Key Evolutionary Transformations in *Homo sapiens*

Guest Editors: Darren Curnoe, Bing Su, Parth Chauhan, and John Gowlett
Key Evolutionary Transformations in *Homo sapiens*
Key Evolutionary Transformations in *Homo sapiens*

Guest Editors: Darren Curnoe, Bing Su, Parth Chauhan, and John Gowlett
Editorial Board

Giacomo Bernardi, USA
Terry Burke, UK
Ignacio Doadrio, Spain
Simon Easteal, Australia
Santiago F. Elena, Spain
Renato Fani, Italy
Dmitry A. Filatov, UK
F. González-Candelas, Spain
D. Graur, USA
Kazuho Ikeo, Japan
Yoh Iwasa, Japan
Henrik J. Jensen, UK
Amitabh Joshi, India
Hirohisa Kishino, Japan
A. Moya, Spain
G. Pesole, Italy
I. Popescu, USA
David Posada, Spain
Jeffrey R. Powell, USA
Hudson Kern Reeve, USA
Y. Satta, Japan
Koji Tamura, Japan
Yoshio Tateno, Japan
E. N. Trifonov, Israel
Eske Willerslev, Denmark
Shozo Yokoyama, USA
# Contents

**Key Evolutionary Transformations in *Homo sapiens***, Darren Curnoe, Bing Su, Parth Chauhan, and John Gowlett  
Volume 2011, Article ID 478690, 1 page

**Neanderthals versus Modern Humans: Evidence for Resource Competition from Isotopic Modelling.**  
Virginie Fabre, Silvana Condemi, Anna Degioanni, and Estelle Herrscher  
Volume 2011, Article ID 689315, 16 pages

**A Chronological Perspective on the Acheulian and Its Transition to the Middle Stone Age in Southern Africa: The Question of the Fauresmith**, Andy I. R. Herries  
Volume 2011, Article ID 961401, 25 pages

Amanuel Beyin  
Volume 2011, Article ID 615094, 17 pages

**Before the Emergence of *Homo sapiens*: Overview on the Early-to-Middle Pleistocene Fossil Record (with a Proposal about *Homo heidelbergensis* at the subspecific level)**, Giorgio Manzi  
Volume 2011, Article ID 582678, 11 pages

**Morphological Integration of the Modern Human Mandible during Ontogeny**, Joshua M. Polanski  
Volume 2011, Article ID 545879, 11 pages

**The Archaeological Record Speaks: Bridging Anthropology and Linguistics**, Sergio Balari, Antonio Benítez-Burraco, Marta Camps, Víctor M. Longa, Guillermo Lorenzo, and Juan Uriagereka  
Volume 2011, Article ID 382679, 17 pages

Volume 2011, Article ID 968012, 8 pages

**The Implications of the Working Memory Model for the Evolution of Modern Cognition**,  
Thomas Wynn and Frederick L. Coolidge  
Volume 2011, Article ID 741357, 12 pages

Volume 2011, Article ID 145262, 11 pages

**A 150-Year Conundrum: Cranial Robusticity and Its Bearing on the Origin of Aboriginal Australians**,  
Darren Curnoe  
Volume 2011, Article ID 632484, 18 pages

**Molecular Adaptation of Modern Human Populations**, Hong Shi and Bing Su  
Volume 2011, Article ID 484769, 8 pages
Editorial

Key Evolutionary Transformations in Homo sapiens

Darren Curnoe, Bing Su, Parth Chauhan, and John Gowlett

1 School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia
2 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology and Kunming Primate Research Centre, Chinese Academy of Sciences, Kunming 650223, China
3 Stone Age Institute, 1392 W. Dittemore Road, Gosport, IN 47433, USA
4 School of Archaeology, Classics and Egyptology, 2-14 Abercromby Square, University of Liverpool, Liverpool L69 7WZ, UK

Correspondence should be addressed to Darren Curnoe, d.curnoe@unsw.edu.au

Received 20 December 2011; Accepted 20 December 2011

The origin of modern humans remains a central question of palaeoanthropology—the discipline devoted to the scientific investigation of human evolution. This research field, like the questions it asks and evidence it draws upon, is multidisciplinary and synthetic in nature. Palaeoanthropology is also in the midst of a transformation, from a once largely descriptive and discovery driven enterprise to a discipline integrating the methods and results of cutting edge discoveries in 3D modeling and morphometrics, genomic sequencing, including ancient DNA and geochronology and isotope geochemistry, among others. Thus, the unfolding story of recent human evolution is as much about the transformation of the history of human evolution research as it is about the evidence and narratives we weave about our origins.

The geographic location and timing of the origin of modern humans are now well understood to have been Sub-Saharan Africa at around 200,000 years ago. Yet, there is still much to be learned about the hypothesised divergence of Homo sapiens and H. neanderthalensis from a common ancestor some 400,000 years ago, the deep-time transformations of archaic humans to modern H. sapiens, and the colonisation of Eurasia by a subset of Sub-Saharan Africans after 100,000 years ago.

Moreover, attempts to define anatomical and behavioural modernity have become increasingly challenging with new research in genetics suggesting the possibility of interbreeding with archaic hominins and discoveries from archaeology blurring the behavioural distinctions between Neanderthals and early modern humans.

Finally, a historical subtext in human origins research deals with explaining the perceived “gulf” between humans and all other life. Yet, this notion is one that sits uncomfortably beside the Darwinian concept of evolutionary continuity through common descent. It is also a subtext built into the very language of human origins research, being evident in widely used terms like “modern” and “archaic.” In reality, such terms lack biological meaning and are arbitrary (descriptive) contrivances—for example, if H. sapiens and H. neanderthalensis actually share a common ancestor, then the latter can be no more archaic than the former; yet H. neanderthalensis is frequently referred to as perhaps the ultimate example of an archaic group. Such terms are clearly loaded and only muddle our attempts to reconstruct human biological history within an evolutionary framework.

This special issue takes a broad brush approach to examining some of the more important topics of contemporary palaeoanthropology. Contributions cover topics ranging from molecular to morphological and population level transformations, and some of the key factors underpinning human evolution and variability, through to important behavioural questions examined using isotopic proxies, ideas from evolutionary psychology, and rich data from the archaeological record. They should find wide relevance across the many disparate fields that comprise this enterprise of human evolutionary research.

We wish to express our sincerest appreciation to all of the authors and reviewers whose efforts have made this special issue such a success. We truly hope it will provide a sense of “where we’re at” with respect to some key questions, answer some old questions, and, most importantly, satisfy curiosity and excite readers about future possibilities.

Darren Curnoe
Bing Su
Parth Chauhan
John Gowlett
Research Article
Neanderthals versus Modern Humans: Evidence for Resource Competition from Isotopic Modelling

Virginie Fabre, 1 Silvana Condemi, 1 Anna Degioanni, 1 and Estelle Herrscher 2

1 Laboratoire d’Anthropologie Bioculturelle, UMR 6578 CNRS/Université de la Méditerranée/EFS, Faculté de Médecine-Secteur Nord, Université de la Méditerranée, CS80011 Bd Pierre Dramard, 13344 Marseille Cedex 15, France
2 Laboratoire Méditerranéen de Préhistoire Europe Afrique (LAMPEA)-UMR 6636-MMSH 5, rue du Château de l’Horloge-BP. 647, 13094 Aix-en-Provence Cedex 2, France

Correspondence should be addressed to Virginie Fabre, mellefabrev@aol.com and Estelle Herrscher, herrscher@mmsh.univ-aix.fr

Received 15 October 2010; Revised 20 March 2011; Accepted 18 July 2011

Academic Editor: John Gowlett

Copyright © 2011 Virginie Fabre et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

During later MOIS3, in Europe two populations were present, autochthonous Neanderthals and modern humans. Ecological competition between these two populations has often been evoked but never demonstrated. Our aim is to establish whether resource competition occurred. In this paper, in order to examine the possibility of ecological competition between these two populations, 599 isotopic data were subjected to rigorous statistical treatment and analysis through mixing models. The aim of this paper was to compare dietary strategies of Neanderthals and modern humans over time. Our conclusions suggest that Neanderthals and modern humans shared dietary habits in the particular environmental context of MOIS3 characterised in Europe by climatic deterioration. In this environmental context, the resource competition between Neanderthals and modern humans may have accelerated the disappearance of the Neanderthal population.

1. Introduction

The Neanderthals are a well-known middle Pleistocene population, which was autochthonous in Europe during MOIS 6, 5, and 4. The European Neanderthals are associated with Mousterian assemblages.

During the later part of MOIS3, in the late Pleistocene, Europe was also populated by modern humans. The presence in Europe of modern humans is inferred, according to some authors, in the oldest Eastern European sites by association with Protoaurignacian or Aurignacian assemblages [1] and, for later periods of MOIS3, by association also with fossil remains [2]. The dates available for the Protoaurignacian, Aurignacian, and late Mousterian sites show that, after the arrival of modern humans, there was a period of coexistence between these two populations in Europe for at least 15,000 years [3]. This period is marked by an increase of Aurignacian sites throughout Europe, the appearance of so-called “transitional assemblage” sites (Chatelperronian, Uluzzian, Szeletian, Lincobian, Ranisian, Jerzmanowician), and the decline of Mousterian sites. After 25,000 years BP, the Mousterian sites and the Neanderthal population completely disappeared in Europe, and only modern humans survived on this continent.

Despite numerous investigations, the debate concerning whether Neanderthals became extinct because of climate change or competition with Modern humans is still unresolved. Some researchers argue that competition alone cannot be the cause of Neanderthal extinction [4–7]. By contrast, other authors support the existence of competitive exclusion for the same niche and argue that competition played a major role in the demise of the Neanderthal population. Some analyses, which are based on mathematical modeling, lack plausibility because they are too theoretical [8]; others, which are based on more integrative simulations [9] or which take into account archeological and ethnologic examples [10], are more convincing.

The modelling approach is used to understand complex systems by working on a simplified model of these systems. Thus, this process involves the choice of certain parameters and variables which, if they are simplified, are nonetheless controlled in such a way that they are capable of representing
the system as a whole. Therefore, the model used in this paper does not attempt to determine the kind of food that Neanderthals and modern humans consumed but to highlight the potential differences in dietary habits characteristic of these two populations.

The aim of this paper is to test the hypothesis that resource competition, analysed through isotopic modelling, was strong between Neanderthals and modern humans. It assumes that if a model shows similar dietary patterns for Neanderthals and modern humans, then these two populations would be in competition for resources. By contrast, if the models show differences in dietary patterns, this would signify that resource competition would be less intense.

For some years, isotopic biochemistry allowed us to improve our knowledge about past human diet using carbon and nitrogen isotopic ratios [11–15]. Since 1990’s these methods have been increasingly used to study palaeontological populations such as Neanderthals or early modern humans, in order to understand their relationship with the local environment (e.g., [11, 16–23]). Therefore, the literature contains a substantial number of isotopic data, mainly on carbon and nitrogen isotopic values measured on bones and dental collagen. The previous studies suggested that prehistoric peoples had a carnivorous diet similar to that of contemporaneous predators, such as cave lions or cave hyenas [19, 24]. These isotopic studies are consistent with zooarchaeological investigations which showed that Neanderthals and modern humans were big game hunters (hunting mainly big ungulates) [25–30].

The isotopic modelling used in this paper presents a new method of investigation that intends to contribute to the debate on resource competition between Neanderthals and modern humans which has often been assumed but never really demonstrated.

2. Material and Methods

2.1. Compilation of the Database. Isotopic data from 51 major archaeological sites in Europe (Figure 1) were compiled from 42 publications. Of these archaeological sites, 14 were attributed to Neanderthal and 37 to modern human settlements. In total, isotopic data from 945 specimens (faunal and human) was assembled from the literature. This paper focuses on the transition between MOIS3 and MOIS2; as such, isotope data from species unavailable during these time periods was eliminated from the data set. Furthermore, the models employed in this paper rely exclusively on three faunal types (reindeer, horse, and bovid) because these were the only remains present at all sites. As a result, only isotope data from 599 specimens were included in this analysis (Table 1 and Tables SI1, SI2, SI3).

As Drucker has shown, local environmental context can influence isotopic signatures of plants and consequently those of consumers [31]. As a result, the first step of our analysis was to verify isotopic modifications for each faunal type through time and space [32]. Thus, in addition to chronology (e.g., late MOIS3 versus MOIS2), data relating to geography and environment were also considered (Table 2). Environmental groupings were in agreement with Allen and
In order to study the transition between late MOIS3 and MOIS2, we created models by grouping faunal types and humans species into three chronological groups: (1) MOIS3 Neanderthals, (2) MOIS3 modern humans, (3) MOIS2 modern humans. In this paper we used the term “MOIS3” in order to nominate the coevolution period of Neanderthals and modern humans in Europe; thus “MOIS3” here represents the later period of MOIS3. These groups were analysed in three different ways: (1) an absence of cluster (global), (2) a geographical cluster, (3) an environmental cluster. Due to limitations related to the faunal isotopic data available for each cluster and the fact that models have to be run with the same characteristics for diachronic comparison, models were limited to: (1) the whole dataset, (2) the data of south-western area, (3) the data relating to cold environments of tundra-steppe and open boreal woodland (Table 3).

2.2. Isotopic Values. Patterns in human and animal food consumption are reconstructed using carbon and nitrogen isotope ratios in bone collagens. Since collagen is protein, the stable isotope ratio of this tissue provides information on the protein component of the diet over approximately the last 10 years of an individual's life [63, 64]. Because plants and animals differ in their carbon and nitrogen isotope ratios it is possible to use their ratio to infer past dietary patterns. Carbon isotope ratios are typically used to differentiate between the consumption of C3 versus C4 plants or marine fish versus fresh water fish [11, 14, 63, 65, 66]. In contrast, nitrogen isotope ratios are indicative of trophic level (i.e., an individual position in the food web) [65, 67]. Stable isotope ratios reflect the type of primary protein sources and are successively enriched in the heavy isotope (13C, 15N) with each step up the food web [65, 68]. Thus, the relative isotopic variability between different organisms of a terrestrial and aquatic trophic web is distributed in a predictive way from plants at the baseline of the food chain through the subsequent levels as herbivorous and carnivorous organisms. For example, the δ13C and δ15N values of collagen from herbivores are approximately 5% and 3–5% higher, respectively, than plants [20, 63]. In a similar way, the δ13C and δ15N values of collagen from carnivores are approximately 0.8–1.3% and 3–5% higher, respectively, than herbivores [19]. According to isotopic data available for the Palaeolithic terrestrial environments, δ13C values of plants range from −35 to −20% with a distinction between open and closed environments, and δ13C values of plants ranged from 0 to 6%. The δ15C and δ15N values of collagen from herbivores range from −30 to −18% and 3 to 8%, respectively. The δ13C and δ15N values of collagen from carnivores range from −24 to −16% and 7 and 13%, respectively. The δ13C and δ15N values of collagen from freshwater fish range from −23 and −19% and 9 and 15%, respectively.

2.3. Modelling Process. Following the predictive fractionation of isotope ratios through the food chain, Phillips and colleagues proposed different isotopic mixing models to quantify the relative contributions of the different dietary sources to an individual [69–71]. IsoError (2001) and IsoConc (2002) based on the isotopic mixing models estimate the proportions for two food sources using a single isotopic element or three sources using two isotopic elements. IsoError mixing model considers the isotopic signature standard deviations in the source and mixture populations and restitutes food proportions with confidence intervals for source proportion estimates. In contrast, the IsoConc mixing model is a concentration-weighted linear mixing model which considers for each element the contribution of source as proportional to the weight of the elemental concentration in that source. The IsoSource mixing model (2005) can estimate more than 3 sources using 2 stable isotopic
elements when food sources are isotopically very different. The different isotopic mixing models have previously been applied in past foodwebs to assess the relative proportions of different sources to a single or population human mixtures [33, 72–78].

Based on isotopic mixing model proposed by Phillips et al., the final aim of the research was not to determine the real contributions of specific food sources to a mixture but to compare the human dietary patterns over time and space within distinct chronological groups. Simulations were run with IsoSource mixing models (version 1.3.1) [69, 79]. The use of mixing models to study European Palaeolithic populations could be limited because the mixing models were designed from datasets composed of living North American animals. However, regardless of chronological period and geographical context, the same principle of isotopic fractionation along the trophic chain is generally applicable to all isotopic studies. Furthermore, the mixing models were applied to compare different cohorts according to different parameters and not to precisely reconstruct the proportions of different resources consumed by past human populations. As such, whatever limitations exist will be similar in each group and should not affect the comparisons.

The use of models necessitates a variety of assumptions. For example, in this study we assume that all the hominids considered consumed the same kinds of resources. Following the recommendation of Phillips et al. [69], only three food sources were considered: fish, meat, and plant resources because they are (i) largely distant from the mixture and (ii) sufficient and reliable to consider the main food items consumed in typical omnivorous diet. Due to the lack of data, plants are not often considered in previous isotopic studies whereas they are necessary for human survival [80]. This resource was included in our simulations permitting to consider one of food items consumed by omnivorous. Due to the isotopic fractionation from diet to consumer (bone collagen), in following simulations, isotopic values of hominid diet, mentioned as “mixing diet” or “mixture”, were considered to be lower than isotopic values of hominid collagens of 0.8–1.3% for carbon and 3–5% for nitrogen [19, 20, 65, 68]. For each cluster datasets, and according to the appropriate fractionation factor requires for running models, isotopic values of plants were estimated from available herbivorous isotopic values (0.8–1.3% for carbon and 3–5% for nitrogen) (see Supplementary Material SI3 available online at doi: 10.4061/2011/689315). Isotopic values of meat and fish resources of hominid diets were derived from the literature and summarized in the database (Table SI1).

For each cluster (global, geographical, and environmental), 3 sets of simulations were performed (one for each chronological groups: MOIS3 Neanderthals, MOIS3 modern humans, and MOIS2 modern humans). To get the best inputs for running IsoSource and to have comparable patterns for each chronological group, a similar geometric construction was realised within the estimated carbon and nitrogen variability of each food source (mean ± SD) (Figure 2). The isotopic variability for herbivorous

source was calculated by averaging the mean of each faunal type isotopic signature to avoid any bias due to sample size. Moreover, concerning the whole dataset and the cold environments dataset, the geometric construction usable for modelling had to be based on the largest isotopic values of sources around the mixture. Concerning the South-western dataset, two sets of simulations have been performed; the first one involves a geometric construction based on the largest isotopic values of sources around the mixture (mean ± 1SD), and the second one involves a geometric construction based on average isotopic values of sources (mean).

2.4. Statistical Analysis. To assess the validity of the different clusters proposed in Section 1, a set of Kruskall-Wallis statistics was applied to the database (Statistica software). Kruskall-Wallis statistics are nonparametric tests generally used to compare the distribution of two independent sets of values. Here the aim is to test (i) if chronological groups are reliable for investigating the homogeneity of stable isotopic signatures of each faunal group across the time and (ii) if all archaeological sites associated in one cluster present homogeneous stable isotopic ratios over space.

Results of simulations are given by proportions of sources expressed as percentages. Although in order to compare, in each hominid group and in each simulation, the contribution of each source to the mixture, Chi-squared test with a Bonferroni correction was used. To use the Chi-squared test, a standardisation of results has been done based on a calculation of a resource unit related to the quantity of protein intake.

\[
\text{Mean } \pm 1\text{SD; the values used for running simulations are described in Table SI3.)}
\]
3. Results

3.1. Isotopic Variability Analyses. A descriptive statistical analysis of the complete isotopic dataset was performed to detect whether isotopic differences existed (i) between the archaeological sites of the considered geographic region, (ii) between the archaeological sites of the considered environments as well as (iii) between chronological subdivisions of MOIS3 and MOIS2. Kruskal-Wallis analyses were implemented on the whole dataset, on the South-western dataset, and on the cold environment dataset to test if chronological groups are efficient and if all archaeological sites associated with a chronological group (MOIS3 Neanderthal, MOIS3 modern humans, or MOIS2 modern humans) could be pooled together in terms of carbon and nitrogen isotopic ratios. P values lower than 0.01 indicate statistically significant differences suggesting an isotopic heterogeneity within groups.

(1) Regarding the analysis of the complete dataset (Table S14), we often observed a particular dietary/isotopic behaviour for the reindeer which seems attributed to local environment. This is why we decided to perform two sets of simulations, a set including the reindeer and a set excluding from the reindeer. Concerning MOIS2 whole dataset, bovids and horse show isotopic variability across archaeological sites. We identified and excluded the divergent archaeological sites for bovids (Gough’s cave and Kendrick’s cave for bovids). Since divergent archaeological sites were not identifiable for horse, in order to avoid any biases, several sets of simulations were conducted with different δ15N values for herbivorous sources (δ15N ± 1SD; Figure 2).

(2) For the South-western area (Table S15), reindeer isotopic variability seems attributed to local environment whereas bovids’ seems to be heterogeneous during MOIS2. To avoid any biases, divergent sites have been excluded from the simulations (Pont d’Ambon, [31]). As mentioned above, additional simulations have also been performed with and without reindeer.

(3) In cold environments (Table S16), reindeer isotopic variability seems also attributed to local environment. Similarly, simulations with and without consideration of the reindeer isotopic variability within the herbivorous group were conducted.

To summarise, based on the results of Kruskal-Wallis analyses, simulations were performed according to each cluster and each chronological group under several conditions: (1) presence of the three faunal types in the herbivorous source (bovid, reindeer, horse), (2) presence of bovids and horse, in the herbivorous source, to take into account the isotopic variability of reindeer across archaeological sites), (3) presence of bovids and reindeer in the herbivorous source, to take into account that horses are nonruminant species [81]. For the environmental cluster only, (4) a fourth simulation was performed considering only the reindeer in the herbivorous source since it was probably the main species consumed under these cold climates [82].

In addition, two sets of simulations were performed to guaranty the most reliable application of isotopic biochemistry in the reconstruction of past diet: (1) one considering the totality of the hominids for each chronological group and (2) a second considering only the hominids associated with fauna. In total, 19 simulations have been performed on the whole dataset, 28 on the South-western dataset, and 16 on the cold environment dataset (Table 4).

3.2. Simulation Results. Simulations were implemented using three food sources selected according to isotopic dataset relative to each cluster and each chronological group. In all simulations, source 1 shows fixed values characterised by high δ15N values and low δ13C values. Source 2 and source 3 isotopic signatures were determined according to isotopic dataset relative to each cluster and each chronological group. In all simulations, source 2 shows intermediate δ15N values and high δ13C values whereas source 3 exhibits low δ13C and δ15N values. The result of the simulations is driven by the relative position of the mixture compared to the three sources.

Regarding the MOIS3 Neanderthals’ dataset (Table 5), in all clusters and under all conditions, the results of the different simulations show the same patterns. The contribution of source 1 is the highest (between 48% and 67% of the mixing food), suggesting a consumption of food with high δ15N values and low δ13C values. The contribution of source 3 (between 28% and 44%) is lower than source 1 but higher than the contribution of source 2 (between 1% and 19%), which confirms the greater consumption of food with low δ13C signatures. More precisely, simulations without considering the reindeer isotopic values tend to increase the contribution of source 2, whereas simulations without considering the horse isotopic values tend to decrease the contribution of source 2. In other words, the more the isotopic value of the herbivorous source is enriched in 13C, the fewer source 2 contributes to the mixture. In the South-western area, even if the contribution of source 2 slightly decreases when simulations are based on average isotopic values of different sources, the differences between the two sets of simulations (regarding a large variability versus an average variability) are not significant.

Concerning the MOIS3 modern humans’ dataset (Table 6), the results of simulations within the South-western area exhibit different dietary patterns relatively to the other two clusters (global and environmental). For the three clusters, the contribution of source 1 is the highest (between 58% and 72%). In the South-western area, the contribution of source 2 is higher than the contribution of source 3 (resp., between 15% and 40% and between 1% and 13%). It is the opposite for the global and environmental clusters, the contribution of source 3 is higher than the contribution of source 2 (resp., between 26% and 32% and between 2% and 15%). These results would suggest a lesser consumption of 13C enriched food. Regardless of the faunal types considered in the herbivorous source, the results globally remain the same for the cold environments and the complete datasets. Nevertheless, when the reindeers are absent, the contribution of source 3 is lower than the
Table 4: Characterization of the simulations performed with IsoSource according to the different geographical groups and the three clusters considered.

<table>
<thead>
<tr>
<th>Conditions tested</th>
<th>Simulations</th>
<th>Simulations</th>
<th>Simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Whole fauna/whole hominids</td>
<td>Whole fauna/associated hominids</td>
<td>Fauna without reindeer/whole hominids</td>
</tr>
<tr>
<td>MOIS3 Neanderthals</td>
<td>Southern Europe area</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Cold environment</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Whole dataset</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MOIS3 modern humans</td>
<td>Southern Europe area</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Cold environment</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Whole dataset</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MOIS2 modern humans</td>
<td>Southern Europe area</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Cold environment</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Whole dataset</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Table 5: Results of simulations by IsoSource for MOIS3 Neanderthals chronological group and under the three clusters; circular diagram represents the different source proportions to the mixture.

<table>
<thead>
<tr>
<th></th>
<th>Cold environment variability (mean ± 1SD)</th>
<th>South-western area variability (mean ± 1SD)</th>
<th>Whole dataset variability (mean ± 1SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>whole fauna/whole hominids</td>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
</tr>
<tr>
<td>whole fauna/associated hominids</td>
<td><img src="image4" alt="Diagram" /></td>
<td><img src="image5" alt="Diagram" /></td>
<td><img src="image6" alt="Diagram" /></td>
</tr>
<tr>
<td>fauna without reindeer/whole hominids</td>
<td><img src="image7" alt="Diagram" /></td>
<td><img src="image8" alt="Diagram" /></td>
<td><img src="image9" alt="Diagram" /></td>
</tr>
<tr>
<td>fauna without reindeer/associated hominids</td>
<td><img src="image10" alt="Diagram" /></td>
<td><img src="image11" alt="Diagram" /></td>
<td><img src="image12" alt="Diagram" /></td>
</tr>
<tr>
<td>fauna without horse/whole hominids</td>
<td><img src="image13" alt="Diagram" /></td>
<td><img src="image14" alt="Diagram" /></td>
<td><img src="image15" alt="Diagram" /></td>
</tr>
<tr>
<td>fauna without horse/associated hominids</td>
<td><img src="image16" alt="Diagram" /></td>
<td><img src="image17" alt="Diagram" /></td>
<td><img src="image18" alt="Diagram" /></td>
</tr>
<tr>
<td>reindeer/whole hominids</td>
<td><img src="image19" alt="Diagram" /></td>
<td><img src="image20" alt="Diagram" /></td>
<td><img src="image21" alt="Diagram" /></td>
</tr>
</tbody>
</table>
Table 5: Continued.

<table>
<thead>
<tr>
<th></th>
<th>Cold environment variability (mean ± 1SD)</th>
<th>South-western area variability (mean ± 1SD)</th>
<th>Whole dataset variability (mean ± 1SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>reindeer/associated hominids</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Results of simulations by IsoSource for MOIS3 modern humans chronological group and under the three clusters; circular diagram represents the different source proportions to the mixture.

<table>
<thead>
<tr>
<th></th>
<th>Cold environment variability (mean ± 1SD)</th>
<th>South-western area variability (mean ± 1SD)</th>
<th>Whole dataset variability (mean ± 1SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>whole fauna/whole hominids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fauna without reindeer/whole hominids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fauna without horse/whole hominids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reindeer/whole hominids</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

contribution of source 2. It is the opposite when horses are not considered in the simulations.

In the South-western area, similar dietary patterns are observed when the reindeers are not considered; the contribution of source 3 clearly decreases whereas the contribution of source 2 clearly increases. It is also the opposite when horses are not considered in simulations. As suggested for MOIS3 Neanderthals, we can suppose that the more the isotopic values of herbivorous source are enriched in $^{13}\text{C}$, the fewer source 2 contributes to
Table 7: Results of simulations by IsoSource for MOIS2 modern humans chronological group and under the three clusters; circular diagram represents the different source proportions to the mixture.

<table>
<thead>
<tr>
<th>Cold environment variability (mean ± 1SD)</th>
<th>South-western area variability (mean)</th>
<th>Whole dataset variability (mean ± 1SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>whole fauna/whole hominids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold environment variability</td>
<td>South-western area variability</td>
<td>Whole dataset variability</td>
</tr>
<tr>
<td>variability</td>
<td>variability</td>
<td>variability</td>
</tr>
<tr>
<td>(mean ± 1SD)</td>
<td>(mean)</td>
<td>(mean ± 1SD)</td>
</tr>
</tbody>
</table>

- whole fauna/whole hominids
- whole fauna/associated hominids
- fauna without reindeer/whole hominids
- fauna without reindeer/associated hominids
- fauna without horse/whole hominids
- fauna without horse/associated hominids
- reindeer/whole hominids
Concerning the MOIS2 modern humans dataset (Table 7), independently of clusters and contrary to the MOIS3 Neanderthals and modern humans, dietary patterns show a slight predominance of source 2 (between 20% and 82%) compared with the other two sources (between 0 and 40% for source 3 and between 30% and 51% for source 1). Some differences have been observed according to faunal types included in the herbivorous source. Overall, the contribution of source 3 slightly decreases when the reindeer are excluded and increases when the horse are excluded. These different dietary patterns according to the inclusion or exclusion of reindeer and horse in herbivorous source are consistent with the observations made for MOIS3 Neanderthals and MOIS3 modern humans. Regarding the
two sets of simulations performed for the South-western area, the contribution of source 3 decreases in conjunction with an increase of source 2 when average isotopic values are considered for sources. An exception is observed when horses are not included in the herbivorous source. The contribution of source 2 also increases when the herbivorous source exhibits high $\delta^{13}C$ values.

Chi-squared tests were conducted according to the three chronological groups. The results of this statistical analysis emphasize previous observations (Table 8). With the Bonferroni correction and two degrees of freedom, the $P$ values are considered significant below 0.017. Dietary patterns of MOIS2 modern humans appear statistically different than the other two chronological groups. Dietary patterns of MOIS3 modern humans and MOIS3 Neanderthals do not exhibit significant differences for global and environmental clusters. For geographical cluster, as observed on raw results, simulations show significant differences in dietary patterns between MOIS3 modern humans and MOIS3 Neanderthals. As mentioned before, this could be explained by the unique nitrogen isotopic value available for MOIS3 South-western modern humans.

### 4. Discussion and Conclusions

Our research on past human diet during the transition from MOIS3 to MOIS2 is based on the modelling of isotopic signatures of a mixture over time and under several clusters (global, geographic, or environmental clusters). Our aim has been to test the hypothesis that resource competition, analysed through isotopic modelling, may have existed between Neanderthals and contemporaneous modern humans. Isotopic analyses, which are generally used for studying local environments, would also seem to be applicable to a population approach. Indeed, the results of our modelling illustrate that, whatever cluster is considered, the dietary behaviour of each chronological group shows similar dietary patterns. In order to compensate for the lack of reliability of certain sets of data, modelling seems to be a relevant approach.

In agreement with the hypotheses underlying our models, we were able to compare MOIS3 modern humans with the two other hominid groups (MOIS3 Neanderthals & MOIS2 modern humans), even if hominid isotopic values for MOIS3 modern humans considered in simulations were...
not associated with the faunal isotopic values. Indeed, results of our modelling show that whatever the conditions considered (hominids with and without associated fauna), the dietary behaviour of each chronological group shows similar patterns.

Concerning the diachronic analysis of past dietary patterns, our study demonstrates the absence of significant differences between Neanderthal diet and that of contemporaneous modern human and highlights the dietary difference among MOIS2 modern human. These conclusions confirm that resource competition may have occurred during MOIS3 between the two hominid populations living in Europe. Some authors have already suggested competition between the two populations by observing a correspondence between the contraction of the Neanderthal ecological niche and the expansion of the ecological niche of modern humans [9]. However, radiocarbon dates do not provide any cases of geological interstratification (shown by Mousterian, Aurignacian, and transitional assemblages), which would support the contemporaneity of Neanderthals and modern humans [3]. It has been argued that much of Europe was almost empty when the modern human expansion occurred [7, 83]. Thus, given the radiocarbon dates, little contact would occur between the two populations in Europe, except in the Southwest of France and in the North of Spain, where encounters may have been more frequent [83]. Therefore, resource competition would have happened only in these areas of contact.

The divergence of MOIS2 modern human behaviour in relation to MOIS3 populations (Neanderthals and modern humans) may be explained through use by the former population of alternative food sources, such as, for example, small fauna. Our results are consistent with the observation of certain prehistorians, who suggest a relative continuity in behavior between Neanderthals and contemporaneous modern humans and a behavioural modification between Aurignacian and Gravettian modern humans [84, 85]. Most publications on modern human dietary spectra have underlined that this population had a more diverse diet than Neanderthals, consuming, beside ungulates, small game prey like fish or small mammals [19, 20, 24, 86–88]. Due to the lack of data, stable isotopic signatures of the small prey were not considered in our simulations, and it is therefore difficult to reach any conclusions concerning potential modifications in the consumption of this type of prey. Nevertheless, if small prey have similar isotopic signatures to those of reindeer, bovids, and horse used here, our results would explain the difference in patterns observed for MOIS2 modern humans (with a higher contribution of source 2) in comparison to MOIS3 Neanderthals and MOIS3 modern humans. On the contrary, the predominance of source 1 in MOIS3 Neanderthals and MOIS3 modern humans could indicate a consumption of food with high $\delta^{15}N$ values and low $\delta^{13}C$ values. Source 1 considered in our simulations was derived from fish isotopic values, although other species with similar isotopic value, such as mammoths [19], may have been taken into account and used instead of fish.

Furthermore, zooarchaeological and genetic studies indicate a decline in the quantity of big game and notably ungulates, beginning at about 50,000 years BP [89–91]. On the basis of our results, Neanderthals did not change their diet during MOIS3, as Richards and Trinkaus have already suggested in 2009 [92]. Prior to the early modern human expansion, it is possible that Neanderthals, which had a small population size [93–98], were not affected by a reduction of the population size of large mammals. By contrast, during the period of coexistence of Neanderthals and modern humans, the MOIS3, we can suppose that faunal contraction, associated with resource competition, might have had more serious consequences. Indeed, Neanderthals were probably close to their carrying capacity due to the decrease of the ungulate populations [99], as is suggested by certain stress markers (dental hypoplasia, etc.) [100, 101]. In such conditions, the arrival of species evolving in a same ecological niche might have led to strong competition and perhaps contributed to the demise of Neanderthals, although competition alone cannot account for Neanderthal extinction. Some authors suggest that early modern humans would have possessed more varied technical abilities [10, 102–104], were able to adjust their hunting toward a more varied prey [5], and would have required less food and fewer foraging returns [105, 106]. All of these factors may have favoured the modern humans’ survival [10].

