time.bins <- read.table(file = "time.bins.txt", header = TRUE)

# Call Graeme T. Lloyd's generalised differencing function:
gen.diff <- function(x, time) {
  # Suppress warning message:
  # if(cor.test(time, x)$p.value > 0.05) print("Warning: variables not
      significantly correlated, generalised differencing not recommended.")
  dt <- x - ((lsfit(time, x)$coefficients[2] * time) + lsfit(time, x)$
      coefficients[1])
  m <- lsfit(dt[1:(length(dt)-1)], dt[2:length(dt)])$coefficients[2]
  gendiffs <- dt[1:(length(dt) - 1)]- (dt[2:length(dt)] * m)
  gendiffs
}
```

```
# Call p.adjust function (from Cleary et al., 2015):
cor = function(x, y, method = c("pearson", "spearman", "kendall"), N) {
  test = cor.test(x, y, method = method)
  adj = p.adjust(test$p.value, "BH", n = N)
  display = c(test$est, test$p.value, adj)
  names(display)[1] = "Coefficient"
  names(display)[2] = "p-value"
  names(display)[3] = "p.adjusted"
  return(display)
}

# Perform generalised differencing function on each variable:
gd.tde <- gen.diff(data$tde, data$time) # TDE
gd.sg.pde <- gen.diff(data$sg.pde, data$time) # SPDE
gd.d15.pde <- gen.diff(data$d15.pde, data$time) # D1PDE
gd.h.pde <- gen.diff(data$h.pde, data$time) # HPDE
gd.d16.pde <- gen.diff(data$d16.pde, data$time) # D2PDE
gd.hbc <- gen.diff(data$hbc, data$time) # hominin collections
gd.hbf <- gen.diff(data$hbf, data$time) # hominin formations
gd.hbl <- gen.diff(data$hbl, data$time) # hominin localities
gd.pbf <- gen.diff(data$pbf, data$time) # primate formations
gd.mbf <- gen.diff(data$mbf, data$time) # mammal formations
gd.arid <- gen.diff(data$aridity.mean, data$time) # aridity (mean)
gd.arid.sd <- gen.diff(data$aridity.sd, data$time) # aridity (SD)
gd.palaeoarea <- gen.diff(data$convex.hull, data$time) # convex hull area
gd.grid.cell <- gen.diff(data$grid.cell.occupancy, data$time) # grid-cell
    occupancy

# Pairwise tests:

# TDE vs. HBC:
cor(gd.tde, gd.hbc, method = "spearman", N = 44)
cor(gd.tde, gd.hbc, method = "kendall", N = 44)

# SPDE vs. HBC:
cor(gd.sg.pde, gd.hbc, method = "spearman", N = 44)
cor(gd.sg.pde, gd.hbc, method = "kendall", N = 44)

# D1PDE vs. HBC:
cor(gd.d15.pde, gd.hbc, method = "spearman", N = 44)
cor(gd.d15.pde, gd.hbc, method = "kendall", N = 44)

# HPDE vs. HBC:
cor(gd.h.pde, gd.hbc, method = "spearman", N = 44)
cor(gd.h.pde, gd.hbc, method = "kendall", N = 44)

# D2PDE vs. HBC:
cor(gd.d16.pde, gd.hbc, method = "spearman", N = 44)
cor(gd.d16.pde, gd.hbc, method = "kendall", N = 44)

# TDE vs. HBF:
cor(gd.tde, gd.hbf, method = "spearman", N = 44)
cor(gd.tde, gd.hbf, method = "kendall", N = 44)

# SPDE vs. HBF:
cor(gd.sg.pde, gd.hbf, method = "spearman", N = 44)
cor(gd.sg.pde, gd.hbf, method = "kendall", N = 44)

# D1PDE vs. HBF:
cor(gd.d15.pde, gd.hbf, method = "spearman", N = 44)
```

```
cor(gd.d15.pde, gd.hbf, method = "kendall", N = 44)

# HPDE vs. HBF:
cor(gd.h.pde, gd.hbf, method = "spearman", N = 44)
cor(gd.h.pde, gd.hbf, method = "kendall", N = 44)

# D2PDE vs. HBF:
cor(gd.d16.pde, gd.hbf, method = "spearman", N = 4)
cor(gd.d16.pde, gd.hbf, method = "kendall", N = 44)

# TDE vs. PBF:
cor(gd.tde, gd.pbf, method = "spearman", N = 44)
cor(gd.tde, gd.pbf, method = "kendall", N = 44)

# SPDE vs. PBF:
cor(gd.sg.pde, gd.pbf, method = "spearman", N = 44)
cor(gd.sg.pde, gd.pbf, method = "kendall", N = 4)

# D1PDE vs. PBF:
cor(gd.d15.pde, gd.pbf, method = "spearman", N = 44)
cor(gd.d15.pde, gd.pbf, method = "kendall", N = 44)

# HPDE vs. PBF:
cor(gd.h.pde, gd.pbf, method = "spearman", N = 44)
cor(gd.h.pde, gd.pbf, method = "kendall", N = 44)

# D2PDE vs. PBF:
cor(gd.d16.pde, gd.pbf, method = "spearman", N = 44)
cor(gd.d16.pde, gd.pbf, method = "kendall", N = 44)

# TDE vs. MBF:
cor(gd.tde, gd.mbf, method = "spearman", N = 44)
cor(gd.tde, gd.mbf, method = "kendall", N = 44)

# SPDE vs. MBF:
cor(gd.sg.pde, gd.mbf, method = "spearman", N = 44)
cor(gd.sg.pde, gd.mbf, method = "kendall", N = 44)

# D1PDE vs. MBF:
cor(gd.d15.pde, gd.mbf, method = "spearman", N = 44)
cor(gd.d15.pde, gd.mbf, method = "kendall", N = 44)

# HPDE vs. MBF:
cor(gd.h.pde, gd.mbf, method = "spearman", N = 44)
cor(gd.h.pde, gd.mbf, method = "kendall", N = 44)

# D2PDE vs. MBF:
cor(gd.d16.pde, gd.pbf, method = "spearman", N = 44)
cor(gd.d16.pde, gd.pbf, method = "kendall", N = 44)

# TDE vs. palaeoarea:
cor(gd.tde, gd.palaeoarea, method = "spearman", N = 44)
cor(gd.tde, gd.palaeoarea, method = "kendall", N = 44)

# SPDE vs. palaeoarea:
cor(gd.sg.pde, gd.palaeoarea, method = "spearman", N = 44)
cor(gd.sg.pde, gd.palaeoarea, method = "kendall", N = 4)

# D1PDE vs. palaeoarea:
cor(gd.d15.pde, gd.palaeoarea, method = "spearman", N = 44)
cor(gd.d15.pde, gd.palaeoarea, method = "kendall", N = 44)
```

```
# HPDE vs. palaeoarea:
cor(gd.h.pde, gd.palaeoarea, method = "spearman", N = 44)
cor(gd.h.pde, gd.palaeoarea, method = "kendall", N = 44)

# D2PDE vs. palaeoarea:
cor(gd.d16.pde, gd.palaeoarea, method = "spearman", N = 44)
cor(gd.d16.pde, gd.palaeoarea, method = "kendall", N = 44)

# TDE vs. aridity:
cor(gd.tde, gd.arid, method = "spearman", N = 44)
cor(gd.tde, gd.arid, method = "kendall", N = 44)

# SPDE vs. aridity:
cor(gd.sg.pde, gd.arid, method = "spearman", N = 44)
cor(gd.sg.pde, gd.arid, method = "kendall", N = 44)

# D1PDE vs. aridity:
cor(gd.d15.pde, gd.arid, method = "spearman", N = 44)
cor(gd.d15.pde, gd.arid, method = "kendall", N = 44)

# HPDE vs. aridity:
cor(gd.h.pde, gd.arid, method = "spearman", N = 44)
cor(gd.h.pde, gd.arid, method = "kendall", N = 44)

# D2PDE vs. aridity:
cor(gd.d16.pde, gd.arid, method = "spearman", N = 44)
cor(gd.d16.pde, gd.arid, method = "kendall", N = 44)

# HBC vs. HBF:
cor(gd.hbc, gd.hbf, method = "spearman", N = 44)
cor(gd.hbc, gd.hbf, method = "kendall", N = 44)

# HBC vs. PBF:
cor(gd.hbc, gd.pbf, method = "spearman", N = 44)
cor(gd.hbc, gd.pbf, method = "kendall", N = 44)

# HBC vs. MBF:
cor(gd.hbc, gd.mbf, method = "spearman", N = 44)
cor(gd.hbc, gd.mbf, method = "kendall", N = 44)

# HBC vs. palaeoarea:
cor(gd.hbc, gd.palaeoarea, method = "spearman", N = 44)
cor(gd.hbc, gd.palaeoarea, method = "kendall", N = 44)

# HBC vs. aridity:
cor(gd.hbc, gd.arid, method = "spearman", N = 44)
cor(gd.hbc, gd.arid, method = "kendall", N = 44)

# HBF vs. PBF:
cor(gd.hbf, gd.pbf, method = "spearman", N = 44)
cor(gd.hbf, gd.pbf, method = "kendall", N = 44)

# HBF vs. MBF:
cor(gd.hbf, gd.mbf, method = "spearman", N = 44)
cor(gd.hbf, gd.mbf, method = "kendall", N = 44)

# HBF vs. palaeoarea:
cor(gd.hbf, gd.palaeoarea, method = "spearman", N = 44)
cor(gd.hbf, gd.palaeoarea, method = "kendall", N = 44)
```

```
# HBF vs. aridity:
cor(gd.hbf, gd.arid, method = "spearman", N = 44)
cor(gd.hbf, gd.arid, method = "kendall", N = 44)

# PBF vs. MBF:
cor(gd.pbf, gd.mbf, method = "spearman", N = 44)
cor(gd.pbf, gd.mbf, method = "kendall", N = 44)

# PBF vs. palaeoarea:
cor(gd.pbf, gd.palaeoarea, method = "spearman", N = 44)
cor(gd.pbf, gd.palaeoarea, method = "kendall", N = 44)

# PBF vs. aridity:
cor(gd.pbf, gd.arid, method = "spearman", N = 44)
cor(gd.pbf, gd.arid, method = "kendall", N = 44)

# MBF vs. palaeoarea:
cor(gd.mbf, gd.palaeoarea, method = "spearman", N = 44)
cor(gd.mbf, gd.palaeoarea, method = "kendall", N = 44)

# MBF vs. aridity:
cor(gd.mbf, gd.arid, method = "spearman", N = 44)
cor(gd.mbf, gd.arid, method = "kendall", N = 44)

# Palaeoarea vs. HBL:
cor(gd.palaeoarea, gd.hbl, method = "spearman", N = 1)
cor(gd.palaeoarea, gd.hbl, method = "kendall", N = 1)

cor(gd.palaeoarea, gd.grid.cell, method = "spearman", N = 1)
cor(gd.palaeoarea, gd.grid.cell, method = "kendall", N = 1)

# Input hominin-bearing horizon counts from Hopley (2017):
horizons <- c(1,1,3,5,5,6,5,4,2,1,2,2,3,2,1,1,2,0,1,0,2,2,1,0,0,1)
tde.hopley <- c(1,2,4,11,25,19,6,6,4,1,6,8,5,2,5,2,2,0,1,0,4,2,1,0,0,1)
time.hopley <- seq(0.625, 6.875, 0.25)

# Perform generalised differencing:
gd.horizons <- gen.diff(horizons, time.hopley)
gd.tde.hopley <- gen.diff(tde.hopley, time.hopley)

# Get test statistic:
cor.test(gd.tde.hopley, gd.horizons, method = "spearman")

# Calculate ghost lineage diversity estimate (GDE) for each diversity estimate
    :
d16.gde <- data$d16.pde - data$tde
h.gde <- data$h.pde - data$tde
d15.gde <- data$d15.pde - data$tde
sg.gde <- data$sg.pde - data$tde

gd.d16.gde <- gen.diff(d16.gde, data$time) # D2GDE
gd.h.gde <- gen.diff(h.gde, data$time) # HGDE
gd.d15.gde <- gen.diff(d15.gde, data$time) # D1GDE
gd.sg.gde <- gen.diff(sg.gde, data$time) # SGDE

# SGDE vs. HBC:
cor(gd.sg.gde, gd.hbc, method = "spearman", N = 20)
cor(gd.sg.gde, gd.hbc, method = "kendall", N = 20)

# D1GDE vs. HBC:
cor(gd.d15.gde, gd.hbc, method = "spearman", N = 20)
```

```
cor(gd.d15.gde, gd.hbc, method = "kendall", N = 20)

# HGDE vs. HBC:
cor(gd.h.gde, gd.hbc, method = "spearman", N = 20)
cor(gd.h.gde, gd.hbc, method = "kendall", N = 20)

# D2GDE vs. HBC:
cor(gd.d16.gde, gd.hbc, method = "spearman", N = 20)
cor(gd.d16.gde, gd.hbc, method = "kendall", N = 20)

# SGDE vs. HBF:
cor(gd.sg.gde, gd.hbf, method = "spearman", N = 20)
cor(gd.sg.gde, gd.hbf, method = "kendall", N = 20)

# D1GDE vs. HBF:
cor(gd.d15.gde, gd.hbf, method = "spearman", N = 20)
cor(gd.d15.gde, gd.hbf, method = "kendall", N = 20)

# HGDE vs. HBF:
cor(gd.h.gde, gd.hbf, method = "spearman", N = 20)
cor(gd.h.gde, gd.hbf, method = "kendall", N = 20)

# D2GDE vs. HBF:
cor(gd.d16.gde, gd.hbf, method = "spearman", N = 20)
cor(gd.d16.gde, gd.hbf, method = "kendall", N = 20)

# SGDE vs. PBF:
cor(gd.sg.gde, gd.pbf, method = "spearman", N = 20)
cor(gd.sg.gde, gd.pbf, method = "kendall", N = 20)

# D1GDE vs. PBF:
cor(gd.d15.gde, gd.pbf, method = "spearman", N = 20)
cor(gd.d15.gde, gd.pbf, method = "kendall", N = 20)

# HGDE vs. PBF:
cor(gd.h.gde, gd.pbf, method = "spearman", N = 20)
cor(gd.h.gde, gd.pbf, method = "kendall", N = 20)

# D2GDE vs. PBF:
cor(gd.d16.gde, gd.pbf, method = "spearman", N = 20)
cor(gd.d16.gde, gd.pbf, method = "kendall", N = 20)

# SGDE vs. MBF:
cor(gd.sg.gde, gd.mbf, method = "spearman", N = 20)
cor(gd.sg.gde, gd.mbf, method = "kendall", N = 20)

# D1GDE vs. MBF:
cor(gd.d15.gde, gd.mbf, method = "spearman", N = 20)
cor(gd.d15.gde, gd.mbf, method = "kendall", N = 20)

# HGDE vs. MBF:
cor(gd.h.gde, gd.mbf, method = "spearman", N = 20)
cor(gd.h.gde, gd.mbf, method = "kendall", N = 20)

# D2GDE vs. MBF:
cor(gd.d16.gde, gd.mbf, method = "spearman", N = 20)
cor(gd.d16.gde, gd.mbf, method = "kendall", N = 20)

# SGDE vs. palaeoarea:
cor(gd.sg.gde, gd.palaeoarea, method = "spearman", N = 20)
cor(gd.sg.gde, gd.palaeoarea, method = "kendall", N = 20)
```

```
# D1GDE vs. palaeoarea:
cor(gd.d15.gde, gd.palaeoarea, method = "spearman", N = 20)
cor(gd.d15.gde, gd.palaeoarea, method = "kendall", N = 20)

# HGDE vs. palaeoarea:
cor(gd.h.gde, gd.palaeoarea, method = "spearman", N = 20)
cor(gd.h.gde, gd.palaeoarea, method = "kendall", N = 20)

# D2GDE vs. palaeoarea:
cor(gd.d16.gde, gd.palaeoarea, method = "spearman", N = 20)
cor(gd.d16.gde, gd.palaeoarea, method = "kendall", N = 20)

# GDE:PDE ratio:
sg.gpratio <- sg.gde/data$sg.pde
d15.gpratio <- d15.gde/data$d15.pde
h.gpratio <- h.gde/data$h.pde
d16.gpratio <- d16.gde/data$d16.pde

gd.sg.gprat <- gen.diff(sg.gpratio, data$time) # Strait & Grine (2004) GDE:PDE
gd.d15.gprat <- gen.diff(d15.gpratio, data$time) # Dembo et al. (2015) GDE:PDE
gd.h.gprat <- gen.diff(h.gpratio, data$time) # Haile-Selassie et al. (2015)
    GDE:PDE
gd.d16.gprat <- gen.diff(d16.gpratio, data$time) # Dembo et al. (2016) GDE:PDE

# SGDE:SPDE vs. HBC:
cor(gd.sg.gprat, gd.hbc, method = "spearman", N = 20)
cor(gd.sg.gprat, gd.hbc, method = "kendall", N = 20)

# D1GDE:D1PDE vs. HBC:
cor(gd.d15.gprat, gd.hbc, method = "spearman", N = 20)
cor(gd.d15.gprat, gd.hbc, method = "kendall", N = 20)

# HGDE:HPDE vs. HBC:
cor(gd.h.gprat, gd.hbc, method = "spearman", N = 20)
cor(gd.h.gprat, gd.hbc, method = "kendall", N = 20)

# D2GDE:D2PDE vs. HBC:
cor(gd.d16.gprat, gd.hbc, method = "spearman", N = 20)
cor(gd.d16.gprat, gd.hbc, method = "kendall", N = 20)

# SGDE:SPDE vs. HBF:
cor(gd.sg.gprat, gd.hbf, method = "spearman", N = 20)
cor(gd.sg.gprat, gd.hbf, method = "kendall", N = 20)

# D1GDE:D1PDE vs. HBF:
cor(gd.d15.gprat, gd.hbf, method = "spearman", N = 20)
cor(gd.d15.gprat, gd.hbf, method = "kendall", N = 20)

# HGDE:HPDE vs. HBF:
cor(gd.h.gprat, gd.hbf, method = "spearman", N = 20)
cor(gd.h.gprat, gd.hbf, method = "kendall", N = 20)

# D2GDE:D2PDE vs. HBF:
cor(gd.d16.gprat, gd.hbf, method = "spearman", N = 20)
cor(gd.d16.gprat, gd.hbf, method = "kendall", N = 20)

# S_GDE:PDE vs. PBF:
cor(gd.sg.gprat, gd.pbf, method = "spearman", N = 20)
cor(gd.sg.gprat, gd.pbf, method = "kendall", N = 20)
```

```
# D15_GDE:PDE vs. PBF:
cor(gd.d15.gprat, gd.pbf, method = "spearman", N = 20)
cor(gd.d15.gprat, gd.pbf, method = "kendall", N = 20)

# H_GDE:PDE vs. PBF:
cor(gd.h.gprat, gd.pbf, method = "spearman", N = 20)
cor(gd.h.gprat, gd.pbf, method = "kendall", N = 20)