In conclusion, our study has adopted the hypothesis that all the hominids we have considered consumed the same kind of resources. Provided that this hypothesis does not involve a major bias, our methodology, based upon modelling, has permitted us to address the question concerning resource competition between Neanderthals and modern humans, to which zooarchaeological approaches, in view of the complexity of faunal assemblages, could not supply a clear answer [5]. Thanks to a substantial isotopic database, this study confirms the occurrence of resource competition between Neanderthals and contemporaneous modern humans living in the same area. It is however not possible to reach a conclusion concerning the role of competition in Neanderthals’ demise. Furthermore, this study underlines the dietary changes that occurred during MOIS2, as proposed by the study of lithic industry and archaeological data of faunal assemblage. It also demonstrates that modelling approaches and dietary assessment are useful for investigating ecological interaction among both present and past populations. In order to answer this last question, we are engaged in a study in progress, which uses complex mathematical models to represent, as plausibly as possible, the trophic web of Neanderthals and the interaction between prey, predators, and hominid groups (Neanderthals and early modern humans).

Acknowledgments

The authors would like to thanks the anonymous reviewers for the careful reading and suggestions, Terry Brown (University of Manchester), Jocelyn Williams (Trent University), and Stephan Naji for helping in English editing.

References

M. Schoeninger, "Changes in human subsistence activities from..."  
[18] H. Bocherens, D. Billiou, A. Mariotti et al., "New isotopic..."  
[34] B. Hayden, B. Chisholm, and H. Schwarzc, "Fishing and Foraging Marine Ressources in the Upper Paleolithic of..."


Review Article

A Chronological Perspective on the Acheulian and Its Transition to the Middle Stone Age in Southern Africa: The Question of the Fauresmith

Andy I. R. Herries

Australian Archaeomagnetism Laboratory, School of Historical and European Studies, Faculty of Humanities and Social Sciences, La Trobe University, Melbourne, VIC 3086, Australia

Correspondence should be addressed to Andy I. R. Herries, andyirherries@gmail.com

Received 8 November 2010; Revised 25 January 2011; Accepted 27 March 2011

Academic Editor: Parth Chauhan

Copyright © 2011 Andy I. R. Herries. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An understanding of the age of the Acheulian and the transition to the Middle Stone Age in southern Africa has been hampered by a lack of reliable dates for key sequences in the region. A number of researchers have hypothesised that the Acheulian first occurred simultaneously in southern and eastern Africa at around 1.7-1.6Ma. A chronological evaluation of the southern African sites suggests that there is currently little firm evidence for the Acheulian occurring before 1.4 Ma in southern Africa. Many researchers have also suggested the occurrence of a transitional industry, the Fauresmith, covering the transition from the Early to Middle Stone Age, but again, the Fauresmith has been poorly defined, documented, and dated. Despite the occurrence of large cutting tools in these Fauresmith assemblages, they appear to include all the technological components characteristic of the MSA. New data from stratified Fauresmith bearing sites in southern Africa suggest this transitional industry maybe as old as 511–435 ka and should represent the beginning of the MSA as a broad entity rather than the terminal phase of the Acheulian. The MSA in this form is a technology associated with archaic H. sapiens and early modern humans in Africa with a trend of greater complexity through time.

1. Introduction

In the most recent reorganisation of the Pleistocene period (2.58 Ma–0.01 Ma [1]), the Ionian is defined as a geological stage between the Brunhes-Matuyama boundary at 781 ka (end of the Calabrian stage 1.81–0.78 Ma) and the beginning of Marine Isotope Stage 5 interglacial period at 126 ka (beginning of the Upper Pleistocene 126–11.7 ka). Just prior to this is a period termed the Mid-Pleistocene Transition (or revolution [2]). This is the transition from what is known as the 41 ka world to the 100 ka world, essentially a switch from 41 ka to 100 ka cyclicity in glacial cycles [2]. This led to major environmental changes in Africa from about 1.0 Ma to 700 ka [2]. The Ionian is also a period when a broad group of potential modern human ancestors attributed to “archaic Homo sapiens” (or more specifically to Homo heidelbergensis, Homo helmei, or Homo rhodesiensis) evolved to become the first anatomically modern humans. Noonan et al. [3] suggest that the split between Neandertal and ancestral H. sapiens populations occurred at ∼370 ka and that divergence from a last common ancestor occurred at ∼706 ka. In Europe, H. heidelbergensis fossils such as those from Sima de los Huesos at 530 ka [4] may represent fossils occurring soon after this initial split from a common ancestor.

In Africa, (Figures 1 and 2), the hominin record is more fragmentary and less well dated. The Kabwe (Broken Hill 1) hominin remains from Zambia may represent a similar post-divergence population in Africa, as might the Elandsfontein hominin remains from South Africa; however, the age of these fossils remains uncertain. The fauna from Kabwe is suggested to be broadly similar to Bed III-IV at Olduvai (>1.07 Ma or >780 ka [5, 6]) although its association to the type specimen of Homo rhodesiensis [7] is questionable and some age estimates are as young as 125 ka. The Elandsfontein hominin remains are contemporary with a “Cornelian Land Mammal Age” faunal assemblages in southern Africa [8], which date to between ∼1.1 and 0.8 Ma at other sites (Cornelia-Uitzoek; Buffalo Cave; [9, 10]) and represent
either *Homo erectus* or *H. rhodesiensis*. In contrast, the Florisbad hominin skull (the type specimen of *Homo helmei* [11]) is associated with a “Florisan Land Mammal Age” (see [8] FLMA; ~780–~10 ka) faunal assemblage and MSA technology and is dated to ~259 ka [12, 13]. The oldest FLMA assemblage comes from Gladysvale Cave at between 780 and 578 ka [14]. Currently, the earliest well-dated *Homo sapiens* remains come from Omo and Herto in Ethiopia between 198 and 147 ka [15, 16]. In southern Africa, the oldest modern human remains come from Border Cave, Klasies River Mouth, and Pinnacle Point and are all dated to less than 184 ka [17–20]. The Border Cave 1 and 2 remains could be as old as 184–143 ka based on ESR and depending on their actual provenience, while BC5 is somewhere between 79 and 69 ka [19]. The Pinnacle Point remains are likely between 170 and 91 ka although again, their provenience is uncertain [17, 18]. The oldest Klasies River Mouth remains from layer LBS are likely around 115–110 ka, while the younger SAS layer fossils are likely around 90–75 ka [20]. Fish Hoek Man (Peers Cave P4), once thought to be an early modern human, has recently been radiocarbon dated to 7.8–7.6 kcal yr BP [21].

The Ionian is also a period when we see the transition from the Early to the Middle Stone Age (MSA) and the end of the Acheulian Industry in Africa. This is generally defined as a change from large cutting tools (LCTs) such as hand axes and cleavers to the use of blades, points, and prepared core technology, specifically the Levallois method [22]. An understanding of both the archaeological and hominin transitions in southern Africa has been hampered by a lack of reliable dated sites in the region. The current data suggest that the earliest MSA in both eastern and southern Africa is not associated with anatomically modern humans but with archaic *Homo sapiens* also referred to as *H. helmei* at sites such as Florisbad [13, 23] and Eyasi Springs [24]. Many researchers [25–28] have suggested the occurrence of a number of transitional industries in southern Africa, namely, the Fauresmith and the Sangoan, covering this transition, but they have generally been poorly defined, documented, and dated. The Fauresmith is generally considered to have aspects of both MSA and ESA technology and is most often defined as a terminal phase of the Acheulian [29]. The Sangoan is characterised by a heavy-duty, less-elegant component of picks and steep, sometimes denticulated scrapers as well as LCTs and MSA elements. Both the Sangoan and the Fauresmith have long been considered to mark the beginning of regional cultural specialisations towards the end of the ESA [28]. The Fauresmith, with its MSA-like features, perhaps holds the key to understanding this transition and
the relationship between the archaeological and hominin record of this transitional time period between the ESA and MSA.

2. History of the Fauresmith Industry and Issues of Terminology

Creating an all-encompassing term for the definitive Acheulian artefacts, the hand axe, and the cleaver is in itself a problem. The term “hand axe” alone does not encompass cleavers, while biface does not take account of size, and many cleavers and hand axes are unifaces, being made from a single large flake and unretouched on the dorsal face. Large cutting tools (LCTs) is perhaps the best term and has been used throughout to denote hand axes and cleavers of Acheulian character unless referring to a specific type of stone tool.

Goodwin [32] first split the Palaeolithic of Africa into three divisions, namely, the Early, Middle, and Later Stone Age (ESA, MSA, and LSA). Industries such as the Still Bay typified the MSA, where the most prominent artefacts were triangular flakes with convergent dorsal scars and faceted butts (“platforms”). The ESA was characterised by LCTs as seen in the Stellenbosch Culture, first found at the type site of Bosman’s Crossing in Stellenbosch (Western Cape) in 1899 [33]. These deposits were later resampled by Hilary Deacon and James Brink, and there is now an attempt to date them by the latter and this author. The Stellenbosch was originally classified as the LSA). While the type site for the Stellenbosch, while van Riet Lowe [45] thought it was a regional variant related to the presence of indurated shale or that it was cultural, due to the arrival of new people into South Africa. While the Fauresmith may be a regional variant confined to the Free State and Northern Cape regions, surface collections from the site of Hopefield (aka Elandsfontein) were also called Cape Fauresmith. The term Southern Rhodesian Fauresmith has also been used [26], suggesting a wider occurrence of this industry. Moreover, the site of van der Elst Donga, near Vereeniging [46], is in Gauteng and extends the Fauresmith laterally across most of South Africa. Attempts are also being made to date these occurrences by the current author. The lack of Fauresmith industries may also be simply related to deposits of that age not being found in other areas.

In the first paper devoted exclusively to the Fauresmith, Van Riet Lowe [25] elaborated more on this culture. He said “…it represents the culminating phase of the Earlier Stone Age and, incidentally, the transition from core to flake technology... With the greater use of flakes, we come in contact with an evolutionary process that shows marked improvement in technique, the introduction of less clumsy artefacts and the beginnings of a variety of attendant tools that were not in evidence”. Goodwin [32] notes that the Fauresmith hand axes are made on a longitudinal flake. In contrast, while earlier Stellenbosch hand axes are also made on flakes, they are laterally struck. Prime examples of this can be seen in the “early to middle Acheulian” hand axes of Cornelia and the Vaal River (personal observation).

Goodwin [32] notes that in the Fauresmith, the flake is always struck from a point situated, where the butt of the hand axe would eventually lie and that it is the beginning of a true flake industry. Instead of flakes that are struck off during the making of the hand axe being made into unconventional implements, true flake cores make their appearance along with true flake implements with longitudinal trimming, including points. In this earlier description by Goodwin [32], there is no mention of the size of the hand axes, something often taken to define the industry [47]. However, Goodwin and van Riet Lowe [26] later state that the hand axes are generally small and a neat almond, sometimes ovate
shape, and that triangular hand axes are exceedingly rare. Fauresmith hand axes are also stated to have a straight edge, often with an s-shaped twist and differing from the zig-zag edge so often apparent in the Stellenbosch [26]. They further state that the hand axes are far finer than the best seen in the Stellenbosch and that the first beginnings of the Levallois technique are potentially seen [26, 48]. Along with these are discs, scrapers, slightly trimmed flakes, and points [26]. Moreover, Fauresmith flakes are suggested to typically have a faceted platform [26], as in the MSA.

In comparison, the MSA was characterised archaeologically in 1928 [49] by the absence of LCTs that characterise the ESA, an increased prevalence of prepared core reduction strategies, the production of projectile points, and a variety of flake-based tools. The oldest MSA industry was considered to be the Still Bay with its lanceolate points [48], now known to date to ~72 ka [50]. Goodwin [32] also notes similarities between the Fauresmith and some MSA industries, for example, that the Glen Grey Industry (Eastern Cape) directly evolved from the Fauresmith. So, here, Goodwin [32] is noting distinct similarities between the MSA and the Fauresmith but chooses to align the latter with the ESA based on the occurrence of LCTs.

Sampson [51] noted that the Fauresmith has been primarily defined as an industry later than the Stellenbosch and containing smaller, more refined hand axes. However, in 1945, van Riet Lowe [52] suggested that “the progress of man’s skill from Stellenbosch to Fauresmith times is thus not so much measured by a series of successive refinements in his hand axes and cleavers as by an abandonment of old methods and by an improved technical skill revealed in the greater refinement and variety of his waste products in the form of smaller and more elaborately prepared cores and flakes”. Moreover, van Riet Lowe states [52] that it “is not the continued presence of hand axes so much as the integrally associated and now considerably improved [fully developed] Levallois technique” and “while many of the earlier Fauresmith hand axes are smaller and more refined than are their Stellenbosch proto-types, this group of implements is generally neither so beautifully made nor so varied as a representative group from the final Stellenbosch”. Here van Riet Lowe sets out the differences between the Stellenbosch and the Fauresmith, which he [52, 53] divided into three stages.

In comparing the ESA and MSA in 1946, Goodwin [27] states simply that the ESA (meaning the Acheulian) is defined by the hand axe, while the MSA is defined by an accentuation of Levallois (notably used by Tryon et al. [54] to define the MSA) and preferably triangular or sector-shaped points. Goodwin suggests that while we are dealing primarily with techniques, each of these techniques is foreshadowed in the previous phase. We only reach each new “age” as the new technique becomes dominant and replaces previous modes [27]. It is not clear which dominates an interpretation in this scheme, the presence of LCTs or the refinement of new techniques. As such, this leans towards the idea of true intermediate periods as proposed by Clark in 1965 [55]. In 1954, Clark [56] noted the similarity of the Kalambo Falls material to the Fauresmith of South Africa. By 1964, Clark [57] had begun to see the Fauresmith as essentially a southern African (with pockets in Eastern Africa) entity that was in essence the survival of earlier Acheulian traditions, south of the Limpopo. In central and central southern Africa (i.e., Zambia), the addition of heavy chopping tools and small denticulates was termed the Sangoan. It seems possible that the Sangoan as described by Clark at Kalambo Falls represents a transitional Acheulian to MSA industry that occurred in more forested areas, whereas the Fauresmith occurred in more open savannah and thornveld environments. Clark suggested that the Sangoan/Fauresmith represents an intermediate period between the final Acheulian and the MSA, which he terms the “first intermediate period”. At the 1955 Panafican Congress, the term ”First Intermediate Period” was adopted to describe this transition period between the ESA and MSA, but the term was then dropped at the Burg Wartenstein symposium of 1965 [58].

Based on his work in the Transvaal, Mason [39] suggested that all the hand axe cultures and developed Oldowan be assumed under the term Acheulian. Mason [39] notes that he “discontinues the use of the term Fauresmith for the last known phase of the Acheulian, to the term MSA and abandoning the term Fauresmith for the final phase of the ESA. As Inskeep notes in 1969 [59], the trend at that time was for a reduction of names referring to different regional variants of stone tool industries. That is, incorporation of the Stellenbosch into the Acheulian and removal of the Victoria West into a technique rather than an industry. But he also notes that “while hand axes and cleavers maybe characteristic they are by no means the only tools of the Acheulian and not the most numerous”. Going further, Humphreys [60] argues that all the tool types found within the Fauresmith are also found in the Acheulian.

Humphreys [60] notes that the distribution of Fauresmith and Stellenbosch sites raises two important points. The first is that most of the Stellenbosch sites are associated with the Vaal River area, while all but two of the Fauresmith sites are well away from that river, perhaps suggesting differences related to activities, raw materials, or even survival in different geological contexts. Humphreys [60] also notes that the correlation of Stellenbosch and Fauresmith occurrences with the different types of raw materials suggested by the distribution of sites seems too clear to be accidental. Humphreys [60] tries to take consideration of this and cites Mason’s [61] work at the Cave of Hearths as an example of a late Acheulian industry contemporary with the Fauresmith that suggests the latter is a regional or raw material-specific variation. Indeed Mason [61] originally called the assemblage Fauresmith, in part based on the occurrence of small hand axes but later changed it to Acheulian [39]. While the Fauresmith occurrences have been shown to be clearly
stratified above Acheulian deposits, Humphreys believes this is perhaps due to a lack of recovery or survival of artefacts in the gravel contexts compared to the sand contexts as at Canteen Kopjie. Excavations by Gibbon [62] and Leader [63] have shown that the small fraction is preserved in Acheulian bearing gravel deposits in the Vaal River, so this question can now begin to be tested with other recent Fauresmith excavations at Canteen Kopjie, Wonderwerk, and Kathu Pan, as will be discussed below.

Returning to van Riet Lowe, in 1945 [52], he suggested that “throughout the Fauresmith we have three principal types of cores, the majority with prepared striking platforms: (a) circular or tortoise cores, (b) triangular flake-cores, and (c) rectangular blade-cores. Giant blades are also noted (foot long). The makers of Fauresmith tools were not only masters of a fully developed Levallois technique but specialized in flake tools of considerable variety. In the final stages the hand-axe became less and less important and was gradually superseded and replaced by other types of tools until...we cannot say whether we are at the end of the Earlier or at the beginning of the Middle Stone Age.” “This final Fauresmith is literally a period of transition and is distinguished only by the presence of hand axes; the remaining tools and debitage being completely Middle Stone Age in form and fineness”. In 1952, van Riet Lowe [53] further suggested that the Fauresmith showed “the earliest local use of the mounted tool in the form of a spear”, something most researchers would equate with the MSA, not the ESA.

Here, van Riet Lowe [53] appears to be stating that the only difference between the late Fauresmith and the MSA is the continued presence of LCTs, and this could be used to suggest that the Fauresmith should be included within the MSA rather than the ESA and that an industry should be defined on new technology rather than the continued use of old technology. The idea that transitional industries such as the Fauresmith and Sangoan have more in common with the MSA is a view since put forward by Davies [64], Van Peer et al. [65], and Beaumont and Vogel [47], the latter of which suggests the Fauresmith should be termed early MSA (EMSA) and assemblages without LCTs later MSA (LMSA). In van Riet Lowe’s 1937 [45] definition of the 3 phases of the Fauresmith, the 1st phase was described as having “hand axes, cleavers, crude scrapers and cores that yielded long, narrow flake”. This is not dissimilar from the Acheulian that we now know and likely represents the Fauresmith first occurs in southern Africa around 1.6 Ma coeval with its occurrence in Eastern Africa. While a review of the dating makes this possible, the refinement of the dating is perhaps too coarse to currently make this statement. Kuman and Clarke [77] have argued for the Acheulian from Sterkfontein Member 5c being as early as 1.7–1.4 Ma based on faunal comparisons. However, a recent comparison of fauna, palaeomagnetic, electron spin resonance (ESR), and uranium-lead (U-Pb) age estimates [67, 78, 79] have suggested that the Acheulian deposits are more likely dated to between 1.3 and 1.1 Ma [80]. The robust nature of the ESR ages is supported by identical age ranges for Member 4 (2.6–2.0 Ma) based on independent ESR, palaeomagnetism, and uranium-lead analyses [79–81] as well as geochronological comparisons at other sites in the cradle (see below). Moreover, the StW 53 infill (aka Member 5a) at Sterkfontein has been shown to be older than the Member 5b and 5c deposits based on stratigraphy and geochronology [79, 81, 82]. The StW 53 infill has been dated to <1.78 Ma based on a U-Pb age for speleothem that formed before the deposition of Member of the Acheulian within the Fauresmith should be the most important factor in (a) defining it as its own transitional entity and (b) including it within the MSA rather than the Acheulian. It is quite possible that some of the Fauresmith being described is material that has since been defined as early MSA. Only a chronological analysis of these sites to assess their contemporaneity would help resolve these issues.

3. The Early Acheulian in Southern Africa

To fully understand the context and difference of the Fauresmith, an overview of the Acheulian record in southern Africa is first needed. Well-dated Acheulian sites are still few and far between in southern Africa mostly due to the difficulty of dating caves and river sequences, where much of this material is preserved. Acheulian sites often occur on deflated landscapes rather than in well-stratified and datable contexts and even then reliable means of dating such contexts were unavailable until recently. Moreover, many Acheulian assemblages come from reworked contexts rather than primary occupations. The most extensive Acheulian bearing deposits are the numerous paleocave sites within the Cradle of Humankind World Heritage Area in Gauteng (CoHK, e.g., Sterkfontein, Swartkrans, and Kromdraai [66, 67]), donga sequences of the Free State (e.g., Vereeniging and Cornelia [10, 38]), the Vaal River sites (e.g., Pneil and Canteen Kopje [68]), a series of Pan deposits around Kimberley (e.g., Kathu Pan and Rooidam [69, 70]), dune sequences of the western Cape coast (e.g., Elandsfontein and Duinefontein II [71]), and a series of isolated caves occurring throughout the country (e.g., Wonderwerk Cave in Northern Cape, the Cave of Hearths in Limpopo Province, and Montagu Cave in Western Cape [40, 47, 72]). Other localities include Amanzi Springs in Eastern Cape [73] and a series of sites in the Mapungubwe National Park along the Limpopo River (Limpopo Province; [74]).

Three phases of the Acheulian have been defined based on typology [45]. Kuman [66], Chazan et al. [75], and Gibbon et al. [76] have all suggested that the early Acheulian first occurs in southern Africa around 1.6 Ma coeval with its occurrence in Eastern Africa. While a review of the dating makes this possible, the refinement of the dating is perhaps too coarse to currently make this statement. Kuman and Clarke [77] have argued for the Acheulian from Sterkfontein Member 5c being as early as 1.7–1.4 Ma based on faunal comparisons. However, a recent comparison of fauna, palaeomagnetic, electron spin resonance (ESR), and uranium-lead (U-Pb) age estimates [67, 78, 79] have suggested that the Acheulian deposits are more likely dated to between 1.3 and 1.1 Ma [80]. The robust nature of the ESR ages is supported by identical age ranges for Member 4 (2.6–2.0 Ma) based on independent ESR, palaeomagnetism, and uranium-lead analyses [79–81] as well as geochronological comparisons at other sites in the cradle (see below). Moreover, the StW 53 infill (aka Member 5a) at Sterkfontein has been shown to be older than the Member 5b and 5c deposits based on stratigraphy and geochronology [79, 81, 82]. The StW 53 infill has been dated to <1.78 Ma based on a U-Pb age for speleothem that formed before the deposition of Member
5a and the reversed magnetic signal of this deposit [79]. ESR ages are again consistent with this and taken together an age between 1.8 and 1.5 Ma is most likely [67, 79]. Given this data, there is no definitive reason based on the current data to discount the ESR ages for the Member 5c Acheulian.

O’Regan and Reynolds [83] suggest that the diversity of carnivore species in the deposit indicate that it is a palimpsest formed over a long period of time. The spread of ESR ages may support this suggestion. Given this suggestion and taking the maximum age range for all the ESR ages into account, rather than a weighted mean age as per [79], a maximum age range of 1.39–0.82 Ma is suggested for the Acheulian, and so, it is not older than 1.4 Ma (The M5b deposit has a maximum ESR range of 1.62–0.83 Ma (minus a single age with an extremely large error) and suggests that the Oldowan is not older than 1.6 Ma at Sterkfontein). The age for the Acheulian is at the lower age range of Kuman and Clarke’s [77] estimate. While some fauna from Sterkfontein Member 5 suggests an older age, the fauna collections are likely quite mixed due to numerous years of excavation into the deposits (see [79]) without understanding their complexity. Vrba [84] previously suggested the fauna from Member 5 was likely around 1.5 Ma, and a reanalysis of the carnivores from Member 5c has suggested that it is likely younger than previously suggested [67, 85]. O’Regan [85] suggests that the only older element in Member 5c, Dinofelis barlowi, likely does not come from Member 5. O’Regan [85] notes that a number of archaic carnivores (Megantereon and Chasmaporithetes) are present in Swartkrans Member 1 but not in Sterkfontein Member 5 despite the fact that they are generally considered contemporary. Herries et al. [67] suggest that Swartkrans Member 1 maybe around 2.1–1.9 Ma, and so, this may be a temporal difference. Pickering et al. [86] have recently suggested that the Hanging Remanent at Swartkrans is dated to sometime between 2.2 and 1.8 Ma. These temporal changes may explain some of the O’Regan [85] hypotheses that this might be related to taphonomic differences, different ages or different environments being sampled. Given that the caves are across the valley from each other the environment would likely be similar if of the same age. Taphonomic reasons can never be ruled out; however, given the new younger ESR ages for Member 5b/c (1.4–1.1 Ma) and the suggested older age for Swartkrans Member 1 [67, 86], this difference, and differences noted by Reynolds [87], is likely to be partly a reflection of age.

Kuman and Clarke’s [77] age assessment is also in part due to the fact that the LCTs from Sterkfontein M5c are both limited in number and very unrefined. The earliest date for the Acheulian is at the lower age range of Kuman and Clarke’s [77] estimate. While some fauna from Sterkfontein Member 5 suggests an older age, the fauna collections are likely quite mixed due to numerous years of excavation into the deposits (see [79]) without understanding their complexity. Vrba [84] previously suggested the fauna from Member 5 was likely around 1.5 Ma, and a reanalysis of the carnivores from Member 5c has suggested that it is likely younger than previously suggested [67, 85]. O’Regan [85] suggests that the only older element in Member 5c, Dinofelis barlowi, likely does not come from Member 5. O’Regan [85] notes that a number of archaic carnivores (Megantereon and Chasmaporithetes) are present in Swartkrans Member 1 but not in Sterkfontein Member 5 despite the fact that they are generally considered contemporary. Herries et al. [67] suggest that Swartkrans Member 1 maybe around 2.1–1.9 Ma, and so, this may be a temporal difference. Pickering et al. [86] have recently suggested that the Hanging Remanent at Swartkrans is dated to sometime between 2.2 and 1.8 Ma. These temporal changes may explain some of the O’Regan [85] hypotheses that this might be related to taphonomic differences, different ages or different environments being sampled. Given that the caves are across the valley from each other the environment would likely be similar if of the same age. Taphonomic reasons can never be ruled out; however, given the new younger ESR ages for Member 5b/c (1.4–1.1 Ma) and the suggested older age for Swartkrans Member 1 [67, 86], this difference, and differences noted by Reynolds [87], is likely to be partly a reflection of age.

Kuman and Clarke’s [77] age assessment is also in part due to the fact that the LCTs from Sterkfontein M5c are both limited in number and very unrefined. The earliest date for the Acheulian is often quoted as ~1.7 Ma at locality KGA of Konso Gardula in Ethiopia [88]. However, the earliest Acheulian deposits occur in deposits dated to somewhere between 1.69 and 1.41 Ma (1.66 ± 0.03 Ma and 1.43 ± 0.02 Ma; [89]). The first Acheulian tools occur closer to 1.4 Ma than 1.7 Ma based on the stratigraphy and the occurrence of a stratigraphic break between the hand axe horizons and the basal age [90]. While the KS4 site from the Nachukui Formation (West Lake Turkana, Kenya) is suggested to have an age slightly older than 1.65 Ma [91], the majority of other early Acheulian sites in Eastern Africa are generally dated to less than 1.5 Ma including Olduvai Gorge, middle and upper Bed II (1.53–1.27 Ma; [92]), to which many of the southern African early Acheulian deposits are often compared [93]. As such, there seems little typological basis for the Sterkfontein Member 5 Acheulian being at the older part of the 1.7–1.4 Ma age range as suggested by Kuman [66, 93]. Moreover, it is not unreasonable to assume that the Acheulian may first occur later in southern Africa than it does in Eastern Africa, as is the case for other parts of Africa or the Levant, where it does not occur until somewhere between 1.5 and 1.2 Ma [94].

Acheulian tools are also known from Swartkrans Member 2 and 3 dated to sometime between 1.65 and 0.6 Ma [67]; however, Swartkrans remains one of the least well-dated site in the CoHK and many of the other sites have relatively undiagnostic stone tool assemblages [66]. Many, including Swartkrans Member 1, have been classified as Developed Oldowan or early Acheulian based on the fact that many have no LCTs, but a large flake size than the preceding Oldowan [66, 95]. However, the recent dating suggests that parts of Swartkrans Member 1 are as old as 2.2–1.8 Ma [67, 86], making an assessment of these stone tools as Acheulian perhaps less likely [67]. That being said, Swartkrans Member 1 may be more complex than previously noted with multiple in-fills of more than one age.

Gibbons et al. [76] have suggested that Acheulian stone tools in the Rietputs (Rietputs 15) formation of the Vaal River are also dated to ~1.6 Ma. These deposits have maximum ages between 2.08 and 1.12 Ma (1.89 ± 0.19 Ma and 1.34 ± 0.22 Ma) and minimum ages between 1.88–1.08 Ma (1.72 ± 0.16 Ma and 1.29 ± 0.21 Ma) suggesting that the LCA gravel deposits date to between ~2.1 and ~1.1 Ma. The LCA gravel deposits are up to 4 m deep in some of the pits sampled and up to 7 m thick overall and appear to have accumulated over perhaps a million years. The stone tools collected from an undated pit and were collected out of context from mining debris piles and sporadically from an excavating conveyor belt over two days of 24 hour operation [76]. As such, the Rietputs Acheulian described by Gibbon et al. [76] could come from anytime during the deposition of the LCA deposits. The later collected Acheulian assemblage described by Leader [63] comes from Pit 5, and so dates to somewhere between 1.63 and 1.11 Ma (1.43 ± 0.23–1.32 ± 0.21 Ma). Note that ~1.6 Ma is at the extreme upper end of this age range, and a more likely age for this assemblage is perhaps 1.4–1.3 Ma. A further collection described by Leader [63] came from an in situ excavation with no current ages. However, the stratigraphy of each of the pits and dating sample location compared to depth of the LCA is never presented in either Leader [63] or Gibb et al. [76], making it difficult to confidently relate the ages to the artefacts collected beyond the age for the LCA as a whole, 2.0–1.1 Ma.

In Pit 2, the top of the LCA was dated to somewhere between 1.42 and 1.08 Ma (1.27 ± 0.15–1.22 ± 0.14 Ma), and the overlying UFA has been dated to somewhere between 1.45 and 1.01 Ma (1.26 ± 0.19–1.16 ± 0.15 Ma) suggesting that the LCA deposits are more likely older than ~1.2 Ma. For some reason, Gibbon et al. [76] and Gibbon [62] included
the sample from the top of the LCA [72, Table 1] within the overlying UFA when calculating the mean burial age for the LCA deposit. This will have had the effect of very slightly increasing the mean maximum age estimate for the LCA deposits from 1.53 to 1.57 Ma. Moreover, this is the maximum mean age for the deposits. Based on erosional history, Gibbon et al. [76] suggest the true age is likely closer to the maximum age estimate; however, this is by no means certain, and an age between the maximum and minimum age estimates is also perfectly possible. As such, some Acheulian deposits from the Vaal River could be as old as 1.6 Ma or older, but the age of the collections described by Leader [63] is closer to ∼1.4–1.3 Ma.

Cosmogenic burial ages for the base of the Canteen Kopjie Acheulian bearing lower coarse alluvium (stratum 2b) range between 1.25 and 0.82 Ma (1.06 ± 0.19 Ma and 1.00 ± 0.18 Ma) in one pit and 1.61 and 1.23 Ma (1.46 ± 0.15 Ma and 1.37 ± 0.14) in a second pit [62]. Gibbon [62] notes that these ages suggest that deposition of the gravels was very complex on a local scale (within 40 m), and as such, ages from one location may not represent the age of deposits on a locality scale. As such, for reliable dates of assemblages from the gravels, the levels need to be dated directly in each case, and as the collection described by Gibbon et al. [76] comes from an undated pit, the age of the stone tools is much less certain.

Chazan et al. [75] also suggest an age of ∼1.6 Ma for the beginning of the Acheulian at Wonderwerk butagain this age is by far from secure. The Acheulian first occurs in sediments (strata 11) that are dated to somewhere between 1.78 and 1.07 Ma based on palaeomagnetism and a basal cosmogenic isotope burial age [75]. A number of depositional breaks appear to occur in the sequence, notably between the layers recording the palaeomagnetic transition at the end of the Oldowan (1.78 Ma; strata 12) and the layers containing the Acheulian (strata 11). This may suggest the material is closer to 1.07 Ma than 1.78 Ma. The Vaal River and Wonderwerk Cave early Acheulian may date to as old, or in theory older than 1.6 Ma, but could be closer in age or contemporary with the Acheulian from Sterkfontein M5c at 1.4–1.1 Ma. As such, there is little definitive evidence for the Acheulian in South Africa older than 1.4 Ma. Hominin fossils are rare, but these earliest Acheulian occurrences are normally associated with Homo ergaster [77].

Recent palaeomagnetic studies at the Cornelia-Uitzoek locality have shown that multiple levels of Acheulian bearing deposits occur there between 1.07 and 0.78 Ma [10]. The oldest Acheulian layers at Elandsfontein may also be of a similar age, and Klein et al. [5] suggest that Elandsfontein dates to somewhere between 1.0 and 0.6 Ma. However, Klein et al. [5] suggest that the site most likely dates closer to 600 ka based on the presence of Rabaticeras and Pelorovis antiquus, the hominin remains, and the typology of the Acheulian artefacts. Klein et al. [5] suggest that the extinct alcelaphine Rabaticeras likely gave rise to the extant hartebeest genus, Alcelaphus which has a first appearance date close to 600 ka, based on its occurrence at Bodo and that the extinct long-horned buffalo Pelorovis antiquus has a first appearance at or soon after 1 Ma based on its evolution from Pelorovis oldowayensis after the deposition of Olduvai Bed IV. Bodo is dated to between 1.6 Ma and 2.3 Ma [91]. However, the Bodo material shows a lot of affinities to material from Olduvai bed IV at >0.78–1.07 Ma [6, 96]. Moreover, Rabaticeras itself actually occurs at Buffalo Cave and Cornelia-Uitzoek, both of which are dated to between 1.1 and 0.8 Ma [9, 10]. The remainder of the fauna from Elandsfontein is also similar to that from Cornelia-Uitzoek and Buffalo Cave, representing the Cornelian Land Mammal Age [10]. Also, Klein et al. [5] also note that the occurrence of Sivathere and the dirk-toothed cat which would have to represent the youngest occurrences of these species by some margin. This suggests that there is little evidence for Elandsfontein being younger than at least 0.8 Ma.

Also, in Western Cape, the Duinefontein II fauna also belongs to the Cornelian Land Mammal Age (Brink, pers. comm.). The site also contains an Acheulian estimated to date to between 347 and 217 ka [71, 97] based on a combination of thermoluminescence (TL), optically stimulated luminescence (OSL), and infrared-stimulated luminescence (IRSL). This is extremely young and of a similar age to some of the earliest MSA industries in South Africa (see below). Moreover, the Cornelian Land Mammal Age is followed by the Florisian Land Mammal Age, which is currently dated back to sometime between 780 and 578 ka [14] at Gladysvale Cave. This, along with the faunal similarities to Elandsfontein, Buffalo Cave, and Cornelia-Uitzoek, may suggest that Duinefontein II should, in fact, be older than 800 ka, at odds with the luminescence age estimates. If the luminescence age estimates are correct, then this would indicate marked regional variation in the timing of the transition from the Cornelian to the Florisian Land Mammal Age and the Acheulian to the MSA. The Acheulian certainly lacks elements of the MSA as seen in Fauresmith assemblages despite Elandsfontein once being called Cape Fauresmith based on the small size and shape of some of its hand axes. However, Feathers [20] outlines the problems with the dating samples from this site. Based on the problems of variable dose rates and shifting dunes through time at this locality, these age estimates maybe underestimations of the true age of the deposits and should perhaps be regarded as minimum ages. These ages were also performed on feldspar whose luminescence signal is known to fade with time, causing younger age estimates [98] and redating using post IR-IRSL could be attempted. However, given the faunal age estimates, the site is likely also beyond the limit of this method. Future potential for the dating of the west coast sites comes in the form of thermally transferred OSL (TT-OSL), which can date back to at least 780 ka if the correct conditions occur [99].

Gladysvale Cave has also yielded a single hand axe dated to older than 780 ka [100, 101]. The age of the Bed 1–3 Cave of Hearths Acheulian is still a matter of debate, but it is certainly younger than 780 ka based on palaeomagnetism [102]. McNabb [103] suggests that the Cave of Hearths assemblage is lacking elements suggestive of the Fauresmith or MSA such as Levallois and other prepared core technology points and blades. A direct ESR date on the mandible from Bed 3 suggests and age of 600–400 ka however there are issues
with the dosimetry for this sample [104]. Given this ESR age and based on the age of confirmed geomagnetic field events [105], a short geomagnetic reversal in the basal deposits [101, 102] most likely dates to either the Big Lost event dated to between 580–560 ka or the Stage 17 excursion at ~670 ka. As such, the Bed 1–3 Acheulian from the Cave of Hearths, including the hominin remains, may date to the period between 700 and 400 ka but most likely towards the older end of this age range. The fossils associated with these sites are often referred to as H. rhodesiensis, H. heidelbergensis, H. erectus, or simply archaic H. sapiens [29]. Taking H. erectus as an Asian origin species and H. heidelbergensis as a European origin species, the use of H. rhodesiensis is perhaps most appropriate for much of the African hominin record (as per [22, 23]) of this time period. However, designation to a specific species should obviously be undertaken on a fossil by fossil basis, as there is always the possibility that species did not just leave Africa but that H. erectus and H. heidelbergensis may have migrated into Africa at various times.

4. The ESA-MSA Transition in Eastern Africa

In the older K3 deposits of the Kapthurin Formation (GnJh-03, GnJh-17; <518 ka) Deino and McBrearty [106] document the use of LCTs, Levallois, and blades and a notable feature is the use of the Levallois technique to make hand axes. Johnson and McBrearty [107] describe a further small set of blades from the base of the K3 deposit (GnJh-42 and GnJh-50) at between 548 (545 ± 3 ka) and 500 ka (509 ± 9 ka) and suggest that the move to blade technology is a feature of the Acheulian period. While the blades are standardised Johnson and McBrearty [107] suggest the earliest blades are not made by the Levallois method. Refinement of the K4 sequence by Byron and McBrearty [108, 109] indicates that >272 ka (284 ± 12 ka) LCTs (cleavers) and Levallois technology occur at the same sites in the southern area, while LCTs produced from Levallois cores (e.g., at LHA) and points occur together in slightly older deposits in the northern area (e.g., at GnJh-17). By ~200 ka (perhaps 250 ka) the site of Koimilot (also in the Kapthurin Formation) shows a range of Levallois flake reduction strategies including convergent Levallois flakes and points and an absence of LCTs [109, 110]. Slightly further north between the Turkana and Baringo basins, much younger MSA assemblages occur in the Kapeto Tuffs [111]. These sites illustrate that by 135–120 ka (132 ± 3 ka and 123 ± 3 ka) a wide range of Levallois flake-based reduction techniques are being undertaken within the MSA [54]. As such, Levallois blade and point technology characteristic of the MSA occur at the same time as characteristic LCTs of Acheulian character. This sequence suggests the early occurrence of blades within the Acheulian before 500 ka. Points begin to occur towards the end of the K3 deposits sometime between 518 and 272 ka. At sites within the base of the K4 deposits, LCTs made from Levallois flakes occur at sites such as LHA and GnJh-17 [106]. Soon after this centripetal Levallois, flakes and cores occur at sites such as Rorop Lingop [106]. LCTs in the form of cleavers are then last seen in the middle of the K4 deposits at some time before 272 ka [106]. After this (~250–200 ka), at the site of Koimilot a range of Levallois flaking occurs including the presences of convergent Levallois flakes and points [108, 109]. Tryon [110] uses this change in Levallois flaking behaviours at Koimilot as the beginning of the MSA despite tools of MSA-like character occurring earlier in the sequence. McBrearty [22] suggests that this sequence indicates an independent evolution of blades, points, and Levallois technology during the transition from the Acheulian to the MSA and that it does not occur as a package of behaviour evolving together. As such, this causes a major issue in defining exactly what the MSA is and when it might begin. As such, the diversification in Levallois technology is suggested as the defining features of the end of the ESA and beginning of the MSA around 250–200 ka in the Kapthurin Formation.

Malewa Gorge in Kenya is another site where MSA (Kenyan Still Bay) has been recovered from beneath a tuff that was originally dated to 240 ka [112]. Clark [55] suggested that if the Still Bay was older than 240 ka the tuff samples dated must have been contaminated. An attempt at redating this tuff by Morgan and Renne [113] gave an age of 102 ± 16 ka, much younger than the original age. Pseudo-Still Bay artefacts were also dated to <557 and >440 ka at Wetherill’s and Cartwright sites and suggests that transitional industries could potentially occur here as early as half a million years [112].

Morgan and Renne [113] describe the dating of MSA and perhaps transitional ESA/MSA localities within the Gademotta Formation in southern-central Ethiopia. At the type site unit 9 of Laury and Albritton [114] lies beneath unit 10 which has been dated to sometime between 280 and 272 ka (276 ± 4 ka; G3; [105], and so, the earliest MSA is older than 272 ka. Unit 10 at the nearby site of Kulkuletti is dated to 280 ± 8 ka, but this age is less precise [113]. However, an age of 193–173 ka (183 ± 10 ka) from here provides an age for younger MSA deposits [113]. Small Acheulian hand axes are found at the base of unit 9 [115], while from the upper part of Unit 9 and upwards, only MSA artefacts are found. These MSA artefacts include medium to large retouched points and scrapers, some of which were made by the Levallois method and many look quite advanced and like younger MSA assemblages in the region [113]. This is suggested to be in part due to the use of obsidian at these sites when compared to others like Kapthurin [112].

These artifacts display enormous variability in the sense of technology and retouched tool forms and neither the frequency of use of the Levallois technique nor tool size showed consistent change within the sequence (>272–173 ka; [114, 115]). Despite the small hand axes in the base of unit 9 the Gademotta and Kulkuletti, MSA is suggested not to have an underlying Acheulian with characteristics of the Fauxsmith or Sangoan [112].

Clark et al. [16] describe the remains of anatomically modern humans from the Upper Herto Member of the Bouri Formation, Middle Awash in Ethiopia dated to between 166 ka (163 ± 3 ka; MA98–25), and 147 ka (154 ± 7 ka; TG 120) that are associated with LCTs. Along with the Omo 1 fossil, these anatomically modern human remains represent some of the earliest in Africa. It should be noted that while 195 ka is the often cited date of the Omo fossils, in reality,
the fossils date to sometime between 198 and 97 ka (196 ± 2 ka; 104 ± 7 ka) although they are argued to lie closer to the older age [15]. Millard [116] suggests an alternative age range based on the data of 191–98 ka but again stresses that the true age is likely closer to the older value. This could make them roughly contemporary with the Herto fossils rather than significantly older than them. Despite the young age, like the Kaphurin Formation, the Herto assemblage includes some LCTs being made from Levallois flakes. MSA tools dated to around 125 ± 7 ka have also been recovered in association with LCTs and “Acheulian cores” at the site of Abdur in Eritrea [117]. Bruggemann et al. [117] suggest that this transitional industry indicates Acheulian and MSA technologies continued to coexist for much longer and that the Acheulian maybe widespread on the Red Sea coast from Egypt to Djibouti at this period. These two occurrences appear to represent the youngest occurrence of LCTs in Africa, but whether these stone tools are intrusive or re-used Acheulian artefacts remains to be seen. No LCTs were recovered from in situ excavations at Herto and all the LCTs come from controlled surface collections.

The Sangoan (like the Fauresmith) has been classed as a final ESA industry, possibly about 300 000 years old [23] although some researchers would classify it as a transitional industry or even an early MSA [30]. Van Peer et al. [65] describes a transition from the Acheulian, through Sangoan to the Lupemban-like MSA at site 8-B-11, Sai Island, Sudan. Maximum age estimates for the Sangoan deposits is based on OSL dating of the underlying aeolian ES sands between 242–204 ka (223 ± 19 ka [65]). Minimum age estimates are provided by an OSL age from above the Middle Sangoan levels at between 202–142 ka (152 ± 10 ka [65]). Van Peer et al. [65] suggest that this site, along with Kalambo Falls, strengthens the case for identifying the long-debated Sangoan as a valid taxonomic entity for the early MSA, with a geographical distribution extending far beyond the Congo basin. Compared to other “transitional” sites, it is perhaps significantly younger at between 242 and 142 ka, when classical MSA assemblages already occur in eastern and southern Africa. Van Peer et al. [65] state that the OSL ages below the lowest Sangoan layer and above the Acheulian layers were quite scattered. Without publication of this data it is impossible to evaluate the reliability of these ages. In certain circumstances with high background radiation concentrations these ages could be considered as minimum age estimates due to saturation of the OSL signal. In the lowest Sangoan assemblage, hand axes are actually absent, but it also has few flake tools and has been designated primarily on the basis of heavy duty tools such as core axes [65]. The only hand axes that do occur come are two from the Middle Sangoan levels. Rots and Van Peer [118] suggest this is a reflection of core-axe manufacture at the site. The core axes are suggested to have been hafted and represent a complex behavioural system suggestive of economic specialization [118]. Rots and Van Peer [118] suggest that the Sangoan should be maintained as a taxonomic unit and that it represents the earliest stage of the MSA, at least in that part of Africa. Other Sangoan sites in eastern Africa include those from the Lake Eyasi Beds in Tanzania which has an interpreted age of >132 ± 7 ka based on uranium series ages from bones within the Mumba Rockshelter [119]. This age should be taken as a minimum age estimate and its stratigraphic correlation to the Sangoan bearing deposits is also likely questionable. Younger simplistic flake industries from the Northeast Bay of Lake Eyasi, which have also been attributed to the MSA and are dated to between 132 and 82 ka although high thorium concentrations in the samples make the dating less than ideal [24]. These MSA artefacts are interestingly associated with an archaic Homo sapiens skull despite this young age.

5. Earlier MSA Sites in Southern Africa

Until recently, there was little evidence for the period between these Acheulian deposits and the occurrence of widespread MSA bearing caves along the coast of South Africa in the last 120 ka [120]. The oldest MSA sites were Florisbad (279 ± 47 ka [12, 19]), Klasies River Mouth (<120 ka [120]; ∼115–107 ka; [20]), and Border Cave [19]. Florisbad is associated with (H. helmei) and the other two sites with modern humans. More recently a series of new potentially early MSA sites have been identified in southern Africa. Including sites in South Africa, Zambia, and Namibia. The new river terrace site of Cafema, along the Lower Cunene River on the border of Namibia and Angola is noticeable in being in an area where little information was previously known [121]. However, current dating of the site is based on sand layers beneath the archaeology which provides a maximum age estimate of ∼220 ka and as such the MSA could be much younger [121].

5.1. Pinnacle Point. Along the coast the lack of older deposits was hypothesised to be due to the fact that the last interglacial high sea level stand at ∼125 ka (MIS 5) would have eroded out any earlier deposits (e.g., MIS 6) from these coastal cave sites, or they would have occurred on the now submerged coastal platform [122]. However, there was little evidence of this. Recent confirmation of this has come in the form of MSA deposits from the LC-MSA lower horizon of Cave 13B at Pinnacle Point at ∼174–153 ka [18]. These have survived as remnants along the wall of the cave and were not washed out by high sea levels due to the caves elevation and due to the cementation of the deposits by speleothem. Blades, points, and Levallois flakes all occur in the LC-MSA Lower along with frequent bladlets [18]. Blades are more common than points, which have similarities to the sequence described from the undated levels at the Cave of Hearths [123]. Schoville [124] suggests that edge damage is consistent with the use of points as knives rather than projectile points and as a whole the assemblage lacks significant retouch or formal tools. Brown et al. [125] also suggest the presence of heat-treated silcrete lithics in these deposits although they are more numerous during the LC-MSA Middle deposits between 130 and 120 ka [126]. Pinnacle Point has also yielded humans remains [17]; however, their age is uncertain given that they were found out of context. An age of between 170 and 91 ka is a broad estimate for the age of the fossils based on the age of the deposits at the site [126].
5.2. Border Cave. All the *H. sapiens* fossil remains from Border Cave (BC1-5) are older than the Howieson’s Poort layers and are likely older than ~65 ka based on the dating of the Howieson’s Poort at numerous sites across S. Africa [19, 50]. This is supported by the direct dating of the youngest BC5 fossil to between 79 and 69 ka using ESR [19, 127]. The ESR dates further suggest that the BC3 burial is around 84–72 ka [19]. Perforated shells similar to the ones found in the burial are known from Blombos and Sibudu Caves in association with the Still Bay around 76–69 ka [50] and in North Africa with the Aterian around ~82 ka [128]. The BC1-2 remains and the oldest MSA lie somewhere between 179 and 80 ka depending on their provenience, while potentially older and relatively unexcavated MSA layer, 6BS, is perhaps as old as 238–217 ka [19]. As at Sterkfontein (see above), the reliability of the ESR ages at Border Cave are high due to the very low levels of uranium in the samples, which simplifies uptake history modelling, as well as their broad agreement with the radiocarbon ages on younger deposits (ESR ages are very slightly younger than the C14 ages) and amino acid age estimates [19, 129, 130].

5.3. Cave of Hearths. The Cave of Hearths has often been considered to be comparable to the older deposits at Border Cave, being termed part of the Pietersberg Industry, also suggested to occur at Wonderwerk Cave [61]. Unfortunately, the Cave of Hearths remains undated radiometrically. Sinclair [131] notes that hominins practiced PCTs to make blades and convergent points with a small number of these flakes being modified further without clear evidence that they were being retouched to a clear design. Sinclair [131] suggests that this minimal sense of patterning in the MSA is consistent throughout the sequence. However, in the earliest levels, blades generally tend to be much longer, while in the younger layers, retouched points tend to be more numerous. In both cases these pieces are minimal. Beaumont and Vogel [47] suggest that bed 4 here represents a Fauresmith Industry due to the occurrence of two possible LCTs from this layer. However, McNabb [103] and Sinclair [131] note no evidence for a Fauresmith like assemblage at the site.

5.4. Florisbad. Florisbad contains a number of MSA layer sand has also yielded remains of *Homo helmei*, dated by ESR to 259 ± 35 ka (294–225 ka [12, 19]). The oldest MSA bearing layers are Units N-P (Peat 1 and Brown Sand) which range in age between 327 and 208 ka based on a combination of ESR and OSL [19]. The ESR and OSL correlate quite well although with large error ranges (ESR: 259 ± 35 ka/OSL 279 ± 47 ka and 281 ± 73 ka [19]). However, there is a significant scatter in ESR ages for some layers and inversion of ages for both ESR and OSL ages. However, the ESR ages are not completely inverted as stated by Millard [116] the depth scale is inverted in the Grün et al. [12] publication. Grün et al. [12] state that this is due to the complex history of the spring formation and intermixing of both fossils and older and younger quartz. As the OSL dating was multigrain rather than the more modern single-grain dating method this is quite likely. Grün [19] states that the main error is the uncertainty in the reconstruction of the radioactive environ-

ment. Most layers have a very low density of stone tools (15–176 artefacts), the exceptions being Unit F (1654 artefacts), dated to 121 ± 6 ka (127–115 ka) using ESR and 133 ± 31 ka (164–102 ka) and 128 ± 22 ka (150–106 ka) using OSL [13]. As such, the MSA occupation is most likely dated to the beginning of MIS 5. This deposits has a large assemblage of expedient, lightly retouched MSA tool types and utilized flakes with few formal tools and little variability [13].

Florisbad Unit M, has a relatively substantial occupation given the size of the trench at this depth (120 artefacts) and is dated by OSL to 157 ± 32 ka (189–125 ka [13, 19]). As such, this is an MIS 6 occupation, a period when some researchers believe early modern humans were only located along the coast of South Africa [132]. The Florisbad dates and perhaps also those from Border Cave suggest this was not entirely the case, unless these MSA assemblages are being made by remnant but soon to be extinct populations of archaic modern humans. Kuman et al. [13] note that unit M has a proportion of heavily retouched artefacts, mostly side scrapers and a number of retouched points and convergent flake blades. The oldest MSA artefacts are dated to between 327 and 208 ka, and Kuman et al. [13] suggests that the oldest MSA is on a wider range of raw materials and is generally quite informal when compared to the younger material. There is less emphasis on prepared core techniques, and there are no heavily retouched pieces and only one point. However, this is with the caveats that this is only the case if the material excavated by Meiring (see [133]; 75 artefacts plus Kuman et al. [13]) were well provenance and representative. Given the formation of the site the hominin skull likely dates to MIS 7 (~145–190 ka). An upper age range for the MSA at Florisbad is probably the beginning of MIS 8–310 ka. The age of the oldest MSA and the hominin remains from Florisbad are therefore, roughly the same age ranges for the oldest MSA in the Kaphurin and Gademotta Formations (see [106, 113, 115] 280–250 ka) and again older than the anatomically modern human remains from Omo and Herto 198–147 ka [15, 16].

5.5. CoHK. A number of new sites have also been excavated in the CoHK. Lincoln Cave has yielded an MSA assemblage dated by uranium series to somewhere between 278–107 ka (252.6 ± 35.6 and 115.3 ± 7.7 ka; [134]). Reynolds et al. [134] note that Acheulian like cores occur in Lincoln cave and explain their occurrence as mixing from Member 5c rather than the late occurrence of these core types. Contemporary MSA bearing deposits (post-Member 6; [135]) from the main quarry at Sterkfontein have also been dated to between 294–210 ka using ESR (252 ± 42 ka [74]). An archaic *Homo sapiens* fossil has been recovered from here [135]. Significant mixing has taken place and *Homo ergaster* fossils have also been recovered from Lincoln Cave [134]. Given that the post-Member 6 deposits are seemingly continuous with Lincoln Cave [135] these ESR ages may suggest that the Lincoln Cave MSA is closer in age to 278 ka than 115 ka. Taken together, this suggests that MSA tools are at least as old as 210 ka in the Sterkfontein area. Younger MSA deposits have also been recovered from Member 4 deposits at Swartkranx (<110.0 ± 2.0 ka [136]) and between 90.3–61.6 ka (62.9 ± 1.3 ka and 88.7 ± 1.6 ka [137]) at Plovers Lake (both dated by uranium
series on speleothem). A direct single isochron ESR date of 81.2–70.0 ka (75.6 ± 5.6 ka) for the Plowers Lake artefact horizon correlates well with the U-Th age [130]. This data, along with that presented by Herries et al. [67] and Herries and Shaw [79] further illustrates the robust nature of the ESR dating of the CoHK sites with ESR correlating well with other geochronological methods over a time range of 2.6 Ma to ∼80 ka. As Grün [19] notes, these sites are ideal for ESR due to their low uranium content. Such comparisons make it highly likely that ESR ages for the ESA bearing Sterkfontein Member 5 deposits are also reliable when discrepancies in the geological context are taken into account [75, 85]. These new dates for the CoHK sites indicate a rich MSA heritage that has until recently been overlooked due to a predominant interest in the early hominin and early stone age bearing layers of these sites. This recent work suggests that MSA deposits may occur in the CoHK between ∼250 and ∼60 ka with the oldest being as old as those from East Africa.

5.6. Twin Rivers (Zambia). Barham et al. [138] describe early MSA technology from block A, and F at Twin Rivers in Zambia. In block A a flowstone at the base of the deposit as well as a piece within the deposits has been dated to >400 ka and is an upper age limit for the deposit. However, Barham et al. [138] suggest a hiatus occurs before deposition of the slurry flow which has seemingly eroded parts of the basal flowstone into the deposit. The oldest MSA deposits from Twin Rivers (Block A) are suggested to date to between ∼266 and ∼172 ka, while those from F block are suggested to date to between ∼200 and ∼140 ka [138]. This A block assemblage includes backed tools that suggest early evidence for composite tool technology as well as bifacial projectile points and tanged points. Unifacial points were not recovered during the 1999 excavations but occurred in very low frequency during Clark’s excavations during the 1950s [138]. Such points do occur in the deposits younger than ∼100 ka [138]. The bifacial points include lanceolates characteristic of the early Lupemban and flaking occurred using both an anvil technique and radial and prepared core technique [138]. Twin Rivers stands as some of the earliest potential evidence for the MSA in southern Africa and particularly the definitive use of ochre [139]. However, many researchers are sceptical over the association of the flowstone to the MSA bearing deposits.

In many instances, flowstone is sampled from the wall or edges of a cave cavity without definitive evidence for its association to the archaeology. At Twin Rivers, for example, the flowstone does not occur as a continuous layer capping the MSA deposits but as remanents on the edge of the cavity. Therefore, the flowstone perhaps only provides a maximum age estimate ([138, Figure 10.15, 10.13]). The fact that younger speleothem dated to between 184–172 ka (178 ± 6 ka) and 141–137 ka (139 ± 2 ka) occurs under a speleothem dated to between 200–190 ka (195 ± 5 ka; 131) in block A. All TL dates from G block are younger than 117 ka (101 ± 16 ka) and lend further suspicion to an extremely complicated stratigraphy and infill. The speleothem dates to between 266 ± 6 ka (272–260 ka) and 172 ± 2 ka (174–170 ka) may also have been eroded out from earlier deposits when the MSA in-filled the cavity. Again younger speleothem samples occur with depth with the 172 ka sample being deeper (3.2 m) than the 192 ± 2 ka sample at the top of A block at 2.2 m. All the MSA in the top 1 m of A block is, therefore, younger than 174 ka, as the speleothem must have formed before it was eroded and incorporated into the breccia and so provides a maximum age. A TL age from over 4 m depth gave an age of 132 ± 31 ka and would suggest that all the deposits are younger than 163 ka, but the dose rate estimates for the deposits from which the sample came were not estimated, and a dose from higher in the sequence was used [138], making the date unreliable. However, a U-Th sample with a date of 160 ± 3 ka also occurs at 2.4 m depth and is the youngest age from block A. This suggests that all the MSA in block A may in fact be younger than 163 ka. If the majority of speleothem represents material eroded into the deposit then the Lupemban from block A would be younger than 141 ka, significantly younger than the 266–170 ka cited by Barham et al. [138], the ∼266 ka or ≥230 ka age estimate that is often cited [140], or the ∼300 ka age cited by Barham [141]. There is little evidence that the MSA deposits are older than 266 ka except for a small deposit at the very bottom of the excavation, and this does not apply to the assemblage as a whole. The F block deposits are also potentially <140 ka based on the U-Th ages and G block is likely <117 ka based on TL ages. An age range of 141–48 ka could be estimated for the various MSA deposits at Twin Rivers based on the uncertainty of the provenience of the speleothem samples (Excavators need to be extremely careful when relating fragments of flowstone to wider archaeological deposits in caves due to their complex depositional history. All publications should show a clear photograph of the association of the speleothem to the deposit been dated or even micromorphological analysis across the contact. Often, dates are presented without any information regarding their reliability or context).

6. Fauresmith and Sangoan Sites in Southern Africa

Fauresmith sites have been noted at numerous sites on the Vaal River (e.g., Canteen Kopjie, Nooitgedacht 2, Roseberry Plain 1, and Powers site), the Orange River, and the Seacow Valley as well as pan sites near Kimberley (Kathu Pan and Rooidam) and the cave site of Wondwerk Cave [140]. Almost all these occur in eastern-central south Africa, the exceptions being at and perhaps Bushman’s rockshelter [142].

6.1. Kalambo Falls (Zambia). Clarke [56] first described the Kalambo Falls site and identified a number of floors. Clarke [56] notes that it is the smaller chert tools (blades and unifacial points with faceted platforms and prepared cores) in addition to the larger hand axes and cleavers that give the Kalambo Falls assemblages, particularly floor 4, a Fauresmith flavor, and make it seemingly distinct from the industries in the Luangwa Valley. These floors were later termed archaeological horizons, but many are now not considered to be in situ accumulations but material redistributed by alluvial action [30]. The locality is extensively described in [30], and four main sites have been defined [A–D]. The
lowest horizons have been defined as Upper to Terminal Acheulian in character and are referred to as the Bwalya Industry, and it has a very high percentage of LCTs (e.g., 67% of the shaped tools in one excavated horizon [30]). Many of the LCTs are small, lanceolate in shape, and could be argued to represent bifacial points evolving in the direction of the later Lupemban points. Small scrapers make up the majority of the light duty tools a few of which are convergent and resemble unifacial triangular points [30].

A series of Sangoan horizons follow that are referred to as the Chipeta Industry [30]. It has a much higher percentage of light duty tools, particularly scrapers, and it is suggested that this stage represents the beginning of the MSA and disappearance of Acheulian LCTs [30]. There is a trend towards more specialised flake forms, with blades [30]. The characteristic large tool is the core axe, which does occur in the lower Acheulian horizons but is in much greater number in the Sangoan levels. Clark [30] states “that the Chipeta Industry, though in some ways transitional, is the initial stages in the MSA trend to increased specialisation”. While lanceolates are absent in the Kalambo Falls Sangoan, Clark [30] notes that they do occur in other Sangoan assemblages. These are also undated and may be younger than that from Kalambo Falls. Lanceolate bifaces are the characteristic tool of the Lupemban levels of the MSA [30] along with an increase in lighter duty tools such as retouched unifacial points and blades. Potentially, utilised wood also come from the Acheulian and Sangoan levels and consist of possible clubs and pointed stakes and tools [30]. Sheppard and Kleindienst [143] concluded that there is little change, at Kalambo Falls, in the basic techniques of blank production or the attributes of the blanks produced from the earliest Acheulian layers to the later MSA layers. The only marked change to occur is the loss of LCTs (hand axes and cleavers) and their replacement by heavy-duty forms of core axes and picks. While lanceolates are absent in the Kalambo Falls Sangoan, some of the LCTS from the Acheulian layers have a similar lanceolate character that suggests some continuity between these earliest levels and the Lupemban. The intermediate Sangoan may represent an industry related to a particular climatic shift in the region, perhaps related to a warmer, more forested interglacial.

A range of radiocarbon ages exist for these early levels (60.3–32.6 ka BP [144]), but they are considered to be infinite age estimates [30]. This was followed by an amino-acid racemization age on wood from the Acheulian levels that suggested an age of ~110 ka [143]. A series of uranium series ages were then undertaken on wood, but because of the open system nature of the samples, they should be considered as minimum age estimates. The Acheulian layers were dated to 182 ± 16 ka and 182 ± 10 ka (198–166 ka) and the Sangoan levels to 76 ± 10 ka (85–65 ka). Hopefully, OSL dates will soon be forthcoming for at least the younger MSA (Lupemban) deposits at Kalambo Falls [145].

6.2. Bundu Farm. Peter Beaumont excavated a number of sites in the Northern Cape Province throughout the 1970s and 1980s and it is these sites that now appear to hold the key to understanding the transition between the Acheulian and the MSA in the region. Kiberd [146] recently described new excavations at the site of Bundu Farm in Northern Cape. The site has yielded only one LBCT from layer G6 but Kiberd [146] defines layers 4–6 as a terminal Acheulian or transitional ESA/MSA industry (i.e., Fauresmith). Layer G4 was dated by Kiberd [146] to 145.7 ± 16.0 Ka, providing a minimum age estimate for the G6 layer of >129.7 ka. Unprovenienced teeth from the site gave ESR age estimates of between 394 and 144 ka. Based on the occurrence and absence of fauna in layers of the site, Kiberd [146] suggests that this fauna most likely came from layers G4 and G5. This suggests that the G6 horizon containing the LCTS is dated to >~300–400 ka based on a linear uptake model (this model is the one that gives consistent ages for the CoHK sites) and ~200 ka based on an early uptake model. Prepared cores occur in all layers but are slightly more abundant in the younger layers that are <161 ka. In contrast, flake blades (this includes points, which are often referred to as convergent flake blades in South Africa) are generally more abundant in the oldest layer (G6), which also contains the one LBCT. Worked bifacial points also occur throughout the sequence. Overall, the sequence shows the occurrence of Levallois technology, blades, and points and in the oldest layer (G6) a single LBCT. Overall, the Bundu Farm sequence appears to contain all the components of the MSA in layers that are most likely dated to sometime between 400 and 200 ka. Brink (pers. Comm. In Kiberd [146]) notes that the Connochaetes gnou fossils are of a similar size to those from Florisbad (294–225 ka; [19]) and that the fossils reflect a period of increased rainfall. Based on the early uptake model for the ESR ages the oldest deposits G4–6 perhaps dates to roughly MIS 7 at ~240–190 ka. If the linear uptake model is correct (as it appears to be for many sites in South Africa), the oldest layers likely date to MIS9 between ~340–310 ka.

6.3. Wonderwerk Cave. Beaumont and Vogel [47] describe the results of excavations at Wonderwerk Cave. It has a sequence of what are suggested to be almost continuous deposits dating back nearly 2 million years [47, 71]. Beaumont and Vogel [47] suggest that MSA assemblages are represented in Major Units (MU) 2 in excavations 1, 3, 5, 6 and 7 and Fauresmith assemblages in MU3–4 in excavations 1 and 2. In contrast, Chazan et al. [75] note that no Fauresmith like material is noted in their reanalysis of the excavation 1 material. The excavation 1 material is, therefore, entirely Acheulian in character, and its age has previously been discussed above. Chazan and Horwitz [147] further suggest that all material from excavation 6 is Fauresmith in character with the co-occurrence of “oversized blades”, prepared core technology and LBCT in contrast to Beaumont and Vogel’s [47] interpretation of these deposits as being MSA. Chazan and Horwitz [147] also note that in Excavation 6, there is no evidence for underlying earlier ESA deposits as are found in Excavation 1 at the front of the cave, the top of which dates to <780 ka [75]. A single U/Th date (U-576) on a fragment of stalagmite [147] recovered from within the Fauresmith deposits of Excavation 6 gave an age of 187 ± 8 ka (195–179 ka) [47] and palaeomagnetism records a normal magnetic polarity [47]. Beaumont and Vogel [47]
state that this is a “minimum age” estimate for the age of the deposit and that it dates the Fauresmith to between $<780$ and $>179$ ka. In actual fact, if a fragment of speleothem occurs within a deposit it must have grown before the deposit was formed, not after it. As such, this is a maximum age estimate for the layer in which the speleothem fragment was found and suggests the Fauresmith assemblage is $<195$ ka, making it only slightly older than the Pinnacle Point MSA at 170–160 ka and perhaps similar to the LBCT assemblage from Herto. If the speleothem capped the underlying deposit, then it would indeed suggest that the deposit below this were older than 179 ka. However, Chazan and Horwitz [147] are not clear on the contextual issue of this sample, the age for which has come from the earlier publication by Beaumont and Vogel [47]. Beaumont and Vogel [47] state that items of calcite, mainly in the form of small ($\sim 5–20$ cm high) stalagmites, were found sporadically in many strata. They do not elaborate on the in situ or ex situ context of the material other than if it is a stalagmite of a stalactite. As stated above, if a fragment of speleothem occurs ex situ within a deposit, that deposit must be younger than the age of the speleothem, as the speleothem must have formed before being incorporated. If, however, a speleothem caps a deposit by forming directly on it and is in situ, then the deposits under it are older than that speleothem and the deposits above it in theory younger. If the speleothem is a flowstone, then it may also represent a false floor under which younger deposits may have been deposited. As previously noted, it is very important to understand and state the relationship of speleothems to the deposits and archaeology being dated and present clear evidence of this in publications.

The other ages for Fauresmith layers at Wonderwerk come from the top of stratum 3 in excavation 2 and give ages of between 315 and 247 ka. As these are soda straws that fell into the deposit after deposition of the Fauresmith deposits, they provide little chronological evidence for the age of the majority of the deposit other than suggesting that the layers in which the straws were found are younger than 247 ka. It does not, as Beaumont and Vogel [47] suggest necessarily mean that the Fauresmith layers date back to between 286 and 276 ka. A stalagmite in the top part of the MSA layers in excavation 2 had a basal age of 234–206 ka and may suggest that the MSA in this area is greater than 206 ka. However, the exact relationship of the stalagmite to the deposits is again not stated. It could simply have been incorporated into the deposits from elsewhere. There appears to be some evidence for the movement of material by water at the site which was once an active cave resurgence. This can be seen by the mixed nature of artefacts in the rear mostly in-filled tunnel and the eroded surface in many sections which slope back towards the rear of the cave. Moreover, a stalagmite that is seemingly lower in the section is dated to between 182 and 154 ka. This inversion of ages in the same excavation does not suggest the speleothems are in situ and would suggest that all of the MSA deposits are in fact younger than 182 ka. Until a better description of context is given for these samples all age assessments using, the data of Beaumont and Vogel [47] should be regarded with extreme caution. Many of the speleothems from the top of the MSA excavations in different trenches date to between 100 and 68 ka and without a context for the samples the deposits could in theory be not much older than that age or even in theory younger. Based on this re-analysis, there appears to be little current evidence for Fauresmith deposits older than $\sim 300–200$ ka.