# D16_GDE:PDE vs. PBF:
cor(gd.d16.gprat, gd.pbf, method = "spearman", N = 20)
cor(gd.d16.gprat, gd.pbf, method = "kendall", N = 20)

# S_GDE:PDE vs. MBF:
cor(gd.sg.gprat, gd.mbf, method = "spearman", N = 20)
cor(gd.sg.gprat, gd.mbf, method = "kendall", N = 20)

# D15_GDE:PDE vs. MBF:
cor(gd.d15.gprat, gd.mbf, method = "spearman", N = 20)
cor(gd.d15.gprat, gd.mbf, method = "kendall", N = 20)

# H_GDE:PDE vs. MBF:
cor(gd.h.gprat, gd.mbf, method = "spearman", N = 20)
cor(gd.h.gprat, gd.mbf, method = "kendall", N = 20)

# D16_GDE:PDE vs. MBF:
cor(gd.d16.gprat, gd.mbf, method = "spearman", N = 20)
cor(gd.d16.gprat, gd.mbf, method = "kendall", N = 20)

# S_GDE:PDE vs. palaeoarea:
cor(gd.sg.gprat, gd.palaeoarea, method = "spearman", N = 20)
cor(gd.sg.gprat, gd.palaeoarea, method = "kendall", N = 20)

# D15_GDE:PDE vs. palaeoarea:
cor(gd.d15.gprat, gd.palaeoarea, method = "spearman", N = 20)
cor(gd.d15.gprat, gd.palaeoarea, method = "kendall", N = 20)

# H_GDE:PDE vs. palaeoarea:
cor(gd.h.gprat, gd.palaeoarea, method = "spearman", N = 20)
cor(gd.h.gprat, gd.palaeoarea, method = "kendall", N = 20)

# D16_GDE:PDE vs. palaeoarea:
cor(gd.d16.gprat, gd.palaeoarea, method = "spearman", N = 20)
cor(gd.d16.gprat, gd.palaeoarea, method = "kendall", N = 20)

# Stratigraphic gaps proxy from Benton et al. (2011):
# strat.gaps <- max(data$pbf) - data$pbf
# gd.strat.gaps <- gen.diff(strat.gaps, data$time)

# Generalised Least Squares (GLS):

# Load function from Cleary et al. (2015) to calculate model weights:
weighted <- function (aic){
  aic.wt <- exp(-0.5 * (aic - min(aic)))/sum(exp(-0.5 * (aic - min(aic))))
  return(aic.wt)
}

D <- data$tde
C <- data$hbc
P <- data$pbf
A <- data$aridity.mean
```

```
# Create all possible models plus null:
gls.null <- gls(D ~ 1, correlation = corARMA(p = 1), method = "ML")
gls.1 <- gls(D ~ C + P + A, correlation = corARMA(p = 1), method = "ML")
gls.2 <- gls(D ~ C + P, correlation = corARMA(p = 1), method = "ML")
gls.3 <- gls(D ~ C + A, correlation = corARMA(p = 1), method = "ML")
gls.4 <- gls(D ~ P + A, correlation = corARMA(p = 1), method = "ML")
gls.5 <- gls(D ~ C, correlation = corARMA(p = 1), method = "ML")
gls.6 <- gls(D ~ P, correlation = corARMA(p = 1), method = "ML")
gls.7 <- gls(D ~ A, correlation = corARMA(p = 1), method = "ML")

weights <- weighted(c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3),
    AICc(gls.4), AICc(gls.5), AICc(gls.6), AICc(gls.7))); weights

AICcscores <- (c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3), AICc(
    gls.4), AICc(gls.5), AICc(gls.6), AICc(gls.7))); AICcscores

anova(gls.null, gls.1, gls.2, gls.3, gls.4, gls.5, gls.6, gls.7)

# Calculate n for generalised R2:
n <- length(data$tde) # here n refers to the length of the time series

# Generalised R2 from Nagelkerke (1991):
1 - exp(-1 * 2/n * (gls.1$logLik - gls.null$logLik))

# EARS analysis:

# Perform generalised differencing function on each variable:
gd.ea.tde <- gen.diff(ea.data$tde, ea.data$time) # Eastern African TDE
gd.ea.hbc <- gen.diff(ea.data$hbc, ea.data$time) # hominin collections
gd.ea.hbf <- gen.diff(ea.data$hbf, ea.data$time) # hominin formations
gd.ea.pbf <- gen.diff(ea.data$pbf, ea.data$time) # primate formations
gd.ea.mbf <- gen.diff(ea.data$mbf, ea.data$time) # mammal formations
gd.lmean <- gen.diff(ea.data$lvi.mean, ea.data$time) # LVI (mean)
gd.lmax <- gen.diff(ea.data$lvi.max, ea.data$time) # LVI (max)
gd.arid <- gen.diff(ea.data$aridity.mean, ea.data$time) # Arabian Sea aridity
gd.arid.west <- gen.diff(ea.data$aridity.west.mean, ea.data$time) # West
    African aridity

# eaTDE vs. LVI:
cor(gd.ea.tde, gd.lmean, method = "spearman", N = 15)
cor(gd.ea.tde, gd.lmean, method = "kendall", N = 15)

# eaTDE vs. Arabian Sea aridity:
cor(gd.ea.tde, gd.arid, method = "spearman", N = 15)
cor(gd.ea.tde, gd.arid, method = "kendall", N = 15)

# eaTDE vs. West African aridity:
cor(gd.ea.tde, gd.arid.west, method = "spearman", N = 15)
cor(gd.ea.tde, gd.arid.west, method = "kendall", N = 15)

# eaTDE vs. eaHBC:
cor(gd.ea.tde, gd.ea.hbc, method = "spearman", N = 15)
cor(gd.ea.tde, gd.ea.hbc, method = "kendall", N = 15)

# eaTDE vs. eaPBF:
cor(gd.ea.tde, gd.ea.pbf, method = "spearman", N = 15)
cor(gd.ea.tde, gd.ea.pbf, method = "kendall", N = 15)

# LVI vs. Arabian Sea aridity:
cor(gd.lmean, gd.arid, method = "spearman", N = 15)
cor(gd.lmean, gd.arid, method = "kendall", N = 15)
```

```
# LVI vs. West African aridity:
cor(gd.lmean, gd.arid.west, method = "spearman", N = 15)
cor(gd.lmean, gd.arid.west, method = "kendall", N = 15)

# LVI vs. eaHBC:
cor(gd.lmean, gd.ea.hbc, method = "spearman", N = 15)
cor(gd.lmean, gd.ea.hbc, method = "kendall", N = 15)

# LVI vs. eaPBF:
cor(gd.lmean, gd.ea.pbf, method = "spearman", N = 15)
cor(gd.lmean, gd.ea.pbf, method = "kendall", N = 15)

# Arabian Sea aridity vs. West African aridity:
cor(gd.arid.west, gd.arid, method = "spearman", N = 15)
cor(gd.arid.west, gd.arid, method = "kendall", N = 15)

# Arabian Sea aridity vs. eaHBC:
cor(gd.arid, gd.ea.hbc, method = "spearman", N = 15)
cor(gd.arid, gd.ea.hbc, method = "kendall", N = 15)

# Arabian Sea aridity vs. eaPBF:
cor(gd.arid, gd.ea.pbf, method = "spearman", N = 15)
cor(gd.arid, gd.ea.pbf, method = "kendall", N = 15)

# West African aridity vs. eaHBC:
cor(gd.arid.west, gd.ea.hbc, method = "spearman", N = 15)
cor(gd.arid.west, gd.ea.hbc, method = "kendall", N = 15)

# West African aridity vs. eaPBF:
cor(gd.arid.west, gd.ea.pbf, method = "spearman", N = 15)
cor(gd.arid.west, gd.ea.pbf, method = "kendall", N = 15)

# eaPBF vs. eaHBC:
cor(gd.ea.pbf, gd.ea.hbc, method = "spearman", N = 15)
cor(gd.ea.pbf, gd.ea.hbc, method = "kendall", N = 15)

# eaTDE vs. LVI (max):
cor(gd.ea.tde, gd.lmax, method = "spearman", N = 10)
cor(gd.ea.tde, gd.lmax, method = "kendall", N = 10)

# eaHBC vs. LVI (mean):
cor(gd.lmean, gd.ea.hbc, method = "spearman", N = 10)
cor(gd.lmean, gd.ea.hbc, method = "kendall", N = 10)

# eaHBC vs. LVI (max):
cor(gd.lmax, gd.ea.hbc, method = "spearman", N = 10)
cor(gd.lmax, gd.ea.hbc, method = "kendall", N = 10)

# eaPBF vs. LVI (mean):
cor(gd.lmean, gd.ea.pbf, method = "spearman", N = 10)
cor(gd.lmean, gd.ea.pbf, method = "kendall", N = 10)

# eaPBF vs. LVI (max):
cor(gd.lmax, gd.ea.pbf, method = "spearman", N = 10)
cor(gd.lmax, gd.ea.pbf, method = "kendall", N = 10)

# LVI (mean) vs. LVI (max):
cor(gd.lmean, gd.lmax, method = "spearman", N = 10)
cor(gd.lmean, gd.lmax, method = "kendall", N = 10)
```

```
# GLS:
y1 <- ea.data$tde
x1 <- ea.data$hbc
x2 <- ea.data$pbf
x3 <- ea.data$aridity.mean
x4 <- ea.data$aridity.west.mean
x5 <- ea.data$lvi.mean

# Create all possible models plus null:
gls.null <- gls(y1~1, correlation = corARMA(p = 1), method = "ML") # 1
gls.1 <- gls(y1~x1+x2+x3+x4+x5, correlation = corARMA(p = 1), method = "ML") #
     2
gls.2 <- gls(y1~x1+x2+x3+x4, correlation = corARMA(p = 1), method = "ML") # 3
gls.3 <- gls(y1~x1+x2+x3+x5, correlation = corARMA(p = 1), method = "ML") # 4
gls.4 <- gls(y1~x1+x2+x4+x5, correlation = corARMA(p = 1), method = "ML") # 5
gls.5 <- gls(y1~x1+x3+x4+x5, correlation = corARMA(p = 1), method = "ML") # 6
gls.6 <- gls(y1~x2+x3+x4+x5, correlation = corARMA(p = 1), method = "ML") # 7
gls.7 <- gls(y1~x1+x2+x3, correlation = corARMA(p = 1), method = "ML") # 8
gls.8 <- gls(y1~x1+x2+x4, correlation = corARMA(p = 1), method = "ML") # 9
gls.9 <- gls(y1~x1+x2+x5, correlation = corARMA(p = 1), method = "ML") # 10
gls.10 <- gls(y1~x1+x3+x4, correlation = corARMA(p = 1), method = "ML") # 11
gls.11 <- gls(y1~x1+x3+x5, correlation = corARMA(p = 1), method = "ML") # 12
gls.12 <- gls(y1~x1+x4+x5, correlation = corARMA(p = 1), method = "ML") # 13
gls.13 <- gls(y1~x2+x3+x4, correlation = corARMA(p = 1), method = "ML") # 14
gls.14 <- gls(y1~x2+x3+x5, correlation = corARMA(p = 1), method = "ML") # 15
gls.15 <- gls(y1~x2+x4+x5, correlation = corARMA(p = 1), method = "ML") # 16
gls.16 <- gls(y1~x3+x4+x5, correlation = corARMA(p = 1), method = "ML") # 17
gls.17 <- gls(y1~x1+x2, correlation = corARMA(p = 1), method = "ML") # 18
gls.18 <- gls(y1~x1+x3, correlation = corARMA(p = 1), method = "ML") # 19
gls.19 <- gls(y1~x1+x4, correlation = corARMA(p = 1), method = "ML") # 20
gls.20 <- gls(y1~x1+x5, correlation = corARMA(p = 1), method = "ML") # 21
gls.21 <- gls(y1~x2+x3, correlation = corARMA(p = 1), method = "ML") # 22
gls.22 <- gls(y1~x2+x4, correlation = corARMA(p = 1), method = "ML") # 23
gls.23 <- gls(y1~x2+x5, correlation = corARMA(p = 1), method = "ML") # 24
gls.24 <- gls(y1~x3+x4, correlation = corARMA(p = 1), method = "ML") # 25
gls.25 <- gls(y1~x3+x5, correlation = corARMA(p = 1), method = "ML") # 26
gls.26 <- gls(y1~x4+x5, correlation = corARMA(p = 1), method = "ML") # 27
gls.27 <- gls(y1~x1, correlation = corARMA(p = 1), method = "ML") # 28
gls.28 <- gls(y1~x2, correlation = corARMA(p = 1), method = "ML") # 29
gls.29 <- gls(y1~x3, correlation = corARMA(p = 1), method = "ML") # 30
gls.30 <- gls(y1~x4, correlation = corARMA(p = 1), method = "ML") # 31
gls.31 <- gls(y1~x5, correlation = corARMA(p = 1), method = "ML") # 32

# Compute and compare AIC weights (AICw) for each model:
weights <- weighted(c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3),
    AICc(gls.4), AICc(gls.5), AICc(gls.6), AICc(gls.7), AICc(gls.8), AICc(gls
    .9), AICc(gls.10), AICc(gls.11), AICc(gls.12), AICc(gls.13), AICc(gls.14),
     AICc(gls.15), AICc(gls.16), AICc(gls.17), AICc(gls.18), AICc(gls.19),
    AICc(gls.20), AICc(gls.21), AICc(gls.22), AICc(gls.23), AICc(gls.24), AICc
    (gls.25), AICc(gls.26), AICc(gls.27), AICc(gls.28), AICc(gls.29), AICc(gls
    .30), AICc(gls.31))); weights

# Compute and compare AICc values for each model:
AICcscores <- (c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3), AICc(
    gls.4), AICc(gls.5), AICc(gls.6), AICc(gls.7), AICc(gls.8), AICc(gls.9),
    AICc(gls.10), AICc(gls.11), AICc(gls.12), AICc(gls.13), AICc(gls.14), AICc
    (gls.15), AICc(gls.16), AICc(gls.17), AICc(gls.18), AICc(gls.19), AICc(gls
    .20), AICc(gls.21), AICc(gls.22), AICc(gls.23), AICc(gls.24), AICc(gls.25)
    , AICc(gls.26), AICc(gls.27), AICc(gls.28), AICc(gls.29), AICc(gls.30),
    AICc(gls.31))); AICcscores
```

```
# Compare all gls models and derive a global p-value using ANOVA:
anova(gls.null, gls.1, gls.2, gls.3, gls.4, gls.5, gls.6, gls.7, gls.8, gls.9,
     gls.10, gls.11, gls.12, gls.13, gls.14, gls.15, gls.16, gls.17, gls.18,
    gls.19, gls.20, gls.21, gls.22, gls.23, gls.24, gls.25, gls.26, gls.27,
    gls.28, gls.29, gls.30, gls.31)

# Calculate n for generalised R2:
n <- length(ea.data$tde) # here n refers to the length of the time series

# Generalised R2 from Nagelkerke (1991):
1 - exp(-1 * 2/n * (gls.1$logLik - gls.null$logLik))

# Count mammal-bearing formations in bins:

formations <- read.xlsx("MBF.xlsx", sheet = 1, colNames = TRUE)
FAD <- as.numeric(formations$Start)
LAD <- as.numeric(formations$End)

# FAD <- as.numeric(timeData[,1]); LAD <- as.numeric(timeData[,2])
int.start <- time.bins[,1]
int.end <- time.bins[,2]
midtimes <- (int.start + int.end)/2

mbf <- sapply(1:length(midtimes), function(x) sum(FAD >= int.end[x]) - sum(LAD
     > int.start[x])); mbf

# Palaeoarea estimation:
source("http://www.graemetlloyd.com/pubdata/functions_2.r")

# latitudes <- c(a, b, c)
# longitudes <- c(x, y, z)

sphpolyarea(vlat = latitudes, vlon = longitudes)

# Using randomised trials to tests the strength of the correlation between
    hominin diversity and hominin-bearing formations (HBF):

# Get mid-point age of each interval:
time <- apply(time.bins, 1, median)

# Enter hominin-bearing formations (HBFs):
hbf <- data$hbf

r <- length(hbf) # number of time bins (midpoint ages)
c <- 40 # number of trials

# Generate random matrix of 0s, 1s, and 2s:
randtax <- matrix(0:2, r, c)

# For each row and each column:
for(i in 1:r){
  for(j in 1:c){

    # Repeat and sum output for time bins of hbf > 1:
    randtax[i, j] <- sum(matrix(sample(c(0:2), hbf[i], replace = TRUE)))

  }

}

# Create empty list:
```

```
gdrandtax <- list()

# For each randomised FFC time series:
for (i in 1:(ncol(randtax))) {

  # Perform generalised differencing:
  gdrandtax[[i]] <- gen.diff(x = randtax[,i], time = time)

}

# Create empty list for Spearman's rho and p-value:
cor.stat <- list()
cor.p <- list()

# For each randomised FFC:
for (i in 1:(length(gdrandtax))) {

  # Perform correlation and extract each rho:
  cor.stat[[i]] <- cor.test(x = gd.hbf, gdrandtax[[i]], method = "spearman")$
      estimate

  # Perform correlation and extract each p-value:
  cor.p[[i]] <- cor.test(x = gd.hbf, gdrandtax[[i]], method = "spearman")$p.
      value

}

max(as.numeric(cor.stat)) # max
min(as.numeric(cor.stat)) # min
mean(as.numeric(cor.stat)) # mean

# Calculate critical value for an alpha of 0.05:
cutoff <- qnorm(0.95, mean(as.numeric(cor.stat)), sd(as.numeric(cor.stat)))

# Calculate Spearman's rho and p for raw data:
spearman <- as.numeric(cor.test(x = gd.tde, y = gd.hbf, method = "spearman")$
    estimate)

# Plot histogram of Spearman's rho values:
hist(as.numeric(cor.stat), xlim = c(0, 1), xlab = "Spearman's rho",
     main = "Spearman's rho for randomised TDE against HBF", col = "black",
         breaks = seq(0,1,0.01))
lines(x = c(cutoff, cutoff), y = c(0, 20), lty = 5)
lines(x = c(spearman, spearman), y = c(0, 20), lty = 5, col = "red")

# Calculate percentage of trials stronger than actual Spearman's value:
(sum(as.numeric(cor.stat) > spearman)/40) * 100 # 70% (i.e., 28/40)

# Using randomised trials to tests the strength of the correlation between
    hominin diversity and primate-bearing formations (PBF):

# Get mid-point age of each interval:
time <- apply(time.bins, 1, median)

# Enter primate-bearing formations (PBFs):
pbf <- data$pbf

r <- length(pbf) # number of time bins (midpoint ages)
c <- 40 # number of trials

# Generate random matrix of 0s, 1s, and 2s:
```

```
randtax <- matrix(0:1, r, c)

# For each row and each column:
for(i in 1:r){
  for(j in 1:c){

    # Repeat and sum output for time bins of hbf > 1:
    randtax[i, j] <- sum(matrix(sample(c(0:1), pbf[i], replace = TRUE)))

  }

}

# Create empty list:
gdrandtax <- list()

# For each randomised FFC time series:
for (i in 1:(ncol(randtax))) {

  # Perform generalised differencing:
  gdrandtax[[i]] <- gen.diff(x = randtax[,i], time = time)

}

# Create empty list for Spearman's rho and p-value:
cor.stat <- list()
cor.p <- list()

# For each randomised FFC:
for (i in 1:(length(gdrandtax))) {

  # Perform correlation and extract each rho:
  cor.stat[[i]] <- cor.test(x = gd.pbf, gdrandtax[[i]], method = "spearman")$
      estimate

  # Perform correlation and extract each p-value:
  cor.p[[i]] <- cor.test(x = gd.pbf, gdrandtax[[i]], method = "spearman")$p.
      value

}

max(as.numeric(cor.stat)) # max
min(as.numeric(cor.stat)) # min
mean(as.numeric(cor.stat)) # mean

# Calculate critical value for an alpha of 0.05:
cutoff <- qnorm(0.95, mean(as.numeric(cor.stat)), sd(as.numeric(cor.stat)))

# Calculate Spearman's rho for raw data:
spearman <- as.numeric(cor.test(x = gd.tde, y = gd.pbf, method = "spearman")$
    estimate)

# Plot histogram:
hist(as.numeric(cor.stat), xlim = c(0, 1), xlab = "Spearman's rho",
     main = "Spearman's rho for randomised TDE against PBF", col = "black",
         breaks = seq(0,1,0.01))
lines(x = c(cutoff, cutoff), y = c(0, 20), lty = 5)
lines(x = c(spearman, spearman), y = c(0, 20), lty = 5, col = "red")

# Calculate percentage of trials over 95% critical value:
(sum(as.numeric(cor.stat) > spearman)/40) * 100 # 7% (i.e., 3/40)
```

```
# Interpolating deMenocal's (1995) Arabian Sea (ODP 721/722) dust flux curve:

# Input data:
dust <- read.xls("climate_data/deMenocal_dust_curve.xls", sheet = 1, header =
    TRUE)

# Transform ages from ka to Ma:
dust$age <- dust$age/1000

# Extract 7-1 Ma data:
dust <- dust[which(dust$age >= 0.99999 & dust$age <= 7), ]

# Calculate minimum time between data points (limit for interpolation):
delta.t <- diff(dust$age); min(delta.t) # max = 210 ka or 0.21 Ma, mean = 1.9
    ka

# Create evenly-spaced 50 ka time vector:
time <- seq(1, 7, 0.05)

# Interpolate the sea-level data to evenly spaced time (ought to be PCHIP):
out <- interp1(x = dust$age, y = dust$Terr.flux, xi = time, method = "cubic")

# Plot:
plot(time, out, type = "l")

# Combined interpolation output into data frame:
interp.dust <- data.frame(time, out)

# Create time bin vector:
bins <- seq(1, 7, by = 0.25)

# Calculating means in bins:
mean.dust <- as.numeric(tapply(interp.dust$out, cut(interp.dust$time, bins),
    mean)); mean.dust

# Calculating means in bins:
sd.dust <- as.numeric(tapply(interp.dust$out, cut(interp.dust$time, bins), sd)
    ); sd.dust

# Interpolating Teidemann et al. (1994)'s West African (ODP 697) dust flux
    curve:

# Input data:
WA.dust <- read.xls("climate_data/Teidemann_dust_curve.xls", sheet = 1, header
     = TRUE)

# Extract 7-1 Ma data:
WA.dust <- WA.dust[which(WA.dust$age >= 0.99903276 & WA.dust$age <= 5.0264798)
    , ]

# Create evenly-spaced 50 ka time vector:
time <- seq(1, 5, 0.05)

# Interpolate the sea-level data to evenly spaced time (ought to be PCHIP):
out <- interp1(x = WA.dust$age, y = WA.dust$flux, xi = time, method = "cubic")

# Plot:
plot(time, out, type = "l")

# Combined interpolation output into data frame:
```

```
interp.dust <- data.frame(time, out)

# Create time bin vector:
bins <- seq(1, 5, by = 0.25)

# Calculating means in bins:
mean.dust <- as.numeric(tapply(interp.dust$out, cut(interp.dust$time, bins),
    mean)); mean.dust

# Calculating means in bins:
sd.dust <- as.numeric(tapply(interp.dust$out, cut(interp.dust$time, bins), sd)
    ); sd.dust

# Code to make consensus tree for Figure 1 (all time series):
# Create consensus PDE:
con.pde <- c
    (2,1,2,2,2,2,2.5,3,3,3.5,4.0,3.5,4.5,4.0,5.5,5.5,6.5,6.5,8.0,5.5,5.5,4,3,2.0)

con.upper <- c
    (4,3,4,4,4,4.5,5,5.5125,6,6.5,7,7,7.5,7.5,8,8,9,8.5,8.5,5.5,5.5,4,3,2)
con.lower <- c(2,1,1,1,1,1,1,1,1,2,2,2,3,3.5,4,3.5,4.4875,4.5,6,4.5,5.5,4,3,2)

# Generalise difference consensus PDE:
gd.con.pde <- gen.diff(con.pde, data$time)

# Compare consensus curve to each empirical PDE:
cor.test(gd.con.pde, gd.d16.pde, method = "spearman")
cor.test(gd.con.pde, gd.h.pde, method = "spearman")
cor.test(gd.con.pde, gd.d15.pde, method = "spearman")
cor.test(gd.con.pde, gd.sg.pde, method = "spearman") # rho ranges from 0.763
    to 0.828 (p << 0.001)

plot.pd <- data.frame(data$time, data$tde, con.pde, con.upper, con.lower, data
    $aridity.mean)

soil <- read.xls("~/git-repo/thesis-scripts/Chapter3/climate_data/
    EstAfrSoilCO3_compiled.xls", sheet = 1)
temp <- read.xlsx("~/git-repo/thesis-scripts/Chapter3/climate_data/Zachos_2001
    _d18O.xlsx")
lakes <- c
    (0,0,0,0,0,0,0,0,0,0.4,0.6,0.2,0.6,0.8,1.8,0.8,0.6,1,0.6,1.8,3.2,4.6,1.6,1.2)

time <- apply(time.bins, 1, median)
lakes.df <- data.frame(time, lakes)

# FIGURE 1:
# a: TDE
m1 <- ggplot(plot.pd, aes(x = data.time, y = data.tde)) +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab("Taxic diversity") +
  scale_y_continuous(breaks = seq(0,6,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

# b: PDE
```

```
m2 <- ggplot(plot.pd, aes(x = data.time, y = con.pde)) +
  geom_ribbon(aes(ymin = con.lower, ymax = con.upper), fill = "#9ecae1") +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab("Phylogenetic diversity") +
  scale_y_continuous(breaks = seq(0,8,1), position = "right") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))


# c: d18O
m3 <- ggplot(temp, aes(x = age, y = d18O)) +
  geom_line(data = temp, aes(x = age, y = d18O), size = 0.6) +
  ylab(expression(paste(delta^18, O ~ ("\u2030"),  sep = ""))) +
  scale_y_continuous(breaks = seq(1,5,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('c') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

# d: dust flux
m4 <- ggplot(dust, aes(x = age, y = Terr.flux)) +
  geom_point(size = 1, col = "chocolate") + #d95f0e
  ylab(expression(paste("Dust flux (", g, ~ cm^-2, ~ ka^-1,")", sep = ""))) +
  scale_y_continuous(breaks = seq(0,2,1), position = "right") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('d') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

# e: soil carbonate d13C
m5 <- ggplot(soil, aes(x = age, y = d13C)) +
  geom_point(size = 1, color = "#276419") +
  ylab(expression(paste(delta^13, C ~ ("\u2030"), sep = ""))) +
  scale_y_continuous(breaks = seq(-15,5,5)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('e') +
  theme(legend.position = "none") +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

# f: lake variability
m6 <- ggplot(lakes.df, aes(x = time, y = lakes)) +
  geom_bar(stat = "identity", fill = "#1D91C0") +
  xlab("Time (Ma)") +
  ylab("Lake variability index") +
  scale_y_continuous(breaks = seq(0,5,1), position = "right") +
```

```
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('f') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# Create multi-plot:
ggarrange(m1, m2, m3, m4, m5, m6, ncol = 1)

# FIGURE 2:
s1 <- ggplot(data, aes(x = time, y = hbc)) +
  geom_bar(stat = "identity", fill = "#56B4E9") +
  ylab("HBC") +
  scale_y_continuous(breaks = seq(0,50,10)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

s2 <- ggplot(data, aes(x = time, y = hbf)) +
  geom_bar(stat = "identity", fill = "#E69F00") +
  ylab("HBF") +
  scale_y_continuous(breaks = seq(0,12,2)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

s3 <- ggplot(data, aes(x = time, y = pbf)) +
  geom_bar(stat = "identity", fill = "#999999") +
  ylab("PBF") +
  scale_y_continuous(breaks = seq(0,12,2)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('c') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

s4 <- ggplot(data, aes(x = time, y = mbf)) +
  geom_bar(stat = "identity", fill = "#56B4E9") +
  xlab("Time (Ma)") +
  ylab("MBF") +
  scale_y_continuous(breaks = seq(0,20,5)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('d') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

# Get percentage of contintent sampled per time bin:
```

```
per.afr <- (data$convex.hull/30240000)*100; per.afr <- round(per.afr, digits =
     1)
per.afr <- c(NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
    ,0.1,0.2,0.4,6.1,0.6,0.1,2.3,1.5,1.6,0.2,2.2,3.3,1.9,0.2)

s5 <- ggplot(data, aes(x = time, y = convex.hull)) +
  geom_bar(stat = "identity", fill = "#E69F00") +
  geom_text(aes(label = per.afr), vjust = -0.25, family = "Myriad Pro") +
  ylab(expression(paste("Convex hull area (", km^2,")", sep = ""))) +
  scale_y_continuous(limits = c(0,2000000)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('e') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

s6 <- ggplot(data, aes(x = time, y = grid.cell.occupancy)) +
  geom_bar(stat = "identity", fill = "#999999") +
  xlab("Time (Ma)") +
  ylab("Grid cell occupancy") +
  scale_y_continuous(breaks = seq(0,14,2)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('f') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

ggarrange(s1, s2, s3, s4, s5, s6, ncol = 1)

###

# FIGURE 3 - TDE against sampling metrics:
pdf("TDEsampling.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))
# A: TDE against HBC:
plot(gd.hbc, gd.tde, pch = 16, xlim = c(-15,20), ylim = c(-2,3), family = "
    Myriad Pro", xlab = expression(paste(Delta,"HBC")), ylab = expression(
    paste(Delta,"TDE")))
# B: TDE against HBF:
plot(gd.hbf, gd.tde, pch = 16, xlim = c(-3,3), ylim = c(-2,3), family = "
    Myriad Pro", xlab = expression(paste(Delta,"HBF")), ylab = expression(
    paste(Delta,"TDE")))
# C: TDE against PBF:
plot(gd.pbf, gd.tde, pch = 16, xlim = c(-3,5), ylim = c(-2,3), family = "
    Myriad Pro", xlab = expression(paste(Delta,"PBF")), ylab = expression(
    paste(Delta,"TDE")))
# D: TDE against MBF:
plot(gd.mbf, gd.tde, pch = 16, xlim = c(-4,4), ylim = c(-2,3), family = "
    Myriad Pro", xlab = expression(paste(Delta,"MBF")), ylab = expression(
    paste(Delta,"TDE")))
par(mfrow = c(1,1))
dev.off()

# Plotting the ghost lineages:
time <- data$time

# Create data frame of data to plot:
```

```
plot.g <- data.frame(time, sg.gde, sg.gpratio, d15.gde, d15.gpratio, h.gde, h.
    gpratio, d16.gde, d16.gpratio)

# FIGURE 4:
g1 <- ggplot(data = plot.g) +
  geom_bar(aes(x = time, y = sg.gde), stat="identity", fill = "#feb24c") +
  ylab("SGDE") +
  scale_y_continuous(breaks = seq(-1,4,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
      panel.grid.major = element_blank(),
      axis.line = element_line(),
      axis.title.x = element_blank(),
      text = element_text(family = "Myriad Pro", size = 16))

g2 <- ggplot(data = plot.g) +
  geom_bar(aes(x = time, y = d15.gde), stat="identity", fill = "#feb24c") +
  ylab("D1GDE") +
  scale_y_continuous(breaks = seq(0,5,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

g3 <- ggplot(data = plot.g) +
  geom_bar(aes(x = time, y = h.gde), stat="identity", fill = "#feb24c") +
  ylab("HGDE") +
  scale_y_continuous(breaks = seq(-1,6,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('c') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

g4 <- ggplot(data = plot.g) +
  geom_bar(aes(x = time, y = d16.gde), stat = "identity", fill = "#feb24c") +
  ylab("D2GDE") +
  xlab("Time (Ma)") +
  scale_y_continuous(breaks = seq(0,4,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('d') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

ggarrange(g1, g2, g3, g4, ncol = 1)

# FIGURE 5 - GDE against sampling metrics:
pdf("xyGDE.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))
# A: PBF/MBF against SGDE:
plot(gd.pbf, gd.sg.gde, pch = 16, xlim = c(-3,5), ylim = c(-2,3), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"Sampling
```

```
    metric")), ylab = expression(paste(Delta,"SGDE")))
points(gd.mbf, gd.sg.gde, pch = 17, col = "#E31A1C")
op <- par(family = "Myriad Pro")
legend("topright", c("PBF", "MBF"), pch = c(16,17), col = c("#56B4E9","#E31A1C
    "), bty = "n")
par(op)
# B: PBF/MBF against D1GDE:
plot(gd.pbf, gd.d15.gde, pch = 16, xlim = c(-3,5), ylim = c(-2,4), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"Sampling
    metric")), ylab = expression(paste(Delta,"D1GDE")))
op <- par(family = "Myriad Pro")
points(gd.mbf, gd.d15.gde, pch = 17, col = "#E31A1C")
legend("topright", c("PBF", "MBF"), pch = c(16,17), col = c("#56B4E9","#E31A1C
    "), bty = "n")
par(op)
# C: PBF/MBF against HGDE:
plot(gd.pbf, gd.h.gde, pch = 16, xlim = c(-3,5), ylim = c(-2,4), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"Sampling
    metric")), ylab = expression(paste(Delta,"HGDE")))
op <- par(family = "Myriad Pro")
points(gd.mbf, gd.h.gde, pch = 17, col = "#E31A1C")
legend("topright", c("PBF", "MBF"), pch = c(16,17), col = c("#56B4E9","#E31A1C
    "), bty = "n")
par(op)
# D: PBF/MBF against D2GDE:
plot(gd.pbf, gd.d16.gde, pch = 16, xlim = c(-3,5), ylim = c(-2,4), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"Sampling
    metric")), ylab = expression(paste(Delta,"D2GDE")))
op <- par(family = "Myriad Pro")
points(gd.mbf, gd.d16.gde, pch = 17, col = "#E31A1C")
legend("topright", c("PBF", "MBF"), pch = c(16,17), col = c("#56B4E9","#E31A1C
    "), bty = "n")
par(op)
par(mfrow = c(1,1))
dev.off()

# FIGURE 6:
gp1 <- ggplot(data = plot.g) +
  geom_bar(aes(x = time, y = sg.gpratio), stat="identity", fill = "#feb24c") +
  ylab("SGDE:SPDE") +
  scale_y_continuous(breaks = seq(-0.2,1,0.2)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

gp2 <- ggplot(data = plot.g) +
  geom_bar(aes(x = time, y = d15.gpratio), stat="identity", fill = "#feb24c")
     +
  ylab("D1GDE:D1PDE") +
  scale_y_continuous(breaks = seq(0,1,0.2)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))
```

```
gp3 <- ggplot(data = plot.g) +
  geom_bar(aes(x = time, y = h.gpratio), stat="identity", fill = "#feb24c") +
  ylab("HGDE:HPDE") +
  scale_y_continuous(breaks = seq(-0.2,1,0.2)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('c') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

gp4 <- ggplot(data = plot.g) +
  geom_bar(aes(x = time, y = d16.gpratio), stat = "identity", fill = "#feb24c"
      ) +
  ylab("D2GDE:D2PDE") +
  xlab("Time (Ma)") +
  scale_y_continuous(breaks = seq(0,1,0.2)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('d') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

ggarrange(gp1, gp2, gp3, gp4, ncol = 1)

# FIGURE 7 - GDE:PDE against sampling metrics:
pdf("xyGDE_PDE.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))
# A: PBF/MBF against SGDE:
plot(gd.pbf, gd.sg.gprat, pch = 16, xlim = c(-3,5), ylim = c(-0.7,0.7), family
     = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"Sampling
    metric")), ylab = expression(paste(Delta,"SGDE:SPDE")))
points(gd.mbf, gd.sg.gprat, pch = 17, col = "#E31A1C")
op <- par(family = "Myriad Pro")
legend("topright", c("PBF", "MBF"), pch = c(16,17), col = c("#56B4E9","#E31A1C
    "), bty = "n")
par(op)
# B: PBF/MBF against D1GDE:
plot(gd.pbf, gd.d15.gprat, pch = 16, xlim = c(-3,5), ylim = c(-0.6,0.6),
    family = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"
    Sampling metric")), ylab = expression(paste(Delta,"D1GDE:D1PDE")))
op <- par(family = "Myriad Pro")
points(gd.mbf, gd.d15.gprat, pch = 17, col = "#E31A1C")
legend("topright", c("PBF", "MBF"), pch = c(16,17), col = c("#56B4E9","#E31A1C
    "), bty = "n")
par(op)
# C: PBF/MBF against HGDE:
plot(gd.pbf, gd.h.gprat, pch = 16, xlim = c(-3,5), ylim = c(-0.6,0.6), family
    = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"Sampling
    metric")), ylab = expression(paste(Delta,"HGDE:HPDE")))
op <- par(family = "Myriad Pro")
points(gd.mbf, gd.h.gprat, pch = 17, col = "#E31A1C")
legend("topright", c("PBF", "MBF"), pch = c(16,17), col = c("#56B4E9","#E31A1C
    "), bty = "n")
par(op)
# D: PBF/MBF against D2GDE:
plot(gd.pbf, gd.d16.gprat, pch = 16, xlim = c(-3,5), ylim = c(-0.6,0.6),
```

```
    family = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"
    Sampling metric")), ylab = expression(paste(Delta,"D2GDE:D2PDE")))
op <- par(family = "Myriad Pro")
points(gd.mbf, gd.d16.gprat, pch = 17, col = "#E31A1C")
legend("topright", c("PBF", "MBF"), pch = c(16,17), col = c("#56B4E9","#E31A1C
    "), bty = "n")
par(op)
par(mfrow = c(1,1))
dev.off()