6.4. Kathu Pan. Previous dating of the site was based on elephant fossils that were more evolved than those from Olduvai Bed IV [62]. This simply gave the site an age of $<1.07$ Ma or $<780$ ka based on the interpretation of the palaeomagnetic data [6]. At Kathu Pan, the MSA layers fall in the time range between 336 and 254 ka ($291 \pm 45$ ka; [148], perhaps during MIS9 (340–310 ka). The layer 4a Fauresmith assemblage at Kathu Pan contains Levallois cores, retouched points, blades and LCBTs and has been dated, to somewhere between 511 and 435 ka based on a combination of OSL (464 ± 47 ka) and ESR (542 + 107/−140 ka) [148]. Porat et al. [148] suggest that the OSL age may represent a minimum age estimate. If so, the layer would date to between 682 and 435 ka (based on ESR alone). An MIS13 age (540–470 ka) might be a good estimate based on this data but certainly older than 417 ka. This suggests that all the tool forms found in the MSA are already in place by at least 417 ka. The retouched points have faceted platforms and are in stark contrast to early MSA assemblages that are suggested to lack formal tools and retouched points [148]. Porat et al. [148] note that the extreme lateral convexity of the lithics distinguishes them from the norm for the Levallois method, despite the fact that they typologically and technologically fit within the Levallois. The age of the Acheulian 4b layers has not been determined other than being older than 4a. Porat et al. [148] note that LCTs in the Fauresmith horizons are made on a wide range of raw materials and are crude and irregular, while those from the Acheulian are exclusively made on banded ironstone and are symmetrical and refined. This may reflect the hominins developing new raw materials for LBCT manufacture during the Fauresmith as part of experimentation of new methods of stone tool manufacture. The occurrence of a Fauresmith industry at Kathu Pan, so close to Wonderwerk Cave, dated to $>417$ ka may lend weight to the Fauresmith at Wonderwerk also being in this time range or at least older than 182 ka as suggested by Chazan and Horwitz [147] unless it occurred over 200 ka in the region and was being produced contemporarily with the MSA.

6.5. Mapungubwe. Kuman et al. [74] describe a series of sites (Keratic Koppie, Hackthorne, and Kudu Koppie) from the Mapungubwe National Park (formerly Vhembe Dongola NP) in northern South Africa on the border with Botswana and Zimbabwe. Kuman et al. [74] suggest the site represents a Charaman like industry as described from Zimbabwe, which itself is perhaps a version of the Sangoan. These two industries were grouped together into Clark’s First Intermediate Period along with the Fauresmith [28]. The sites have tools that are again characteristic of the Acheulian and the MSA. Although they show marked typological differences, the parallels between the Sangoan and the Fauresmith are striking. Kuman et al. [74] suggest that after 200 ka classic MSA technology occurs that is well
developed, yet this does not reflect a change so much in the actual technology as a shift toward a lighter, more mobile toolkit which incorporates hafting. Wilkins et al. [149] note that the prepared core reduction strategies at Kudu Koppie suggest that both the late ESA and MSA toolmakers employed the Levallois Volumetric Concept, but they often exploited a nodule’s natural convexities and form. The MSA toolmakers use a greater variety of prepared core methods and more intensively exploited cryptocrystalline and microcrystalline nodules, the scarcity of which may have resulted in a more “formalized” application of the Levallois Volumetric Concept. Kempson [150] suggests that the Hackthorne site may represent mixing of ESA and MSA deposits, while Keratic Koppie preserves an assemblage with a mixture of Acheulian and MSA components. Kempson [150] argued the assemblage is a post-Acheulian industry with a major component of woodworking tools, suggestive of the Sangoan Industrial Complex. Davies [64] and Kuman [66] also suggest the occurrence of Sangoan-like tools along the KwaZulu-Natal coast including coastal dunes near Port Edward, which if confirmed would truly throw out the idea that these are environmentally specific or regional entities. This is close to the site of Amanzi Springs which Deacon [73, 151] suggested might be later Acheulian.

6.6. Rooidam. The alluvial sites of Rooidam 1 and 2 were excavated by Fock [152]. Rooidam 1 contained a rich Stratrum 9, which sits 3.9 m below the surface and yielded 90% of the 19,000 artefacts recovered from the site [47, 70, 152]. Previously, the occurrence of small hand axes as well as a small amount of cleavers, blades and convergent points suggested this was a Fauresmith layer [70]. Beaumont and Vogel [47] later assigned this layer to the final Acheulian, stating that while prepared cores and blades were present, there were no convincing Levallois points. Here, Beaumont and Vogel [47] suggest that the late Acheulian can be distinguished on the basis of true blades and that Levallois points distinguish the Fauresmith. Here, the Fauresmith is reduced to an MSA assemblage that simply contains LCT. The upper MSA levels were also reclassified as Fauresmith, which also occurs at Rooidam 2. Szabo and Butzer [153] conducted U-Th dates on two calcrites sandwiched between the late Acheulian and Fauyremish occupations and overlying the main stratrum 9 occupation. Szabo and Butzer’s [153] unit C correlates with Unit 9 of Beaumont [47, 70]. The U-Th ages are inverted for the sequence. The Unit C deposit of Szabo and Butzer [153] covering stratrum 9 gave an age of 108 ± 40/−20 ka (148–88 ka). The stratigraphically higher unit G gave an age of 174 ± 35 ka (209–139 ka). Szabo and Butzer [153] argue that the younger lower age is due to the open system nature of the system in Unit C and recrystallization. They state that recrystallization does not appear to have occurred in unit G and so the age is accurate for that unit. It seems likely that both ages are minimum age estimates due to the open nature of the system, and so, the age of the stratrum 9 late Acheulian deposit is the site can only be said to be older than 139 ka. The upper Fauresmith deposits in theory date to less than 209 ka and might suggest that the Fauresmith occurs relatively late at the site; however, if recrystalisation had also effected the lower sample to any degree, then these deposits might also be significantly older than this.

6.7. Taung. Kuman [154] describes a series of undated sites (DB3 and 4) near Taung where LCTs are associated with convergent points and prepared core technology and also with a higher representation of end struck rather than side struck flakes, diagnostic features of the Fauresmith. At DB3 Kuman [154] notes the occurrence of Levallois and Victoria West cores in the same context. Hence, despite the dominance of simple flaking features, core types, faceted platforms and dorsal scar patterns together show a range of relatively advanced flaking techniques that are seen in the assumingly later MSA of the region [154]. At DB3, the majority of LCT are side struck, while most small tools are end struck, which is out of character for LCT of Fauresmith type as described by Goodwin [32]. This appears to show the very mixed character of the assemblage. The DB3 Acheulian assemblage contains three types of flaking technique (radial, convergent, and parallel opposed), which are documented more consistently in the ensuing Middle Palaeolithic and MSA [154].

6.8. Bushman’s Rockshelter. Bushman’s Rockshelter was excavated by Louw [142] who describes a number of LCTs or LCT-like stone tools from layer 41. This has an uncalibrated C14 age of 47.5 BP [155] and is no doubt an infinite age. Louw [142] suggests that the hand axes may have been picked up by MSA people and reused, but it is also possible that this is a transitional industry, perhaps the Fauresmith, as the deposits do not seem to have Sangoan elements like the Limpopo sites. If so, then this would suggest a greater regional spread of the Fauresmith, but the deposit is as yet undated.

7. Discussion and Conclusions

As outlined in the opening discussion, there are a number of theories concerning the terminology of the ESA to MSA transition and it is likely far from a simple transition that occurred at exactly the same time or the same way in different areas of Africa. The age of all the sites is outlined in Table 1. The sites of the Kaphurin and Gademotta Formation in eastern Africa indicate that the MSA in its classical sense occurs at roughly the same time between 300 and 250 ka. The Kathu Pan MSA at 336–254 ka is potentially older and suggests the earliest MSA occurs synchronously in eastern and southern Africa. Further potential evidence for this comes from Florisbad and Sterkfontein-Lincoln Cave sometime between 300 and 200 ka. The sites of the Kapthurin Formation also indicate that many characteristic elements of the MSA occur alongside elements of the ESA including LCTs from roughly 500 to 250 ka. These have been termed MSA, Sangoan, Fauresmith, and Acheulian depending on the groups of tools found at different sites and has been used to suggest independent evolution of all the elements (points, Levallois, etc.) that come together to form the MSA as an entity and in essence augments previous Acheulian technology [22]. Blades come into use first around 545–509 ka, followed by points, large prepared cores, and then centripetal
<table>
<thead>
<tr>
<th>Country</th>
<th>Locality</th>
<th>Site/layers</th>
<th>Dating methods</th>
<th>Upper age</th>
<th>Lower age</th>
<th>Best age estimate</th>
<th>Industry</th>
<th>Hominin association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kenya</td>
<td>Nachukui Formation, West Turkana</td>
<td>KS4</td>
<td>Ar-Ar</td>
<td>&gt;1.65 Ma</td>
<td>∼1.65 Ma</td>
<td>Acheulian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Konso Gardula</td>
<td>KGA</td>
<td>Ar-Ar</td>
<td>1.69 Ma</td>
<td>1.41 Ma</td>
<td>∼1.5 Ma</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>Olduvai Gorge</td>
<td>Upper Bed II</td>
<td>Ar-Ar/Palaeomag.</td>
<td>1.53 Ma</td>
<td>1.27 Ma</td>
<td>1.5-1.3 Ma</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Reitpuits Formation [23], Vaal River</td>
<td>CBD</td>
<td>Cosmogenics</td>
<td>2.08 Ma</td>
<td>1.08 Ma</td>
<td>1.4–1.2 Ma</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Canteen Kopjie</td>
<td>Stratum 2b lower</td>
<td>Cosmogenics</td>
<td>1.61 Ma</td>
<td>0.82 Ma</td>
<td>&lt;1.6–&gt;0.8 Ma</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Wonderwerk Cave</td>
<td>Excavation 1</td>
<td>Pmag.</td>
<td>1.78 Ma</td>
<td>0.78 Ma</td>
<td>&lt;1.6–&gt;1.1 Ma</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Swartkrans</td>
<td>Member 2</td>
<td>Fauna/U-Pb [bone]</td>
<td>1.65 Ma</td>
<td>1.07 Ma</td>
<td>1.6–1.1 Ma</td>
<td>Acheulian</td>
<td>H. ergaster</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Sterkfontein</td>
<td>Member 5c</td>
<td>Pmag/ESR</td>
<td>1.39 Ma</td>
<td>0.82 Ma</td>
<td>1.4–1.1 Ma</td>
<td>Acheulian</td>
<td>H. ergaster</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Cornelia</td>
<td>Uitzoek</td>
<td>Pmag/fauna</td>
<td>1.07 Ma</td>
<td>0.78 Ma</td>
<td>1.1–0.8 Ma</td>
<td>Acheulian</td>
<td>early Homo</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Elandsfontein</td>
<td>all</td>
<td>Pmag/Fauna</td>
<td>1.07 Ma</td>
<td>0.78 Ma</td>
<td>1.1–0.8 Ma</td>
<td>Acheulian</td>
<td>H. erectus/H. rhodesiensis</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Swartkrans</td>
<td>Member 3</td>
<td>Fauna/U-Pb [bone]</td>
<td>1.04 Ma</td>
<td>0.62 Ma</td>
<td>1.0–0.6 Ma</td>
<td>Acheulian</td>
<td>early Homo</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Gladysvale</td>
<td>internal</td>
<td>P-mag/ESR</td>
<td>&lt;990 ka</td>
<td>&gt;780 ka</td>
<td>∼800 ka</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Cave of Hearths</td>
<td>Bed 1–3</td>
<td>P-Mag</td>
<td>&lt;780 ka</td>
<td>400 ka</td>
<td>670–560 ka</td>
<td>Acheulian</td>
<td>H. rhodesiensis</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Kathu Pan</td>
<td>STR4a</td>
<td>ESR/OSL</td>
<td>647 ka</td>
<td>435 ka</td>
<td>540–470 ka</td>
<td>Fauresmith</td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>Kapthurin Formation</td>
<td>Koimillot stratigraphy</td>
<td>K-Ar</td>
<td>&lt;557 ka</td>
<td>&gt;440 ka</td>
<td>?</td>
<td>Pseudo-Stillbay</td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>Kapthurin Formation</td>
<td>K3</td>
<td>Ar-Ar</td>
<td>548 ka</td>
<td>500 ka</td>
<td>548–500 ka</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>Kapthurin Formation</td>
<td>K4</td>
<td>Ar-Ar</td>
<td>&lt;518 ka</td>
<td>&gt;272 ka</td>
<td>&lt;518 ka</td>
<td>Acheulian</td>
<td>with points</td>
</tr>
<tr>
<td>Kenya</td>
<td>Kapthurin Formation</td>
<td>K4 (Rorop Lingop)</td>
<td>Ar-Ar</td>
<td>&lt;518 ka</td>
<td>&gt;272 ka</td>
<td>&gt;272 ka</td>
<td>Acheulian/MSA/Sangoan/Fauresmith</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Bundu Farm</td>
<td>G4–6</td>
<td>ESR/Fauna</td>
<td>&gt;394 ka</td>
<td>∼200 ka</td>
<td>Fauresmith</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Rooidam 2</td>
<td>Unit to G</td>
<td>U-Th</td>
<td>?</td>
<td>&gt;154 ka</td>
<td>&gt;154 ka</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>Sudan</td>
<td>Sai Island</td>
<td>YG</td>
<td>OSL</td>
<td>?</td>
<td>&gt;204 ka</td>
<td>&gt;204 ka</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Rooidam 2</td>
<td>Unit A-B</td>
<td>U-Th</td>
<td>&gt;154 ka</td>
<td>?</td>
<td>Fauresmith</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Kathu Pan</td>
<td>STR3</td>
<td>OSL</td>
<td>336 ka</td>
<td>254 ka</td>
<td>336–254 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Wonderwerk Cave</td>
<td>Excavation 2, 6</td>
<td>U-Th</td>
<td>315 ka</td>
<td>&lt;195 ka</td>
<td>315–&lt;195 ka</td>
<td>Fauresmith</td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>Kapthurin Formation</td>
<td>Koimillot stratigraphy</td>
<td>K-Ar</td>
<td>&lt;296 ka</td>
<td>&lt;237 ka</td>
<td>250–200 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Florisbad</td>
<td>Peat 1</td>
<td>ESR/OSL</td>
<td>295 ka</td>
<td>225 ka</td>
<td>295–225 ka</td>
<td>MSA</td>
<td>H. helmei archaic H. sapiens</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Sterkfontein</td>
<td>M6/post M6</td>
<td>ESR</td>
<td>294 ka</td>
<td>210 ka</td>
<td>294–210 ka</td>
<td>MSA</td>
<td>H. ergaster?</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Lincoln Cave</td>
<td>south/north</td>
<td>U-Th</td>
<td>278 ka</td>
<td>107 ka</td>
<td>278–107 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Gademotta Formation</td>
<td>Type site Unit 9</td>
<td>Ar-Ar</td>
<td>?</td>
<td>&gt;272 ka</td>
<td>∼280 ka</td>
<td>MSA</td>
<td></td>
</tr>
</tbody>
</table>
Table 1: Continued.

<table>
<thead>
<tr>
<th>Country</th>
<th>Locality</th>
<th>Site/layers</th>
<th>Dating methods</th>
<th>Upper age</th>
<th>Lower age</th>
<th>Best age estimate</th>
<th>Industry</th>
<th>Hominin association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethiopia</td>
<td>Gademotta Formation</td>
<td>Kulkuletti</td>
<td>Ar-Ar</td>
<td>?</td>
<td>&gt;272 ka</td>
<td>~280 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>Zambia</td>
<td>Twin Rivers (Zambia)</td>
<td>Block A</td>
<td>U-Th</td>
<td>&lt;266 ka</td>
<td>&lt;132 ka</td>
<td>&lt;132 ka</td>
<td>MSA</td>
<td>(Lupemban)</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Gademotta Formation</td>
<td>Unit 11</td>
<td>Ar-Ar</td>
<td>&lt;280 ka</td>
<td>&gt;173 ka</td>
<td>280–173 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>Sudan</td>
<td>Sai Island</td>
<td>BLG/TLG/UG</td>
<td>OSL</td>
<td>&lt;243 ka</td>
<td>&gt;142 ka</td>
<td>200–160 ka</td>
<td>Sangoan</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Border Cave</td>
<td>6BS</td>
<td>ESR</td>
<td>238 ka</td>
<td>217 ka</td>
<td>238–217 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>Zambia</td>
<td>Kalambo Falls</td>
<td>Acheulian Layers</td>
<td>U-Th/AAR</td>
<td>&gt;198 ka</td>
<td>&gt;166 ka</td>
<td>&gt;166 ka</td>
<td>Acheulian</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Twin Rivers</td>
<td>Block F</td>
<td>U-Th</td>
<td>&gt;195 ka</td>
<td>&lt;140 ka</td>
<td>&lt;140 ka</td>
<td>MSA</td>
<td>(Lupemban)</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Border Cave</td>
<td>5BS/5WA</td>
<td>ESR/AAR</td>
<td>183 ka</td>
<td>141 ka</td>
<td>183–141 ka</td>
<td>MSA</td>
<td>H. sapiens?</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Wonderwerk Cave</td>
<td>Excavation 2, 3, 5</td>
<td>U-Th</td>
<td>182 ka</td>
<td>118 ka</td>
<td>182–118 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Pinnacle Point 13B</td>
<td>LC-MSA Lower</td>
<td>OSL</td>
<td>174 ka</td>
<td>153 ka</td>
<td>~164 ka</td>
<td>MSA</td>
<td>H. sapiens?</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Bouri Formation</td>
<td>Upper Herto Member</td>
<td>Ar-Ar</td>
<td>166 ka</td>
<td>147 ka</td>
<td>166–147</td>
<td>MSA</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Pinnacle Point 13B</td>
<td>DB Sand 4b</td>
<td>OSL</td>
<td>166 ka</td>
<td>152 ka</td>
<td>166–152 ka</td>
<td>MSA</td>
<td>H. sapiens?</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Bundu farm</td>
<td>G2-3</td>
<td>ESR/Fauna</td>
<td>162 ka</td>
<td>130 ka</td>
<td>162–130 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>Sudan</td>
<td>Sai Island</td>
<td>Nile silts</td>
<td>OSL</td>
<td>&lt;162 ka</td>
<td>?</td>
<td>&lt;162 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>Zambia</td>
<td>Kalambo Falls</td>
<td>Floor 4</td>
<td>U-Th</td>
<td>&gt;85 ka</td>
<td>&gt;65 ka</td>
<td>&gt;65 ka</td>
<td>Sangoan</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Pinnacle Point 13B</td>
<td>LC-MSA Middle/Upper</td>
<td>OSL/U-Th</td>
<td>130 ka</td>
<td>115 ka</td>
<td>130–115 ka</td>
<td>MSA</td>
<td>H. sapiens?</td>
</tr>
<tr>
<td>Kenya</td>
<td>Central Rift (Kaperto)</td>
<td>Kapedo Tuffs</td>
<td>Ar-Ar</td>
<td>135 ka</td>
<td>120 ka</td>
<td>135–120 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>Eritrea</td>
<td>Abdur</td>
<td>Ar-Ar</td>
<td>OSL</td>
<td>132 ka</td>
<td>118 ka</td>
<td>125–118 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Klapes River Mouth</td>
<td>LBS</td>
<td>OSL</td>
<td>120 ka</td>
<td>90 ka</td>
<td>120–90 ka</td>
<td>MSA</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Border Cave</td>
<td>4WA</td>
<td>ESR/AAR</td>
<td>122 ka</td>
<td>111 ka</td>
<td>122–111 ka</td>
<td>MSA</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>S. Africa</td>
<td>Swartkrans</td>
<td>Member 4</td>
<td>U-Th</td>
<td>?</td>
<td>&lt;108 ka</td>
<td>&lt;108 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Twin Rivers</td>
<td>Block G</td>
<td>TL</td>
<td>&lt;117 ka</td>
<td>48 ka</td>
<td>117–48 ka</td>
<td>MSA</td>
<td></td>
</tr>
<tr>
<td>S. Africa</td>
<td>Pinnacle Point 13B</td>
<td>Western Excavation</td>
<td>OSL/U-Th</td>
<td>114 ka</td>
<td>91 ka</td>
<td>114–91 ka</td>
<td>MSA</td>
<td>H. sapiens?</td>
</tr>
<tr>
<td>Kenya</td>
<td>Malewa Gorge</td>
<td>Malewa Gorge</td>
<td>Ar-Ar</td>
<td>?</td>
<td>&gt;86 ka</td>
<td>~118–86 ka</td>
<td>MSA</td>
<td></td>
</tr>
</tbody>
</table>

Levallois cores and flakes all the while occurring with LCTs. It is not until 250–200 ka that convergent Levallois flakes and cores occur in the absence of LCTs that the MSA is seen to start. Of course, LCTs may occur much later (<160 ka) in some contexts in Etiopia and Ethiopia. Interestingly, the >272 ka MSA in the Gademotta Formation in Ethiopia seems not to be underlain by such transitional industries [22].

Potentially similar trends are seen at Kalambo Falls with a transition from the Acheulian, through the Sangoan and into the Lupemban-MSA, but this site remains unreliably dated. At Kalambo Falls, there seems to be a definite trend through time but with the Sangoan being more of a distinct entity, perhaps relating to a more forested, interglacial period. The only reliably dated Sangoan assemblage comes from Sai Island in the Sudan at sometime between 242–142 ka. This is contemporary with the early MSA assemblages in the Kaphurin and Gademotta Formation as well as several sites in southern Africa (Border Cave, Florisbad, and Sterkfontein-Lincoln Cave). At Kalambo Falls, the Sangoan directly underlies the Lupemban-MSA as at Sai Island. Another age for the Sangoan comes from TL dates (305–203 ka) that overlies Sangoan deposits at the site of Béti 1 in the Ivory Coast [156]. This suggests a potentially older origin for the Sangoan. However, Barham and Mitchell [29] note
that the Bété 1 site has a high percentage of LCTs than other Sangoan assemblages such as Kalambo Falls and may indicate mixing of this open deposit. The Sai Island Sangoan is also the same age as the Lupemban-MSA deposits at Twin Rivers in Zambia if the dates of 266–172 ka are correct. This appears to show that the Sangoan and early MSA are contemporary in different areas of Africa with Lupemban-like MSA at Sai Island also being later than seen in Zambia. It is possible that the Sangoan here is of a similar age to that from Sai Island perhaps representing a spread from a more southern central area of Africa with the Kalambo and Kaptthurin Formation Sangoan being older. However, if the Twin Rivers Lupemban is younger than 160–130 ka, as potentially seems the case based on the inverted speleothemm ages, then it is of a similar age to the Lupemban-like MSA at Sai Island (<162 ka) This is the same age as the Pinnacle Point 13B MSA in South Africa and material from Border Cave, Wonderwerk Cave, and Florisbad but much younger than the earliest Fauresmith from Kathu Pan at 417 ka. If this is the case then there is perhaps little reason to see the Sangoan at Kalambo as much older than its 182 ka minimum age. If it is related to a climate shift then it may be MIS7 in age (240–190 ka).

It has often been suggested that the Sangoan is a wood-working specific industry [56] in part due to its discovery in heavily forested areas such as Zambia and the Congo. This is perhaps even more influenced by the association of wood at Kalambo Falls, one of the few sites to preserve such material. A number of undated Sangoan occurrences have also been described from South Africa, both along the costal dune systems of KwaZulu-Natal and the very northern Limpopo River region. If confirmed this suggests a distribution, if sparse, across almost all of Africa and appears to argue against either the Sangoan being a regional entity, or perhaps an environmentally specific one. The occurrence of the Sangoan in the Sudan further argues against this. Moreover, Rots and Van Peer [118] suggest the core axes are being used as digging rather than wood working tools.

Based on known occurrences, the Sangoan is, however, not widespread across most of South Africa, where the related transitional industry, the Fauresmith, has also been described. This too has been suggested to be a regionally or environmentally specific industry, but again, a lack of dates has hampered its comparison to other industries such as the Sangoan. The majority of the Fauresmith occurrences have been described from the River sequences of the Northern Cape and Free State. However, a number of potential Fauresmith sites; namely, Vereeniging and particularly Bushman’s rock shelter occur at the very extremes of the Fauresmith range [38, 142]. Only reanalysis and dating of these deposits and tools will help confirm if the Sangoan and Fauresmith are different regional entities or if the Fauresmith is perhaps an earlier entity than the Sangoan as the Sai Island dates suggest. The Acheulian levels at Kalambo Falls have some qualities distinctive of the Fauresmith, in terms of LCTs, although it doesn’t seem to include many of the light duty tools often associated with it.

In South Africa, the Fauresmith industry has recently been dated in stratified deposits at Kathu Pan to at least 417 ka and perhaps as old as 647 ka and at Wonderwerk Cave to sometimes between 305 and 179 ka. This is towards the older end of the Kaptthurin Formation sequence. While the Kaptthurin Formation is suggested to demonstrate independent evolution of all the traits that are used to classify the MSA [22] all these traits appear to occur at Kathu Pan during the Fauresmith. Of course, Kathu Pan is only one site and the Kaptthurin Formation is many that indicate variation across both time and the landscape. Given that Kathu Pan and Wonderwerk are very close to each other the age of the Fauresmith and MSA at both sites suggests firstly that the Fauresmith is a relatively long lived entity, perhaps 200 ka and secondly that Fauresmith and MSA assemblages may overlap in the time range between 300–200 ka. Either that or the MSA assemblages from Kathu Pan represent the Fauresmith but simply do not have the elements that would be used to classify it as such.

At Wonderwerk, Chazan and Horwitz [147] and Beaumont and Vogel [47] certainly have different views on which layers are Fauresmith and which are MSA. This in itself asks the question regarding the difference between the Fauresmith and the MSA and whether as some researchers have suggested that the Fauresmith and other transitional industries should, in fact, represent the start of the MSA and that the presence of LCTs should be seen as merely a remnant of mode 2 technology in otherwise mode 3 dominant assemblages. Bruggemann et al. [117] argue that the Acheulian biface persisted at Abdur, possibly because it acquired a new, more profitable use for the exploitation of aquatic resources by early modern humans. If this is the case, then it provides further evidence that LCTs are not distinctive only of the Acheulian and their persistence in some assemblages should not be used to equate them with the Acheulian but simply as a surviving ESA element in an otherwise MSA assemblage.

Prepared core technology, particularly Levallois, is inextricably linked to the MSA, but later stages of the ESA also document the use of prepared core reduction [30, 47, 58, 154, 157]. Good chronostratigraphic records are rare for this time period, but outside South Africa, there is some evidence that prepared core reduction may even date back to almost 800 ka [158]. Giant cores [159] also referred to as Levallois-like [158] are known from the II–6 archaeological unit from Gesher Benot Ya’aqov in Israel. Goren-Inbar and Saragusti [158] suggest that biface production involved at least two well-established and different techniques: the Levallois and the Kombewa. These techniques produced predetermined, large-sized flakes that were modified into tools by a minimal amount of retouch [158]. A Technological and morphometric comparison between tools manufactured by the different techniques does not demonstrate any bimodal patterning of the end products [158].

Rink and Schwarz [160] suggest the II–6 unit dates to around 681–623 ka based on a range of combined U-Th/ESR ages. Older ages have been suggested based on the location of the Brunhes-Matuyama boundary at 780 ka in the lower unit II–7 and depositional rate estimates. However, given that the ESR ages are undertaken on fossils within the II–6 sand unit, these ages should perhaps be taken as the true age of the site unless there is significant uncertainty in the ages. Many researchers have seemingly dismissed the ESR method as
unreliable (see [161]). However, in the correct circumstances (i.e., low uranium concentrations as is the case of GBY), the method is reliable when a linear uptake model is used based on comparisons with other geochronological methods (see [19, 79]). The palaeomagnetic reversal occurs in the base of clay unit II-7 and as such the entire of II-7 was deposited after the transition as well as a change of sedimentation and potential hiatus between units II-7 and II-6. As such, the II-6 unit could not be as old as ∼780 ka as is most often quoted [162]. There is also a suggestion from the fauna [163] that the site has a number of species not seen in Europe until closer to 600 ka. Rink and Schwarcz [160] suggest that a hiatus in deposition may be responsible for this difference in palaeomagnetic and ESR age estimates. However, Rink and Schwarcz’s [160] study suffers from a common problem in ESR studies (see discussion in [67, 79] for other examples), where all the ages are averaged together to create a mean value for different units or teeth with vastly different ages where the reasons and mechanisms for the differing ages has not been established. If the ESR dates (linear uptake, 40% moisture) of Rink and Schwarcz [160] are separated by units; the lower II-7 deposit has an optimal age of between 718 and 555 ka (max age range 767–534 ka) and the upper II-6 unit has an optimal age of 708–559 ka (max age range of 718–492 ka). The upper age range for the top of the II-7 deposit, which contains the ∼780 ka Brunhes-Matuyama boundary in its base, is 767 ka, which is not much younger than 780 ka (moreover, recent estimates put the boundary at 773 ka [164]). This suggests that the II-6 archaeology most likely dates to sometime between ∼700 and ∼550 ka, although a slightly older age cannot be ruled out.

These cores are not too dissimilar to both the large LCT cores described for the Kaphthurin Formation [106, 109] between 509--272 ka or the Victoria West Cores of South Africa. Sharon and Beaumont [43] have suggested that Victoria West cores were also developed as a prepared core technique within the Acheulian as a means of producing ready-made blanks for LCTs. Part of the problem in southern Africa is that no Victoria West core bearing sites have been accurately dated and few are well stratified. Victoria West technology has been recovered from stratified Acheulian contexts in stratum 2a at Canteen Kopjie in South Africa [68]. The stratum 2a contains an Acheulian assemblage that is overlain by Fauresmith material in stratum 1 (Hutton sands), where it is then followed by an MSA assemblage which has dates of 120–80 ka [165]. As such, the Victoria West technology appears to be associated with terminal Acheulian artefacts and is older than the Fauresmith. While the Victoria West layers at Canteen Kopjie are undated, if the ages for Kathu Pan are extrapolated, then they should be older than 540–470 ka and perhaps not too dissimilar in age from those from GBY at 700–550 ka. This points to a widespread use of this giant core technology around 700–300 ka in Africa and the Levant in late Acheulian and transitional industries. Whether these early prepared core technologies for making large blanks for LCTs are proto-Levallois and evolved directly to MSA Levallois technology or para-Levallois is beyond the scope of this paper (see [41, 43]). However, in the Kaphthurin Formation, there seems to be a relative progression from these giant cores to centripetal and then convergent Levallois cores and flakes [109]. Lycett [41] suggests that Victoria West cores show independent development of prepared core technology from that of Levallois technology, and as such it is “para-Levallois” rather than “proto-Levallois”. However, the Victoria West is chronologically older than the Levallois technology of the Fauresmith and similar progression as just described for the Kaphthurin Formation could also be argued for South Africa.

Blades also seem to be a component of either the late Acheulian or the beginning of this transitional phase as shown by their first occurrence in the Kaphthurin Formation between 545 and 509 ka and in the Fauresmith at Kathu Pan. Large blades were one of the earliest indicators of the Fauresmith and these are noted in the Wonderwerk occurrence [147]. As noted before, Tryon et al. [54] and Tryon and McBrearty [109] suggest that diversification of Levallois technology as seen at Koimiloy by ∼250 ka and the occurrence of convergent Levallois points and cores stands as the marker of the early MSA. Tryon et al. [54] note that Levallois technology for making LCTS occurs by 350–300 ka based on ESR age estimates at the Grotte des Rhinocéros and by OSL at Cap Chatelier [166] in North-west Africa, and Levallois-like cores make from LCTs at Kharga Oasis (Egypt) between 400–300 ka based on U-series dating of tufa [167]. McBrearty [22] also suggests that the fundamental change from the ESA to the MSA is the end of LCTs and a shift to projectile point technology. Of course, it should be noted that Acheulian bearing hominins in Europe were utilising an entirely wooden projectile technology for hunting as shown by the occurrence of the Schöningen spears at either ∼400 (MIS 11 [157]) or ∼310 ka (MIS 9d-e; [168]) but were seemingly still disarticulating their kill with LCTs. Whether a similar wooden projectile technology was being used by hominins in Africa is almost impossible to tell given the almost complete lack of preservation of such organic remains in most MSA sites. The exceptions are two wooden tools from Floor 1 at Kalombo Falls in Zambia [30, 56] and one from Florisbad in South Africa [13]. Other sites where large pieces of wood have been recovered include the Acheulian sites of Amanzi Springs [73] and Gesher Benot Ya’aqov [169]. Despite the discovery of significant amounts of wood from these deposits, no tools have been noted. The Kalombo falls tools are reminiscent in some ways of the European “spears” and are associated with large well-formed cleavers from the Acheulian bearing Floor 2, below the Sangoan. Given their context these wood tools might be older than those from Europe and might point to a wooden projectile point technology in the late Acheulian, complimenting the earlier LCT technology. At most sites, the only clue would be in finding injury patterns on faunals remains indicative of such activities. In a similar vein, the co-occurrence of LCTs and projectile point technology in the Sangoan and Fauresmith may reflect similar activity patterns, or as McBrearty [22] suggests that the mix of technologies may, in fact, represent different hominins using different technologies at the same time in the same regions of Africa.

All this has somewhat blurred the distinction between the terminal Acheulian, the transitional industries, and the
earliest MSA and highlights the need to perhaps define a transitional phase as per Clark [28] or to perhaps redefine the MSA as an entity as suggested by Beaumont and Vogel [47] and Van Peer et al. [65]. Certainly, the end of an industrial complex should not be defined on the last appearance of its classic tool form but by the introduction of new technology. However, in some views, this change should only be defined when this new form has become dominant [27], creating a rather arbitrary line. Goodwin and van Riet Lowe [26] defined the MSA as being intermediate between the ESA and LSA, both temporally and technologically. It was in part defined by a lack of LCTS at its upper age limit and lack of microliths at its younger age limit. MSA industries such as the Howieson’s Poort have shown that microlithic technology does occasionally occur within the later phases of the MSA, perhaps in response to distinct climatic fluctuations (68–57 ka [50]). The MSA was later recognized as being blade based, rather than flake based, and that it included the common use of the Levallois technique [120, 170]. The sites described above show that all these markers of the MSA, that is, blades, Levallois technology begin to occur contemporaneously with LCTS between 548 and 272 ka in both eastern and southern Africa [107, 108, 148]. Kathu Pan and Bundu Farm suggest that even at this early time period LCTS seem to make up a much smaller component of the stone tool assemblages than at earlier Acheulian sites.