# FIGURE 8 - East African correlations:
pdf("eaTDEclimate.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))
# A: TDE against HBC:
plot(gd.ea.hbc, gd.ea.tde, pch = 16, xlim = c(-12,16), ylim = c(-1,2), family
    = "Myriad Pro", xlab = expression(paste(Delta,"HBC"[EA])), ylab =
    expression(paste(Delta,"TDE"[EA])))
# B: TDE against PBF:
plot(gd.ea.pbf, gd.ea.tde, pch = 16, xlim = c(-2.2,3.2), ylim = c(-1,2),
    family = "Myriad Pro", xlab = expression(paste(Delta,"PBF"[EA])), ylab =
    expression(paste(Delta,"TDE"[EA])))
# C: TDE against aridity:
plot(gd.arid, gd.ea.tde, pch = 16, xlim = c(-0.3,0.2), ylim = c(-1,2), family
    = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"Dust flux"
    )), ylab = expression(paste(Delta,"TDE"[EA])))
op <- par(family = "Myriad Pro")
points(gd.arid.west, gd.ea.tde, pch = 17, col = "#E31A1C")
legend("bottomright", c("East", "West"), pch = c(16,17), col = c("#56B4E9","#
    E31A1C"), bty = "n")
par(op)
# D: TDE against LVI:
plot(gd.lmean, gd.ea.tde, pch = 16, xlim = c(-1.6,2.1), ylim = c(-1,2), family
     = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"Lake
    variability")), ylab = expression(paste(Delta,"TDE"[EA])))
op <- par(family = "Myriad Pro")
points(gd.lmax, gd.ea.tde, pch = 17, col = "#E31A1C")
legend("bottomright", c("mean", "maximum"), pch = c(16,17), col = c("#56B4E9",
    "#E31A1C"), bty = "n")
par(op)
par(mfrow = c(1,1))
dev.off()

# FIGURE 9 - COLLECTOR CURVES:
# Create collector curves:
african.year <- c(1920,1930,1940,1950,1960,1970,1980,1990,2000,2010)
african.taxa <- c(1,2,2,3,5,7,8,12,16,18)
african.form <- c(1,3,4,5,6,13,16,23,27,27)

# Combine African data for plotting:
df.african <- data.frame(african.year, african.taxa, african.form)

cc1 <- ggplot(df.african, aes(x = african.year, y = african.taxa)) +
  geom_point(size = 4) +
  xlab("Research time") +
  ylab("Cumulative number of\nhominin species") +
  scale_y_continuous(breaks = seq(0,20,5)) +
  scale_x_continuous(breaks = seq(1860,2010,10)) +
  geom_smooth(method = "glm", method.args = list(family = gaussian(link = "log
      ")), colour = "red", se = FALSE) +
  ggtitle("a") +
```

```
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.text.x  = element_text(angle = 90),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 18))

cc2 <- ggplot(df.african, aes(x = african.year, y = african.form)) +
  geom_point(size = 4) +
  xlab("Research time") +
  ylab("Cumulative number\nof HBF") +
  scale_y_continuous(breaks = seq(0,30,5)) +
  scale_x_continuous(breaks = seq(1860,2010,10)) +
  geom_smooth(method = "glm", method.args = list(family = gaussian(link = "log
      ")), colour = "red", se = FALSE) +
  ggtitle("b") +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.text.x  = element_text(angle = 90),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 18))

cc3 <- ggplot(df.african, aes(x = african.form, y = african.taxa)) +
  geom_point(size = 4) +
  xlab("Cumulative number of HBF") +
  ylab("Cumulative number\nhominin ofspecies") +
  scale_y_continuous(breaks = seq(0,20,5)) +
  scale_x_continuous(breaks = seq(0,30,5)) +
  geom_smooth(method = "lm", colour = "red", se = FALSE) +
  ggtitle("c") +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 18))

ggarrange(cc1, cc2, cc3, ncol = 3)

# FIGURE 10: Map of sampled area
require(ggmap)
map <- get_map(maptype = "watercolor")
ggmap(map) +
  geom_point(data = df, aes(x = longitudes, y = latitudes, fill = "black"),
      size = 4, shape = 20) +
  xlab(expression(paste('Longitude (',degree,')', sep = ''))) +
  ylab(expression(paste('Latitude (',degree,')', sep = ''))) +
  guides(fill = FALSE, alpha = FALSE, size = FALSE) +
  theme(text = element_text(family = "Myriad Pro", size = 14))

# FIGURE 11: Sampled area boxplot
# Re-label epoch vector:
data$epoch = factor(data$epoch, levels = 1:3,
                    labels = c("Miocene", "Pliocene", "Pleistocene")
)

b1 <- ggplot(data, aes(x = epoch, y = convex.hull)) +
  geom_boxplot(aes(group = epoch)) +
  ylab(expression(paste("Convex hull area (", km^2,")", sep = ""))) +
  scale_y_continuous(limits = c(0,2000000)) +
  # scale_y_continuous(breaks = seq(0,2000000,500000)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
```

```
        axis.line = element_line(), axis.title.x = element_blank(), axis.text.
            x = element_text(),
        text = element_text(family = "Myriad Pro", size = 24))

#
b2 <- ggplot(data, aes(x = epoch, y = grid.cell.occupancy)) +
  geom_boxplot(aes(group = epoch)) +
  ylab("Grid cell occupancy") +
  scale_y_continuous(breaks = seq(0,16,2)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(), axis.title.x = element_blank(), axis.text.
            x = element_text(),
        text = element_text(family = "Myriad Pro", size = 24))

ggarrange(b1, b2, ncol = 2)

shapiro.test(data$pbf)
shapiro.test(data$hbc)
shapiro.test(data$mbf)
shapiro.test(data$convex.hull)
shapiro.test(data$grid.cell.occupancy)
shapiro.test(data$aridity.mean)

y1 <- data$pbf
x1 <- log10(data$hbc + 1)
x2 <- data$mbf
x3 <- log10(data$convex.hull + 1)
x4 <- log10(data$grid.cell.occupancy + 1)
x5 <- data$aridity.mean

# Create all possible models plus null:
gls.null <- gls(y1~1, correlation = corARMA(p = 1), method = "ML") # 1
gls.1 <- gls(y1~x1+x2+x3+x4+x5, correlation = corARMA(p = 1), method = "ML") #
    2
gls.2 <- gls(y1~x1+x2+x3+x4, correlation = corARMA(p = 1), method = "ML") # 3
gls.3 <- gls(y1~x1+x2+x3+x5, correlation = corARMA(p = 1), method = "ML") # 4
gls.4 <- gls(y1~x1+x2+x4+x5, correlation = corARMA(p = 1), method = "ML") # 5
gls.5 <- gls(y1~x1+x3+x4+x5, correlation = corARMA(p = 1), method = "ML") # 6
gls.6 <- gls(y1~x2+x3+x4+x5, correlation = corARMA(p = 1), method = "ML") # 7
gls.7 <- gls(y1~x1+x2+x3, correlation = corARMA(p = 1), method = "ML") # 8
gls.8 <- gls(y1~x1+x2+x4, correlation = corARMA(p = 1), method = "ML") # 9
gls.9 <- gls(y1~x1+x2+x5, correlation = corARMA(p = 1), method = "ML") # 10
gls.10 <- gls(y1~x1+x3+x4, correlation = corARMA(p = 1), method = "ML") # 11
gls.11 <- gls(y1~x1+x3+x5, correlation = corARMA(p = 1), method = "ML") # 12
gls.12 <- gls(y1~x1+x4+x5, correlation = corARMA(p = 1), method = "ML") # 13
gls.13 <- gls(y1~x2+x3+x4, correlation = corARMA(p = 1), method = "ML") # 14
gls.14 <- gls(y1~x2+x3+x5, correlation = corARMA(p = 1), method = "ML") # 15
gls.15 <- gls(y1~x2+x4+x5, correlation = corARMA(p = 1), method = "ML") # 16
gls.16 <- gls(y1~x3+x4+x5, correlation = corARMA(p = 1), method = "ML") # 17
gls.17 <- gls(y1~x1+x2, correlation = corARMA(p = 1), method = "ML") # 18
gls.18 <- gls(y1~x1+x3, correlation = corARMA(p = 1), method = "ML") # 19
gls.19 <- gls(y1~x1+x4, correlation = corARMA(p = 1), method = "ML") # 20
gls.20 <- gls(y1~x1+x5, correlation = corARMA(p = 1), method = "ML") # 21
gls.21 <- gls(y1~x2+x3, correlation = corARMA(p = 1), method = "ML") # 22
gls.22 <- gls(y1~x2+x4, correlation = corARMA(p = 1), method = "ML") # 23
gls.23 <- gls(y1~x2+x5, correlation = corARMA(p = 1), method = "ML") # 24
gls.24 <- gls(y1~x3+x4, correlation = corARMA(p = 1), method = "ML") # 25
gls.25 <- gls(y1~x3+x5, correlation = corARMA(p = 1), method = "ML") # 26
gls.26 <- gls(y1~x4+x5, correlation = corARMA(p = 1), method = "ML") # 27
```

```
gls.27 <- gls(y1~x1, correlation = corARMA(p = 1), method = "ML") # 28
gls.28 <- gls(y1~x2, correlation = corARMA(p = 1), method = "ML") # 29
gls.29 <- gls(y1~x3, correlation = corARMA(p = 1), method = "ML") # 30
gls.30 <- gls(y1~x4, correlation = corARMA(p = 1), method = "ML") # 31
gls.31 <- gls(y1~x5, correlation = corARMA(p = 1), method = "ML") # 32

# Call function from Cleary et al. (2015) to calculate model weights:
weighted <- function (aic){
  aic.wt <- exp(-0.5 * (aic - min(aic)))/sum(exp(-0.5 * (aic - min(aic))))
  return(aic.wt)
}

# Compute and compare AIC weights (AICw) for each model:
weights <- weighted(c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3),
    AICc(gls.4), AICc(gls.5), AICc(gls.6), AICc(gls.7), AICc(gls.8), AICc(gls
    .9), AICc(gls.10), AICc(gls.11), AICc(gls.12), AICc(gls.13), AICc(gls.14),
     AICc(gls.15), AICc(gls.16), AICc(gls.17), AICc(gls.18), AICc(gls.19),
    AICc(gls.20), AICc(gls.21), AICc(gls.22), AICc(gls.23), AICc(gls.24), AICc
    (gls.25), AICc(gls.26), AICc(gls.27), AICc(gls.28), AICc(gls.29), AICc(gls
    .30), AICc(gls.31))); weights

# Compute and compare AICc values for each model:
AICcscores <- (c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3), AICc(
    gls.4), AICc(gls.5), AICc(gls.6), AICc(gls.7), AICc(gls.8), AICc(gls.9),
    AICc(gls.10), AICc(gls.11), AICc(gls.12), AICc(gls.13), AICc(gls.14), AICc
    (gls.15), AICc(gls.16), AICc(gls.17), AICc(gls.18), AICc(gls.19), AICc(gls
    .20), AICc(gls.21), AICc(gls.22), AICc(gls.23), AICc(gls.24), AICc(gls.25)
    , AICc(gls.26), AICc(gls.27), AICc(gls.28), AICc(gls.29), AICc(gls.30),
    AICc(gls.31))); AICcscores

# Compare all gls models and derive a global p-value using ANOVA:
anova(gls.null, gls.1, gls.2, gls.3, gls.4, gls.5, gls.6, gls.7, gls.8, gls.9,
     gls.10, gls.11, gls.12, gls.13, gls.14, gls.15, gls.16, gls.17, gls.18,
    gls.19, gls.20, gls.21, gls.22, gls.23, gls.24, gls.25, gls.26, gls.27,
    gls.28, gls.29, gls.30, gls.31)

# Calculate generalised R2:
1-exp(-1*2/24*(gls.1$logLik-gls.null$logLik))

# Isolate best-supported model and create new data frame for plotting:
pbf.gls <- data.frame(data$pbf, gls.15$fitted)

# # FIGURE 12: PBF agaisnt fitted data
ggplot(pbf.gls, aes(x = gls.15$fitted, y = data$pbf)) +
  geom_point(size = 4) +
  xlab("MBF + grid cell occupancy + aridity (fitted)") +
  ylab("PBF") +
  scale_x_continuous(breaks = seq(-1,10,1)) +
  scale_y_continuous(breaks = seq(0,12,2)) +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.text.x  = element_text(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 18))

# End of Chapter 3 script
```

Chapter 4

```
# R code for Chapter 4: The completeness of the early hominin fossil record
```

```
# Set working directory:
setwd("~/git-repo/thesis-scripts/Chapter4")

# Load packages:
library(ggplot2) # if (!require("ggplot2")) install.packages("ggplot2")
library(labeling) # if (!require("labeling")) install.packages("labeling")
library(openxlsx) # if (!require("openxlsx")) install.packages("openxlsx")
library(plyr) # if (!require("plyr")) install.packages("plyr")
library(randtests) # if (!require("randtests")) install.packages("randtests")
library(egg)
library(RColorBrewer) # if (!require("RColorBrewer")) install.packages("
    RColorBrewer")
require(extrafont) # if (!require("extrafont")) install.packages("extrafont")
require(extrafontdb) # if (!require("extrafontdb")) install.packages("
    extrafontdb")
require(nlme)
require(qpcR)
require(tseries) # for Jarque-Bera test
require(lmtest) # for Breusch-Pagan test

# Input data:
data <- read.xlsx("Chapter4_data.xlsx", sheet = 1)

kruskal.test(data$ccm2, data$epoch)
kruskal.test(data$scm2, data$epoch)

# Call Graeme T. Lloyd's generalised differencing function:
gen.diff <- function(x, time) {
  # Suppress warning message:
  # if(cor.test(time, x)$p.value > 0.05) print("Warning: variables not
      significantly correlated, generalised differencing not recommended.")
  dt <- x - ((lsfit(time, x)$coefficients[2] * time) + lsfit(time, x)$
      coefficients[1])
  m <- lsfit(dt[1:(length(dt)-1)], dt[2:length(dt)])$coefficients[2]
  gendiffs <- dt[1:(length(dt) - 1)]- (dt[2:length(dt)] * m)
  gendiffs
}

# Perform generalised differencing function on each variable:
gd.ccm1 <- gen.diff(data$ccm1, data$time) # CCM1
gd.ccm2 <- gen.diff(data$ccm2, data$time) # CCM2
gd.ccm1max <- gen.diff(data$ccm1.max, data$time) # max CCM1
gd.ccm2max <- gen.diff(data$ccm2.max, data$time) # max CCM2
gd.scm1 <- gen.diff(data$scm1, data$time) # SCM1
gd.scm2 <- gen.diff(data$scm2, data$time) # SCM2
gd.skull <- gen.diff(data$skull, data$time) # SCM2_skull
gd.scm1max <- gen.diff(data$scm1.max, data$time) # max SCM1
gd.scm2max <- gen.diff(data$scm2.max, data$time) # max SCM2
gd.tde <- gen.diff(data$tde, data$time) # Taxic diversity
gd.hbc <- gen.diff(data$hbc, data$time) # Hominin-bearing collections (HBC)
gd.hbf <- gen.diff(data$hbf, data$time) # Hominin-bearing formations (HBF)
gd.hbl <- gen.diff(data$hbl, data$time) # Hominin-bearing localities (HBL)
gd.pbf <- gen.diff(data$pbf, data$time) # Primate-bearing formations (PBF)
gd.mbf <- gen.diff(data$mbf, data$time) # Mammal-bearing formations (MBF)
gd.bonanza <- gen.diff(data$bonanza, data$time) # ratio of HBC:HBF
gd.spec <- gen.diff(data$spec, data$time) # Specimen count
gd.arid <- gen.diff(data$arid, data$time) # deMenocal's aridity curve

# Call p.adjust function (from Cleary et al., 2015):
cor = function(x, y, method = c("pearson", "spearman", "kendall"), N) {
  test = cor.test(x, y, method = method)
```

```
  adj = p.adjust(test$p.value, "BH", n = N)
  display = c(test$est, test$p.value, adj)
  names(display)[1] = "Coefficient"
  names(display)[2] = "p-value"
  names(display)[3] = "p.adjusted"
  return(display)
}

### Pairwise tests - mean completeness scores:

# SCM1 vs. SCM2:
cor(gd.scm1, gd.scm2, method = "spearman", N = 6)
cor(gd.scm1, gd.scm2, method = "kendall", N = 6)

# SCM1 vs. CCM1:
cor(gd.scm1, gd.ccm1, method = "spearman", N = 6)
cor(gd.scm1, gd.ccm1, method = "kendall", N = 6)

# SCM1 vs. CCM2:
cor(gd.scm1, gd.ccm2, method = "spearman", N = 6)
cor(gd.scm1, gd.ccm2, method = "kendall", N = 6)

# SCM2 vs. CCM1:
cor(gd.scm2, gd.ccm1, method = "spearman", N = 6)
cor(gd.scm2, gd.ccm1, method = "kendall", N = 6)

# SCM2 vs. CCM2:
cor(gd.scm2, gd.ccm2, method = "spearman", N = 6)
cor(gd.scm2, gd.ccm2, method = "kendall", N = 6)

# CCM1 vs. CCM2:
cor(gd.ccm1, gd.ccm2, method = "spearman", N = 6)
cor(gd.ccm1, gd.ccm2, method = "kendall", N = 6)

### Pairwise tests - maximum completeness scores:

# SCM1 vs. SCM2:
cor(gd.scm1max, gd.scm2max, method = "spearman", N = 6)
cor(gd.scm1max, gd.scm2max, method = "kendall", N = 6)

# SCM1 vs. CCM1:
cor(gd.scm1max, gd.ccm1max, method = "spearman", N = 6)
cor(gd.scm1max, gd.ccm1max, method = "kendall", N = 6)

# SCM1 vs. CCM2:
cor(gd.scm1max, gd.ccm2max, method = "spearman", N = 6)
cor(gd.scm1max, gd.ccm2max, method = "kendall", N = 6)

# SCM2 vs. CCM1:
cor(gd.scm2max, gd.ccm1max, method = "spearman", N = 6)
cor(gd.scm2max, gd.ccm1max, method = "kendall", N = 6)

# SCM2 vs. CCM2:
cor(gd.scm2max, gd.ccm2max, method = "spearman", N = 6)
cor(gd.scm2max, gd.ccm2max, method = "kendall", N = 6)

# CCM1 vs. CCM2:
cor(gd.ccm1max, gd.ccm2max, method = "spearman", N = 6)
cor(gd.ccm1max, gd.ccm2max, method = "kendall", N = 6)

### Pairwise tests - skull completeness scores:
```

```
# SCMskull vs. CCM1:
cor(gd.skull, gd.ccm1, method = "spearman", N = 4)
cor(gd.skull, gd.ccm1, method = "kendall", N = 4)

# SCMskull vs. CCM2:
cor(gd.skull, gd.ccm2, method = "spearman", N = 4)
cor(gd.skull, gd.ccm2, method = "kendall", N = 4)

# SCMskull vs. CCM1maximum:
cor(gd.skull, gd.ccm1max, method = "spearman", N = 4)
cor(gd.skull, gd.ccm1max, method = "kendall", N = 4)