Klein [171] notes that many researchers (e.g., [39, 40, 172]) have moved beyond Goodwin’s 1928 [49] classification of the MSA to include any industries which are characterized principally by retouched and unretouched flakes of various kinds and/or which are chronologically intermediate between the hand axe industries of the ESA and the largely microlithic industries of the LSA. Klein notes that the prominence of triangular flakes with convergent dorsal scars and faceted butts is no longer an important criterion as they do not occur in all assemblages, many of which would be considered “classic MSA”. The example Klein [171] provides is the Mossel Bay Industry from Cape St. Blaze Cave. Unfortunately, this is an industry whose age is still questionable. If a similar age to material from the nearby caves at Pinnacle Point it could be of MIS 6 age (200–120 ka) but there are definite differences [123], so perhaps it is even older or represents another part of MIS 6 variation. As most lithic specialists would concede Goodwin emphasized in his early papers that the MSA was not homogeneous in time and space and in this sense many researchers would see the classification of the Fauresmith into the ESA or MSA a semantic one or an attempt at pigeonholing. However, in archaeology, as with geology, classifications, frameworks, and the boundaries between these entities need to be established somewhere. With the ever increasing age for the first LSA industries based on the presence of microlithics and mode 4 and 5 technology and the retention of LCTS (particularly hand axes) as the defining character of the Acheulian, the MSA is beginning to be squeezed into an ever decreasing age range with little potential significance and no association with the hominin record.

Cornelia-Uitzoek and arguably Elandsfontein and Dui-
nefontein II are Acheulian assemblages that date to around 1.1–0.8 Ma, and all three do not contain prepared core technology or other technology reminiscent of the MSA. A such, they seem to represent what might be termed a middle Acheulian period between the earliest Acheulian found prior to 1.1 Ma at Sterkfontein and the oldest Vaal River deposits and the final Acheulian period that contains Victoria West Giant prepared core technology such as Canteen Kopjie. It is clear that the MSA first occurs before the advent of anatomically modern humans, and yet, there is also potential evidence that some of these modern human are also still utilising ESA technology. The co-occurrence of MSA technologies at Florisbad with H. helmei at ~260 ka and H. sapiens with LCTS at ~160 ka is the opposite of what might be expected if these industries were being made by different hominins based on the exclusive co-occurrence of the MSA with H. sapiens fossils after 120 ka. Many researchers see the beginning of the Acheulian as a clear change in hominin behaviour related to the appearance of H. ergaster [29] and the middle Acheulian site of Elandsfontein, without prepared core technology is associated with H. erectus or very early H. rhodesiensis. The first occurrence of MSA-like characters in the Fauresmith is a time period where archaic forms of Homo sapiens begin to occur in Africa in the form of Homo rhodesiensis and then H. helmei. Unfortunately, the Middle Pleistocene hominin record of southern Africa is very spartan and still not well dated. Certainly, the traditional start of the MSA between 300 and 200 ka associates it with H. helmei fossils such as the Florisbad cranium and not H. sapiens. It seems likely that the beginnings of MSA style technology in the Fauresmith began with H. rhodesiensis and gradually evolved through time to incorporate a greater range of technology, including diversity in projectile point, composite stone tool, and Levallois technology by the time H. helmei is first seen soon after 300 ka. By the time the Fauresmith is noted, it has all the features of the MSA with small prepared cores and points that may very well represent the beginning of composite tool technology and also projectile points. A number of the Kathu Pan tools certainly look like they were made as projectile points (see [148, Figure 6]. The Sangoan has certainly been suggested to contain some of the first hafted stone tools, and it is perhaps this major technological change along with the classic MSA forms that occur in the Fauresmith that makes these two entities either regional variants at the beginning of the MSA or perhaps temporally distinct industries during a transitional phase between the ESA and MSA. The switch is in essence one of scale with the period represented by LCTS and prepared cores for making LCTS representing the beginnings of a transitional period in both stone tool industries and hominin species. As the Acheulian is inextricably linked to H. ergaster and H. erectus, so too, the beginnings of MSA technology appear to occur in the Fauresmith with the advent of archaic H. sapiens (H. rhodesiensis/H. helmei) and come to full flourishion before the advent of the first modern humans at ~200–150 ka. Only better dating of the Fauresmith and Sangoan sites will help answer what will most likely turn out to be an extremely complex process of transition from the ESA to the MSA.
Acknowledgments

The authors thanks Darren Curnoe for inviting him to write this paper for this special issue of IJEB and all the researchers he has worked with in South Africa over the last 15 year and who have granted him access to work or visit the various sites (David Braun, James Brink, Ron Clarke, Kevin Kuykendall, Curtis Marean, David Morris, Kaye Reed, and Lyn Wadley). A special thanks to the late Tim Partridge who strived to understand the age of the early hominin bearing sites of southern Africa. He thanks also the staff of the University of Liverpool Geomagnetism Laboratory where he has undertaken much of his palaeomagnetic work. This work was supported by ARC Discovery Grant no. DP0877603 and research money from the UNSW Faculty of Medicine. He thanks two anonymous referees for extremely useful comments on an earlier draft.

References


occupations using stratigraphy and geochronology of fossil-
spring tufas, Kharga Oasis, Egypt,” Geoarchaeology, vol. 19, 

Germany: Is there anything out there?” Quaternary Interna-

[169] E. Werker, “780,000-Year-old wood from Gesher Benot 


[171] R. G. Klein, “Problems in the Study of the Middle Stone Age 
of South Africa,” The South African Archaeological Bulletin, 

[172] C. G. Sampson, “The Middle Stone Age industries of the 
Orange River Scheme area,” Memoirs of the National Museum 
(Bloemfontein), vol. 4, 1969.
Review Article

Upper Pleistocene Human Dispersals out of Africa: A Review of the Current State of the Debate

Amanuel Beyin

Turkana Basin Institute, Stony Brook University, SBS Building 5th Floor, Stony Brook, NY 11794, USA

Correspondence should be addressed to Amanuel Beyin, abeyin@ic.sunysb.edu

Received 15 October 2010; Revised 22 January 2011; Accepted 24 February 2011

Academic Editor: Darren Curnoe

Copyright © 2011 Amanuel Beyin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Although there is a general consensus on African origin of early modern humans, there is disagreement about how and when they dispersed to Eurasia. This paper reviews genetic and Middle Stone Age/Middle Paleolithic archaeological literature from northeast Africa, Arabia, and the Levant to assess the timing and geographic backgrounds of Upper Pleistocene human colonization of Eurasia. At the center of the discussion lies the question of whether eastern Africa alone was the source of Upper Pleistocene human dispersals into Eurasia or were there other loci of human expansions outside of Africa? The reviewed literature hints at two modes of early modern human colonization of Eurasia in the Upper Pleistocene: (i) from multiple *Homo sapiens* source populations that had entered Arabia, South Asia, and the Levant prior to and soon after the onset of the Last Interglacial (MIS-5), (ii) from a rapid dispersal out of East Africa via the Southern Route (across the Red Sea basin), dating to ∼74–60 kya.

1. Introduction

Central questions to the study of human evolution are when and where modern humans first evolved, and how they colonized areas outside their place of origin [1–5]. Fossil, genetic, and archaeological data currently accumulating from sub-Saharan Africa supports an African origin of modern humans sometime between 200 and 150 kya [6–10], but the timing and the geographic routes along which they migrated out of Africa have never been resolved. Northeast (NE) Africa, Southwest Asia (the Levant), and the Arabian Peninsula have been major foci of research on this topic [5, 11–17]. So far, the oldest cultural traces and fossil remains of human ancestors have been discovered in East Africa, specifically in Ethiopia [18–20]. With the advent of intensive research on modern human mitochondrial DNA (mtDNA), paleoanthropology attained a turning point; among other things, the genetic data reinforced the Out-of-African origin theory and set possible scenarios for human dispersals [21–23]. African populations display greater genetic diversity, implying that Africa was populated by early human ancestors longer than any other region [10, 21]. From this, some geneticists hypothesize that all modern human genes descended from a single population—dubbed “African Eve”—that lived in sub-Saharan Africa around 150 kya [ibid.]. The Levant and the Arabian Peninsula play a central role in the current debate about hominin (humans and their ancestors) dispersal history due to their strategic location at the main gateways of biographic movements between Africa and Eurasia. Both regions have yielded rich Paleolithic and paleontological data that should be useful in any discussions of prehistoric migrations out of Africa and vice versa.

This paper reviews existing archaeological and genetic data from NE Africa (that part of East Africa north of the Equator) and its adjacent regions, to assess the timing and geographic background of Upper Pleistocene (128–12 kya) human colonization of Eurasia. The underlying question is whether eastern Africa alone was the source of Upper Pleistocene human dispersals into Eurasia or were there other loci of human expansions outside of Africa? This analysis differs from that of African versus multiregional origin-related discussions in that this paper focuses specifically on the Upper Pleistocene colonization process of Eurasia by early modern humans (*Homo sapiens*) as opposed to dealing with the geographic origin of *Homo sapiens* as a species. The theoretical premise here is that *Homo sapiens* originated in Africa after 200 kya. The lingering issue is how soon they formed effective founder populations outside of
Africa. In the context of this paper, the term “dispersal” is used synonymously with the term “migration,” and early human dispersal is regarded as a range expansion or niche broadening mechanism in search for stable habitats that, once encountered, become preferred habitats [24, page 251]. The paper concludes with a brief discussion of paleoclimate and coastal habitats in light of their potential role in stimulating prehistoric human dispersals.

The question of what specific conditions stimulated human migration out of Africa has not been satisfactorily answered, but it is important to examine the mechanisms by which early humans successfully colonized Eurasia. Climate change [25, 26], population expansion [3, 27], increased alliance network [28, 29], projectile technology [29, 30], and the emergence of complex behavior associated with the use of language and symbolism [31–33] have been ample tools used upon to explain Upper Pleistocene dispersals out of Africa, but researchers remain divided in testing the contribution of each variable. Some of these traits lack concrete archaeological traces (e.g., language), while others (lithic technology and symbolic relics) are inconsistently represented across regions. The fact that dispersal is not a unique phenomenon to humans makes it difficult to identify a single cause or to establish a universal model for early modern human dispersals. For instance, paleontological studies show that numerous terrestrial mammals (e.g., several carnivores, the straight-tusked elephant, hippo, and some primates) have dispersed from Africa to Eurasia in the Plio-Pleistocene span [34, 35]. In essence, if dispersal is a response shared by other organisms, it is likely the case that the forces that stimulated dispersals in other species may have triggered early modern human migrations out of Africa.

2. Upper Pleistocene Human Dispersals out of Africa: A Theoretical Review

Granted that Homo sapiens first emerged in Africa, the question of whether they left Africa in a single dispersal event or in multiple episodes has never been resolved. Within the Out-of-Africa theory, there are three principal models for human dispersals out of Africa (as reviewed by Ambrose [28]). The first one, commonly referred to as the “Strong Garden of Eden” [23], or the “African Eve” [36], model posits that after modern humans first emerged in Africa between 200–150 kya, they dispersed to other parts of the world after 100 kya and replaced all archaic hominins without any genetic admixture. The second model known as “Weak Garden of Eden” [23] asserts that anatomically modern humans evolved in a restricted area of Africa sometime after 200 kya and dispersed to separate regions around 100 kya replacing all archaic forms. According to the latter model, each human population is thought to have passed through a bottleneck in their respective destinations and then recovered after favorable conditions returned. Advocates of this model claim that around 50 kya, Later Stone Age/Upper Paleolithic (UP) cultural innovations triggered population expansion within isolated groups in Africa and Eurasia. The third dispersal model associated with the Out-of-Africa origin theory argues for multiple dispersals of early modern humans out of Africa. Originally posited by Lahr and Foley [1], this hypothesis recognizes three dispersal events in the Upper Pleistocene, and perhaps a fourth one recently proposed by Armitage [5].

2.1. Eastward Dispersal from NE Africa to Arabia (150–130 kya). Just proposed on the basis of a recently discovered Paleolithic site (Jebel Faya) in the United Arab Emirates (UAE), dating to as old as 127 kya [5], this event is thought to have involved human crossing of the Red Sea basin via the Strait of Bab al Mandab during the terminal phase of Marine Isotope Stage (MIS)-6, in the time range of 150–130 kya (date range by the present author). The observed lithic affinity between Jebel Faya and contemporaneous Middle Stone Age (MSA) assemblages in NE Africa suggested to researchers [5] that a direct route of migration may have existed between the two regions during low sea level events associated with MIS-6, a glacial phase dating ~180–125 kya [37]. This event is believed to have resulted in long-term human occupation of Arabia from which later expansions to other parts of Eurasia may have commenced [5]. However, whether this dispersal represents the first expansion of early modern humans out of Africa and if indeed the Jebel Faya remains were made by modern humans are unclear at this point because no characteristic hominin fossil remains have been discovered at the site yet. At best, the evidence indicates the presence of hominins in eastern Arabia with a similar lithic technology as the makers of the Middle Stone Age (MSA) tradition in Africa. A relevant evidence in connection to this route is the recent discovery of modern human fossil remains in Southern China, at the cave site of Zhirendong [38]. Dated to >100 kya, the Zhirendong evidence signifies the presence of a successful hominin dispersal into Southeast Asia prior to 100 kya, possibly through the Bab al Mandab during MIS-6.

2.2. Northward Dispersal to Southwest Asia (130–90 kya). Before the recent discovery at Jebel Faya [5], the modern looking fossil remains associated with Middle Paleolithic (MP) industries at the Levantine cave sites of Jebel Qafzeh and Skhul, dating in the range of 130–90 kya [39–41], were considered to be the first dispersal of early modern humans from Africa [1]. As such, this event has been variously described as an “early dispersal of anatomically modern humans” by Finlayson [26] or “temporary dispersal of anatomically modern populations” by Mellars [3]. However, the identity and evolutionary significance of the Qafzeh/Skhul populations remain less well understood [42]. It is unclear whether they represent dispersed African hominins or in situ evolution from pre-existing Levantine archaic forms. Another intriguing question is what role did the Qafzeh/Skhul populations play in the Upper Pleistocene colonization of Eurasia? Perhaps, some lineages of this group specifically adapted to the Levantine environments may have served as source populations for later expansions to Europe and East Asia. Alternatively, the Qafzeh/Skhul hominins may have migrated to Arabia [12, 28] or to North African coastal refugia [43] in response to overpopulation or climate changes associated with MIS-4, a glacial episode dating between 75 and 60 kya [44] (Figure 1(b)).
2.3. Rapid Dispersal from NE Africa to Southeast Asia (80–60 kya). Initially posited by Kingdon [45], advanced by Lahr and Foley [1], and lately by geneticists and archaeologists [3, 11, 28, 46], this dispersal event is thought to be responsible for the colonization of Southeast Asia and Australia around 60 kya. The key assumption behind this hypothesis is that when northward movement from East Africa into the Levant was blocked due to the expansion of the Sahara Desert during MIS–4, some East African early human groups launched an eastward expansion via the Strait of Bab al Mandab (southern Red Sea) before they successfully dispersed to Europe and western Asia. Dubbed “rapid dispersal,” this event is typically associated with coastal and estuarine adaptations along the circum-Indian Ocean [4, 45, 47]. There is limited fossil and archaeological evidence supporting this event (the Zhirendong evidence [38] might fit into this route, but that belongs to earlier dispersals). The rapid dispersal event is mainly supported by mtDNA analyses of some aboriginal populations in Southeast Asia and Australia [11, 46, 48, 49]. For instance, Macaulay et al. [11] recently identified mtDNA types M21 and M22 in their Malaysian sample, and Thangaraj et al. [50] discovered M31 mtDNA type among indigenous Andamanese tribes. All these subclades stem from haplogroup M, which is believed to have originated as a founder around 63 kya [11] or around 65 kya [51]. Haplogroup M is a direct descendant of the ancestral L3
haplogroup, which first appeared as a founder in Africa ∼85 kya [11, 51]. The short time gap between the coalescence date of L3 (around 85 kya) and the arrival of its descendants in South Asia shortly afterwards (∼65 kya) suggested to researchers that the initial colonization of South and East Asia involved rapid coastal dispersal of human groups from NE Africa via the Strait of Bab al Mandab andArabo-Indian littorals (Figure 3). According to Oppenheimer [46] and Macaulay et al. [11] the initial exodus into Arabia dates to roughly 85–70 kya, resulting in the initial colonization of Australia at about 60 kya, with a dispersal rate of 0.7 km/year.

2.4. Reoccupation of Southwest Asia around 50 kya. Two possible source areas have been suggested for the reoccupation of Southwest Asia by anatomically modern humans around 50 kya: (i) expansion of Later Stone Age populations from NE Africa toward the eastern Mediterranean Levant through the Nile corridor during MIS-3 (59–27 kya) warm intervals [1, 14], (ii) a south-north expansion from Arabian Peninsula to the Levant by human groups who previously entered the Arabian Peninsula either through the Bab al Mandab or through the Levant [4, 12, 52]. Essentially, humans who successfully occupied Southwest Asia and Europe after 50 kya are believed to have encountered Neanderthals [53]. Mellars [54] describes two phases of modern human dispersals from Southwest Asia into Europe: (i) initial occupation of southeastern Europe around 43 kya, (ii) followed by a westward movement along the Danube Valley that led to the occupation of central and southern Europe around ∼40 kya.

3. How Many Dispersal Routes?

Currently, there are two widely accepted dispersal routes for early modern humans out of Africa [1, 11, 13, 45, 55, 56]. These are the Northern Route (NR) via the Nile-Sinai-Land Bridge and the Southern Rout (SR) through the Strait of Bab al Mandab at the southern end of the Red Sea (Figure 1). What follows below is a brief discussion of the two proposed routes and associated evidence in the context of Upper Pleistocene span.

3.1. The Northern Route. Advocates of this route propose successive dispersals of hominids and early modern humans from NE Africa into the East Mediterranean Levant via the Nile-Sinai Corridor [17, 45, 56–58] (Figure 1(a)). The NR is often linked with the arrival of the Qafzeh and Skhul hominins into Southwest Asia during MIS-5 (an interglacial phase dating ∼130–74 kya) [44]. Although the presence of a dry land bridge across Sinai makes this route most accessible, considering environmental barriers associated with the aridity of the Saharan Desert, this route may have been most preferable during wetter climatic conditions [13, 59, 60]. Attempts to link the Nile Valley MP archaeology with the Levantine record do not always indicate a strong cultural connection in the Upper Pleistocene [16, 61]. After the Qafzeh and Skhul groups, no modern human presence has been detected in the Levant until after 50 kya [53]. By this time, humans had already reached Australia [62, 63]. This long gap in the fossil record for modern human reappearance in the Levant may imply an interruption of the East African-Levantine connection. One possible explanation for this is that the extreme cooling of the North Atlantic during MIS-4 glacial episode may have caused the Intertropical Convergence Zone and its rainfall belt to move southward closer to the Equator [25]. In this regard, Carto et al. [25, page 149] note that “the more open and arid landscape in the Sahel region may have acted as a major barrier to early H. sapiens dispersal at this time, restricting movement into North Africa.”

Outside of the Nile Corridor, two additional dispersal corridors from East Africa into the Levant seem to have been present in the Upper Pleistocene: one to the west through central Sahara to North Africa [59] and the other to the east along the western coastal margins of the Red Sea basin [45, 65] (Figure 1(a)). Using geochemical data from wadis and by locating several fossil river channels in the Libyan Desert, Osborne et al. [59] recently demonstrated that humid corridors existed through the central Sahara during the Last Interglacial (specifically in the time range of 130–117 kya). On the basis of this evidence and the discovery of some discrete lithic traditions in the region (specifically Aterian sites), the authors propose that the central Sahara/Libyan Desert may have served as an alternative dispersal route for early humans from central and NE Africa into the Mediterranean coast. Populations that took the central Saharan route may have ultimately dispersed to Southwest Asia along the Mediterranean coast (Figure 1). Perhaps, the Qafzeh and Skhul remains may represent an off-shoot of this dispersal route.

A less explored, but seemingly a vital region in the dispersal history of early humans is the western coastal littoral of the Red Sea basin [60, 65] (Figure 1(a)). So far, there is only one well-dated Upper Pleistocene site from the entire western coastal strip of the Red Sea—the Abdur site along the Gulf of Zula, on the Eritrean coast [66]. The site produced lithic artifacts indicating MSA affinities and some bifacial tools found within an emerged coral reef terrace of the Last Interglacial period. The artifact bearing Reef Limestone at Abdur dates to 125 ± 7 kya [ibid.]. While there are currently some questions regarding the context of the artifacts, the chronological placement of the Abdur Reef broadly coincides with the purported northward migration of Homo sapiens into the Levant (e.g., Qafzeh and Skhul). Recently, a surface MSA assemblage has been recorded along the Eritrean Red Sea coast at a site called Asfet, about 20 km southwest of Abdur [67, 68]. The site is located at less than 1 km from the present shoreline. While the Asfet assemblage lacks an absolute age, it broadly reflects MSA adaptation. Homo sapiens that successfully adapted to the Eritrean coast in prehistoric times may have continued moving northward up to the Sinai land bridge following a coastal route. Even if there is limited archaeological evidence from the Sudanese and Egyptian coasts (largely due to inadequate research), there is no conceivable obstacle for northward dispersals of early humans along the Eritrean-Sudanese-Egyptian coastal landscapes at any point in time.

3.2. The Southern Route. This route proposes that Upper Pleistocene foragers specifically adapted to coastal habitats
in NE Africa dispersed into Southern Arabia via the Strait of Bab al Mandab during low sea level events associated with major glacial episodes [1, 4, 5, 11, 45, 46, 55, 60] (Figures 1 and 3). Once they entered southern Arabia, those human groups may have dispersed eastward using a coastal route along the Yemen-Oman littorals up to the Persian Gulf and South Asia, or they might have traveled northward up to the Levant. Most advocates of this route link it with the initial colonization of East Asia and Australia [ibid.]. The SR is mainly supported by genetic data [11, 48, 50], discussed
above. There are some technotypological similarities between Paleolithic sites in NE Africa and the Arabian Peninsula indicating sporadic Upper Pleistocene cultural contacts [5, 69, 70]. Generally, however, the role of the SR to long-term human occupation of Arabia and Southeast Asia remains unclear due to the lack of fossils and secured dating for the sites.

Indirect evidence suggesting episodic cultural connections between Africa and Arabia comes from a recent study concerning hamadryas baboon phylogeographic history [71, 72]. The hamadryas baboon (Papio hamadryas hamadryas) is found exclusively in East Africa and western Arabia, and is the only free-ranging non-human primate in Arabia [71, 73]. Previously, it has been hypothesized that hamadryas baboons colonized Arabia in the Holocene [73]. However, a recent study of mtDNA variation among Saudi Arabian and East African (Eritrean) hamadryas populations shows that these baboons did not colonize Arabia in the recent past nor did they use a northerly route via the Sinai land bridge to enter Arabia Peninsula. The data suggests that hamadryas baboons reached Arabia via temporary land bridges formed during glacial maxima along the Strait of Bab al Mandab. Given the fact that the most recent common ancestor of hamadryas baboons lived in Arabia around 85–119 kya [ibid.], the likely time for hamadryas entrance to the Peninsula has been estimated ~86–220 kya [ibid.]. If such a recent dispersal of hamadryas baboons was possible across the Red Sea (Strait of Bab al Mandab), this route must have been equally accessible to early modern humans during low sea level events in the Upper Pleistocene, such as MIS 6, 4, and 2 [74]. There is no evidence for a prolonged interruption of water exchange between the Red Sea and the Indian Ocean throughout the last 470 kya years [75]. Therefore, any faunal or hominin exchange along this route should have occurred during temporary land bridges. Otherwise, human migration across the Red Sea during humid periods must have involved crossing a body of water using floating objects or by swimming, which is difficult to prove archaeologically.

4. Single or Multiple Centers of Upper Pleistocene Human Dispersals?

Archaeological and Genetic Perspectives

4.1. Archaeological Perspectives. The archaeological assessment here focuses on the general aspects of MSA/MP lithic industries from NE Africa, the Levant, and Arabia Peninsula. The degree to which lithic techno-typology could be informative about hominin dispersal history requires some clarification. In the older literature, it has been suggested that typological similarities signal population movement, while technological parallels may reflect similar adaptive behavior between comparative regions [76, 77]. But, such criteria may not work consistently across broad geographic areas because lithic technology is contingent upon the local habitats, ecological factors (raw material), the intended use of the artifacts, and idiosyncratic choices of the knappers [78]. For instance, when we look at the cultural traces found in association with the modern looking fossil remains in Australia, they do not show any affinity with Middle or
Later Stone Age (LSA) traditions of Africa or Europe [4], signifying that there is not always direct association between lithic traditions and specific hominin groups.

By and large, the Upper Pleistocene human dispersals out of Africa correlate with the MSA tradition in Africa or MP in Eurasia, broadly dated between 250 and 40 kya [8, 29, 81]. The MSA is best characterized by the use of Levallois technique and point production [29, 82]. The distinctive features of the Levallois technique are striking platform preparation and predetermination of the resulting flakes [83]. The relative paucity of scrapers and the presence of a large number of foliate points distinguish the African MSA from the MP of Eurasia [29]. Paleolithic scholars working in Africa identify numerous behavioral innovations associated with the MSA cultural entity. These include (i) technological standardization in producing diverse point styles, (ii) abstract thinking and symbolic behavior manifested in the use of ochre, geometric engraving design, and use of body adornments, (iii) planning and extensive alliance networks reflected in long distance raw material transportation, and (iv) use of aquatic resources [29, 84–86]. These behavioral complexes are thought to have stimulated economic diversification and dispersals within and out of Africa in the Upper Pleistocene [ibid.]. Of these, projectile technology is considered to be an important innovation that enabled modern humans to broaden their spatial and dietary niches [29, 30, 87, 88].

As stated above, East Africa has yielded the oldest modern looking fossil hominins [6, 7], and the earliest MSA industries characterized by regionally diversified point technology produced using façonnage technique, and a large range of flake tools from Levallois and blade cores [8, 20, 29, 82, 89]. Two main technocomplexes: the “Nubian” and “Lower Nile Valley” characterize the MP tradition in the Nile Valley area. The relationship between these complexes and those of the Levantine MP and East African MSA remains poorly understood [16, 17, 78]. In general, the façonnage technique distinguishes the NE African MSA industries from the Levantine and the Nile Valley MP assemblages. Façonnage method involves tool production from a core by invasive and usually bifacial flaking [90]. Middle Paleolithic assemblages with façonnage technique have been recorded in some MP sites in Arabia [5, 15], but not from the western region—the area expected to receive the most east African MSA cultural influence. The lack of strong technological similarity between east Africa and western Arabia has become an obstacle for researchers trying to glean evidence for early human contacts via the Strait of Bab al Mandab. In this regard, Marks [69, page 302] writes “given the proximity between the western Yemeni littoral and the eastern coast of East Africa, it is strange that no clearly African or even African related assemblages have been found.”

The role of the Levant in clarifying early human migration out of Africa has long been recognized [16, 58, 91]. Indeed, the region has produced rich paleontological and paleoanthropological evidence that should be informative about hominin migration history to and from Africa [92–95]. Initially, the Levantine record was vital in supporting the gradual emergence of modern humans out of the Neanderthal line [96]. However, with recent improvements in dating techniques [97, 98], it has become evident that Neanderthals and early modern humans in the Levant were contemporaneous, both manipulating Levallois technique, sharing similar habitats [99] and possibly interbreeding [64]. The MP archaeological complexes in the Levant are collectively referred to as the Levantine Mousterian, which are believed to have evolved from local Late Acheulian and Mugharan industries that were mainly confined to the northern Levant [16, 100–102]. Laminar and recurrent Levallois techniques for the production of blades and triangular to subtriangular flakes characterize the Levantine assemblages [ibid.]. They are mainly distinguished from the African MSA assemblages by the absence of bifacial/façonnage reduction. One problem with the Levantine MP is that at least two species of Upper Pleistocene Homo had been associated with it. For instance, the Levallois technique, which is one of the defining features of MSA/MP assemblages, was produced by Neanderthals and early modern humans in this region [99]. This makes it difficult to discern local technological innovations from those brought by the dispersing humans.

Although Arabia had a marginal mention in human evolutionary discussions in the past, past and recent research activities in the region have produced a plethora of prehistoric data spanning from the Lower Palaeolithic up to the Neolithic times ([103] and references therein). When it comes to the MP record of the region, Marks [69] identifies three variants corresponding to three ecogeographic zones: (i) the western and south-western regions mainly characterized by convergent Levallois and blade technology, (ii) the South (mainly Oman), featuring unidirectional hard hammer, large blade production from cores with flat platform, and the production of bifacial foliates using façonnage, (iii) sites on the eastern region, exhibiting a number of different reduction strategies, such as façonnage to produce both small handaxes and foliates, blade production from multiple platform and unidirectional cores, and flake production from crude Levallois and centripetal radial strategy. Despite the existence of abundant Paleolithic sites in Arabia, the findings lack reliable dates and hominin fossil association. Some recently discovered MP industries on the western and southern parts of Arabia contain characteristic Levantine technological features, such as heavy emphasis on unidirectional Levallois technique for the production of convergent points and blades [69, 104, 105]. The Levantine-related Levallois method is poorly represented in Oman and in those areas east of the Rub‘al Khali Desert; instead the sites there display a mosaic of local technological developments and some East African MSA affinities [69, 104].

Somewhat appealing evidence concerning Upper Pleistocene cultural connections between NE Africa and the Arabian Peninsula comes from the recently excavated MP assemblage at the site of Jebel Faya 1, UAE [5]. Radiometrically dated to ~127–95 kya, one stratified assemblage at the site (Assemblage C) contains both small handaxes and foliate points produced by façonnage technology, revealing close technological similarity with NE African MSA industries [5, 69]. Two younger assemblages from the site (A and B) indicate no apparent technotypological affinity with...
the Levantine, NE African, or other assemblages in the Arabian Peninsula [69, page 305]. Interestingly, a surface MSA assemblage similar took kit to that of Jebel Faya Assemblage C has recently been documented on the Eritrean Red Sea coast, at a site called Asfet [67]. Although the Asfet material lacks an absolute date, the presence of bifacial foliates and small handaxes in association with various kinds of blades/flakes produced by Levallois and parallel method signifies similar adaptive behavior and/or culture-historical relationship between the inhabitants of Asfet and Jebel Faya. The author hopes to undertake detailed comparative investigation of the two assemblages in the near future.

Overall, the Arabian MP lithic entities, especially those on the western margin, do not show strong affinity with their east African MSA counterparts. The findings from the western side of the Peninsula suggest stronger cultural connection with the Levant. However, whether this is due to population diffusion from the Levant into western Arabia and vice versa, or due to adaptation to similar environmental settings, remains unclear. It is worth noting that the western side of Arabia and southern Levant exhibit similar environments, featuring coastal plains and high plateaus [69]. Thus, the vast strip of land stretching from southern Levant up to southwestern Arabia may have served as an ecological refugium and a conduit of continuous prehistoric human contacts [69, 106].

On the basis of new paleoenvironmental, archaeological, and genetic evidence from the Arabian Peninsula and southern Iran, Rose [12] has recently formulated an enticing model dubbed the “Gulf Oasis,” in which he proposes “early modern humans were able to survive periodic hyper-arid oscillations by contracting into environmental refugia around the coastal margins of the peninsula” [ibid., page 849]. Based on this model, the author hypothesizes that the Persian Gulf may have been a source area for Upper Pleistocene human diffusion to East Asia, to Southwest Asia, and possibly to East Africa. At the center of the model lies the assumption that sea level decline during arid periods (associated with glacial episodes) enhanced the discharge of fresh water from terrestrial aquifers along the continental shelves, thereby creating favorable habitats along the gulf basin when much of the hinterlands were dry [107]. The presence of subterranean freshwater upwells beneath the Persian Gulf has been confirmed by geological studies (see detailed literature review on this topic in Rose 2010 [12]). While the Gulf Oasis model sets a plausible scenario for Upper Pleistocene hominin diffusion from Arabia, it can be critiqued based on the lack of evidence (so far) for in situ origin of any of the two ancestral mtDNA—haplogroups M and N—in Arabia [108, 109].

Outside of Arabia to the east, several MP sites have been documented in India [110, 111]. However, the cultural-historical relationship of the region to the African continent has not been systematically investigated yet. An ongoing archaeological investigation around the Jurreru Valley (India) has documented cultural traces in a deposit predating the Toba eruption ~74 kya, but the exact cultural context of the artifacts has yet to be determined [112]. Recently, Mellars [3, 4] reported the presence of a strong resemblance between Later Stone Age (LSA) assemblages in India and Sri Lanka (dated to ~34 kya) and some derived MSA tools and symbolic objects from sub-Saharan Africa (e.g., crescents, geometric engravings, and shell beads). According to Mellars [ibid.,], these similarities imply a rapid wave of modern human dispersal from eastern Africa carrying diverse cultural innovations, thereby supporting the “single dispersal” hypothesis for Upper Pleistocene human colonization of South and East Asia. In this regard, Mellars suggests that MSA behavioral innovations that first appeared in southern and eastern Africa played a key role in fueling human dispersals into Eurasia. Although Mellars’ assertions appear plausible, the apparent lack of such artifacts at other (relatively older) fossil sites in the region, such as Niah Cave [113], and Lake Mungo [114], calls into question the role of the purported cultural novelties in Africa in fueling human dispersals into Eurasia. After all, it is not clear if, indeed, the Indian LSA sites were derived from the rapid dispersal event discussed above. Perhaps, some of the fossil and archaeological sites in this region may represent local evolution from preceding modern human settlements, such as the one reported from Zhireshondong [38].