# SCMskull vs. CCM2maximum:
cor(gd.skull, gd.ccm2max, method = "spearman", N = 4)
cor(gd.skull, gd.ccm2max, method = "kendall", N = 4)

### Runs tests:
runs.test(data$ccm1, threshold = median(data$ccm1))
runs.test(data$ccm2, threshold = median(data$ccm2))
runs.test(data$scm1, threshold = median(data$scm1))
runs.test(data$scm2, threshold = median(data$scm2))

### Pairwise tests - diversity, sampling, & climate:

# CCM1 vs. TDE (diversity):
cor(gd.ccm1, gd.tde, method = "spearman", N = 20)
cor(gd.ccm1, gd.tde, method = "kendall", N = 20)

# CCM2 vs. TDE (diversity):
cor(gd.ccm2, gd.tde, method = "spearman", N = 20)
cor(gd.ccm2, gd.tde, method = "kendall", N = 20)

# SCM1 vs. TDE (diversity):
cor(gd.scm1, gd.tde, method = "spearman", N = 20)
cor(gd.scm1, gd.tde, method = "kendall", N = 20)

# SCM2 vs. TDE (diversity):
cor(gd.scm2, gd.tde, method = "spearman", N = 20)
cor(gd.scm2, gd.tde, method = "kendall", N = 20)

# CCM1 vs. HBC (collections):
cor(gd.ccm1, gd.hbc, method = "spearman", N = 20)
cor(gd.ccm1, gd.hbc, method = "kendall", N = 20)

# CCM2 vs. HBC (collections):
cor(gd.ccm2, gd.hbc, method = "spearman", N = 20)
cor(gd.ccm2, gd.hbc, method = "kendall", N = 20)

# SCM1 vs. HBC (collections):
cor(gd.scm1, gd.hbc, method = "spearman", N = 20)
cor(gd.scm1, gd.hbc, method = "kendall", N = 20)

# SCM2 vs. HBC (collections):
cor(gd.scm2, gd.hbc, method = "spearman", N = 20)
cor(gd.scm2, gd.hbc, method = "kendall", N = 20)

# CCM1 vs. PBF (formations):
cor(gd.ccm1, gd.mbf, method = "spearman", N = 20)
cor(gd.ccm1, gd.mbf, method = "kendall", N = 20)
```

```
# CCM2 vs. PBF (formations):
cor(gd.ccm2, gd.mbf, method = "spearman", N = 20)
cor(gd.ccm2, gd.mbf, method = "kendall", N = 20)

# SCM1 vs. PBF (formations):
cor(gd.scm1, gd.mbf, method = "spearman", N = 20)
cor(gd.scm1, gd.mbf, method = "kendall", N = 20)

# SCM2 vs. PBF (formations):
cor(gd.scm2, gd.mbf, method = "spearman", N = 20)
cor(gd.scm2, gd.mbf, method = "kendall", N = 20)

# CCM1 vs. bonanza (HBC:HBF):
cor(gd.ccm1, gd.bonanza, method = "spearman", N = 20)
cor(gd.ccm1, gd.bonanza, method = "kendall", N = 20)

# CCM2 vs. bonanza (HBC:HBF):
cor(gd.ccm2, gd.bonanza, method = "spearman", N = 20)
cor(gd.ccm2, gd.bonanza, method = "kendall", N = 20)

# SCM1 vs. bonanza (HBC:HBF):
cor(gd.scm1, gd.bonanza, method = "spearman", N = 20)
cor(gd.scm1, gd.bonanza, method = "kendall", N = 20)

# SCM2 vs. bonanza (HBC:HBF):
cor(gd.scm2, gd.bonanza, method = "spearman", N = 20)
cor(gd.scm2, gd.bonanza, method = "kendall", N = 20)

# CCM1 vs. aridity:
cor(gd.ccm1, gd.arid, method = "spearman", N = 20)
cor(gd.ccm1, gd.arid, method = "kendall", N = 20)

# CCM2 vs. aridity:
cor(gd.ccm2, gd.arid, method = "spearman", N = 20)
cor(gd.ccm2, gd.arid, method = "kendall", N = 20)

# SCM1 vs. aridity:
cor(gd.scm1, gd.arid, method = "spearman", N = 20)
cor(gd.scm1, gd.arid, method = "kendall", N = 20)

# SCM2 vs. aridity:
cor(gd.scm2, gd.arid, method = "spearman", N = 20)
cor(gd.scm2, gd.arid, method = "kendall", N = 20)

# CCM1 max. vs. TDE (diversity):
cor(gd.ccm1max, gd.tde, method = "spearman", N = 20)
cor(gd.ccm1max, gd.tde, method = "kendall", N = 20)

# CCM2 max. vs. TDE (diversity):
cor(gd.ccm2max, gd.tde, method = "spearman", N = 20)
cor(gd.ccm2max, gd.tde, method = "kendall", N = 20)

# SCM1 max. vs. TDE (diversity):
cor(gd.scm1max, gd.tde, method = "spearman", N = 20)
cor(gd.scm1max, gd.tde, method = "kendall", N = 20)

# SCM2 max. vs. TDE (diversity):
cor(gd.scm2max, gd.tde, method = "spearman", N = 20)
cor(gd.scm2max, gd.tde, method = "kendall", N = 20)

# CCM1 max. vs. HBC (collections):
```

```
cor(gd.ccm1max, gd.hbc, method = "spearman", N = 20)
cor(gd.ccm1max, gd.hbc, method = "kendall", N = 20)

# CCM2 max. vs. HBC (collections):
cor(gd.ccm2max, gd.hbc, method = "spearman", N = 20)
cor(gd.ccm2max, gd.hbc, method = "kendall", N = 20)

# SCM1 max. vs. HBC (collections):
cor(gd.scm1max, gd.hbc, method = "spearman", N = 20)
cor(gd.scm1max, gd.hbc, method = "kendall", N = 20)

# SCM2 max. vs. HBC (collections):
cor(gd.scm2max, gd.hbc, method = "spearman", N = 20)
cor(gd.scm2max, gd.hbc, method = "kendall", N = 20)

# CCM1 max. vs. PBF (formations):
cor(gd.ccm1max, gd.pbf, method = "spearman", N = 20)
cor(gd.ccm1max, gd.pbf, method = "kendall", N = 20)

# CCM2 max. vs. PBF (formations):
cor(gd.ccm2max, gd.pbf, method = "spearman", N = 20)
cor(gd.ccm2max, gd.pbf, method = "kendall", N = 20)

# SCM1 max. vs. PBF (formations):
cor(gd.scm1max, gd.pbf, method = "spearman", N = 20)
cor(gd.scm1max, gd.pbf, method = "kendall", N = 20)

# SCM2 max. vs. PBF (formations):
cor(gd.scm2max, gd.pbf, method = "spearman", N = 20)
cor(gd.scm2max, gd.pbf, method = "kendall", N = 20)

# CCM1 max. vs. bonanza (HBC:HBF):
cor(gd.ccm1max, gd.bonanza, method = "spearman", N = 20)
cor(gd.ccm1max, gd.bonanza, method = "kendall", N = 20)

# CCM2 max. vs. bonanza (HBC:HBF):
cor(gd.ccm2max, gd.bonanza, method = "spearman", N = 20)
cor(gd.ccm2max, gd.bonanza, method = "kendall", N = 20)

# SCM1 max. vs. bonanza (HBC:HBF):
cor(gd.scm1max, gd.bonanza, method = "spearman", N = 20)
cor(gd.scm1max, gd.bonanza, method = "kendall", N = 20)

# SCM2 max. vs. bonanza (HBC:HBF):
cor(gd.scm2max, gd.bonanza, method = "spearman", N = 20)
cor(gd.scm2max, gd.bonanza, method = "kendall", N = 20)

# CCM1 max. vs. aridity:
cor(gd.ccm1max, gd.arid, method = "spearman", N = 20)
cor(gd.ccm1max, gd.arid, method = "kendall", N = 20)

# CCM2 max. vs. aridity:
cor(gd.ccm2max, gd.arid, method = "spearman", N = 20)
cor(gd.ccm2max, gd.arid, method = "kendall", N = 20)

# SCM1 max. vs. aridity:
cor(gd.scm1max, gd.arid, method = "spearman", N = 20)
cor(gd.scm1max, gd.arid, method = "kendall", N = 20)

# SCM2 max. vs. aridity:
cor(gd.scm2max, gd.arid, method = "spearman", N = 20)
```

```
cor(gd.scm2max, gd.arid, method = "kendall", N = 20)

### Generalised Least Squares:

shapiro.test(data$ccm1) # non-normal
shapiro.test(data$ccm1.max) # non-normal
shapiro.test(data$ccm2) # non-normal
shapiro.test(data$ccm2.max) # non-normal
shapiro.test(data$scm1) # non-normal
shapiro.test(data$scm1.max) # non-normal
shapiro.test(data$scm2) # non-normal
shapiro.test(data$scm2.max) # non-normal
shapiro.test(data$tde) # non-normal
shapiro.test(data$pbf) # normal (histogram looks non-normal)
shapiro.test(data$hbc) # non-normal
shapiro.test(data$bonanza) # normal-ish
shapiro.test(data$aridity) # normal

y1 <- log(data$ccm1 + 1)
y2 <- log(data$ccm2 + 1)
y3 <- log(data$scm1 + 1)
y4 <- log(data$scm2 + 1)
y5 <- log(data$ccm1.max + 1)
y6 <- log(data$ccm2.max + 1)
y7 <- log(data$scm1.max + 1)
y8 <- log(data$scm2.max + 1)
x1 <- log(data$tde + 1) # diversity
x2 <- log(data$pbf + 1) # formations
x3 <- log(data$hbc + 1) # collections
x4 <- log(data$bonanza + 1) # HBC:HBF ratio
x5 <- data$aridity # aridity

# Create all possible models plus null:
gls.null <- gls(y1~1, correlation = corARMA(p = 1), method = "ML") # 1
gls.1 <- gls(y1~x1+x2+x3+x4+x5, correlation = corARMA(p = 1), method = "ML") #
    2
gls.2 <- gls(y1~x1+x2+x3+x4, correlation = corARMA(p = 1), method = "ML") # 3
gls.3 <- gls(y1~x1+x2+x3+x5, correlation = corARMA(p = 1), method = "ML") # 4
gls.4 <- gls(y1~x1+x2+x4+x5, correlation = corARMA(p = 1), method = "ML") # 5
gls.5 <- gls(y1~x1+x3+x4+x5, correlation = corARMA(p = 1), method = "ML") # 6
gls.6 <- gls(y1~x2+x3+x4+x5, correlation = corARMA(p = 1), method = "ML") # 7
gls.7 <- gls(y1~x1+x2+x3, correlation = corARMA(p = 1), method = "ML") # 8
gls.8 <- gls(y1~x1+x2+x4, correlation = corARMA(p = 1), method = "ML") # 9
gls.9 <- gls(y1~x1+x2+x5, correlation = corARMA(p = 1), method = "ML") # 10
gls.10 <- gls(y1~x1+x3+x4, correlation = corARMA(p = 1), method = "ML") # 11
gls.11 <- gls(y1~x1+x3+x5, correlation = corARMA(p = 1), method = "ML") # 12
gls.12 <- gls(y1~x1+x4+x5, correlation = corARMA(p = 1), method = "ML") # 13
gls.13 <- gls(y1~x2+x3+x4, correlation = corARMA(p = 1), method = "ML") # 14
gls.14 <- gls(y1~x2+x3+x5, correlation = corARMA(p = 1), method = "ML") # 15
gls.15 <- gls(y1~x2+x4+x5, correlation = corARMA(p = 1), method = "ML") # 16
gls.16 <- gls(y1~x3+x4+x5, correlation = corARMA(p = 1), method = "ML") # 17
gls.17 <- gls(y1~x1+x2, correlation = corARMA(p = 1), method = "ML") # 18
gls.18 <- gls(y1~x1+x3, correlation = corARMA(p = 1), method = "ML") # 19
gls.19 <- gls(y1~x1+x4, correlation = corARMA(p = 1), method = "ML") # 20
gls.20 <- gls(y1~x1+x5, correlation = corARMA(p = 1), method = "ML") # 21
gls.21 <- gls(y1~x2+x3, correlation = corARMA(p = 1), method = "ML") # 22
gls.22 <- gls(y1~x2+x4, correlation = corARMA(p = 1), method = "ML") # 23
gls.23 <- gls(y1~x2+x5, correlation = corARMA(p = 1), method = "ML") # 24
gls.24 <- gls(y1~x3+x4, correlation = corARMA(p = 1), method = "ML") # 25
gls.25 <- gls(y1~x3+x5, correlation = corARMA(p = 1), method = "ML") # 26
gls.26 <- gls(y1~x4+x5, correlation = corARMA(p = 1), method = "ML") # 27
```

```
gls.27 <- gls(y1~x1, correlation = corARMA(p = 1), method = "ML") # 28
gls.28 <- gls(y1~x2, correlation = corARMA(p = 1), method = "ML") # 29
gls.29 <- gls(y1~x3, correlation = corARMA(p = 1), method = "ML") # 30
gls.30 <- gls(y1~x4, correlation = corARMA(p = 1), method = "ML") # 31
gls.31 <- gls(y1~x5, correlation = corARMA(p = 1), method = "ML") # 32

# Call function from Cleary et al. (2015) to calculate model weights:
weighted <- function (aic){
  aic.wt <- exp(-0.5 * (aic - min(aic)))/sum(exp(-0.5 * (aic - min(aic))))
  return(aic.wt)
}

# Computing and comparing the weightings and AICc values for all models:
wi <- weighted(c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3), AICc(
    gls.4),
           AICc(gls.5), AICc(gls.6), AICc(gls.7), AICc(gls.8), AICc(gls.9),
               AICc(gls.10),
           AICc(gls.11), AICc(gls.12), AICc(gls.13), AICc(gls.14), AICc(gls
               .15), AICc(gls.16),
           AICc(gls.17), AICc(gls.18), AICc(gls.19), AICc(gls.20), AICc(gls
               .21), AICc(gls.22),
           AICc(gls.23), AICc(gls.24), AICc(gls.25), AICc(gls.26), AICc(gls
               .27), AICc(gls.28),
           AICc(gls.29), AICc(gls.30), AICc(gls.31))); wi

AICcscores <- (c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3), AICc(
    gls.4), AICc(gls.5),
                 AICc(gls.6), AICc(gls.7), AICc(gls.8), AICc(gls.9), AICc(gls
                     .10), AICc(gls.11),
                 AICc(gls.12), AICc(gls.13), AICc(gls.14), AICc(gls.15), AICc(
                     gls.16), AICc(gls.17),
                 AICc(gls.18), AICc(gls.19), AICc(gls.20), AICc(gls.21), AICc(
                     gls.22), AICc(gls.23),
                 AICc(gls.24), AICc(gls.25), AICc(gls.26), AICc(gls.27), AICc(
                     gls.28), AICc(gls.29),
                 AICc(gls.30), AICc(gls.31))); AICcscores

# Compare all gls models and derive a global p-value using ANOVA:
anova(gls.null, gls.1, gls.2, gls.3, gls.4, gls.5, gls.6, gls.7, gls.8, gls.9,
     gls.10, gls.11, gls.12,
       gls.13, gls.14, gls.15, gls.16, gls.17, gls.18, gls.19, gls.20, gls.21,
           gls.22, gls.23, gls.24,
       gls.25, gls.26, gls.27, gls.28, gls.29, gls.30, gls.31)

# Plotting:
# FIGURE 2:
taxa <- c("Sahelanthropus", "Orrorrin", "Ardipithecus", "Ardipithecus", "
    Kenyanthropus",
          "Australopithecus", "Australopithecus", "Australopithecus", "
              Australopithecus","Australopithecus", "Australopithecus",
          "Australopithecus", "Paranthropus", "Paranthropus", "Paranthropus",
              "Homo", "Homo", "Homo")

taxon.ccm2 <- c(28.34,5.12,6.63,17.05,17.25,20.93,71.22,5.58,
  8.66,84.82,20.08,28.88,51.45,81.01,69.96,84.17,59.95,86.48)
taxon.scm2 <- c(13.77,8.74,5.61,49.97,12.57,16.10,84.79,1.57,
  4.76,69.95,3.98,52.23,13.33,23.56,33.52,30.57,12.44,80.33)
time.since <- c(16,17,17,24,17,23,40,23,3,93,19,8,50,59,80,54,32,43)
taxon.scores <- data.frame(taxa, taxon.ccm2, taxon.scm2, time.since)

# Re-order taxa by approximate FAD:
```

```
taxon.scores$taxa <- factor(taxon.scores$taxa, levels = c("Sahelanthropus","
    Orrorrin","Ardipithecus","Australopithecus",
  "Kenyanthropus","Paranthropus","Homo"))

u1 <- ggplot(taxon.scores, aes(x = taxa, y = taxon.ccm2)) +
  geom_boxplot(aes(group = taxa)) +
  ylab("CCM2 (%)") +
  scale_y_continuous(limits = c(0,100)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        axis.text.x = element_text(face = "italic"),
        text = element_text(family = "Myriad Pro", size = 22))

u2 <- ggplot(taxon.scores, aes(x = taxa, y = taxon.scm2)) +
  geom_boxplot(aes(group = taxa)) +
  ylab("SCM2 (%)") +
  scale_y_continuous(limits = c(0,100)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        axis.text.x = element_text(face = "italic"),
        text = element_text(family = "Myriad Pro", size = 22))

ggarrange(u1, u2, ncol = 1)

# Input data:
compmet <- read.csv(file = "Com_timeseries_data.csv", header = TRUE, sep = ","
    )

# Input script for calculating SE (from Cleary et al., 2015):
summarySE <- function(data = NULL, measurevar, groupvars = NULL, na.rm = TRUE,
                      conf.interval = 0.95, .drop = TRUE) {
  require(plyr)
  length2 <- function (x, na.rm = FALSE) {
    if (na.rm) sum(!is.na(x))
    else       length(x)
  }
  datac <- ddply(data, groupvars, .drop = .drop,
                 .fun = function(xx, col, na.rm) {
                   c( N    = length2(xx[,col], na.rm = na.rm),
                      mean = mean(xx[,col], na.rm = na.rm),
                      median = median(xx[,col], na.rm = na.rm),
                      sd   = sd(xx[,col], na.rm = na.rm)
                   )
                 },
                 measurevar,
                 na.rm
  )
  datac <- rename(datac, c("mean" = measurevar))
  datac$se <- datac$sd / sqrt(datac$N)
  ciMult <- qt(conf.interval/2 + 0.5, datac$N - 1)
  datac$ci <- datac$se * ciMult
  return(datac)
}
pd <- position_dodge(.1)
```