4.2. Genetic Perspectives. Genetic studies of modern human populations show that the most recent common ancestor of all humans—referred to as the “Mitochondrial Eve” or “Real Eve”—lived in Africa between 200 and 150 kya [21, 115–117]. The mitochondrial deoxyribonucleic acid (mtDNA) has been particularly useful for studies of human evolution and prehistoric migrations [49, 118, 119]. Its high copy number, maternal inheritance, lack of recombination, and high mutation rate have made mtDNA an important molecule in tracing human descent back to several thousand years [118]. The basic premise of the genetic approach is that similarities in mtDNA are due to shared descent. Harpending et al. [117] described the main tenets of the genetic approach to human evolutionary studies as follows: (i) all humans originated from a population that was effectively a separate species for at least 1 or 2 million years, (ii) human genetic variation is relatively low indicating that the number of effective population size was roughly on the order of 10,000 breeding individuals throughout the Middle to Upper Pleistocene or 10,000–20,000 throughout the Pleistocene [120], (iii) a specific population ancestral to modern humans underwent demographic expansion in Africa, subsequently occupying Eurasia and the New World, and (iv) the best available estimates of mtDNA mutation rates imply that the expansion occurred between 100,000 and 50,000 kya.

Mitochondrial DNA variation among present-day African lineages suggests that Africa saw a “remarkable expansion” of two mtDNA haplogroups, namely, L2 and L3, between 85 and 55 kya [11, 49, 51]. Subsequently, a small population carrying L3 haplotype is thought to have left East Africa via the SR ~70–65 kya (Figure 3), giving rise to two founding sister haplogroups outside of Africa, namely, M and N [49]. This means that all contemporary non-African humans have descended from either M or N haplogroups, both derived from the ancestral African haplogroup L3. The age of haplogroup L3 is estimated to be ~85 kya and that of
haplogroups M and N ∼63 kya [11, 121]. The two descendant haplogroups have somewhat distinct geographic distribution with the M lineages mainly concentrated among eastern African, South Arabian, Indian, and Southeast Asian populations, whereas descendants of haplogroup N are mainly found in western and central Eurasia [49, 122]. Due to its presumed older age and widespread presence in eastern Africa and southeast Asia, haplogroup M or its African type clade M1 has been considered a vital genetic marker for tracking Upper Pleistocene human dispersals [123]. Moreover, the absence of haplogroup M clades among Southwest Asian populations has suggested to some researchers that the Arabian Peninsula may have served as the sole conduit of human dispersals out of Africa, thus lending support to the SR hypothesis [123, 124]. A recent genetic study by Macaulay et al. [11] and Thangaraj et al. [50] discovered ancestral lineages of macrohaplogroup M in some native populations of Malaysian Peninsula and Andaman Islands, respectively (discussed above). From this, the authors propose a single wave of early modern human dispersal from East Africa to Southeast Asia via a southern coastal route sometime between 65 and 60 kya. Despite the emerging rebuttals (see below), the authors claim that the southern coastal route was the only successful migration that resulted in a long-term occupation of South and East Asia.

4.3. Recent Critiques and Alternative Views. The most viable critique to the single rapid dispersal hypothesis comes from four concurrent lines of evidence: (i) a growing body of recent genetic evidence that shows in situ origin of haplogroups M in South Asia [50, 125–129], (ii) the recent discoveries of a well-stratified MSA/MP assemblage in Eastern Arabia dating to 85 kya [69, 130] and modern human fossil remains from Zhirendong cave in southern China dating to >100 kya [38], (iii) a recent multivariate analysis of morphometric data that shows close genetic affinity between early modern humans from the Levant and terminal Pleistocene/early Holocene fossils from Australasia [131], and (iv) a more recent genetic study on the Neanderthal nuclear genome that shows that Neanderthals shared more genetic variants with present-day humans in Eurasia than with present-day humans in sub-Saharan Africa [64].

By examining genetic diversity among Asian and North African extant populations, Olivieri et al. [125] and González et al. [126] propose an Asian origin of haplogroup M and its back migration via the Sinai land bridge into Africa sometime between 45 and 40 kya. In this regard, González et al. [ibid., page 223] write, “The coalescence age of the African haplogroup M1 is younger than those for other M Asiatic clades. In contradiction to the hypothesis of an eastern Africa origin for modern human expansions out of Africa, the most ancestral M1 lineages have been found in Northwest Africa and in the Near East, instead of in East Africa.” Rowold et al. [132] provide additional evidence for back migration of Upper Paleolithic populations through the Sinai land bridge. The possibility of back migration from Asia to Africa has also been confirmed by other genetic studies on human Y-chromosome haplotypes [128, 129]. Further indirect evidence for an Asian origin of haplogroup M comes from recent studies on the genetic diversity of the Arabian populations, which show stronger influence of South Asian N and M haplogroups in the founding lineages of the Arabian mtDNA [109, 122, 127]. Interestingly, much of the existing genetic ties between NE Africa and Arabia have been attributed to historic contacts between Afro-Arabian kingdoms [122]. In light of the growing genetic data in favor of an Asian origin of haplogroup M (one of the two founding mtDNA haplogroups for all modern humans outside of Africa), there appears to have been an Asiatic locus/loci of Upper Pleistocene human expansions.

As discussed above, a diagnostic MSA/MP assemblage (Assemblage C) has recently been identified at the site of Jebel Faya, UAE [5]. Its date falls in the range of MIS-5e (127–95 kya) and is considered to be the oldest evidence of modern human dispersals out of Africa prior to the purported rapid coastal migration [11]. The period between 100 and 50 kya is poorly represented by fossil hominins and well-dated Paleolithic assemblages outside of Africa. In this regard, some scholars [3, 26, 46] hypothesize that the initial migration of modern humans from Africa (represented by Qafzeh and Skhul fossils) did not result in a widespread human expansion outside the Levant. The identity and contribution of the Jebel Faya hominins to the Upper Pleistocene colonization of Eurasia is unknown, but given its pivotal location en route between NE Africa and South Asia, the evidence is of a paramount significance in assessing the timing and the geographic contexts of early modern human dispersals. Early modern humans that originated in Africa around 190 kya may have subsequently migrated to Arabian Peninsula during MIS-6 through the SR. From Arabia, some modern human groups may have launched an eastward migration along the Arabian-Iranian-Indian coastal refugia reaching southern China around 100 kya [38]. Subsequently, hominins specifically adapted to those putative refugia may have served as founder populations for later human expansions into East Asia and Australia (Figure 1).

The Australian genetic and fossil records have been important in calibrating Upper Pleistocene human dispersal events at various points in time [1, 48, 114, 131, 133]. The oldest archaeological traces of early modern humans in Australia have been dated to ∼50 kya at the sites of Lake Mungo [114] and Devil’s Lair [63]. Geneticists extend the arrival of humans in Australia up to 60 kya or slightly earlier [11, 46, 133]. A number of geneticists also claim close mtDNA affinity between East Africans and Australasian aboriginals, in so doing they propose a single wave of human expansion from eastern Africa after 85 kya [ibid.]. Nonetheless, a recent multivariate analysis of cranial morphometric data has shown close genetic affinity between early modern humans from the Levant (Qafzeh and Skhul) and terminal Pleistocene/early Holocene fossils from Australasia [131]. According to the study, an early dispersal (∼100–76 kya) from Africa by a more ancient lineage of modern humans may have contributed to the initial colonization of Australasia [ibid., page 814], thereby suggesting a direct contribution of the Qafzeh and Skhul groups in peopling Eurasia during the Upper Pleistocene.
The fourth source of disagreement with the single dispersal hypothesis comes from a recent genetic study on the Neanderthal genome, which shows that Neanderthals shared more genetic variants with present-day humans in Eurasia than with present-day humans in sub-Saharan Africa [64]. From this, the authors of the study conclude that “gene flow from Neanderthals into the ancestors of non-Africans occurred before the divergence of Eurasian groups from each other” [ibid., page 710]. Although it is too early to speculate much about this finding, the study offers a plausible scenario for deep genetic roots in Eurasia and hints that the initial dispersal of modern humans into Southwest Asia associated with the Qafzeh and Skhul groups was not a failed one (Figures 1 and 2). Perhaps, that expansion may have ultimately resulted in early modern human and Neanderthal contacts/interbreeding in Southwest Asia. Modern human groups with Neanderthal genetic material may have survived in Southwest Asia throughout the succeeding millennia giving rise to some ancestors of present-day Europeans and Asians (Figure 2). In this regard, Ridl [134] proposes two possible scenarios as to why parts of the Neanderthal genome may have been preserved in present-day Europeans and Asians but not in Africans:

…the first explanation is that there was an ancient polymorphism among archaic populations in Africa. In that case, Neanderthals and all humans outside of Africa would stem from the same population whose other descendants have become lost (or minor) in Africa. The second and probably more parsimonious scenario involves interbreeding between Neanderthals and modern humans. This would mean that there was a single population of modern humans whose ancestors met (and mated) Neanderthals somewhere in southwest Asia and whose descendants subsequently colonized the rest of the world outside of Africa [page 873].

In a nutshell, the recent study on the Neanderthal genome joins the other lines of evidence discussed above [5, 50, 125–129, 131] in suggesting that several dispersals or multiple loci of human expansions (the Levant included) may have existed outside of Africa during the Upper Pleistocene.

5. Paleoclimate, Coastal Environments, and Human Dispersals

The effects of past climate changes on biogeographic distribution, extinction, and migration of species have long been recognized [26, 49, 135, 136]. In his recent review of the biogeography and evolution of the genus Homo, Finlayson [26, page 457] categorically argued that “climate-driven ecological change has been, as with many other taxa, the driving force in the geographic range dynamics of the genus Homo.” Similarly, Forster [49, page 255] notes that “most of our own mtDNA studies during the past decade strongly imply a major role for paleoclimate in determining conditions for prehistoric migrations and demographic expansions.” The period associated with the origin and dispersals of early humans (200–50 kya) saw recurrent glacial events that caused prolonged sea level decline (Figure 4). Two of such events are MIS-6 and 4, dating in the range of 180–125 and 75–60 kya, respectively, [37, 44]. Potential adaptive responses to climatic aridity may include migration, population fragmentation, and subsistence on low-rank resources.

On the basis of global climate model simulation, Carto et al. [25, page 140] recently reported that “Heinrich events which occurred episodically throughout the last glacial cycle, led to abrupt changes in climate that may have rendered large parts of North, East, and West Africa unsuitable for hominin occupation, thus compelling early Homo sapiens to migrate out of Africa.” Heinrich events are short cold periods that occur when freshwater fluxes associated with ice sheet surges into the North Atlantic caused southward shift of the Intertropical Convergence Zone over Africa [25, page 139]. During such cold times, human movements may have been restricted to coastal environments. Rose [104] in formulating his “Arabian Corridor Migration Model,” linked human dispersals into Arabia with the expansion of the savanna ecogeographic zone during wet (pluvial) episodes. Ideally, human dispersals out of Africa may have occurred during cold and warm conditions in response to population expansion, increasing competition and some autocatalytic factors whereby the discovery of new attractive areas led to the expectation of more vacant habitats [137].

There is increasing evidence showing that coastal habitats played an important role in the course of human evolution as stable refugia and corridors of biogeographic expansion [45, 66, 138–140]. Decades ago, Sauer ([140] page 311–112) proposed that “the dispersal of early man took place most readily by following along the seashore; coastwise there was a scarcely a barrier to the spread of early man through tropical and subtropical latitudes.” Sauer’s idea was later advanced by an Africanist biogeographer Jonathan Kingdon, who related early human migration out of Africa to mammalian dispersal patterns from East Africa to Southeast Asia [45]. Kingdon proposed a circum-Indian Ocean coastal dispersal for the colonization of Southeast Asia and Australia by early humans, arguing that the main prerequisites to human dispersal out of Africa were adaptation to coastal environments and raft building technology. Recently, Bulbeck [47] upon adopting Kingdon’s view hypothesized an estuarine based model for the eastward dispersal of early humans from Africa to Southeast Asia up to Australia. Accordingly, adaptation to resource rich estuaries is thought to be the “main impetus for the migratory movement” of Homo sapiens toward Southeast Asia [ibid., page 315]. The Indian Ocean coastal rims may have been particularly attractive to mobile foragers due to the availability of a mix of estuarine, coastal, and terrestrial resources. Using a Geographic Information System generated model, Field and Lahr [60, page 1] recently proposed that, “rapid dispersals along coastlines and river valleys would have occurred upon the initial expansion out of Africa, but slowed as populations expanded demographically into South Asia and the Sunda Shelf.”

During glacial times, a vast amount of freshwater is locked up in the polar ice sheets, causing sea levels to
Figure 4: Sea level changes and early modern human dispersal events in the Upper Pleistocene [3, 46, 79]. Sea level chart redrawn from Siddall et al. ([80, page 143], Figure 2).

Drop as low as 130 m below their present height [80, 141]. Such events may have created vast coastal territories for early humans to exploit. Coastal margins may have been particularly attractive when sea levels were low due to the formation of freshwater springs along the newly exposed coastal gradients [107]. According to the Coastal Oasis model proposed by Faure et al. [ibid.], the flow of groundwater along the coastal gradients increases when sea level drops during major Ice Ages. Sea level is at a high interglacial position now, thus relics of coastal adaptation associated with Ice Age coastal oases may be found underwater now. Human movement along coastal margins can be faster and safer because once humans developed the strategy to exploit one coastal territory, there is little need to device new technology to exploit another coastal habitat [47, 79, 140]. Despite all these potentially viable scenarios, there is limited coastal archaeological evidence from Southeast Asia. This could be due to successive sea level transgressions that may have inundated most prehistoric sites formed near coastal landscapes. It is worth noting that a few sites that suggest human use of tropical refugia have been documented in East Asia [47, 113].

A topic of particular interest related to this discussion is the occupation of Australia by early modern humans as early as 50 kya or slightly earlier [63, 114, 142]. The occupation of Australia is intriguing because it did not only involve coastal movement, but crossing a vast body of water possibly using floating rafts [143]. Lake Mungo, from where the oldest fossil remains of early modern humans have been recovered [144], is located about 2700 km from the northern coast of Australia, which is the likely entry point for early humans into Australia. The site dates to ~50 kya [114]. The accepted model is that humans first travelled along the western coastal rims all the way up to the southern tip of Australia before reaching Lake Mungo [144]. Another similar scenario is human colonization of North America during the terminal Pleistocene, which involved a rapid dispersal to South America along the Pacific coast before occupying the interior habitats [145]. In both cases, humans embarked on coastal colonization prior to moving into the interior landscapes.

The earliest well-dated evidence for coastal adaptation by anatomically modern humans has been reported from South Africa, at the site of Pinnacle Point dating to 164 kya [86, 146]. South Africa has also produced rich coastal evidence for the time range 120–70 kya [147–151]. The presence of a plethora of later Pleistocene coastal sites in Southern Africa may signify that the region served as an important refugium during glacial times from where later human expansions to the rest of Africa commenced when favorable conditions prevailed [146, 147]. The Abdur site on the Eritrean coast (~125 kya) offers the only evidence for Upper Pleistocene human coastal adaption in NE Africa [66]. Moreover, a recent study concerning the evolutionary history of the Giant Clam (Tridacna costata) along the Sinai Peninsula (northern Red Sea) reported a sharp decline in proportion and shell size in this species starting at ~130 kya [152]. This is interpreted as “the earliest depletion reported so far of a shallow-water megafaunal invertebrate associated with the dispersal of anatomically modern humans out of Africa into the Red Sea and adjacent regions 110,000 to 90,000 years ago” [ibid., page 2]. Although there is limited coastal evidence from eastern
Africa (largely due to lack of sustained research), any direct hominin contacts between East Africa and Arabian Peninsula must have involved coastal adaptation, and possibly raft building technology.

6. Summary and Conclusions

In spite of continued research progress in all fields of paleoanthropology, the present generation is particularly confronted with a complex set of long accumulated conflicting views about the origin and dispersal history of early modern humans. This paper is an attempt to shed some light on the current state of the scientific debate on Upper Pleistocene (128–12 kya) human dispersals out of Africa and their successive colonization of Eurasia. Researchers have often relied on archaeological, genetic, and fossil data to examine patterns of early human dispersals [11, 13, 15, 131]. However, these approaches have continued to suffer from a lack of consistent evidence from the contributing regions, such as the Nile Valley, the Levant, and the Arabian Peninsula [12, 16, 153]. Some, but not all archaeological and genetic publications reviewed here, hint at multiple loci of Upper Pleistocene human expansions outside of Africa. The main conclusions of the current paper are succinctly presented in Figures 1 and 2.

The existing MSA/MP archaeological record does not show strong technotypological parallels between NE Africa and the two adjacent regions—Arabia and the Levant. The Levantine MP industries have local origin from the late Acheulean tradition, and they indicate minimum incursions of Afro-Arabian technologies into the region. Scholars working in the Arabian Peninsula identify regional lithic varieties, some possessing local origin, while few others demonstrating close ties to MSA industries in the neighboring regions [69, 104, 105]. Although broad technotypological similarities have been recognized among some MP/MSA sites in Arabia (e.g., Jebel Faya in the UAE, Bir Khasfa, and Jabal Ardif 3 in Oman) and NE Africa (e.g., Station One in Sudan, Kulkuletti and Porc Epic in Ehtiopia) [13, 69, 154, 155], overall, the existing evidence hints at rather weak cultural contacts between the two regions (assuming that lack of lithic similarity denotes absence of cultural exchange, although that may not be always the case). The recently reported Assemblage C from Jebel Faya remains the prima facie evidence for potential cultural connections between Arabia and NE Africa. However, in the absence of corroborative evidence from the western and southwestern margins of Arabia (areas closer to Africa), it is difficult to ascertain the origin and migration routes for the makers of the Jebel Faya Industry. Much of the existing MP evidence from the western and southern Arabia shows closer affinity with the Levantine MP than with East African MSA [69]. From the recently formulated “Gulf Oasis” model [12], human presence in the Arabian peninsula appears to have persisted since the onset of the Last Interglacial (MIS-5) by relying on the expanding grassland habitat during wet phases and on environmental refugia during dry conditions [12]. The present author is in agreement with Rose’s assertion that the Arabian Peninsula may have served as a potential center for Upper Pleistocene human expansion to East Asia, to Southwest Asia or back to Africa.

The reviewed literature hints at two modes of early modern human colonization of Eurasia in the Upper Pleistocene: (i) from multiple Homo sapiens source populations that had entered Arabia, South Asia, and the Levant prior to and soon after the onset of the Last Interglacial (MIS-5), (ii) from a rapid dispersal out of East Africa via the Southern Route (across the Red Sea basin) during MIS-4. A wealth of genetic data accumulating from South and Southwest Asia [50, 64, 122, 125-127] supports in situ origin of one of the ancestral mtDNA lineages—haplogroup M in Asia, signifying an Asianic source populations for the colonization of Eurasia in the Upper Pleistocene. This view contrasts with the “single dispersal” hypothesis [11, 50], which posits that the founding lineages (M and N lineages) of South and East Asian populations were products of a single wave of modern human expansion from eastern Africa during MIS-4, dating to 74–60 kya. In conclusion, two potential routes of human migration out of Africa (the Northern and Southern Routes) can be recognized, but there does not appear to be a clear consensus on the exact timing and source populations for Upper Pleistocene human colonization of Eurasia: the genetic data suggesting multiple source populations outside of Africa and the archaeological evidence indicating weak cultural relationships among those regions presumed to be key players in human dispersals (NE Africa, Arabia, and the Levant).

Acknowledgments

Thanks are due to Richard Leakey and Lawrence Martine for providing the author with a postdoctoral fellowship in the Turkana Basin Institute which enabled him to undertake this work. The author expresses his gratitude to Parth Chuan for inviting him to submit this paper to this special issue of International Journal of Evolutionary Biology. For review and editorial assistance, the author is particularly grateful to Darren Curnoe, an anonymous reviewer, Karen Baab, and John Shea. Several publications [5, 12, 64, 69, 131] helped the author to streamline the contents of the paper, and he is thankful to the authors for their inspiring ideas and reports.

References


[40] H. Valladas, J. L. Reys, J. L. Joron, G. Valladas, O. Bar-Yosef, and B. Vandermeersch, “Thermoluminescence dating...


[52] M. Metspalu, T. Kivisild, E. Metspalu et al., “Most of the extent mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans,” *BMC Genetics*, vol. 5, no. 1, article 26, 2004.


Review Article

Before the Emergence of Homo sapiens: Overview on the Early-to-Middle Pleistocene Fossil Record (with a Proposal about Homo heidelbergensis at the subspecific level)

Giorgio Manzi

Dipartimento di Biologia Ambientale, Università di Roma “La Sapienza”, Piazzale Aldo Moro 5, 00185 Roma, Italy

Correspondence should be addressed to Giorgio Manzi, giorgio.manzi@uniroma1.it

Received 15 September 2010; Revised 5 January 2011; Accepted 24 February 2011

Academic Editor: Darren Curnoe

Copyright © 2011 Giorgio Manzi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The origin of Homo sapiens has deep roots, which include two crucial nodes: (1) the emergence and diffusion of the last common ancestor of later Homo (in the Early Pleistocene) and (2) the tempo and mode of the appearance of distinct evolutionary lineages (in the Middle Pleistocene). The window between 1,000 and 500 thousand years before present appears of crucial importance, including the generation of a new and more encephalised kind of humanity, referred to by many authors as Homo heidelbergensis. This species greatly diversified during the Middle Pleistocene up to the formation of new variants (i.e., incipient species) that, eventually, led to the allopatric speciation of Homo neanderthalensis and Homo sapiens. The special case furnished by the calvarium found near Ceprano (Italy), dated to 430–385 ka, offers the opportunity to investigate this matter from an original perspective. It is proposed to separate the hypodigm of a single, widespread, and polymorphic human taxon of the Middle Pleistocene into distinct subspecies (i.e., incipient species). The ancestral one should be Homo heidelbergensis, including specimens such as Ceprano and the mandible from Mauer.

1. Introduction

The origin of anatomically and genetically modern humans (H. sapiens) from a small population of “archaic” Homo is an event reasonably well set in sub-Saharan Africa around 200 thousand years before present (or ka) [1–3]. Nevertheless, this event has deep roots in the Middle Pleistocene, primarily at the time of the divergence between the evolutionary lineage of our own species and that of the Neanderthals—between approximately 800 and or 520 ka, according to Briggs and coworkers [4] (compare [5]), or between 538 and 315 ka, according to Endicott and colleagues [6]—and even earlier, in the late Early Pleistocene, when the common ancestor of both H. sapiens and H. neanderthalensis emerged and began to spread geographically.

This paper aims at investigating such a new frontier for paleoanthropology. It will focus on topology, chronology, tempo, and mode of the main evolutionary nodes before the appearance of H. sapiens.

When considering all the available data, we are confronted with a comprehensive scenario about the deep roots of our species. At the same time, it becomes possible to approach the issue from regional and/or local perspectives [7]. The special case study provided by a well-known fossil specimen from Italy—that is, the calvarium from Ceprano (for a review, see [8])—may help to see the remote origins of Homo sapiens from an interesting and helpful perspective.

2. Old and New Paradigms

2.1. The Midcentury View. Views about the origin of our own species greatly changed during the last couple of decades, involving our interpretation of the evolution of the genus Homo as a whole. The paradigm of mid-20th century, shared among many scholars until recently, was based on the existence of a single human species that evolved gradually and sequentially during the entire Pleistocene: it is the so-called “single-species hypothesis,” influenced by...
the gradualist perspective of the “modern synthesis” [9, 10]. Consistently with this perspective, the current human diversity would have been the result of small and constant changes among populations and within the species as a whole (viewed as a single, extremely large population) taking place from the original geographical diffusion of the genus Homo, where the roots of the present human diversity would be found.

According to this view, successive stages and/or regional variants were implied within a widespread archaic species—H. erectus—and subsequently among H. sapiens. The result was a model of either progressive or regional changes that, with the persistence of a single polymorphic humanity in each span of geological time, was evolving towards variants of the modern species and, ultimately, to the present human “races” [11]. As a consequence, the taxon H. sapiens included the extant humankind, but also extinct morphotypes like the Neanderthals in Europe and the Near East as well as, more in the depth of the Middle Pleistocene, diverse archaic humans in Africa, and in eastern Asia, collectively referred to as “archaic H. sapiens.” Thus, under the same specific name, H. sapiens, each of these morphs was ascribed to a different subspecies, with the adoption of a trinomial nomenclature that gave rise to nomina such as H. sapiens neanderthalensis and H. sapiens sapiens.

The theory called “multiregional evolution” is a good example of this view. It is well known that according to the model introduced by Wolpoff and colleagues [12] the anatomically modern populations of Africa, Asia, and Europe have been viewed as emerging in continuity with the preceding archaic humans of the same geographical area. This hypothesis was based on the observation that a certain degree of “regional continuity” characterizes the morphologies of archaic and modern populations within each geographical area. This observation, however, has been contradicted by a number of works [13, 14] and, more in general, the model is regarded as not in agreement with most of the paleogenetic data (since the seminal work by Cann et al. [15]).

To a large extent, this view has now been abandoned, after a debate that lasted for more than two decades [16]. Some of the issues of this scientific revolution (sensu Kuhn [17]) may be summarised with reference to Figure 1 and are briefly discussed below.

### 2.2. A New Scenario

One of the consequences of the middle-century paradigm on the interpretation of the fossil record was that a single species, H. erectus, included a largely distributed, archaic-looking, and polymorphic humankind that was ancestral to modern humans. It was considered to be the first hominid species that left Africa, equipped with a brain of about 1,000 ml and skilled enough to produce the elaborated stone tools of the Acheulean, and able to face different climatic and environmental conditions because of its behavioural and technological—someone says “cultural”—aptitudes. In brief, H. erectus was viewed as the quasi-modern “hero” who “conquered” the Old World: an authentic “dawn of humanity.”

However, an increasing number of data—including the evidence coming from the Georgian site of Dmanisi [18, 19]—suggests a different scenario. Those that were responsible for the first out-of-Africa diffusion were not derived, encephalised, and technologically advanced humans, but more archaic hominins, with a brain just above the threshold of 500–600 ml (maximum encephalic volume of the australopithecines) and morphologically close to the changing definition and hypodigm of H. habilis (after Leakey et al. [20]). Driven by ecological rather than behavioural or “cultural” motives, these hominins had a tendency—which was new with respect to Australopithecus, Paranthropus, and the Primates in general—to diffuse and adapt to variable, nontropical environments. The same corpus of data suggests that this process should have started well before the appearance of either H. erectus or the Acheulean (which are now viewed as separated and independent phenomena), that is roughly between 2,000 and 1,500 ka.

Within this new approach, H. erectus may be viewed as a Far Eastern (Java and China) species only (contra Asfaw and colleagues [21]) whereas their African counterparts are considered as a distinct species, H. ergaster [22, 23]. Alternatively, these two geographical variants are grouped under the definition H. erectus sensu lato, while H. erectus sensu stricto would be the Asian deme of such a multiregional taxon. At the same time, other species have been named or old nomina have been reconsidered, and they include (according to the chronological order of references that are pivotal for the present debate): H. rudolfensis [22], H. heidelbergensis [24], H. antecessor [25], H. helmei [26], H. mauritanicus [27], H. rhodesiensis [28], H. georgicus [29], H. cepranensis [30], and H. floresiensis [31]. However, many of these taxa are debated and/or not widely acknowledged; thus, the generally accepted scenario is far less “speciose” than it might appear from such a tentative list.

Although the identification of all these different species clearly implies an overestimation of interspecific diversity—that in many cases was more probably intraspecific (see below)—this plethora of nomina is in accordance with a scenario that foresees the geometry of an “adaptive radiation,” describing the generation from a common stem of a great variability, in space as well as in time. Moreover, this confers a clearer, more intelligible significance to human varieties that were formerly hidden, being referred either to H. erectus (sensu lato) or to the informal and confusing entity called until recently “archaic H. sapiens” (which has been abandoned and has almost disappeared from the literature).

Looking at Europe, the relatively best known regional example, at least two distinct waves of immigrants seem to be recognizable between the late Early and the early Middle Pleistocene. In terms of fossil record, the former wave is documented at present only in Spain and is referred to H. antecessor. As a matter of fact, this species occurs in two sites of the Sierra de Atapuerca, near Burgos: in the layer TE9 of the Sima del Elefante, dated to about 1,200 ka [32], and in the layer TD6 of the Gran Dolina, dated to more than 780 ka [25]. These humans are in association with the so-called “Oldowan” (or Mode 1 of the Lower Paleolithic [33]), largely diffused in a number of sites of Mediterranean and
continental regions [34]. In addition, their settlements could have been discontinuous until about 600 ka, being strongly influenced by ecological conditions [35].

Seemingly, the latter possible dispersal into Europe was more recent than 700–600 ka and related to morphologically derived hominids, with clear signs of further encephalization, which are well known from a number of sites. The most notable assemblage of fossil material is again in the Sierra de Atapuerca [36], namely, at the site with the evocative name of Sima de los Huesos, or SH, whose extraordinarily rich and well-preserved human sample has been pushed back to about 600 ka [37]. These fossils are generally included within the species H. heidelbergensis, which is usually, but not always, associated with Mode 2 (or “Acheulean”) Paleolithic assemblages.

H. antecessor and H. heidelbergensis compete, one against the other, for the same phylogenetic position in current evolutionary trees of the genus Homo, viewed as provisional, alternative models of human evolution. H. antecessor is regarded by the Spanish workers (after [25]) as the stem species that was ancestral to the evolutionary divergence within the late natural history of the genus Homo, leading to the evolution of the Neanderthals in Europe and to the emergence of our species in sub-Saharan Africa. Alternatively, it is H. heidelbergensis to be claimed for the same crucial role [24]. In a sense, the Sierra de Atapuerca provides the evidence that could explain this ambivalence, either in the former or in the latter direction. At the same time, however, it should be remarked that the material from Sima de los Huesos is clearly Neanderthal oriented, being characterized by a number of features that, later in the Pleistocene, will become typical of the Neanderthals [38, 39]. Looking at the material from Atapuerca SH, therefore, H. heidelbergensis acquires the identity of a European regional chronospecies in continuity with H. neanderthalensis, which appears inappropriate as also ancestral to the African lineage leading to H. sapiens. There are, anyway, in Europe other fossil specimens—such as the calvarium from Ceprano, Italy (see discussion below)—that are penecontemporaneous with the material from Atapuerca SH and might represent (far better than the Spanish sample) a possible ancestral morphotype for H. heidelbergensis, if we look at this species as the evolutionary stem before the divergence between Neanderthals and modern humans (compare [6]).

Moving to Africa, it has been shown [8, 23] that specimens dated to about 1,000 ka or slightly less—such as Daka, Buia, and Olorgesaille—share phenetic affinities closer to H. ergaster than to Middle Pleistocene African hominins—like Bodo and Kabwe—which are referred to H. heidelbergensis or, alternatively, to H. rhodesiensis. Hence, these sub-Saharan specimens of the late Early Pleistocene signal a morphological discontinuity with the subsequent fossil record and should be considered as late representatives of H. ergaster. In addition, as emphasized in Figure 1, the
period between 900 and 600 ka is unfortunately very poor of fossil evidence in sub-Saharan Africa, and not only there. Thus, the differences occurring between H. ergaster (until 900 ka) and H. heidelbergensis (after 600 ka) might also suggest an allochthonous origin of the latter species.

More in general, late H. ergaster representatives are also distinct from the Middle Pleistocene fossil record from either Europe—including the Italian calvarium from Ceprano (contra Mallegni and coworkers [30]) as well as the assemblage from Atapuerca SH—or mainland Asia, looking at specimens like Narmada, Dali, and Jinniushan. These, in turn, are also respectively distinguishable from H. antecessor, in Europe, and from H. erectus sensu stricto in China (e.g., the fossil sample from Zhoukoudian near Beijing) and Java. We still do not know where, but it seems that something happened between about 1,000 ka and 500 ka that was capable to generate—in sub-Saharan Africa, Europe, and mainland Asia—a new and more encephalised kind of humanity, which was different from the variability coming from the earliest human dispersal out of Africa: that is, H. ergaster, H. antecessor and H. erectus (leaving aside the singular, intriguing case of H. floresiensis [31]).

To sum up, at present, the chronology, topology, and phylogenetic dynamics related to the geographically dispersed and rather synchronous appearance of Middle Pleistocene humans—or H. heidelbergensis, viewed as a single widespread species—are still unclear. Ultimately, the provenience of those archaic humans (see question mark in Figure 1) that, in turn, were ancestral to the origin of both Neanderthals and modern humans is not evident [1, 40, 41].

A possible answer comes from the results recently obtained with the analysis of the complete mitochondrial DNA (mtDNA) sequence retrieved from the isolated human phalanx of the Denisova cave in the Altai mountains, southern Siberia. In the context of episodic occupations of this site in the Late Pleistocene, the layer where the phalanx was found has been dated to 48–30 ka, in association with an archaeological assemblage including both Middle and Upper Palaeolithic elements. The mtDNA evidence surprisingly points to humans that were different from both H. neanderthalensis and H. sapiens, but share with them a common ancestor around 1,000 ka [42]. As a working hypothesis, this suggests that the Denisova phalanx may represent an unknown type of hominin that reached mainland Asia during a diffusion that begun (where, we do not know) about 1 million years ago; this humankind was in close phylogenetic relationship with the last common ancestor of Neanderthals and modern humans and survived in north-eastern Eurasia at least until about 40 ka. As visible in Figure 1, the origin of these three clades corresponds well to the morphological discontinuity occurring in the fossil record before the appearance of H. heidelbergensis.

Researchers opted to wait for their data to provide a clearer picture of the relationship with Neanderthals and modern humans before giving the hypothetical unknown species a formal name. Nevertheless, it is already possible to speculate that the Denisova hominins were in relation with the “non-erectus” occupation of mainland Asia during the Middle Pleistocene. Assuming in fact that H. erectus was one of the evolutionary outcomes of the earliest diffusion in Eurasia, prior to 1,500 ka (see Figure 1)—and excluding, therefore, this taxon from the scope of possibilities—we have to look elsewhere and focus on Asian Mid-Pleistocene specimens such as Dali and Jinniushan, in the past ascribed to H. sapiens daliensis [43] and currently considered by some authors as the easternmost variant of H. heidelbergensis (after Rightmire [24, 40]). Now, should it be appropriate to ascribe the phalanx from Denisova and its precious molecular content to a late variant of the same taxon as Dali Jinniushan and other “non-erectus” specimens from the late Middle Pleistocene of Mainland Asia?