```
# Processing data:
scm1 <- summarySE(compmet, measurevar = "scm1", groupvars = c("time"))
scm2 <- summarySE(compmet, measurevar = "scm2", groupvars = c("time"))
scmts <- summarySE(compmet, measurevar = "scmts", groupvars = c("time"))
ccm1 <- summarySE(compmet, measurevar = "ccm1", groupvars = c("time"))
ccm2 <- summarySE(compmet, measurevar = "ccm2", groupvars = c("time"))
ccmts <- summarySE(compmet, measurevar = "ccmts", groupvars = c("time"))

# CCM1:
a <- ggplot(ccm1, aes(x = time, y = ccm1)) +
  geom_ribbon(aes(ymin = ccm1 - se, ymax = ccm1 + se), fill = "#feb24c") +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab("CCM1") +
  scale_y_continuous(limits = c(0,100), position = "left") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('A') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# CCM2:
b <- ggplot(ccm2, aes(x = time, y = ccm2)) +
  geom_ribbon(aes(ymin = ccm2 - se, ymax = ccm2 + se), fill = "#feb24c") +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab("CCM2") +
  scale_y_continuous(limits = c(0,100), position = "left") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('B') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# SCM1:
c <- ggplot(scm1, aes(x = time, y = scm1)) +
  geom_ribbon(aes(ymin = scm1 - se, ymax = scm1 + se), fill = "#feb24c") +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab("SCM1") +
  scale_y_continuous(limits = c(0,100), position = "left") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('C') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# SCM2:
d <- ggplot(scm2, aes(x = time, y = scm2)) +
  geom_ribbon(aes(ymin = scm2 - se, ymax = scm2 + se), fill = "#feb24c") +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  xlab("Time (Ma)") +
  ylab("SCM2") +
  scale_y_continuous(limits = c(0,100), position = "left") +
```

```
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('D') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# Create multiplot using R package egg (installed from GitHub):
# require(devtools)
# devtools::install_github("baptiste/egg")
# require(egg)
# ggarrange(a, b, c, d, nol = 1) # NOT WORKING!

multiplot <- function(..., plotlist = NULL, file, cols = 1, layout = NULL) {
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])

  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                      layout.pos.col = matchidx$col))
    }
  }
}

# FIGURE 3:
multiplot(a, b, c, d, cols = 1)

# FIGURE 4:
k1 <- ggplot(data, aes(x = time, y = ccm1.max)) +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab(expression(paste(CCM1[max]))) +
  scale_y_continuous(limits = c(0,100), position = "left") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
```

```
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

k2 <- ggplot(data, aes(x = time, y = ccm2.max)) +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab(expression(paste(CCM2[max]))) +
  scale_y_continuous(limits = c(0,100), position = "left") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

k3 <- ggplot(data, aes(x = time, y = scm1.max)) +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab(expression(paste(SCM1[max]))) +
  scale_y_continuous(limits = c(0,100), position = "left") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('c') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

k4 <- ggplot(data, aes(x = time, y = scm2.max)) +
  geom_line(size = 0.6, colour = "black") +
  geom_point(colour = "black", size = 4, shape = 21, fill = "white") +
  ylab(expression(paste(SCM2[max]))) +
  xlab("Time (Ma)") +
  scale_y_continuous(limits = c(0,100), position = "left") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('d') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

ggarrange(k1, k2, k3, k4, ncol = 1)

# FIGURE 5: plotting diversity and sampling metrics:
# TDE:
m <- ggplot(data, aes(x = time, y = tde)) +
  geom_bar(stat = "identity", fill = "#56B4E9") +
  ylab("Taxic diversity") +
  scale_y_continuous(breaks = seq(0,6,1)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        text = element_text(family = "Myriad Pro", size = 16))

# HBC:
```

```
n <- ggplot(data, aes(x = time, y = hbc)) +
  geom_bar(stat = "identity", fill = "#999999") +
  ylab("HBC") +
  scale_y_continuous(breaks = seq(0,50,10), position = "right") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# PBF:
o <- ggplot(data, aes(x = time, y = pbf)) +
  geom_bar(stat = "identity", fill = "#E69F00") +
  xlab("Time (Ma)") +
  ylab("PBF") +
  scale_y_continuous(breaks = seq(0,12,2)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('c') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# Bonanza variable:
p <- ggplot(data, aes(x = time, y = bonanza)) +
  geom_bar(stat = "identity", fill = "#56B4E9") +
  xlab("Time (Ma)") +
  ylab("Bonanza variable") +
  scale_y_continuous(breaks = seq(0,10,2), position = "right") +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('d') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# Specimen count:
q <- ggplot(data, aes(x = time, y = spec)) +
  geom_bar(stat = "identity", fill = "#999999") +
  xlab("Time (Ma)") +
  ylab("Specimen abundance") +
  scale_y_continuous(breaks = seq(0,600,200)) +
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('e') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.title.x = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# [Need to run sub-script in Chapter 3 first]
# Aridity
r <- ggplot(dust, aes(x = age, y = Terr.flux)) +
  geom_point(size = 1, col = "#E69F00") +
  xlab("Time (Ma)") +
  ylab(expression(paste("Dust flux (", g, ~ cm^-2, ~ ka^-1,")", sep = ""))) +
  scale_y_continuous(breaks = seq(0,2,1), position = "right") +
```

```
  scale_x_reverse(breaks = seq(1,7,1)) +
  ggtitle('f') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))

# Create multiplot:
ggarrange(m, n, o, p, q, r, ncol = 1)

# FIGURE 6: CCM1-2:
pdf("xyCCM.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))

# A: CCM1/2 against TDE:
plot(gd.ccm1, gd.tde, pch = 16, xlim = c(-30,40), ylim = c(-2,3), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"CCM")), ylab
    = expression(paste(Delta,"TDE")))
points(gd.ccm2, gd.tde, pch = 17, col = "#E31A1C")
op <- par(family = "Myriad Pro")
legend("topleft", c("CCM1", "CCM2"), pch = c(16,17), col = c("#56B4E9","#
    E31A1C"), bty = "n")
par(op)
# B: CCM1/2 against HBC:
plot(gd.ccm1, gd.hbc, pch = 16, xlim = c(-30,40), ylim = c(-20,20), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"CCM")), ylab
    = expression(paste(Delta,"HBC")))
op <- par(family = "Myriad Pro")
points(gd.ccm2, gd.hbc, pch = 17, col = "#E31A1C")
legend("topleft", c("CCM1", "CCM2"), pch = c(16,17), col = c("#56B4E9","#
    E31A1C"), bty = "n")
par(op)
# C: CCM1/2 against PBF:
plot(gd.ccm1, gd.pbf, pch = 16, xlim = c(-30,40), ylim = c(-3,5), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"CCM")), ylab
    = expression(paste(Delta,"PBF")))
op <- par(family = "Myriad Pro")
points(gd.ccm2, gd.pbf, pch = 17, col = "#E31A1C")
legend("topleft", c("CCM1", "CCM2"), pch = c(16,17), col = c("#56B4E9","#
    E31A1C"), bty = "n")
par(op)
# D: CCM1/2 against bonanza effect:
plot(gd.ccm1, gd.bonanza, pch = 16, xlim = c(-30,40), ylim = c(-4,4), family =
     "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"CCM")),
    ylab = expression(paste(Delta,"Bonanza variable")))
op <- par(family = "Myriad Pro")
points(gd.ccm2, gd.bonanza, pch = 17, col = "#E31A1C")
legend("topleft", c("CCM1", "CCM2"), pch = c(16,17), col = c("#56B4E9","#
    E31A1C"), bty = "n")
par(op)

par(mfrow = c(1,1))
dev.off()

# FIGURE 7: SCM1-2:
pdf("xySCM.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))

# A: SCM1/2 against TDE:
```

```
plot(gd.scm1, gd.tde, pch = 16, xlim = c(-40,40), ylim = c(-2,3), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"SCM")), ylab
    = expression(paste(Delta,"TDE")))
points(gd.scm2, gd.tde, pch = 17, col = "#E31A1C")
op <- par(family = "Myriad Pro")
legend("topleft", c("SCM1", "SCM2"), pch = c(16,17), col = c("#56B4E9","#
    E31A1C"), bty = "n")
par(op)
# B: SCM1/2 against HBC:
plot(gd.scm1, gd.hbc, pch = 16, xlim = c(-40,40), ylim = c(-20,20), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"SCM")), ylab
    = expression(paste(Delta,"HBC")))
op <- par(family = "Myriad Pro")
points(gd.scm2, gd.hbc, pch = 17, col = "#E31A1C")
legend("topleft", c("SCM1", "SCM2"), pch = c(16,17), col = c("#56B4E9","#
    E31A1C"), bty = "n")
par(op)
# C: SCM1/2 against PBF:
plot(gd.scm1, gd.pbf, pch = 16, xlim = c(-40,40), ylim = c(-3,5), family = "
    Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"SCM")), ylab
    = expression(paste(Delta,"PBF")))
op <- par(family = "Myriad Pro")
points(gd.scm2, gd.pbf, pch = 17, col = "#E31A1C")
legend("topleft", c("SCM1", "SCM2"), pch = c(16,17), col = c("#56B4E9","#
    E31A1C"), bty = "n")
par(op)
# D: SCM1/2 against bonanza effect:
plot(gd.scm1, gd.bonanza, pch = 16, xlim = c(-40,40), ylim = c(-4,4), family =
     "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"SCM")),
    ylab = expression(paste(Delta,"Bonanza variable")))
op <- par(family = "Myriad Pro")
points(gd.scm2, gd.bonanza, pch = 17, col = "#E31A1C")
legend("topleft", c("SCM1", "SCM2"), pch = c(16,17), col = c("#56B4E9","#
    E31A1C"), bty = "n")
par(op)

par(mfrow = c(1,1))
dev.off()

# FIGURE 8: CCM1-2maximum:
pdf("xyCCMmax.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))

# A: CCM1/2maximum against TDE:
plot(gd.ccm1max, gd.tde, pch = 16, xlim = c(-40,40), ylim = c(-2,3), family =
    "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"CCM" ["max"
    ])), ylab = expression(paste(Delta,"TDE")))
points(gd.ccm2max, gd.tde, pch = 17, col = "#E31A1C")
op <- par(family = "Myriad Pro")
legend("topleft", c(expression(paste("CCM1" ["max"])), expression(paste("CCM2"
     ["max"]))), pch = c(16,17), col = c("#56B4E9","#E31A1C"), bty = "n")
par(op)
# B: CCM1/2maximum against HBC:
plot(gd.ccm1max, gd.hbc, pch = 16, xlim = c(-40,40), ylim = c(-20,20), family
    = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"CCM" ["max
    "])), ylab = expression(paste(Delta,"HBC")))
op <- par(family = "Myriad Pro")
points(gd.ccm2max, gd.hbc, pch = 17, col = "#E31A1C")
legend("topleft", c(expression(paste("CCM1" ["max"])), expression(paste("CCM2"
     ["max"]))), pch = c(16,17), col = c("#56B4E9","#E31A1C"), bty = "n")
```

```
par(op)
# C: CCM1/2maximum against PBF:
plot(gd.ccm1max, gd.pbf, pch = 16, xlim = c(-40,40), ylim = c(-3,5), family =
    "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"CCM" ["max"
    ])), ylab = expression(paste(Delta,"PBF")))
op <- par(family = "Myriad Pro")
points(gd.ccm2max, gd.pbf, pch = 17, col = "#E31A1C")
legend("topleft", c(expression(paste("CCM1" ["max"])), expression(paste("CCM2"
     ["max"]))), pch = c(16,17), col = c("#56B4E9","#E31A1C"), bty = "n")
par(op)
# D: CCM1/2maximum against bonanza effect:
plot(gd.ccm1max, gd.bonanza, pch = 16, xlim = c(-40,40), ylim = c(-4,4),
    family = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"CCM
    " ["max"])), ylab = expression(paste(Delta,"Bonanza variable")))
op <- par(family = "Myriad Pro")
points(gd.ccm2max, gd.bonanza, pch = 17, col = "#E31A1C")
legend("topleft", c(expression(paste("CCM1" ["max"])), expression(paste("CCM2"
     ["max"]))), pch = c(16,17), col = c("#56B4E9","#E31A1C"), bty = "n")
par(op)

par(mfrow = c(1,1))
dev.off()

# FIGURE 9: SCM1-2maximum:
pdf("xySCMmax.pdf", height = 7, width = 7)
par(mfrow = c(2,2))
par(mar = c(4,4.1,1,1))

# A: SCM1/2maximum against TDE:
plot(gd.scm1max, gd.tde, pch = 16, xlim = c(-50,30), ylim = c(-2,3), family =
    "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"SCM" ["max"
    ])), ylab = expression(paste(Delta,"TDE")))
points(gd.scm2max, gd.tde, pch = 17, col = "#E31A1C")
op <- par(family = "Myriad Pro")
legend("topleft", c(expression(paste("SCM1" ["max"])), expression(paste("SCM2"
     ["max"]))), pch = c(16,17), col = c("#56B4E9","#E31A1C"), bty = "n")
par(op)
# B: SCM1/2maximum against HBC:
plot(gd.scm1max, gd.hbc, pch = 16, xlim = c(-50,30), ylim = c(-20,20), family
    = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"SCM" ["max
    "])), ylab = expression(paste(Delta,"HBC")))
op <- par(family = "Myriad Pro")
points(gd.scm2max, gd.hbc, pch = 17, col = "#E31A1C")
legend("topleft", c(expression(paste("SCM1" ["max"])), expression(paste("SCM2"
     ["max"]))), pch = c(16,17), col = c("#56B4E9","#E31A1C"), bty = "n")
par(op)
# C: SCM1/2maximum against PBF:
plot(gd.scm1max, gd.pbf, pch = 16, xlim = c(-50,30), ylim = c(-3,5), family =
    "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"SCM" ["max"
    ])), ylab = expression(paste(Delta,"PBF")))
op <- par(family = "Myriad Pro")
points(gd.scm2max, gd.pbf, pch = 17, col = "#E31A1C")
legend("topleft", c(expression(paste("SCM1" ["max"])), expression(paste("SCM2"
     ["max"]))), pch = c(16,17), col = c("#56B4E9","#E31A1C"), bty = "n")
par(op)
# D: SCM1/2maximum against bonanza effect:
plot(gd.scm1max, gd.bonanza, pch = 16, xlim = c(-50,30), ylim = c(-4,4),
    family = "Myriad Pro", col = "#56B4E9", xlab = expression(paste(Delta,"SCM
    " ["max"])), ylab = expression(paste(Delta,"Bonanza variable")))
op <- par(family = "Myriad Pro")
points(gd.scm2max, gd.bonanza, pch = 17, col = "#E31A1C")
```

```
legend("topleft", c(expression(paste("SCM1" ["max"])), expression(paste("SCM2"
    ["max"]))), pch = c(16,17), col = c("#56B4E9","#E31A1C"), bty = "n")
par(op)

par(mfrow = c(1,1))
dev.off()

# FIGURE 10:
jBrewColors <- brewer.pal(n = 10, name = "Paired")

ggplot(data, aes(x = autapomorphies, y = CCM2, colour = factor(genus))) +
  geom_point(size = 4) +
  xlab("# of autapomorphies") +
  ylab("CCM2 (%)") +
  scale_x_continuous(limits = c(0,20)) +
  scale_y_continuous(limits = c(20,100)) +
  scale_color_manual(values = jBrewColors) +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.text.x  = element_text(),
        axis.line = element_line(),
        legend.position = "bottom",
        legend.title = element_blank(),
        legend.text = element_text(face = "italic"),
        text = element_text(family = "Myriad Pro", size = 18))

# FIGURE 11:
data$epoch = factor(data$epoch, levels = 1:3,
                    labels = c("Miocene", "Pliocene", "Pleistocene")
)

# Box plot - CCM2:
b1 <- ggplot(data, aes(x = epoch, y = ccm2)) +
  geom_boxplot(aes(group = epoch)) +
  ylab("CCM2 (%)") +
  scale_y_continuous(limits = c(0,100)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(), axis.title.x = element_blank(), axis.text.
            x = element_text(),
        text = element_text(family = "Myriad Pro", size = 16))

# Box plot - SCM2:
b2 <- ggplot(data, aes(x = epoch, y = scm2)) +
  geom_boxplot(aes(group = epoch)) +
  ylab("SCM2 (%)") +
  scale_y_continuous(limits = c(0,100)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(), axis.title.x = element_blank(), axis.text.
            x = element_text(),
        text = element_text(family = "Myriad Pro", size = 16))

ggarrange(b1, b2, ncol = 2)

# FIGURE 12:
a <- ggplot(data, aes(x = bins, y = ccm2)) +
  geom_boxplot(aes(group = bins)) +
  ylab("CCM2 (%)") +
```

```
  scale_x_continuous(breaks = seq(1920,2020,10)) +
  scale_y_continuous(limits = c(0,100)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        axis.text.x=element_text(angle=45, hjust=1),
        text = element_text(family = "Myriad Pro", size = 16))

b <- ggplot(data, aes(x = bins, y = scm2)) +
  geom_boxplot(aes(group = bins)) +
  ylab("SCM2 (%)") +
  scale_x_continuous(breaks = seq(1920,2020,10)) +
  scale_y_continuous(limits = c(0,100)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.x = element_blank(),
        axis.text.x=element_text(angle=45, hjust=1),
        text = element_text(family = "Myriad Pro", size = 16))

ggarrange(a, b, ncol = 2)

# FIGURE 13:
ggplot(pt.col, aes(x = reorder(taxon, hbc), y = hbc)) +
  geom_bar(stat = "identity") +
  ylab("Per-taxon collection count") +
  coord_flip() +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        axis.title.y = element_blank(),
        axis.text.y = element_text(face = "italic"),
        text = element_text(family = "Myriad Pro", size = 20))

# Enter number of years since first publication (from 2018) and completeness
    scores:
time.since <- c(93,80,59,40,43,54,23,24,50,17,32,3,17,17,23,19,16,8)
ccm2 <- c(84.82,69.96,81.01,71.22,86.48,84.17,20.93,17.05,51.45,6.63,
  59.95,8.66,17.25,5.12,5.58,20.08,28.34,28.88)
scm2 <- c(69.95,33.52,23.56,84.79,80.33,30.57,16.10,49.97,13.33,5.61,
  12.44,4.76,12.57,8.74,1.57,3.98,13.77,52.23)

pt.col <- data.frame(taxon, hbc, time.since, ccm2, scm2)

pt.col$taxon <- factor(pt.col$taxon, levels = taxon)

# FIGURE 14:
ggplot(data = pt.col) +
  geom_point(aes(x = time.since, y = hbc), size = 4, shape = 16) + # col =
      "#56B4E9",
  ylab("Per-taxon collection count") +
  xlab("# of years since first publication") +
  scale_x_continuous(breaks = seq(0,100,10)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 16))
```