Further analyses on the Denisova material—including exceptionally preserved nuclear DNA from the phalanx and the discovery of an upper molar—drove the same group of researchers to publish additional data [44], which appeared when the present paper was under revision. They confirmed that the Denisova individuals, and the population they belonged to, exhibit molecular (mtDNA) as well as morphological (dental) features that appear extremely archaic.

By contrast, the picture that emerges from the analysis of the nuclear genome suggests that this human group has close affinities with the Neanderthals, larger than expected from the mtDNA. Eventually, the model they suggest is that the Denisova hominins are a sister group to Neanderthals, “with a population divergence time of one-half to two-thirds of the time to the common ancestor of Neanderthals and modern humans” (Reich et al. [44, page 1057]). At the same time, Reich and colleagues [44, page 1057] admit that “other, more complex models could also explain the data.” Actually, in my view (as it will be reported in Figure 2), the same corpus of data may be explained with the occurrence of gene flow across Eurasia between the ancestors of both the Neanderthals and the Denisova hominins along the mid part of the Middle Pleistocene.

Paleogenetic data also indicate that trajectories of human evolution leading in Europe to the Neanderthals and in Africa to modern humans coalesced around 500 ka [4–6]. This substantiates previous conclusions based on morphology and palaeogeography, which suggested isolation and divergence between the European and African lineages during the Middle and the early Late Pleistocene (after Santa Luca [38]). As a matter of fact, looking at the hypodigm of H. heidelbergensis as a whole, it is clear that a considerable amount of variability characterises this putative species—bearing regional features in Africa, Asia, and Europe, respectively, [47]—referable to a phenomenon known as “isolation by distance” [48]. At the same time, phenotypic variation has to be noted also at the regional level, such as within the European fossil record of the Middle Pleistocene, now greatly expanded by the recently revised chronology of the calvarium from Ceprano [49].

3. The Case Study of Ceprano

3.1. A Cranium for the Earliest Europeans? At present, there is a general consensus in assuming that humans spread towards western Europe during the late Early Pleistocene,
probably earlier than 1,200 ka. This is demonstrated by recent findings in Spain [32, 50], and possibly in southern France [51] and Italy, where the earliest settlements are suggested by the stone tools found at Pirro Nord, near Apricena in Puglia (southern Italy) [52, 53], in association with a rich paleontological assemblage biostratigraphically referred to a Faunal Unit of the Early Pleistocene (about 1,400 ka; [54, 55]). Of interest are also sites dated at around 1,000 ka such as Monte Poggio di [56] or a number of localities in the Ceprano basin and surroundings that have been recently object of a new season of excavations and analyses [57]. Between 950 and 700 ka, these humans proved to be capable to reach and adapt to higher latitudes, as demonstrated by sites in southern England [58].

Less than 20 years ago, at the beginning of the 1990s, the available data were not so clear, and a model of “short chronology” for the earliest inhabitants of Europe was put forward [59], indirectly supported by the finding of the human tibia from the Acheulean site of Boxgrove [60]. In brief, the “short chronology” hypothesis postulated that no humans were present in Europe before the time limit of approximately 500 ka, a date that is close to the putative chronology of the mandible from Mauer (holotype of H. heidelbergensis), which remained for about one century the most ancient fossil hominin north of the Mediterranean Sea. This date is also consistent with the earliest clear documentation of Acheulean assemblages spread in various part of the continent from at least 600 ka [61].

One year after the discovery at Boxgrove, hard evidence from Italy (Ceprano, March 1994) and Spain (Atapuerca TD6, July 1994) chronologically referred to more than 700–800 ka [62, 63], suddenly falsified the hypothesis of a “short chronology.” Thus, together with the Gran Dolina material, the calvarium found near Ceprano was considered for more than a decade part of the most ancient fossil evidence of the human presence in Mediterranean Europe, suggesting the tentative attribution of Ceprano to H. antecessor [27] or, alternatively, the occurrence of two different human species in Europe during the late Early Pleistocene [30, 62, 64–66].

The Italian specimen was discovered in several fragments in a field known as Campogrande, near the town of Ceprano, in southern Lazio, less than 100 km south-east of Rome. Its discovery represents the result of systematic field activities conducted for decades in southern Lazio by the Italian Institute of Human Palaeontology (under the supervision of the Soprintendenza Archeologica del Lazio), and particularly by I. Biddittu. On March 13th 1994, Biddittu found a first cranial fragment during a survey along the trench excavated for a new road while other portions of the same cranium were still included in the nearby section created by the excavators. Subsequently, all the fragments (about fifty) were carefully extracted and sieved from the clayey sediments. The reconstruction of the cranium required more than one attempt, the intervention of a composite team, and, overall, about five years [62, 64, 67].

For the purpose of a chronological reference, the geologist A. G. Segre [62, 67] suggested a compilation stratigraphic column at a microregional scale, mainly based on previous geopalaeontological knowledge. This describes two main complexes; the layer where the human calvarium was found belongs to the lower portion of the upper stratigraphic complex, pointing at a tentative age of about 800–900 ka. Consistently, the archaic features of the calvarium were considered in association with Mode 1 technocomplexes coming from sites scattered in the Ceprano basin [57], although a number of Acheulean assemblages are also well known at Campogrande and surroundings and are now submitted to an accurate reappraisal [49].

3. A Tantalizing Specimen of the Middle Pleistocene. With these premises, a project of surveys and excavations started in 2001 under the direction of I. Biddittu and G. Manzi, with a threefold aim: (1) a better comprehension of the Pleistocene stratigraphy of the Ceprano basin; (2) validation of the geochronological model set by A.G. Segre; (3) improvement of the palaeontological and archaeological records. After ten years, the results obtained through a multidisciplinary approach—including stratigraphic and palynological data, combined with sedimentology, geochemistry, soil-micromorphology, taphonomy, and the archaeological evidence—showed that the Ceprano calvarium is more recent than previously believed, pointing at a time range close to about 400 ka and, more precisely, to the interval at the beginning of marine isotopic stage (MIS) 11 bracketed between 430 and 385 ka [49]. This result is also consistent with the normal geomagnetic polarity recorded in the area of discovery down to a depth of about 50 metres [68].

These unexpected results and the consequent new chronology of the fossil specimen in the mid of the Middle Pleistocene led Manzi and colleagues [49, page 584] to guess that “the morphology of the human calvarium from Ceprano—which lacks Neanderthal traits and does not have a real counterpart among the continental penecontemporaneous fossil record—now appears tantalizing. It adds an amount of unexpected diversity into the range of variation known for the European populations of the Middle Pleistocene,” and to suggest the evaluation of “more complex scenarios of human evolution in Europe than previously believed, involving either the occurrence of a considerable intraspecific diversity (with archaeologically distinct settlements) or, alternatively, the coexistence of different lineages (with their own respective archaeological traditions) during part of the Middle Pleistocene.”

This also calls for a taxonomic re-evaluation of the Italian specimen. Originally, Ceprano was attributed to “late H. erectus” [62, 64]. Five years later, two papers criticized the H. erectus affinities that were originally claimed [27, 67], reaching the conclusion that less than two-thirds of the character states in this calvarium are in accordance (and not always unequivocally) with those commonly encountered in H. erectus sensu stricto while others appear peculiar and/or progressive. Further studies included a cladistic approach, with the questionable proposal of a new species name [30]. Moreover, the CT scanning of the specimen [65] and other phenetic data [8, 23, 66] produced additional elements that were useful to better understand the specimen in a comparative framework.
On the whole, these researches largely support conclusions preliminarily reached by Manzi and colleagues [27], which may be summarised as follows. First of all, though some metric and architectural features of Ceprano approach those shared by fossils referred to *H. erectus*, the variability of this taxon appears greatly enlarged when Ceprano is added. Second, there are discrete features that detach Ceprano from its general “erectus-like” appearance and may be viewed as derived, suggesting a connection with the Middle Pleistocene fossil record from Africa and Europe. Third, Ceprano does not display any Neanderthal trait, while it shows affinities with the African penecontemporaneous fossil record, closer than the affinities it has with its European counterparts. A possible conclusion is that Ceprano may be regarded as a mosaic morphological link between the clade composed by the group of species referred to as *H. erectus* sensu lato, on the one hand, and samples commonly referred to as *H. heidelbergensis*, on the other. This in turn suggests that Ceprano might document “the occurrence of an ancestral stock of *H. heidelbergensis/rhodesiensis*” [66, page 365], whose cranial morphology was lost in part along the subsequent trajectory of human evolution in Europe, but preserved elsewhere (including Africa and, possibly, mainland Asia). In addition, it has to be remarked that Mounier and colleagues [47, 69], in the wide framework of a recent reappraisal of the fossil record pertaining to *H. heidelbergensis* and related species, provide robust and independent support to these conclusions.

### 4. Mediterranean Perspectives

Viewed in a wide paleoecological scenario, the earliest dispersal of human groups towards the western Mediterranean regions was likely part of the progressive faunal renewal that involved the diffusion of some large mammals of African and Asian origin during the Early Pleistocene [7, 35, 70]. This diffusion was also favoured by the opportunistic nature of hominins that were *archaic* both in their morphology—highlighted by the affinities observed between the *H. antecessor* specimen from Atapuerca TE9, and the fossil sample from Dmanisi [32]—and for their inferred behaviour, as documented by Mode 1 assemblages [34]. An even more favourable window for human presence in Europe likely opened around 1,000 ka and later, when more consistent settlements were probably related to the major faunal renewal that characterised the Early to Middle Pleistocene transition. This transition—the so-called “Mid-Pleistocene Revolution” (approximately from 1,200 to 600 ka [71])—marks a dramatic change in the Earth’s climate system and represents a major episode in mammal fauna reorganization in the course of the Cenozoic, not only in Europe [72–74]. Thus, since at least 1,300 ka, dispersals of taxa and turnover phases led to a progressive reconstruction of mammalian faunal complexes in Europe that was complete after the beginning of the Middle Pleistocene [75].

The unique hominin hard evidence in Europe for this time period is represented so far at Atapuerca TE9 and TD6, but the presence of human populations is documented by a number of Mode 1 archaeological sites. It is possible that diffusion waves, presumably scattered in time and space, led to the arrival of archaic humans in western Eurasia until MIS 16, one of the worst glaciations of the last million years, with an ice sheet extension below 50° latitude in Eastern Europe [76]. This climatic collapse probably constituted a strong environmental barrier and might not be by chance that it preceded the appearance of the Acheulean tool technology (Mode 2) in the continent. In other words, it is in the framework of the changed environmental scenario implied by the Mid-Pleistocene Revolution, but only after MIS 16, that we have clear evidence in Europe of a second main dispersal of hominins: those bearing the Mode 2 technocomplexes (for a recent review, see [1]).

The exact origin of these humans is still not clear, though it may be assumed that they ultimately emerged from Africa [42, 77]. As already stressed in the first part of this paper, these Acheulean-bearing humans exhibit a clear discontinuity in morphology with the hominins previously diffused into Europe, that is, with *H. antecessor* and/or related species. As a result of this second main diffusion towards western Eurasia, in the middle part of the Middle Pleistocene, we find in the continent a variety of human fossil samples, which are dispersed from northern (e.g., Swanscombe in England, Mauer, and Bilzingsleben in Germany, etc.) to southern latitudes (e.g., Atapuerca SH in Spain, Arago in southern France, Petralona in Greece, etc.), Italy included (e.g., Ceprano, Venosa, and Visogliano). On the whole, this fossil record is considered by many authors as part of the hypodigm of the species *H. heidelbergensis* created on one of these specimens, that is the Mauer mandible [40, 47, 78].

Subsequently, the observed pattern of evolution in Europe during the Middle Pleistocene is consistent with a long period of isolation for humans north of the Mediterranean Sea, which seems to be supported on both morphological and genetic grounds [3, 5, 38, 45, 79]. These populations are characterised by an apparent increase of Neanderthal features; an increase that, in turn, was probably related to the dramatic glacial periods of MIS 12–6, that might have produced demographic crashes among human populations, which resulted in population bottlenecks, likely favouring either genetic drift or adaptations to cold climatic conditions. Consequently, according to the so-called “accretion model” [1, 39, 80], as suggested also by studies on the Saccopastore Neanderthals and other Mid-to-Late Pleistocene fossil specimen [81–83], the European evolutionary lineage eventually led to the origin of *H. neanderthalensis* throughout successive genetic bottlenecks.

On the other hand, however, more recent evidence on the European fossil record of the Middle Pleistocene hardly supports the hypothesis of a linear and gradual process of change [84–86]. Just to give an example, endocranial metric variations fail to demonstrate the occurrence of sequential discrete steps along this hypothetical anagenetic process [87]. Moreover, the Neanderthals—even in their earliest representatives, such as those from Saccopastore—seem to be
characterized by a well-defined brain morphology, emphasizing the phylogenetic independence of *H. neanderthalensis* from *H. heidelbergensis* [83]. This suggests a distinction between two different chronospecies as well as an event of speciation that occurred in Europe towards the end of the Middle Pleistocene, which appear consistent with the paleogenetic data [4, 6].

It is reasonable that something similar—although not identical—happened with the locally evolving populations of late *H. heidelbergensis* in Africa and in Asia. An example might be the emergence in Africa of *H. sapiens*. However, the pattern in this case was peculiar in terms of both evolutionary modalities and changes in morphology. Various scholars have argued, and partly proved, that the allopatric speciations involved in the phenotypic and genetic distinction between Neanderthals and modern humans would reflect crucial differences in the respective ontogenetic processes [88–90]. Although similar trends of encephalization characterize (quantitatively) the two derived species, they diverge in many respects, particularly when we look at the shape more than at the dimensions. As a matter of fact, while Neanderthal morphology combines a fundamental archaic shape of the cranial vault with enlarged brain dimensions, the modern architecture appears completely redesigned in terms of “globularity” [91]. It has been observed, for instance, that the Neanderthals share with more archaic humans the same endocranial model, based on a single allometric trend whereas the modern range of variability implies a peculiar morphological pattern, with a larger amount of parietal development [92]. In this light, we may argue that a certain level of “allometric stasis” has occurred during the evolution of the genus *Homo* and that the exception is represented by the transition to the new phenotypic and ontogenetic equilibrium associated with the emergence of *H. sapiens*, which appears as a sharp disruption within the pattern of evolution of the genus *Homo* taken as a whole.

5. Summary and Conclusions

In this paper, we dealt with arguments concerning the evolution of the genus *Homo* and attempted to put recent and less recent studies in such a broader context.

As stated in the introduction, the aim was to investigate a new frontier for paleoanthropology. This is represented in my view by the discovery of the deep roots for the origin of modern humans in the Early and Middle Pleistocene, respectively, when the common ancestor of both *H. sapiens* and *H. neanderthalensis* began to spread geographically, and when the divergence between the evolutionary lineage of our own species and that of the Neanderthals occurred. In this framework, I speculated that any scenario dealing with the evolution of the genus *Homo* between the Early and Middle Pleistocene must envisage that the fossil specimen from Ceprano, Italy, does not appropriately fit into the known ranges of variability of recognized hominin species, particularly (but not exclusively) in the light of its dating to less than 430 ka. By contrast, it represents a puzzling mosaic of plesiomorphic and apomorphic features, which might be explained according to the following “keywords”: (a) Europe; (b) Middle Pleistocene; (c) morphology “intermediate” between *H. erectus* sensu lato and *H. heidelbergensis*; (d) affinities with penecontemporaneous samples from both sub-Saharan Africa and mainland Asia.

I also observed that something that was crucial for the evolution of the genus *Homo* happened around the beginning of the Middle Pleistocene, between about 1,000 ka and 500 ka. Looking at the fossil record in Africa and Eurasia, in fact, there are differences that occur between the late representatives of the earliest spread of the genus *Homo* (e.g., *H. ergaster, H. erectus*, and *H. antecessor*), on one hand, and *H. heidelbergensis* viewed as a largely diffused species of more derived humans, on the other.

Thus, when connecting all the elements described in this brief overview on the Early-to-Middle Pleistocene fossil evidence preceding the emergence of *H. sapiens*, I point out that there is a single cranial specimen among the potential hypodigm of *H. heidelbergensis* [69]—that is, the calvarium from Ceprano—that is sufficiently archaic and peculiar with respect to the penecontemporaneous fossil record to represent the ancestral morphotype of this species. Moreover, it should be added that the hypodigm of *H. heidelbergensis* shows a considerable amount of variability, in which regional features may be recognised [47, 69] while phenotypic variations are observed also locally such as within the European fossil record [49]. This supports the conclusion that Ceprano would document at present, better than other specimens, a possible ancestral stock of *H. heidelbergensis*, close in time [6] to the evolutionary divergence between Neanderthals and modern humans.

These and other arguments (compare Section 2.2 for discussion and references) greatly expand, at least tentatively, the phylogenetic significance, the geographic distribution, and the time span covered by *H. heidelbergensis* viewed as a taxon that was both geographically widespread and morphologically diversified (Figure 2). Actually, it was so widespread and so diverse that inappropriately it would be referred to a single taxon, without further internal distinctions. At the same time, however, there are several signs that inbreeding among different demes of such a large biological entity were possible.

In conclusion, it seems to me that the time is ripe to introduce a trinomial nomenclature for this species. Furthermore, as stated for instance by Mayr [93, page 155]: “every species that developed through geographic speciation had to pass through the subspecies stage.” *H. heidelbergensis* includes in fact the clear occurrence of regional incipient species, which prelude to allopatric speciations such as those of *H. neanderthalensis* (in Europe) and *H. sapiens* (in Africa). Also, from this perspective, the use of subspecific ranks within *H. heidelbergensis* appears more than appropriate, as well as practical in distinguishing between demes.

My proposal is to consider the single widespread species that was ancestral to both Neanderthals and modern humans (and the “Denisovans” [44]) as fractioned in chronological subsequent and/or geographically distinguished subspecies. With reference to Figure 2 and using already available
names—according to the rules of the International Code of Zoological Nomenclature (http://www.nhm.ac.uk/hosted-sites/iczn/code/; for a pivotal reference of its application in paleoanthropology, as well as for the use of sub-species, see the seminal paper by Campbell [94])—the most proper denominations for these subspecies respectively should be:

(1) *H. heidelbergensis heidelbergensis* [95]: the ancestral and still largely unknown variety of the species, including the name-bearing type from Mauer (Germany) and other specimens that are either demonstrably archaic or not involved in the respective regional lineages; it would include fossil crania such as Arago, Bodo, Ceprano, and possibly Petralona (for the inclusion in this group of a rather composite fossil record, which combines so far mandibles and crania, see [47]; for a consistent approach to postcranial remains, see [96]); among this sample, Ceprano represents at present the best available expression of the ancestral morphology for this taxon, as far as the braincase districts are concerned;

(2) *H. heidelbergensis steinheimensis* [97]: for the European lineage of the Middle Pleistocene leading to the Neanderthals, including the type specimen from Steinheim (Germany) and other samples such as, most notably, the assemblage from Atapuerca SH [36];

(3) *H. heidelbergensis rhodesiensis* [98]: for the African fossil record of the Middle Pleistocene preceding the appearance of modern humans, including the type specimen from Kabwe (Zambia) and possibly all the late Middle Pleistocene material from various parts of the continent formerly included within the informal group “archaic *H. sapiens*” [99], that is with the exclusion of penecontemporaneous anatomically modern specimens;

(4) *H. heidelbergensis daliensis* [43]: for the Asian non-erectus sample bracketed, at least, between Dali (China, type specimen of this subspecies) and the diminutive, but very informative fossil material from Denisova [42, 44].

Figure 2: Evolutionary tree of *H. heidelbergensis* distinguished in subspecies, as suggested in this paper: (1) *H. h. heidelbergensis*; (2) *H. h. steinheimensis*; (3) *H. h. rhodesiensis*; (4) *H. h. daliensis*. The main evolutionary trajectories (dashed-bold lines) and the maintenance of gene flow between populations of distinct lineages (GF) are in accordance with a combination of paleogenetic data reported by Krause and colleagues [42] and by Reich and colleagues [44]. According to this scenario, Ceprano would represent one of the latest representatives of the most archaic variant of *H. heidelbergensis* (1) whereas more derived subspecies, respectively, in Europe (2) and in Africa (3), led to the allopatric speciation of *H. neanderthalensis* and *H. sapiens* (both schematically represented in the graph by triangles). Eventually, *H. h. daliensis* (4) would include all the “non-erectus” archaic humans distributed in mainland Asia before the diffusion of *H. sapiens*. Localised interspecific hybridization between *H. heidelbergensis* or *H. neanderthalensis* versus *H. sapiens* is admitted by the model (according to [44–46]). Penecontemporaneous human lineages and/or species—such as *H. erectus* and *H. floresiensis* (compare Figure 1)—are not represented in this diagram.
Acknowledgments

The contribution of various people and institutions involved in field activities and researches in the Ceprano basin—with special reference to the Soprintendenza Archeologica del Lazio—is gratefully acknowledged. Thanks are also due to David Caramelli, Silvana Condemi, Fabio Di Vincenzo, Jacopo Moggi-Cecchi, and Aurelien Mounier for useful discussions on earlier drafts of the paper. This paper was supported by resources of the Università di Roma “La Sapienza” and was cofinanced by the Italian Ministry of University and Research (M.I.U.R.), including “PRIN grants.” The author is a member of the directive board of the Istituto Italiano di Paleontologia Umana (ISIPU), Roma.

References


Research Article

Morphological Integration of the Modern Human Mandible during Ontogeny

Joshua M. Polanski

Department of Math & Science, Kirkwood Community College, Iowa City Campus, Iowa City, IA 52240, USA

Correspondence should be addressed to Joshua M. Polanski, joshua.polanski@kirkwood.edu

Received 16 September 2010; Revised 3 January 2011; Accepted 24 February 2011

Academic Editor: Darren Curnoe

Copyright © 2011 Joshua M. Polanski. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Craniofacial integration is prevalent in anatomical modernity research. Little investigation has been done on mandibular integration. Integration patterns were quantified in a longitudinal modern human sample of mandibles. This integration pattern is one of modularization between the alveolar and muscle attachment regions, but with age-specific differences. The ascending ramus and nonalveolar portions of the corpus remain integrated throughout ontogeny. The alveolar region is dynamic, becoming modularized according to the needs of the mandible at a particular developmental stage. Early in ontogeny, this modularity reflects the need for space for the developing dentition; later, modularity is more reflective of mastication. The overall pattern of modern human mandibular integration follows the integration pattern seen in other mammals, including chimpanzees. Given the differences in craniofacial integration patterns between humans and chimpanzees, but the similarities in mandibular integration, it is likely that the mandible has played the more passive role in hominin skull evolution.

1. Introduction

Morphological integration of the cranium has become a dominant research paradigm in biological anthropology. While it began with the early work of Weidenreich [1, 2] and the functional matrix model of Moss [3, 4], morphological integration became a driving force in anthropological research through work on macaques and New World monkeys by Cheverud and colleagues [5–11]. Building upon this earlier work, recent years have seen new techniques defined to quantify and assessing morphological integration being used to address the question of hominin evolution, with a particular emphasis on the evolution of modern humans [12–19].

In virtually all hominin studies involving morphological integration, the cranium serves as the major anatomical unit under investigation. The reasons for this are obvious. Craniodental remains are the most predominant morphology in the hominin and primate fossil records; also, taxonomic and evolutionary relationships tend to be defined primarily by cranial morphology. Because of this, a well-defined understanding of the relationships between cranial elements in both living and fossil forms is of extraordinary use. However, mandibular remains are also prevalent within the fossil record. The question then must be asked: why has the mandible been left behind in the field of hominin morphological integration? The mandible serves as an integral component of the skull and has been used extensively in numerous anthropological research questions. Mandibles have served, at one time or another, as type specimens for various hominin species: *Australopithecus anamensis* [20]; *A. afarensis* [21]; *A. bahrelghazali* [22]; *Paranthropus crassidens* [23]; *Homo ergaster* [24]; *H. heidelbergensis* [25].

Although a complete survey of the entire biomechanical anthropological literature is beyond the scope of this paper, the mandible has served as the backbone for numerous studies on biomechanics and stress-loading studies in both hominins and the rest of the primate order, particularly in, but not limited to, the work of Hylander and colleagues [26–31]. The mandible has also been analyzed in relation to the evolutionary history of the chin and questions concerning its utility as a defining characteristic of human anatomical modernity [32–38]. The mandible and its associated dentition have also factored into investigations into the dietary
adaptations of primates from the great apes [39–42] to the creation of the “gracile” and “robust” clades of the australopiths [43–49].

Outside of hominin paleontology, morphological integration of the mandible has been studied in mice [50–52], baboons [50] and chimpanzees and gorillas [53]. Given the emphasis on cranial form in the evolution of modern humans, and the extensive literature concerning nonhuman mandibular integration, the decreased focus of morphological integration research in hominins involving the mandible is surprising. One of the defining features of modern human evolution has been the marked decrease in facial prognathism [17, 54–57], a trend which is mirrored in the evolution of the mandible. Despite changes in the hominin face and the associated alterations of the mandibular form, little to no research into the patterns of integration in the mandible has been undertaken. Also, evolutionary biology has begun to understand the need to investigate evolution from a developmental perspective (evo-devo). This trend is reflected in biological anthropology with the inclusion of ontogeny in various studies, including those involving morphological integration [12, 17, 19]. This paper sets out to rectify this situation by investigating morphological integration of the mandible during ontogeny. More specifically, it will test the null hypothesis of no change in the pattern of integration in the mandible from early childhood to adulthood.

2. Materials and Methods

The modern human sample used in this analysis was derived from the Iowa Facial Growth study, created by and housed at the University of Iowa Dental School. The Iowa Facial Growth Study ran from 1946 to 1960 and was a true longitudinal study; after the original volunteer group reached age five no new participants were allowed into the study. Several years after the end of the Iowa Facial Growth study, participants were asked to be radiographed once more to study the facial skeleton after growth had ceased [58]. True longitudinal studies have an advantage over cross-sectional studies in that the data gathered from longitudinal studies allows for the study of actual growth; cross-sectional studies must use approximations of growth and can risk missing or excluding certain trends. However, longitudinal studies often suffer from a drain of available participants throughout the length of the study.

The subjects were ninety-seven percent Caucasian, all United States-born. Over ninety percent had at least three-quarters of their grandparents descended from northern European immigrants (British Isles, Germany, France, Scandinavia, or The Netherlands), while the remainder possessed at least two grandparents from either central or southeast Europe. Subjects’ families were of above average socioeconomic status; over fifty percent had fathers who worked in managerial/professional vocations, forty percent had fathers who owned small businesses or were skilled traders, and the remaining ten percent came from families where the fathers engaged in semiskilled trade [58].

The major criterion for inclusion in the Iowa Facial Growth Study was neither ethnic nor economic; the researchers were interested in subjects who lived in and around Iowa City at the beginning of the study and whose families were likely to remain in the same area for the duration of the subjects’ development, as well as their willingness (or their parents’ willingness) to participate in a long-term study. All subjects were voluntary ones, no subjects were compensated for their participation [58]. While the primary criterion was not dental occlusion, subjects with non-normal occlusions were not included in the study. The nearly unanimous Caucasian nature of the participants, as well as the above average socioeconomic status, is likely the result of the influence that the University of Iowa (then the State University of Iowa) exerted on the demographics of Iowa City in the mid-1940s.

The Iowa Facial Growth Study consists of radiographs taken in both norma lateralis and norma frontalis; because lateral radiographs are more commonly used in anthropological and developmental studies, only the lateral radiographs were utilized in this particular analysis, allowing for comparison of these results to those in other studies. These lateral and posteroanterior radiographs were originally taken at three-month intervals until the subjects reached age five, after which they were taken semiannually until age twelve, when the radiographs were taken annually. Subjects were examined, once the annual examinations began, within five days of their birthdays.

The sample used in this study is a mixed-sex sample. The Iowa Facial Growth Study was large at its inception, but shrank precipitously over its run; not enough individuals are in the study at the end to allow for separation of males and females. Rather than focus on determining the pattern of integration at every stage of growth, this study instead decided to focus on the pattern of integration at the beginning of mandibular growth (or at the very least the youngest age available in the Iowa Facial Growth study), the pattern of integration at the onset of puberty, and the pattern of mandibular integration after the cessation of mandibular growth.

The sample in this study consisted of 70 individuals (36 males, 34 females) drawn from the Iowa Facial Growth Study. Criteria were only that the individuals had completed the entire study and had returned to take place in the latter follow-up study. Three developmental stages were created. Group no. 1 consisted of radiographs of the 70 individuals from the time period when the subjects were between the ages of 4 and 5 and one-half years old. Group no. 2 consisted of radiographs of those same 70 individuals from the time period when they were between 10 and 12 years old. Group no. 3 consisted of radiographs of those same individuals taken from the time period when they were between 20 and 29 years of age. Only one lateral radiograph from any one subject was used in each grouping, that is, in group no. 1, if one individual had radiographs at four years, four and one half years, five years, and so forth, the best quality radiograph was chosen.

The developmental groups should encompass most of the postbirth mandibular growth, while not overlapping enough that certain groups would be more similar in their integration patterns because they have near-overlapping
ages. While the developmental groups do have different numbers of years encompassed within them, this should not adversely affect the results. The groups are distant enough in time to encapsulate different periods of growth (i.e., 1st molar, 2nd molar, etc.). Also, because this is a true longitudinal study, any differences between these groups are likely to be reflective of real growth, rather than an a combination of different age ranges involving samples from different populations.

Six osteometric landmarks were located on the lateral radiographs for all individuals in all three developmental groups, and from these landmarks five standard mandibular linear measurements were taken (Table 1; Figures 1(a), 1(b), and 1(c)). Because the proposed research deals with modern humans at different ontogenetic stages, the issue of landmark and linear measurement repeatability must be addressed.

The most important criterion for choosing measurements in an ontogenetic study is repeatability between age or developmental groups [59]. To avoid potential difficulties, only skeletal landmarks and derived linear measurements that could be found at all stages throughout ontogeny were selected; these measurements were taken from human mandibular developmental literature. The measurements were chosen to encompass the entire mandible without being overly redundant in any one dimension (height or length), as this could skew the results, making certain measurements from the same dimension appear more integrated than they are in reality [18, 19].

It should be noted that incision (I) is generally taken at the tip of the medial incisor. However, as the incisors at the beginning of this study are deciduous, and those at the end are adult, this point would not be truly homologous. Instead, incision is taken at the upper margin of the alveolar bone directly in front of the incisor (sort of a mandibular prosthion); this renders this landmark homologous.

These criteria do limit the number of available measurements. Standard metrics of the mandible can include condyle-gnathion, condyle-pogonion, condyle-point B, all of which are considered mandibular length and all of which are redundant to one another to varying degrees. The number of measurements in this study is small; however, the measurement set encompasses mandibular variation and over-redundancy was avoided.

This paper also uses upper mandibular length as a proxy for true alveolar length. Determining the posterior end of the dental arcade so that it may include all the developing and erupting dentition, but none of the ramus, was not something that could be done and replicated on all individuals at all developmental stages. As such, true alveolar length would not be homologous across all individuals, and upper mandibular length was substituted as a reasonable, if imperfect, proxy.

As this study involves deriving linear measurements from radiographs, issues of parallax and magnification must be addressed. Parallax was minimized in the Iowa Facial Growth Study through the combination of a long distance between the skull of the subject and the anode (X-ray source) and a short distance between the subject's skull and the X-ray film; additionally, this same combination is what minimizes enlargement in these radiographs [60, 61]. In the few instances where significant parallax existed and became an issue, midpoint averaging techniques were employed to create an average of the right and left sides of the mandible as they would have appeared in the midsagittal plane [60, 61].

Normally when dealing with linear measurements taken on radiographs magnification must be taken into account; the most common solution is to determine the percent of magnification which exists in the radiograph collection and scale the linear measurements accordingly. Because morphological integration in this study is quantified using conditional independence modeling (via correlation factors), this step is unnecessary. All radiographs in age group no. 1 have the same percent magnification, as do all the radiographs in age group no. 2 and age group no. 3, respectively. Because of this, all the measurements in each age group are scaled by the same factor (percent magnification) and therefore the correlations are the same whether the magnification is corrected for or not. As such, it is unnecessary to correct for magnification as magnification does not play a role in influencing the results of the analysis.

The sample utilized in this study is a mixed-sex sample; mixed-sex samples are nothing new in anthropology, and make up quite a number of the morphological integration literature [6, 10, 12, 13, 18, 19]. While various methodologies exist to deal with sexual dimorphism in linear measurements for primates, the most commonly applied methodology when studying morphological integration is the method of “scaling” one sex so that its measurements more closely resemble those of the other sex. Using this method, an average value for each measurement for both sexes was calculated, and the difference between the male and female averages for each measurement was derived. This difference was then subtracted from the male raw values for each measurement for every individual male; in a sense, the males have been scaled to be females [6, 10, 18, 19]. Modern humans possess a low degree of sexual dimorphism, however, and the analysis was run on both raw (unscaled) and scaled (to resemble females) data; differences in the pattern of integration were negligible, but the results of the scaled data are presented here.

This scaling technique should also alleviate any potential puberty issues in developmental group no. 2 of the sample. While females enter puberty, on average, at younger ages than males, the male sample has been scaled up to be females in puberty. The females sample are all of the same general background and socioeconomic status [58], and as such are likely entering puberty at roughly the same time; slight differences in the timing between individual

<table>
<thead>
<tr>
<th>Variable</th>
<th>Measurement</th>
<th>Abbr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandibular length</td>
<td>Condyle-gnathion</td>
<td>Cd-Gn</td>
</tr>
<tr>
<td>Ascending ramus height</td>
<td>Condyle-gonion</td>
<td>Cd-Go</td>
</tr>
<tr>
<td>Corpus length</td>
<td>Gonion-pogonion</td>
<td>Go-Pg</td>
</tr>
<tr>
<td>Upper mandibular length</td>
<td>Condyle-infraentale</td>
<td>Cd-I</td>
</tr>
<tr>
<td>Anterior corpus height</td>
<td>Menton-infraentale</td>
<td>Me-I</td>
</tr>
</tbody>
</table>

Table 1: Landmarks and derived measurements.
females in the commencement of puberty is unlikely to have a deleterious effect on the analyses of group no. 2. Numerous methodologies exist which have been used to investigate morphological integration [6, 8, 15, 51, 59, 62–64]. Conditional independence modeling was chosen for this study because it has the ability to quantify both the pattern and relative degree of morphological integration in systems. Conditional independence modeling operates on the assumption of relative independence between systems; that is, nonintegrated (or modularized) systems are defined as possessing no statistically significant interaction between component variables, in this case linear measurements. The interaction between any two measurements in this study is referred to as an “edge,” that is, the edge between mandibular length and corpus length or the edge between anterior corpus height and upper mandibular length.