```
# FIGURE 15:
pt1 <- ggplot(pt.col, aes(x = hbc, y = ccm2)) +
  geom_point(size = 4) +
  xlab("Per-taxon collection count") +
  ylab("CCM2 (%)") +
  scale_x_continuous(breaks = seq(0,40,10)) +
  scale_y_continuous(limits = c(0,100)) +
  ggtitle('a') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.text.x  = element_text(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 18))

pt2 <- ggplot(pt.col, aes(x = hbc, y = scm2)) +
  geom_point(size = 4) +
  xlab("Per-taxon collection count") +
  ylab("SCM2 (%)") +
  scale_x_continuous(breaks = seq(0,40,10)) +
  scale_y_continuous(limits = c(0,100)) +
  ggtitle('b') +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        axis.text.x  = element_text(),
        axis.line = element_line(),
        text = element_text(family = "Myriad Pro", size = 18))

ggarrange(pt1, pt2, ncol = 2)

# End of Chapter 4 script
```

Appendix C

List of hominin-bearing horizons

Taxon	Country	Stratigraphy	Midpoint age (Ma)	Source
S. tchadensis	Chad	Anthracotheriid Unit	7.00	Lebatard *et al.* (2008)
O. tugenensis	Kenya	Lukeino Fm.	6.10	Pickford & Senut (2001)
O. tugenensis	Kenya	Lukeino Fm.	6.00	Pickford & Senut (2001)
O. tugenensis	Kenya	Lukeino Fm.	5.80	Pickford & Senut (2001)
O. tugenensis	Kenya	Lukeino Fm.	5.70	Pickford & Senut (2001)
cf. *Ar. kadabba*	Ethiopia	Adu-Asa Fm.	6.30	Kleinsasser *et al.* (2008)
Ar. kadabba	Ethiopia	Asa Koma Member, Adu-Asa Fm.	5.80	WoldeGabriel *et al.* (2001)
Ar. kadabba	Ethiopia	Asa Koma Member, Adu-Asa Fm.	5.70	WoldeGabriel *et al.* (2001)
Ar. kadabba	Ethiopia	Asa Koma Member, Adu-Asa Fm.	5.60	WoldeGabriel *et al.* (2001)
Ar. kadabba	Ethiopia	Asa Koma Member, Adu-Asa Fm.	5.50	WoldeGabriel *et al.* (2001)
Ar. kadabba	Ethiopia	Adu-Asa Fm.	5.40	Kleinsasser *et al.* (2008)
cf. *Ar. kadabba*	Ethiopia	Kuserale Member, Sagantole Fm.	5.20	WoldeGabriel *et al.* (2001)
Ar. ramidus	Ethiopia	As Duma Member, Sagantole Fm.	4.51	Semaw *et al.* (2005)
Ar. ramidus	Ethiopia	Lower Aramis Member, Sagantole Fm.	4.40	White *et al.* (1994)
Ar. ramidus	Ethiopia	As Duma Member, Sagantole Fm.	4.30	Semaw *et al.* (2005)
K. platyops	Kenya	8 m below Tulu Bor Tuff, Kataboi Member, Nachukui Fm.	3.54	Leakey *et al.* (2001)
K. platyops	Kenya	17 m above Tulu Bor Tuff, lower Lomekwi Member, Nachukui Fm.	3.35	Leakey *et al.* (2001)
K. platyops	Kenya	~5 m above Tulu Bor Tuff, lower Lomekwi Member, Nachukui Fm.	3.35	Leakey *et al.* (2001)
K. platyops	Kenya	~22 m above Tulu Bor Tuff, lower Lomekwi Member, Nachukui Fm.	3.35	Leakey *et al.* (2001)
Au. anamensis	Ethiopia	Adgantole Member, Sagantole Fm.	4.20	White *et al.* (2006)
Au. anamensis	Ethiopia	Adgantole Member, Sagantole Fm.	4.15	White *et al.* (2006)
Au. anamensis	Ethiopia	Adgantole Member, Sagantole Fm.	4.15	White *et al.* (2006)
Au. anamensis	Ethiopia	Adgantole Member, Sagantole Fm.	4.15	White *et al.* (2006)
Au. aff. *afarensis*	Ethiopia	Harr Fm. (Fejej)	4.10	Kappelman *et al.* (1996)
Au. anamensis	Kenya	Kanapoi Fm.	4.10	Leakey *et al.* (1995)
Au. anamensis	Kenya	Lonyumun Member, Koobi Fora Fm.	3.90	Leakey *et al.* (1995)
Au. anamensis	Kenya	Moiti Member, Koobi Fora Fm.	3.90	Leakey *et al.* (1995)
Au. aff. *afarensis*	Ethiopia	Belohdelie Member, Sagantole Fm.	3.90	Clark *et al.* (1984)
Au. afarensis	Tanzania	Upper Laetoli Beds (LH 10)	3.81	Harrison (2011*b*)

Taxon	Country	Stratigraphy	Midpoint age (Ma)	Source
Au. afarensis	Tanzania	Upper Laetoli Beds (LH 12)	3.79	Harrison (2011*b*)
Au. afarensis	Tanzania	Upper Laetoli Beds (LH 14)	3.70	Harrison (2011*b*)
Au. afarensis	Tanzania	Upper Laetoli Beds (LH 4)	3.66	Harrison (2011*b*)
Au. afarensis	Tanzania	Upper Laetoli Beds (EP 162/00)	3.63	Harrison (2011*b*)
Au. afarensis	Tanzania	Upper Laetoli Beds (EP 2400/00)	3.63	Harrison (2011*b*)
Au. afarensis	Ethiopia	Woranso-Mille study area (Korsi Dora)	3.58	Haile-Selassie *et al.* (2010)
Au. cf. *afarensis*	Kenya	Kaiyumung Member, Nachukui Fm.	3.50	Leakey & Walker (2003)
Au. afarensis	Kenya	Kantis Fossil Site (KFS)	3.45	Mbua *et al.* (2016)
Au. cf. *afarensis*	Kenya	Lomekwi Member, Nachukui Fm.	3.40	Ward *et al.* (1999*b*)
Au. afarensis	Ethiopia	5 m <SHT, Basal Member, Hadar Fm., Dikika	3.40	Wynn *et al.* (2006)
Au. afarensis	Ethiopia	Maka Sands, Matabaietu Fm.	3.40	White *et al.* (2000)
Au. afarensis	Ethiopia	5 m >SHT, Sidi Hakoma Member, Hadar Fm.	3.38	Kimbel & Delezene (2009)
Au. afarensis	Ethiopia	20–20 m >SHT, Sidi Hakoma Member, Hadar Fm.	3.34	Kimbel & Delezene (2009)
Au. afarensis	Kenya	Tulu Bor Member, Koobi Fora Fm.	3.30	Kimbel (1988)
Au. afarensis	Ethiopia	Woranso-Mille study area (Nefuraytu)	3.4	Haile-Selassie *et al.* (2016)
Au. afarensis	Ethiopia	Woranso-Mille study area (Nefuraytu)	3.4	Haile-Selassie *et al.* (2016)
Au. afarensis	Ethiopia	5 m >KHT-2, Kada Hadar Member, Hadar Fm.	3.19	Kimbel & Delezene (2009)
Au. afarensis	Ethiopia	12 m <KHT, Denen Dora Member, Hadar Fm.	3.20	Kimbel & Delezene (2009)
Au. afarensis	Ethiopia	10.5 m <BKT-2, Kada Hadar Member, Hadar Fm.	3.00	Kimbel & Delezene (2009)
Au. afarensis	Ethiopia	15 m <BKT-2, Kada Hadar Member, Hadar Fm.	3.00	Kimbel & Delezene (2009)
Au. afarensis	Ethiopia	Member B, Shungura Fm.	3.00	Hunt & Vitzthum (1986)
Au. bahrelghazali	Chad	Koro Toro (KT) unit	3.58	Lebatard *et al.* (2008)
Au. deyiremeda	Ethiopia	Woranso-Mille study area (Burtele)	3.40	Haile-Selassie *et al.* (2015)
Au. deyiremeda	Ethiopia	Woranso-Mille study area (Waytaleyta)	3.40	Haile-Selassie *et al.* (2015)
Au. africanus	South Africa	Member 3, Makapansgat	3.00	Herries *et al.* (2009)
Au. africanus	South Africa	Taung	2.80	Hopley *et al.* (2013)
Au. africanus	South Africa	Member 4, Sterkfontein	2.70	Herries & Shaw (2011)
Au. africanus	South Africa	Gladysvale	2.40	Herries *et al.* (2009)
Au. garhi	Ethiopia	Hatayae Member, Bouri Fm.	2.50	de Heinzelin *et al.* (1999)
Au. sediba	South Africa	Malapa	1.98	Pickering *et al.* (2011)

Taxon	Country	Stratigraphy	Midpoint age (Ma)	Source
P. aethiopicus	Ethiopia	Member C, Shungura Fm.	2.67	Wood & Leakey (2011)
P. aethiopicus	Tanzania	Upper Ndolanya Beds	2.66	Wood & Leakey (2011)
P. aethiopicus	Ethiopia	Member C, Shungura Fm.	2.60	Wood & Leakey (2011)
P. aethiopicus	Ethiopia	Member C, Shungura Fm.	2.60	Wood & Leakey (2011)
P. aethiopicus	Ethiopia	Member C, Shungura Fm.	2.55	Wood & Leakey (2011)
P. aethiopicus	Kenya	Upper Lokalalei Member, Nachukui Fm.	2.52	Brown & Feibel (1985)
P. aethiopicus	Ethiopia	Member D, Shungura Fm.	2.46	Wood & Leakey (2011)
P. aethiopicus	Kenya	Lokalalei Member, Nachukui Fm.	2.41	Brown & Feibel (1985)
P. aethiopicus	Ethiopia	Member F, Shungura Fm.	2.36	Wood & Leakey (2011)
P. aethiopicus	Ethiopia	Member E, Shungura Fm.	2.38	Wood & Leakey (2011)
P. aethiopicus	Ethiopia	Member E, Shungura Fm.	2.38	Wood & Leakey (2011)
P. aethiopicus	Ethiopia	Member F, Shungura Fm.	2.35	Wood & Leakey (2011)
P. aethiopicus	Ethiopia	Member D, Shungura Fm.	2.46	Wood & Leakey (2011)
P. boisei	Malawi	Unit 3A, Chiwondo Beds	2.40	Kullmer (2008)
P. boisei	Ethiopia	Member E, Shungura Fm.	2.34	Wood & Leakey (2011)
P. boisei	Ethiopia	Member F and G, Shungura Fm.	2.30	Wood & Leakey (2011)
P. boisei	Ethiopia	Member G, Shungura Fm.	2.23	Wood & Leakey (2011)
P. boisei	Ethiopia	Member G, Shungura Fm.	2.10	Wood & Leakey (2011)
P. boisei	Ethiopia	Member G, Shungura Fm.	1.90	Wood & Leakey (2011)
P. boisei	Kenya	Upper Burgi Member, Koobi Fora Fm.	1.89	Wood & Leakey (2011)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.88	Wood & Leakey (2011)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.87	Wood & Leakey (2011)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.85	Wood & Leakey (2011)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.85	Wood & Leakey (2011)
P. boisei	Tanzania	Bed I, Olduvai Beds	1.82	Deino (2012)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.78	Wood & Leakey (2011)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.77	Wood & Leakey (2011)
P. boisei	Kenya	Kaitio Member, Nachukui Fm.	1.77	Wood & Leakey (2011)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.74	Wood & Leakey (2011)
P. boisei	Tanzania	Upper Bed I and Lower Bed II, Olduvai Beds	1.73	Deino (2012)

Taxon	Country	Stratigraphy	Midpoint age (Ma)	Source
P. boisei	Kenya	Kaitio Member, Nachukui Fm.	1.72	Wood & Leakey (2011)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.69	Wood & Leakey (2011)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.67	Wood & Leakey (2011)
P. boisei	Kenya	Okote Member, Koobi Fora Fm.	1.60	Wood & Leakey (2011)
P. boisei	Kenya	KBS Member, Koobi Fora Fm.	1.58	Wood & Leakey (2011)
P. boisei	Kenya	Okote Member, Koobi Fora Fm.	1.58	Wood & Leakey (2011)
P. boisei	Kenya	Okote Member, Koobi Fora Fm.	1.52	Wood & Leakey (2011)
P. boisei	Kenya	Okote Member, Koobi Fora Fm.	1.49	Wood & Leakey (2011)
P. boisei	Tanzania	Upper Bed II, Olduvai Beds	1.43	Deino (2012)
P. boisei	Ethiopia	Member K, Shungura Fm.	1.43	Wood & Leakey (2011)
P. boisei	Ethiopia	Konso Fm.	1.42	Katoh *et al.* (2000)
P. boisei	Kenya	Chemoigut Fm.	1.42	Hooker & Miller (1979)
P. boisei	Tanzania	Humbu Fm.	1.40	Domínguez-Rodrigo *et al.* (2009)
P. boisei	Tanzania	BK Level 4, Upper Bed II, Olduvai Fm.	1.34	Domínguez-Rodrigo *et al.* (2013)
P. robustus	South Africa	Member 1, Swartkrans Fm.	2.00	Pickering *et al.* (2011)
P. robustus	South Africa	GD1 area, Gondolin	1.80	Herries *et al.* (2006)
P. robustus	South Africa	Drimolen	1.75	Herries *et al.* (2009)
P. robustus	South Africa	Member 3, Kromdraai B	1.70	Herries & Adams (2013)
P. robustus	South Africa	Member 5b, Sterkfontein Fm.	1.50	Pickering *et al.* (2011)
P. robustus	South Africa	Member 2, Swartkrans Fm.	1.40	Pickering *et al.* (2011)
P. robustus	South Africa	Cooper's Cave D	1.45	Pickering *et al.* (2011)
P. robustus	South Africa	Member 3, Swartkrans Fm.	1.00	Pickering *et al.* (2011)
H. rudolfensis	Kenya	Upper Burgi Member, Koobi Fora Fm.	1.92	Wood & Leakey (2011)
H. rudolfensis	Kenya	Upper Burgi Member, Koobi Fora Fm.	2.03	Wood & Leakey (2011)
H. rudolfensis	Kenya	Upper Burgi Member, Koobi Fora Fm.	1.93	Wood & Leakey (2011)
H. rudolfensis	Kenya	KBS Member, Koobi Fora Fm.	1.78	Wood & Leakey (2011)
H. habilis	Ethiopia	Kada Hada Member, Hadar Fm.	2.33	Kimbel *et al.* (1997)
H. habilis	South Africa	Member 5, Sterkfontein Fm.	2.30	Hughes & Tobias (1977)
H. habilis	South Africa	Member 1, Swartkrans Fm.	2.19	Clarke *et al.* (1970)
H. habilis	Ethiopia	Member G, Shungura Fm.	2.00	Wood & Leakey (2011)

Taxon	Country	Stratigraphy	Midpoint age (Ma)	Source
H. habilis	Kenya	Upper Burgi Member, Koobi Fora Fm.	1.89	Wood & Leakey (2011)
H. habilis	Tanzania	Bed I, Olduvai Fm. (OH 62)	1.84	Johanson *et al.* (1987)
H. habilis	Ethiopia	Member G, Shungura Fm.	1.84	Wood & Leakey (2011)
H. habilis	Tanzania	Bed I, Olduvai Fm.	1.86	Leakey *et al.* (1971)
H. habilis	Tanzania	Bed I, Olduvai Fm.	1.82	Blumenschine *et al.* (2003)
H. habilis	Kenya	Upper Burgi Member, Koobi Fora Fm.	1.65	Wood & Leakey (2011)
H. habilis	Kenya	Upper Burgi Member, Koobi Fora Fm.	1.75	Wood & Leakey (2011)
H. habilis	Kenya	KBS Member, Koobi Fora Fm.	1.75	Wood & Leakey (2011)
H. habilis	Tanzania	Bed II, Olduvai Fm.	1.65	Deino (2012)
H. erectus	Kenya	Upper Burgi Member, Koobi Fora Fm.	1.90	Wood & Leakey (2011)
H. erectus	Ethiopia	Busidima Fm.	1.80	Simpson *et al.* (2008)
H. erectus	Kenya	KBS Member, Koobi Fora Fm.	1.70	Wood & Leakey (2011)
H. erectus	Ethiopia	Unit B2, Level B, Melka Kunture Fm.	1.65	Chavaillon *et al.* (1977)
H. erectus	Kenya	KBS Member, Koobi Fora Fm.	1.62	Wood & Leakey (2011)
H. erectus	Kenya	Okote Member, Koobi Fora Fm.	1.60	Wood & Leakey (2011)
H. erectus	Kenya	KBS Member, Koobi Fora Fm.	1.60	Wood & Leakey (2011)
H. erectus	Kenya	KBS Member, Koobi Fora Fm.	1.55	Wood & Leakey (2011)
H. erectus	Kenya	Okote Member, Koobi Fora Fm.	1.55	Wood & Leakey (2011)
H. erectus	Ethiopia	Level E, Melka Kunture Fm.	1.50	Condemi (2004)
H. erectus	Kenya	Natoo Member, Nachukui Fm.	1.50	Antón & Swisher (2004)
H. erectus	Kenya	Okote Member, Koobi Fora Fm.	1.50	Wood & Leakey (2011)
H. erectus	Tanzania	Upper Bed II, Olduvai Beds	1.50	Antón (2003)
H. erectus	Ethiopia	Konso Fm.	1.45	Suwa *et al.* (2007)
H. erectus	Ethiopia	Konso Fm.	1.42	Suwa *et al.* (2007)
H. erectus	Ethiopia	Member K, Shungura Fm.	1.40	Wood & Leakey (2011)
H. erectus	Ethiopia	Konso Fm.	1.35	Suwa *et al.* (2007)
H. erectus	Ethiopia	Konso Fm.	1.30	Suwa *et al.* (2007)
H. erectus	Ethiopia	Konso Fm.	1.25	Suwa *et al.* (2007)
H. erectus	Eritrea	DL5, Alat Fm.	1.00	Bigazzi *et al.* (2004)
H. erectus	Ethiopia	Dakanihylo Member, Bouri Member	1.00	Asfaw *et al.* (2002)

Taxon	Country	Stratigraphy	Midpoint age (Ma)	Source
H. erectus	Kenya	Member 7, Olorgesailie Fm.	0.94	Potts *et al.* (2004)
H. erectus	Ethiopia	Melka Kunture Fm.	0.85	Chavaillon *et al.* (1974)
H. erectus	Tanzania	Lower Bed IV, Olduvai Beds	0.78	Antón (2003)
H. erectus	Algeria	Tighenif	0.70	Antón (2003)
H. erectus	Tanzania	Lower Member, Manyara Beds	0.70	Kaiser *et al.* (2010)

Appendix D

Latitude and longitude of all hominin-bearing localities

Time bin (Ma)	Country	Locality	Latitude	Longitude
7.0–6.8	Chad	Toros-Menalla (TM 266)	16.15	17.29
6.8–6.5	–	–	–	–
6.5–6.3	Ethiopia	Asbole Dora Locality 1, GPRP, Afar	11.10	40.24
6.3–6.0	Kenya	Aragai, Tugen Hills	0.57	35.83
6.0–5.8	Ethiopia	Asa Korma Locality 3, Western Margin, Middle Awash, Afar	10.30	40.26
	Kenya	Chepboit, Tugen Hills	0.79	35.86
	Kenya	Kapsomin, Tugen Hills	0.75	35.88
5.8–5.5	Ethiopia	Alayla Locality 2, Western Margin, Middle Awash, Afar	10.30	40.26
	Ethiopia	Saitune Dora Locality 2, Western Margin, Middle Awash	10.27	40.32
	Kenya	Kapcheberek, Tugen Hills	0.75	35.87
5.5–5.3	Ethiopia	Digiba Dora Locality 1, Western Margin, Middle Awash	10.30	40.26
	Ethiopia	Escarpment Locality 2 and 3, GPRP, Afar	11.25	40.22
	Ethiopia	Escarpment Locality 8, GPRP, Afar	11.22	40.23
5.3–5.0	Ethiopia	Amba Locality 1, Eastern Margin, Central Awash Complex, Afar	10.43	40.45
5.0–4.8	–	–	–	–
4.8–4.5	Ethiopia	Gona West Margin Locality 3, Afar	11.13	40.33
4.5–4.3	Ethiopia	Aramis Locality 1, 6, and 17, Middle Awash, Afar	10.43	40.45
	Ethiopia	Aramis Locality 7, 10, and 11, Middle Awash, Afar	10.47	40.43
	Ethiopia	Gona West Margin Locality 5, Afar	11.13	40.33
	Ethiopia	Kuseralee Dora Locality 2, Aramis, Middle Awash, Afar	10.43	40.45
	Ethiopia	Sagantole Locality 4 and 7, Aramis, Middle Awash, Afar	10.43	40.45
	Kenya	Tabarin, Tugen Hills	1.25	35.60
4.3–4.0	Ethiopia	Aramis Locality 14, Middle Awash, Afar	10.47	40.43
	Ethiopia	Asa Issie Locality 2 and 5, Middle Awash, Afar (including Hana Hari Locality 1)	10.39	40.41
	Ethiopia	Fejej	4.73	36.32
	Ethiopia	Galili, Galili Area, Somali Region	9.75	40.55
	Kenya	Kanapoi	2.30	36.08
4.0–3.8	Ethiopia	Belohdelie Locality 1	10.56	40.57
	Kenya	Allia Bay 261, East Turkana	3.71	36.26
	Tanzania	Laetoli (Upper Laetoli Beds)	−3.22	35.14

Time bin (Ma)	Country	Locality	Latitude	Longitude
3.8–3.5	Chad	Bahr el Ghazal, Koro Toro Locality 12, 13, and 40	16.25	17.48
	Ethiopia	Korsi Dora Locality 1, Woranso-Mille, Afar	11.48	40.51
	Kenya	Lomekwi, West Turkana	3.90	35.74
	Kenya	Lothagam, West Turkana	2.92	36.07
	Kenya	South Turkwel, West Turkana	3.13	35.84
	Tanzania	Laetoli (Upper Laetoli Beds)	−3.22	35.14
3.5–3.3	Ethiopia	Hadar (Sidi Hakoma Member)	11.11	40.58
	Ethiopia	Dikika	11.10	40.61
	Ethiopia	Maka Locality 1	10.56	40.57
	Ethiopia	Burtele Locality 1, Woranso-Mille, Afar	11.46	40.53
	Kenya	Il Naibar Lowlands 117, East Turkana	4.08	36.33
	Kenya	Lothagam, West Turkana	2.92	36.07
	Kenya	South Turkwel, West Turkana	3.13	35.84
	Kenya	Kantis Fossil Site	−1.33	36.75
	Kenya	Lomekwi, West Turkana	3.90	35.74
3.3–3.0	Ethiopia	Hadar (Denen Dora Member)	11.13	40.58
	Ethiopia	Hadar (Kada Hadar Member)	11.13	40.60
	Ethiopia	Brown Sands	5.29	36.17
	Ethiopia	White Sands	5.36	36.18
	Kenya	Lomekwi, West Turkana	3.90	35.74
3.0–2.8	Ethiopia	Hadar (Kada Hadar Member)	11.13	40.60
	Ethiopia	Brown Sands	5.29	36.17
	Ethiopia	White Sands	5.36	36.18
	Ethiopia	Ledi-Geraru, Lee Adoyta region	11.35	40.85
	South Africa	Makapansgat	−24.14	29.19
2.8–2.5	Ethiopia	Omo Shungura (Member C)	4.94	35.99
	Kenya	Lomekwi, West Turkana	3.90	35.74
	South Africa	Sterkfontein	−26.02	27.73
	South Africa	Taung	−27.62	24.63
	Tanzania	Laetoli (Upper Ndolanya Beds)	−3.22	35.14
2.5–2.3	Ethiopia	Bouri Locality 12, Middle Awash	10.27	40.54

Time bin (Ma)	Country	Locality	Latitude	Longitude
	Ethiopia	Omo Shungura (Members D, E, F, G)	4.94	35.99
	Ethiopia	Omo Shungura F locality	4.80	35.98
	Ethiopia	Omo Shungura L locality	5.13	36.04
	Ethiopia	Hadar, Afar Locality 666	11.15	40.55
	Kenya	Kangatukuseo Locality 2, West Turkana	4.13	35.90
	Malawi	Uraha	−10.02	33.93
	South Africa	Gladysvale	−25.90	27.75
	South Africa	Sterkfontein	−26.02	27.73
2.3–2.0	Ethiopia	Omo Shungura (Member G)	4.94	35.99
	Kenya	Karari Ridge 131, East Turkana	3.95	36.21
	South Africa	Swartkrans	−26.02	27.72
2.0–1.8	Ethiopia	Fejej	4.73	36.32
	Ethiopia	BSN49, Gona	11.06	40.47
	Ethiopia	Omo Shungura (Members G and H)	4.94	35.99
	Ethiopia	Omo Shungura (L locality)	5.13	36.04
	Kenya	Koobi Fora, East Turkana	4.08	36.33
	Kenya	Ileret, East Turkana	4.31	36.27
	Kenya	Koobi Fora, East Turkana	4.00	36.25
	Kenya	Karari Ridge, East Turkana	4.07	36.36
	Kenya	Bura Hasuma, East Turkana	4.19	36.44
	Kenya	Il Naibar Lowlands, East Turkana	4.08	36.33
	Kenya	Il Dura, East Turkana	4.27	36.30
	South Africa	Drimolen	−25.96	27.75
	South Africa	Gondolin	−25.41	27.86
	South Africa	Swartkrans	−26.02	27.72
	Tanzania	Olduvai Gorge (including MK and DK)	−2.98	35.37
	Tanzania	Olduvai Gorge (FLK)	−2.98	35.33
1.8–1.5	Ethiopia	Melka Kunture	8.71	38.59
	Ethiopia	BSN12, Gona	11.06	40.47
	Kenya	Koobi Fora Ridge, East Turkana	4.08	36.33
	Kenya	Ileret, East Turkana	4.31	36.27

Time bin (Ma)	Country	Locality	Latitude	Longitude
	Kenya	Koobi Fora Ridge, East Turkana	4.08	36.33
	Kenya	Kokiselei, West Turkana	4.13	35.89
	Kenya	Naiyena Engol, West Turkana	4.13	35.89
	Kenya	Nariokotome III, West Turkana	3.92	35.85
	South Africa	Sterkfontein	−26.02	27.73
	South Africa	Kromdraai B	−26.00	27.75
	Tanzania	Peninj, Lake Natron	−2.29	35.91
	Tanzania	Olduvai Gorge (including FLK and HWK)	−2.98	35.33
1.5–1.3	Ethiopia	Konso	5.43	37.39
	Ethiopia	Omo Shungura (Locality 204)	4.94	35.99
	Ethiopia	Melka Kunture	8.71	38.59
	Ethiopia	Omo Shungura (Locality 996)	4.94	35.99
	Kenya	Chesowanja area 2	0.65	36.20
	Kenya	Ileret, East Turkana	4.31	36.27
	Kenya	Koobi Fora Ridge, East Turkana	4.08	36.33
	Kenya	Nariokotome III, West Turkana	3.92	35.85
	South Africa	Cooper's Cave	−26.01	27.75
	South Africa	Swartkrans	−26.02	27.72
	South Africa	Sterkfontein	−26.02	27.73
	Tanzania	Olduvai Gorge (including BK, MNK, and SC)	-2.98	35.32
1.3–1.0	Ethiopia	Konso	5.43	37.39
	South Africa	Swartkrans	−26.02	27.72
	Tanzania	Olduvai Gorge (VEK)	−2.98	35.33

Appendix E

Detailed list of human bone weights used to calculate the Skeletal Completeness Metric

Skeletal region	Mean weight (g)	Region by %	Study
Skeleton	4975.78	–	Ingalls (1931)
Femur (2)*	919.61	18.48	Ingalls (1931)
Cranium	652.03	13.10	Ingalls (1931)
Parietal (2)	203.15	4.08	This study
Frontal	104.36	2.10	This study
Temporal (2)	85.95	1.73	This study
Occipital	83.88	1.69	This study
Maxilla and dentition	77.32	1.55	This study
Maxilla (2)	67.86	1.36	This study
Maxillary I1 (2)	0.94	0.02	Cheyne & Oba (1943)
Maxillary I2 (2)	0.64	0.01	Cheyne & Oba (1943)
Maxillary C (2)	1.01	0.02	Cheyne & Oba (1943)
Maxillary P3 (2)	0.97	0.02	Cheyne & Oba (1943)
Maxillary P4 (2)	0.89	0.02	Cheyne & Oba (1943)
Maxillary M1 (2)	1.95	0.04	Cheyne & Oba (1943)
Maxillary M2 (2)	1.71	0.03	Cheyne & Oba (1943)
Maxillary M3 (2)	1.36	0.03	Cheyne & Oba (1943)
Sphenoid	57.72	1.16	This study
Zygomatic (2)	14.47	0.29	This study
Ethmoid	12.43	0.25	This study
Palatine (2)	4.97	0.10	This study
Nasal concha (2)	3.71	0.07	This study
Vomer	1.77	0.04	This study
Lacrimal (2)	1.40	0.03	This study
Nasal (2)	1.00	0.02	This study
Pelvis	564.02	11.34	Ingalls (1931)
Innominate	450.77	9.06	Ingalls (1931)
Sacrum	111.48	2.24	Ingalls (1931)

*Number in parentheses refers to the number of elements. Where 2 is given this refers to a left and right side.

Skeletal region	Mean weight (g)	Region by %	Study
Coccyx	1.77	0.04	Ingalls (1931)
Tibia (2)	534.18	10.74	Ingalls (1931)
Vertebrae (24)	463.51	9.32	Ingalls (1931)
C1	12.68	0.25	Lowrance & Latimer (1967)
C2	13.85	0.28	Lowrance & Latimer (1967)
C3	8.18	0.16	Lowrance & Latimer (1967)
C4	8.68	0.17	Lowrance & Latimer (1967)
C5	9.01	0.18	Lowrance & Latimer (1967)
C6	9.68	0.19	Lowrance & Latimer (1967)
C7	11.02	0.22	Lowrance & Latimer (1967)
T1	17.22	0.35	Lowrance & Latimer (1967)
T2	15.77	0.32	Lowrance & Latimer (1967)
T3	14.32	0.29	Lowrance & Latimer (1967)
T4	14.11	0.28	Lowrance & Latimer (1967)
T5	14.73	0.30	Lowrance & Latimer (1967)
T6	16.19	0.33	Lowrance & Latimer (1967)
T7	17.02	0.34	Lowrance & Latimer (1967)
T8	18.26	0.37	Lowrance & Latimer (1967)
T9	19.71	0.40	Lowrance & Latimer (1967)
T10	21.58	0.43	Lowrance & Latimer (1967)
T11	23.04	0.46	Lowrance & Latimer (1967)
T12	25.32	0.51	Lowrance & Latimer (1967)
L1	28.05	0.56	Lowrance & Latimer (1967)
L2	32.89	0.66	Lowrance & Latimer (1967)
L3	36.75	0.74	Lowrance & Latimer (1967)
L4	37.53	0.75	Lowrance & Latimer (1967)
L5	37.91	0.76	Lowrance & Latimer (1967)
Humerus (2)	355.39	7.14	Ingalls (1931)
Ribs (24)	330.18	6.64	Ingalls (1931)
Tarsals (14)	208.85	4.20	Ingalls (1931)
Calcaneus (2)	88.62	1.78	Ingalls (1932)
Talus (2)	58.11	1.17	Ingalls (1932)
Navicular (2)	16.84	0.34	Ingalls (1932)
Cuboid (2)	16.80	0.34	Ingalls (1932)
Medial cuneiform (2)	14.68	0.30	Ingalls (1932)
Intermediate cuneiform (2)	5.92	0.12	Ingalls (1932)
Lateral cuneiform (2)	7.88	0.16	Ingalls (1932)
Scapula (2)	154.07	3.10	Ingalls (1931)
Ulna (2)	132.35	2.66	Ingalls (1931)
Fibula (2)	114.63	2.30	Ingalls (1931)
Radius (2)	106.40	2.14	Ingalls (1931)
Metacarpals and manual phalanges	104.38	2.10	Ingalls (1931)
Metacarpals (10)	55.40	1.11	Pyle (1935)
Metacarpal I (2)	9.94	0.20	Pyle (1935)
Metacarpal II (2)	15.18	0.31	Pyle (1935)
Metacarpal III (2)	14.22	0.29	Pyle (1935)
Metacarpal IV (2)	8.66	0.17	Pyle (1935)
Metacarpal V (2)	7.40	0.15	Pyle (1935)

Skeletal region	Mean weight (g)	Region by %	Study
Manual phalanges (28)	48.97	0.98	Pyle (1935))
Proximal phalanx I (2)	4.99	0.10	Pyle (1935)
Proximal phalanx II (2)	7.08	0.14	Pyle (1935)
Proximal phalanx III (2)	8.84	0.18	Pyle (1935)
Proximal phalanx IV (2)	6.75	0.14	Pyle (1935)
Proximal phalanx V (2)	4.11	0.08	Pyle (1935)
Intermediate phalanx II (2)	2.68	0.05	Pyle (1935)
Intermediate phalanx III (2)	4.00	0.08	Pyle (1935)
Intermediate phalanx IV (2)	3.23	0.06	Pyle (1935)
Intermediate phalanx V (2)	1.65	0.03	Pyle (1935)
Distal phalanx I (2)	1.94	0.04	Pyle (1935)
Distal phalanx II (2)	1.01	0.02	Pyle (1935)
Distal phalanx III (2)	1.10	0.02	Pyle (1935)
Distal phalanx IV (2)	0.88	0.02	Pyle (1935)
Distal phalanx V (2)	0.68	0.01	Pyle (1935)
Metatarsals and pedal phalanges	96.76	1.94	Ingalls (1931)
Metatarsals (10)	73.52	1.48	Ingalls (1932)
Metatarsal I (2)	24.93	0.50	Ingalls (1932)
Metatarsal II (2)	13.03	0.26	Ingalls (1932)
Metatarsal III (2)	11.38	0.23	Ingalls (1932)
Metatarsal IV (2)	11.14	0.22	Ingalls (1932)
Metatarsal V (2)	13.05	0.26	Ingalls (1932)
Pedal phalanges (28)	23.24	0.47	Ingalls (1932)
Phalanges (hallux)[†]	11.35	0.23	Ingalls (1932)
Phalanges (II to V)[‡]	11.89	0.24	Ingalls (1932)
Mandible and dentition	91.44	1.84	Ingalls (1931)
Mandible (edentulous)	81.98	1.65	Ingalls (1931)
Mandibular I_1 (2)	0.48	0.01	Cheyne & Oba (1943)
Mandibular I_2 (2)	0.55	0.01	Cheyne & Oba (1943)
Mandibular C (2)	0.95	0.02	Cheyne & Oba (1943)
Mandibular P_3 (2)	0.84	0.02	Cheyne & Oba (1943)
Mandibular P_4 (2)	0.95	0.02	Cheyne & Oba (1943)
Mandibular M_1 (2)	2.01	0.04	Cheyne & Oba (1943)
Mandibular M_2 (2)	1.97	0.04	Cheyne & Oba (1943)
Mandibular M_3 (2)	1.73	0.03	Cheyne & Oba (1943)
Clavicle (2)	53.54	1.08	Ingalls (1931)
Sternum	32.50	0.65	Ingalls (1931)
Patella (2)	32.07	0.64	Ingalls (1931)
Carpals (16)	29.87	0.60	Ingalls (1931)
Capitate (2)	6.05	0.12	Pyle (1935)
Scaphoid (2)	4.96	0.10	Pyle (1935)
Hamate (2)	4.92	0.10	Pyle (1935)
Lunate (2)	3.76	0.08	Pyle (1935)
Trapezium (2)	3.67	0.07	Pyle (1935)

[†]Ingalls (1932) did not provide the weight of each hallucal pedal phalanx, so here this value is divided equally between distal and proximal phalanges, and each side (i.e., 4), such that each bone contributes approximately 0.06%.

[‡]Ingalls (1932) did not provide the weight of each second to fifth pedal phalanx, so here this value is divided equally between distal, intermediate, and proximal phalanges, and each side (i.e., 24), such that each bone contributes 0.01%.

Skeletal region	Mean weight (g)	Region by %	Study
Triquetral (2)	2.62	0.05	Pyle (1935)
Trapezoid (2)	2.54	0.05	Pyle (1935)
Pisiform (2)[§]	1.36	0.03	Pyle (1935)

[§]Ingalls (1931, 1932) and Pyle (1935) studied The Hamann-Todd Human Osteological Collection housed at the Cleveland Museum of Natural History.

Appendix F

Relative proportion of each cranial bone derived from a Bone Clones Inc., magnetic osteological teaching skull™

Ingalls (1931) provided only the weight of the composite cranium. In order to estimate the weight of each of the cranium's twenty-two bones, and thus produce a more reliable estimate of skeletal completeness for fragmentary fossil remains, cranial bone weights were estimated using a Bone Clones Inc., 22-piece magnetic osteological teaching skull. The relative proportion of each cranial bone is shown in the table below.

Skeletal region	Model weight (g)	Relative proportion	Bone weight (g)
Cranium	560.55*	1	652.03
Parietal	174.65	0.31157	203.15
Frontal	89.72	0.16006	104.36
Temporal	73.89	0.13182	85.95
Occipital	72.11	0.12864	83.88
Maxilla	66.47	0.11858	77.32
Sphenoid	49.62	0.08852	57.72
Zygomatic	12.44	0.02219	14.47
Ethmoid	10.69	0.01907	12.43
Palatine	4.27	0.00762	4.97
Nasal concha	3.19	0.00569	3.71
Vomer	1.52	0.00271	1.77
Lacrimal	1.2	0.00214	1.40
Nasal	0.86**	0.00153	1.00

*Measurements were made using an Ohaus Pioneer 4102 precision balance (reliability ±0.01 g).
**Given the small size of the magnets used to affix each bone together, the estimated error in the weight of each cranial bone is ±0.2 g.