In order to calculate edge-exclusion values, first a standard raw correlation matrix (S) is converted into a “concentration” matrix (Ω) via matrix inversion (Ω = S⁻¹). Following this initial step, the concentration matrix is rescaled, such that the diagonal elements equal 1.0, as in a normal correlation matrix, while the off-diagonal elements are equal to the negatives of the partial correlation coefficients for all the pairwise variables. The following formula can be used to calculate the partial correlation coefficients:

\[
p_{ij \mid K} = -\frac{\Omega_{ij}}{\left(\Omega_{ii} \Omega_{jj}\right)^{0.5}}.
\]

An edge is tested for statistical significance via the calculation of an “edge-exclusion value,” which ranges from zero (complete modularity) to infinity (total integration). While edge-exclusion values can theoretically range to infinity, a brief survey of the literature demonstrates that most values do not exceed 180 [18, 19, 65]. Edge-exclusion values are calculated from these scaled partial correlation values, using the formula

\[
-N \cdot \ln \left(1 - p_{ij \mid K}^2\right)
\]

in which N is the sample size for each group being tested, in this case the number of individuals in each of the three groups. Each edge-exclusion value is tested for statistical significance against a fully integrated asymptotic χ² distribution, with one degree of freedom. This yields a methodologically determined edge-exclusion cutoff value of 3.84 [18, 19, 65].

Edge-exclusion values below 3.84 are not statistically significantly integrated and therefore are modularized, while edge-exclusion values above 3.84 are statistically significantly integrated. If an edge is integrated, then the relationship between the two measurements that comprise it is integrated; if the edge is modularized, the relationship between the measurements is modularized. Additionally, edge-exclusion values determine the relative strength of interaction (integration) between any two measurements; values closer to zero are more modularized, whereas an edge-exclusion value of 100 indicates stronger morphological integration than a value of 4.3, though both are statistically integrated [18, 19, 65].

Besides possessing a methodologically determined statistically significant cutoff value for determining integration, conditional independence modeling also differs from other more common integration methodologies in that it requires no a priori need to assign measurements into functional units (i.e., the corpus and the ascending ramus of the mandible). And while conditional independence modeling is designed to quantify only the degree of interaction between any two variables, by studying the overall pattern of edge-exclusion values, it is possible to determine if different functional units (F-sets) are integrated; strong interunit integration would be indicated by integration of the edge-exclusion values for measurements from the different F-sets [18].

3. Results

Before detailing differences and similarities in the patterns of integration between the three developmental groups, it is advantageous to compare the overall level of integration between juvenile, adolescent, and adult modern human
mandibles. When edges are modularized in the developmental groups, the edge-exclusion values tend to be near zero, indicating strong separation of the regions (Tables 2–4). However, when edges are integrated, there is disparity between the two groups. Juvenile and adolescent modern humans have larger edge-exclusion values for their integrated edges than do the adults, meaning that when integration occurs, it is stronger in the earlier developmental stages than in adults (Tables 2–4). However, edge-exclusion values are designed for testing whether specific interactions between variables are statistically significant, and less for determining a general overview of integration. It has been demonstrated elsewhere [12, 13, 18] that the patterning of integration versus modularity is more informative than the overall level; as such, the pattern of integration shall be detailed and discussed in-depth.

3.1. Group No. 1 (4-5 Years). The overall pattern of integration in this age group (Table 2) is one of the strongly integrated edges in certain regions and strongly modularized edges in others. Mandibular length is characterized mainly by integration in its edge-exclusion values. Strong integration can be seen in the edges between mandibular length and ascending ramus height and corpus length; the former edge is characterized by a slightly more robust degree of integration. Moderate morphological integration is seen in the edge-exclusion value for the mandibular length edge involving upper mandibular length. Modularity is seen only in the edge-exclusion value for the edge between mandibular length and anterior corpus height; this edge displays very strong modularity in its edge-exclusion value.

Integration characterizes the majority of the edges involving ascending ramus height. In particular, strong integration exists between this measurement and both mandibular length and corpus length, as can be seen in the respective edge-exclusion values. The edge-exclusion value for the edge between ascending ramus height and anterior corpus height displays fairly weak integration. Modularity in the edges involving ascending ramus height is found only in the edge with upper mandibular length, with an edge-exclusion value indicating that the modularity found here is quite robust.

Corpus length is characterized by integration in most of its edges. Strong integration shows up in the edge-exclusion values for the corpus length edges involving mandibular length and ascending ramus height; the edge-exclusion value for the corpus length to ascending ramus height edge is the more integrated of the two. A weakly integrated edge-exclusion value characterizes the edge between corpus length and anterior corpus height. Finally, the edge between corpus length and upper mandibular length shows strong modularity in its edge-exclusion value, the only modularized edge involving corpus length.

Modularity makes up the majority of the edges involving upper mandibular length. Integration can only be seen in the edge-exclusion value for the edge involving mandibular length; the edge displays moderate integration in its edge-exclusion value. The edge-exclusion values for the edges between upper mandibular length and ascending ramus height and corpus length both are characterized by strong modularity, with near identical edge-exclusion values. More moderate modularity is seen in the edge-exclusion value between upper mandibular length and anterior corpus height.

Anterior corpus height demonstrates an even split between modularity and integration. Integration can be seen in the edges involving ascending ramus height and corpus length; the edge-exclusion values for both edges indicate weak integration, with the former edge being slightly more integrated than the latter. The edge between anterior corpus height and mandibular length has a strongly modularized edge-exclusion value. Moderate modularity exists in the edge between anterior corpus height and upper mandibular length.

3.2. Group No. 2 (10–12 Years). As with group no. 1, group no. 2 is characterized by strong integration in certain regions and strong modularity in others (Table 3). Mandibular length has strongly integrated edges with ascending ramus height, corpus length, and upper mandibular length. It is strongly modularized in its relationship with anterior corpus height. Ascending ramus height is integrated strongly in its edges with mandibular length and corpus length, but shows modularity in the edges involving upper mandibular length and anterior corpus height, with the former the more modularized of the two edge-exclusion values. Corpus length is integrated relatively strongly in the edges involving mandibular length and ascending ramus height, but is

### Table 2: Edge-exclusion values group 1.  
<table>
<thead>
<tr>
<th></th>
<th>Cd-Gn</th>
<th>Cd-Go</th>
<th>Go-Pg</th>
<th>Cd-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cd-Go</td>
<td>42.07</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Go-Pg</td>
<td>37.5</td>
<td>44.1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cd-I</td>
<td>25.7</td>
<td>0.15</td>
<td>0.12</td>
<td>—</td>
</tr>
<tr>
<td>Me-I</td>
<td>0.007</td>
<td>9.60</td>
<td>6.64</td>
<td>1.04</td>
</tr>
</tbody>
</table>

*Bold values are statistically significantly integrated. See text for explanation of cutoff value.*

### Table 3: Edge-exclusion values group 2.  
<table>
<thead>
<tr>
<th></th>
<th>Cd-Gn</th>
<th>Cd-Go</th>
<th>Go-Pg</th>
<th>Cd-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cd-Go</td>
<td>37.25</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Go-Pg</td>
<td>21.29</td>
<td>19.06</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cd-I</td>
<td>40.03</td>
<td>0.44</td>
<td>0.43</td>
<td>—</td>
</tr>
<tr>
<td>Me-I</td>
<td>0.66</td>
<td>1.89</td>
<td>0.22</td>
<td>0.043</td>
</tr>
</tbody>
</table>

*Bold values are statistically significantly integrated. See text for explanation of cutoff value.*

### Table 4: Edge-exclusion values group 3.  
<table>
<thead>
<tr>
<th></th>
<th>Cd-Gn</th>
<th>Cd-Go</th>
<th>Go-Pg</th>
<th>Cd-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cd-Go</td>
<td>29.51</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Go-Pg</td>
<td>24.11</td>
<td>26.25</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cd-I</td>
<td>11.49</td>
<td>3.02</td>
<td>7.10</td>
<td>—</td>
</tr>
<tr>
<td>Me-I</td>
<td>7.75</td>
<td>0.05</td>
<td>0.005</td>
<td>0.18</td>
</tr>
</tbody>
</table>

*Bold values are statistically significantly integrated. See text for explanation of cutoff value.*
modularized strongly in its relationship to upper mandibular length and anterior corpus height. Upper mandibular length is integrated only with mandibular length, and this relationship between these two measurements is characterized by strong integration; its edge-exclusion values with the remaining measurements are all strongly modularized. Finally, anterior corpus height shows complete modularity with the other measurements, with none of its edge exclusion values being integrated.

3.3. Age Group No. 3 (20–29 Years). The adult mandible has strongly modularized edges in certain regions, while its integrated regions do not show the strength of association as was seen in the earlier groups (Table 4). Mandibular length displays integration in all of its associated edges; however, the degree of integration for these edges does differ. The edge-exclusion values for the mandibular length to ascending ramus height edge and the mandibular length–corpus length edge both display stronger integration than is seen in the edges with upper mandibular length and anterior corpus height.

Ascending ramus height possesses an equal number of both integrated and modularized edges. Moderate integration characterizes the edge-exclusion values for edges involving both mandibular length and corpus length. Modularity characterizes the remaining edges, although differences exist in the strength of this modularity. The mandibular length to upper mandibular length edge has very weak modularity in its edge-exclusion value; this edge-exclusion value is quite near the methodologically determined cutoff value (3.84) for being statistically significantly integrated. This likely is indicative of a biologically important relationship, even though it does not achieve true statistical significance. The remaining edge, mandibular length to upper mandibular length, has a strongly modularized edge-exclusion value.

Corpus length is characterized by mainly integrated edges, which display varying degrees of strength. Moderate integration can be found in the edge-exclusion values for both the corpus length to mandibular length and the corpus length to ascending ramus height edges. Weaker integration exists in the corpus length to upper mandibular length edge, as evidenced by its edge-exclusion value. Modularity only exists in the relationship between corpus length and anterior corpus height; the edge-exclusion value is indicative of strong modularity.

Upper mandibular length shows a mixture of integrated and modularized edges. Integration is seen in the edges involving mandibular length and corpus length. The edges involving mandibular length and corpus length are both integrated, although the former has a slightly stronger relationship. The edge between upper mandibular length and ascending ramus height, as mentioned above, is weakly modularized. Finally, the edge involving anterior corpus height shows strong modularity in its edge-exclusion value.

Anterior corpus height is characterized by modularity. Only the edge between anterior corpus height and mandibular length is integrated; its edge-exclusion value indicates weak integration. Modularity defines the remaining edges. All (ascending ramus height, corpus length, upper mandibular length) are characterized by strong modularity; the least modularized edge involves upper mandibular length, while the most modularized edge is with corpus length.

4. Discussion

The overall degree of integration seen in the modern human mandible during ontogeny decreases, with integrated edges possessing lower edge-exclusion values, indicating weaker integration, in the adult group than in either the juvenile or adolescent groups.

More important than the overall level of integration is the change in the pattern of integration seen across ontogeny. Juvenile modern humans (Figure 2(a)) are characterized by strong integration between the posterior and inferior regions, exhibited by the strongly integrated edge exclusion value between the ascending ramus height and corpus length. Further, there is a weaker but meaningful amount of integration between these two regions and the anterior portion of the mandible, seen in the edge-exclusion values for the edge involving anterior corpus height and ascending ramus height and corpus length. Finally, the juvenile mandible shows separation (or modularity) between the superior aspect and the anterior, posterior and inferior regions of the mandible; this separation is demonstrated by the upper mandibular length’s modularity in its edges with all these measurements.

While the adolescent group (Figure 2(b)) is characterized by an overall increase in the number of modularized edges, it shares the pattern seen in the juvenile group of strong integration between the posterior and inferior margins of the mandible and separation of upper mandibular length from the other aspects of the mandible. The major differences between these two stages of ontogeny can be seen in the lack of integration between anterior corpus height and ascending ramus height and corpus length in the adolescent stage of growth.

Like the patterns seen in the preceding ontogenetic stages, the adult modern human mandible (Figure 2(c)) shows integration in the relationships between the posterior (ascending ramus height) and inferior (corpus length) regions. Despite these similarities, the adult modern human mandible differs from both the juvenile and adolescent mandibular patterns of integration in several ways. Unlike the juvenile and adolescent stages, the adult stage is characterized increased integration of upper mandibular length in its edges with other portions of the mandible. This can be seen in the relatively weakly integrated edge-exclusion value between upper mandibular length and corpus length.

Additionally, the upper mandibular length is extremely weakly modularized in its relationship with ascending ramus height; this edge-exclusion value near the methodologically determined cutoff value for statistical significance indicates that while this relationship is not statistically significant, it is likely biologically meaningful and can therefore be viewed as integrated. The adult stage of growth for modern humans differs from the juvenile stage in also being modularized in the edges involving anterior corpus height, specifically its edges with ascending ramus height and corpus length.
Overall, human mandibular ontogenetic integration can be characterized as maintained integration between the posterior and inferior portions of the mandible, paired with an increasing amount of modularity of the anterior portion of the corpus and an increasing amount of integration of the upper mandibular length with the remainder of the mandible.

What is the likely cause of this change in integration pattern throughout ontogeny? It is important to remember that while the mandible is often regarded as one monolithic structure, it does in fact serve two purposes, albeit both related to the same overall function: mastication. First, the mandible serves as the structure for the attachment of the chewing musculature: the temporalis (ascending ramus) and the portions of the medial pterygoid and masseter muscles (corpus length). Secondly, the mandible houses both the developing and erupted dentition [50, 51, 66, 67]. It is likely that the interaction of these two primary purposes shapes the patterns of mandibular integration seen in both throughout ontogeny.

The modern human mandible displays integration between the ascending ramus height and the corpus length in all of the quantified ontogenetic stages. Both these mandibular structures serve as attachment sites for various chewing musculatures, musculatures that must function together in order for the mandible to work properly. Despite overall changes in the size of the mandible throughout ontogeny, the primary function of the chewing muscles is not altered; they raise and lower the mandible and must do so in concert with one another. In addition to their shared function, these two regions also arise from the same embryonic stem cells [68]. As such, an establishment of a well-integrated relationship between these two regions early on in ontogeny makes sense, as does the retention of this integrated relationship throughout childhood and into adulthood. Predictably, this is what is seen in the patterns of integration in the results of this study.

Changes in the pattern of integration between juvenile and adolescent and adult modern human stages are primarily found in the relationships of upper mandibular length and anterior upper mandibular length shows strong modularization from the remaining mandibular measurements, implying that this area exists as a separate morphological module at these stages of development. Given what this measurement encompasses, the length of the alveolar portion of the corpus, it is likely that this separation is related to aspects of dentition.

At the juvenile and adolescent stages of development the human mandible is undergoing numerous changes related to the dentition. The mandible requires enough length to house all the dentition, both developing and erupted, during this period. Specifically, it needs to be large enough to house all the human deciduous dentition, as well as the developing forms of that same dentition and the buds of two new molars and eventually a third [69] in various stages of development or eruption. All of this must happen while allowing the chewing muscles to continue to function properly. As such, the separation of the alveolar portion of the mandible from those structures which serve as attachment sites for muscles makes sense; it would allow for the continued proper muscle function while at the same time ensuring there is enough flexibility for the developing dentition. The alveolus contains odontogenic cells not seen in the corpus [66], an embryological division between these two areas that is likely reflected in the increased modularity between them throughout ontogeny.

Anterior corpus height involves aspects of the alveolar region, as does upper mandibular length. The relationship between anterior corpus height and ascending ramus height and corpus length is integrated in the juvenile ontogenetic stage. In both the adolescent and adult stages, anterior corpus height is modularized within its edges with those two measurements. Anterior corpus height encompasses the alveolar bed, where dentition develops prior to eruption. In Group no. 1 (4-5 years), the tooth buds for most of the adult teeth are still forming in the alveolar bone, while in Group no. 2 (10–12 years) only the third molars are still forming and most of the adult dentition has erupted, and in Group no. 3 (20–29 years) all dentition has erupted [69]. If tooth buds sense a lack of room, they either fail to develop or become

**Figure 2:** Integrated measurements. Lines indicate measurements whose edges are integrated with one another. A dashed line indicates weaker integration of that measurement with the others shown. See Figure 1 for explanation of (a), (b), and (c) groupings.
impacted [70–73]. It may be that the integration of anterior corpus height in its edges in the juvenile stage is indicative of a need of cooperation within the mandible to ensure proper dental development. Conversely, once all adult teeth have developed, as in the latter stages, this integrated relationship is no longer necessary, and so anterior corpus height becomes modularized in its edges with the other measurements, and the alveolar bed separates from the rest of the mandible.

If the integration of the edges involving anterior corpus height with the measurements representing other portions of the mandible in juveniles is related to dental development, the modularization of this area in adolescents and adults may be related to occlusion. The mandible needs to be high enough so that the dentition can occlude with its corresponding maxillary dentition correctly. If this is the case, then it would be expected that anterior corpus height would integrate with portions of the maxilla in modern humans. It is worth noting that the aspects of the lower maxilla in modern humans from this same Iowa Facial Growth Study are more modularized in adults than in juveniles [19]. This result could support this notion of increased modularity due to the need for occlusion, and is an avenue of research worth exploring.

Upper mandibular length in adults has become more integrated in its edges with most of the other measurements. The edge with corpus length has become statistically significantly integrated. The relationship between upper mandibular length and ascending ramus height also has changed. In juveniles and adolescents, these two measurements are strongly modularized in their edge; in adults, this relationship is still modularized, but very weakly. As such, it is likely that the upper mandibular length-ascending ramus height edge is biologically meaningful, although not statistically significant. It is possible, however, that this increase in integration between upper mandibular length and corpus length and ascending ramus height is influenced by upper mandibular length incorporating aspects of the ascending ramus. Given that upper mandibular length is modularized in its edges from corpus length and ascending ramus height in earlier ontogenetic stages (but still incorporates those same areas of the ramus) this is unlikely, but it cannot be completely discounted either.

Additionally, anterior corpus height is weakly integrated in its edge with overall mandibular length, a relationship not seen in the earlier ontogenetic stages. These relationships together in the adult stage indicate that, once dental development and eruption are complete [69] and mastication becomes the primary function of the mandible, the mandible becomes a weakly integrated superstructure, with strong aspects of modularity contained within.

While the juvenile stage in human mandibular ontogeny is characterized by a pattern of integration favoring dental development and the adult stage by a pattern characterized by chewing, the adolescent stage is characterized by an increase in overall modularity and the maintenance of several edges that are integrated throughout ontogeny. The adolescent stage marks the beginning of puberty in humans, a time of marked skeletal growth. It may be that this increase in modularity is necessary to allow the mandible to be flexible enough to function during a time of rapid skull growth. This same phenomenon is seen in the human cranium during ontogeny. Starting at the developmental stage that encapsulates puberty there is an increase in the amount of modularity seen in the cranium, with the retention of a few key relationships that are always integrated. This increased modularity in the cranium ends at the cessation of puberty and new integrated edges are seen [19], similar to what has been documented in the mandible.

Modularization of the alveolus and the dentition from the other portions of the mandible are a basic mammalian bauplan [66], one which is extraordinarily ancient [68]. Given this, how then does the pattern of integration in the modern human mandible compare to integration in other mammalian mandibles?

This same dissociation of the alveolar region from the ascending ramus was also noted in the mandibles of mice; the mouse mandible showed modularization between these two regions, albeit an incomplete form of modularization, indicating that a relationship still exists between these two areas [51, 52]. Modularity in the mouse mandible is a matter of degrees, not absolutes [52], and because the mandible must function as a whole, the minor degree of interaction between modules seen in the mouse mandible is expected and required [51]. This functioning of distinct modules within the overall bauplan of the mandible can also be seen in the modern human mandible. Mandibular length generally possesses integrated edges with the other mandibular measurements, even when those other measurements come from different modules.

The mouse mandible is also characterized by strong modularity between the anterior and posterior portions of the mandible [51], likely the result of the anterior-posterior chewing pattern seen in mice [50]. It is possible that the increasing modularization seen in the edges involving anterior mandibular height is caused by a similar separation of the anterior and posterior portions of the human mandible (a similar anterior-posterior dichotomy is seen in baboon mandibles [50]).

Baboon mandibles have been shown to possess the same generalized mammalian mandibular integration bauplan seen in mice and modern humans. Like those two groups, baboons are characterized by modularity between the alveolar region and the ascending ramus, as well as modularity between the anterior (incisor) and posterior (molar) alveolar regions. Baboons are distinct from the mouse pattern of mandibular integration in having modules within the ascending ramus, specifically in the coronoid and alveolar processes [50]. If this increased number of distinct modules within the ascending ramus is in fact caused by the medial-lateral chewing motion of baboons [50], then further quantification of the modern human upper ascending ramus may yield similar results, given that modern humans and baboons share a similar chewing motion.

Daegling [53] reported on growth in the African ape mandible, portions of which dealt with integration and allow for a limited comparison between ape and modern human mandibles. He found that the growth in different regions is decoupled (modularized) to a very large degree in ape
mandibles. This same result can be seen in the modern human mandible, where those edges that are modularized tend to be very strongly modularized (Tables 2–4). Also, in African apes, it appears that corpus growth is regulated mainly by the need to house the dentition, whereas growth in the ascending ramus tends to be driven by the musculature [53].

The modern human mandible, during the juvenile and adolescent stages, shares an integration pattern with African apes of modularity between regions of the ascending ramus and corpus [53]. In the adult stage, modern humans are still characterized by modularity between the ascending ramus and the alveolar portions of the corpus, a pattern shared not only by African apes, but by mice [50–52] and baboons as well [50]. These similarities support the hypothesis that modern human mandibles are following a mammalian-wide pattern of mandibular integration, one that is well established and largely invariant.

While the mammalian mandibles appear to share a similar pattern of integration, mammalian crania do not. Modern humans and African apes are characterized by different patterns of integration in the cranium, both as adults [13, 18, 19] and during ontogeny [12, 19], with strong differences in the facial skeleton related to lower facial prognathism and upper facial height [18, 19]. The strong differences in the pattern of integration in the cranium coupled with the strong similarities in the pattern of integration in the mandible between apes and modern humans (in the midsagittal plane) are supportive of the mandible being a more passive structure in the evolutionary history of the African hominoid skull. Instead, differences in the mandibles of modern humans relative to other hominoids are probably secondary consequences of the changes seen in the cranium, particularly the decreased prognathism and dentognathic size [55, 56].

The results of these studies would seem to indicate that growth and integration in the mandible follow Moss’s functional matrix hypothesis, which states that if different regions of a system perform different functions, there is no necessary relationship between size, shape, spatial relationship [3, 74] or even patterns of integration of said regions. Even in a confined morphological unit such as the mandible, there is room for flexibility in various regions at different developmental stages depending upon the function required of those regions at that time.

5. Conclusion

The pattern of integration in the modern human mandible during ontogeny is dynamic. The overall trend is one of increasing modularity of the alveolar portion of mandible from the corpus and ascending ramus from childhood to adulthood, with those latter regions remaining integrated throughout ontogeny. Puberty is marked by a sudden increase in the amount of modularity in the mandible (a condition that is mirrored in the modern human cranium at the same time), which may be the result of a need to increase the adaptability of this region during a time of rapid facial growth.

The overall pattern of integration seems reflective of the dual nature of the hominin mandible: to house dentition and serve as muscle attachment sites. Regions of the mandible which serve these different functions arise from different embryological origins [66]; the modularization of these two regions from one another (ascending ramus and the alveolar portions of the mandible) throughout the development of the mandible is both expected and seen in this analysis.

The areas of the mandible concerned with the attachment of the chewing muscles remain relatively stable during ontogeny with respect to the pattern of integration. The major changes in the pattern of integration in the mandible occur in those regions associated with the dentition. Changes in the pattern of integration in the alveolar regions of the mandible are likely reflective of the differing emphasis placed on the dentition at different stages of growth. Early on in development there exists a greater need for enough space to house all the erupted deciduous dentition as well as the developing permanent dentition; this is reflected in the modularization of the superior portion of the mandible. In adulthood, there is a greater need for interaction between the maxillary and mandibular dentition, seen in the modularization of the anterior corpus region.

The human mandible has a pattern of integration similar to what has been documented in the African apes, a separation of the alveolar and ramus regions [53]; this pattern follows the embryological origins of these regions in the mammalian mandible as a whole [66], and these origins began early in mammalian evolution [68]. The hominin cranium must house the brain, the sight, smell and hearing organs, as well as aspects of the respiratory system; the majority of which must function for survival. The mandible only needs enough muscle function and occlusal contact to allow minimal mastication. Given the differences in cranial integration between humans and chimpanzees during ontogeny [12, 13, 18, 19], and the similarities in mandibular integration, it seems that the mandible plays “follow the leader” with the cranium during hominin evolution.

Acknowledgments

The author would like to thank Tom Southard and the Department of Orthodontics at the University of Iowa for their help in gaining access to the Iowa Facial Growth Study. He would also like to thank Hannah E. Marsh and Steven F. Miller for their insightful comments on this paper.

References


The Archaeological Record Speaks: Bridging Anthropology and Linguistics

Sergio Balari,1 Antonio Benítez-Burraco,2 Marta Camps,3 Víctor M. Longa,4 Guillermo Lorenzo,5 and Juan Uriagereka6

1 Departament de Filologia Catalana and Centre de Lingüística Teòrica, Universitat Autònoma de Barcelona, Edifici B, 08193 Barcelona, Spain
2 Departamento de Filología Española y sus Didácticas, Universidad de Huelva, Campus de El Carmen, 21071 Huelva, Spain
3 Department of Anthropology, Center for the Advanced Study of Human Paleobiology, The George Washington University, Washington, DC 20052, USA
4 Departamento de Literatura Española, Teoría da Literatura e Lingüística Xeral, Universidade de Santiago de Compostela, Campus Norte, 15782 Santiago de Compostela, Spain
5 Departamento de Filología Española, Universidad de Oviedo, Campus El Milán, 33011 Oviedo, Spain
6 Department of Linguistics, University of Maryland, 1102 Marie Mount Hall, College Park, MD 20742, USA

Correspondence should be addressed to Guillermo Lorenzo, glorenzo@uniovi.es

Received 15 September 2010; Accepted 31 January 2011

Academic Editor: John Gowlett

Copyright © 2011 Sergio Balari et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper examines the origins of language, as treated within Evolutionary Anthropology, under the light offered by a biolinguistic approach. This perspective is presented first. Next we discuss how genetic, anatomical, and archaeological data, which are traditionally taken as evidence for the presence of language, are circumstantial as such from this perspective. We conclude by discussing ways in which to address these central issues, in an attempt to develop a collaborative approach to them.

1. Introduction

The emergence of human language is generally seen as one of the major transitions in the evolution of the organic world [1]: the defining characteristic of the human species [2–4] or at any rate a crucial twist within it [5–7]. The issue is not without controversy, as linguistic abilities have been argued to be present in other species of hominids. Relevant evidence ranges from genetic data [8, 9] to the presence/absence of some sort of “symbolic culture” [10–13]—and considerations concerning the anatomy of the organs of speech and hearing abound as well [14–22]. Far from attempting to settle the chronological issue, our goal is to put forth some theoretical considerations that may be useful in evaluating the existing evidence, suggesting new avenues of research. Our perspective, often referred to as biolinguistic, goes back to ideas by Eric Lenneberg, Noam Chomsky, Massimo Piattelli-Palmarini, and others [23–27], which extend from Linguistics to other areas of Cognitive Science [28, for an overview].

In Section 2, we present the biolinguistic conception of language. To avoid a dreadful terminological matter, we will refrain from using the common term “language” and will use instead the expression Faculty of Language when referring to the object that may have evolved, in roughly the sense that organs evolve within organisms. In Sections 2.1 through 2.3, we review what is customarily taken to be evidence for the presence of language, demonstrating how this is questionable when interpreted from the perspective of a mental faculty. We argue that the biolinguistic perspective, with its conception of the Faculty of Language, may help Evolutionary Anthropology in the quest for our origins—especially those of our unique cognitive capacities. More specifically, in Section 3 we offer an example of how the biolinguistic perspective may contribute to progress in research, showing how a shift in focus helps us make significant headway.
2. Language: The Biolinguistics’ View

The Faculty of Language may be defined as a natural system of computation that resides in the mind/brain of all members of the human species. Our definition uses the term “human” simply because (so far) no conclusive evidence exists of the presence of the Faculty of Language in any other extant or extinct species. As a result, the Faculty of Language (as present in humans) provides the only frame of reference for us to test scientific hypotheses concerning one putative such faculty in Neanderthals, or any other species. Being computational, this faculty must be studied as an information processing system based on the application of certain symbol manipulating capabilities [29, 30]. Being natural, it should be understood as an organ that is part of a larger organ system—the nervous system [31]. This faculty is also contingently taken to interface other parts of the mind/brain, intuitively related to “meaning” and “sound” (or “gesture”) [32].

To emphasize this point, we see as accidental properties of the Faculty of Language the fact that, in humans, it interfaces other cognitive systems, in particular a Conceptual-Intentional and a Vocal-Auditory component. This collection of systems (the Faculty of Language + Conceptual-Intentional components + Vocal-Auditory components) is often labeled as “the Faculty of Language in the broad sense,” following Hauser et al.’s convention [33]. Factually, how central the Conceptual-Intentional interface may be to the human Faculty of Language is open to debate, but the interface with the Vocal-Auditory system is certainly contingent [34, 35]. As decades of research into human sign languages demonstrate, these share the structural properties of human vocal languages [36–38]. So the externalization of “linguistic thought” does not privilege the Vocal-Auditory system, and it can also interface Gestural-Visual components.

The fact that (some) Faculty of Language interfaces are contingent is important. Our definition does not exclude the (metaphysical) possibility of a Faculty of Language system that shares definitional properties as outlined above but only some of the contingent properties we normally associate to the faculty in the human sense. Our conception allows, also, for a natural computational system with powers analogous to the linguistic ones to be studied here, but which is interfaced with entirely different mental subsystems. In the hypothetical, the faculty in question might implement “functions” that are entirely different from those traditionally attributed to human language. Underlying these assumptions is Chomsky’s distinction between “competence” and “performance” [32]: a system of knowledge (the Faculty of Language) versus the (various) ways in which this system may be put to use. Inasmuch as these properties of the Faculty of Language are contingent, the natural system of computation itself should be seen as functionally unspecific [31].

Those qualifications constrain the range of hypotheses to test with respect to the origins of the Faculty of Language. In particular, considerations about “systems of communication”—or comparable such “functions” ascribed to language in a vague sense—become orthogonal to the faculty itself. Language as humans experience it serves multiple purposes: to communicate thoughts, to be sure, but also to assert the mere presence of an interlocutor, to lie, to joke, to express beauty, to frighten into submission, to “talk to oneself,” to call a distant star or a number that affects no imaginable communicative act, to describe instances of nonenumerable expressions in mathematics, and surely many other purposes that any reader can fathom. Any of those is a “function of language”, though none of them seems more natural than the others. Importantly for our purposes, the idea that language is “tailored to communication” (or any of the other “functions” alluded to) has provided no particular insight into the Faculty of Language as understood by linguists of our persuasion.

The qualifications above also entail that language, when rigorously understood, is far from a skill to be learned by repetitive training, like skiing or scuba-diving. This is particularly the case for “first languages,” as acquired by children up to puberty. It is an open question whether a “second language” acquired by an adult—often through explicit teaching and rarely to native fluency—may indeed be a skill, subject to variations in ease and rapidity of acquisition, final performance, improvement, decay without practice, and so forth. This unfortunately confounds the matter of interest here. For it may well be that English, definable in some abstract sense as a set of instructions one could (ideally) get in training school for immigrants, has relatively little to do with English in the mind of a native English-speaking preschooler. For perspective, one can train a human to hang glide, by taking advantages of the air currents a condor uses. However, it would seem unjustified to assimilate the mental faculty behind the condor’s (natural) flight to the skills necessary to succeed at human (artificial) hang gliding (this is not to imply that learning a second language does not employ the Faculty of Language in some sense)—after all, only humans learn second languages, even if they do not do it to perfection. The point is raised simply to emphasize the difference between an observable behavior and its underlying causes. If we are interested in the evolution of a natural entity like the Faculty of Language, we cannot satisfy ourselves with vague considerations about whatever mental capacity allows humans to learn languages as adults—we have to go after the natural system in children.

That said, it is also important to distinguish the process of acquiring English (any first language) from the innate developmental process leading to the Faculty of Language in an individual. To use another bird analogy, one thing is for a zebra finch to acquire his (paternal) song (as opposed to a different acquirable song by another conspecific) and a very different thing is for that same zebra finch to develop the mind/brain “circuity” that makes the achievement possible. Now just as the development of acquisition and performance brain circuits is a well-understood prerequisite for successful bird-song behavior [39], so too the development of the Faculty of Language seems to be a precondition for the acquisition of a given human language.

Having set aside contingent properties of the Faculty of Language, it is worth emphasizing its intrinsic characteristic: combinatory power. In short, the Faculty of Language, as a natural computational system, appears to be roughly equivalent to the family of systems traditionally classified...
as (mildly) context-sensitive in the Chomsky Hierarchy of grammars [40–43]. In a nutshell, this means that the system is capable of constructing complex expressions with (i) a hierarchical structure and (ii) dependencies among nodes in the hierarchy that are not expressible as hierarchical nodes, thus requiring a more powerful computation (see Section 3 on this). It is commonly held that no other natural computational system has the same power of the Faculty of Language [33]. Hauser et al. refer to this system as “the Faculty of Language in the narrow sense,” meaning both that it is the core component of the Faculty of Language in the Broad Sense and that it seems to be an evolutionary novelty [33, 44].

Given this set of assumptions, we would like to examine next the evidence that is customarily adduced for the presence of linguistic abilities in a given species. Our interpretation of the data differs substantially from the traditional one in Evolutionary Anthropology and suggests a slight shift in focus. We will concentrate on the sorts of evidence that are most often discussed in relation to language origins: genetic, anatomical, and archeological considerations, in that order.

2.1. Genes for Language?

Molecular biology currently benefits from methodological tools capable of elucidating the differences, at the molecular level, between human beings and other organisms. As a consequence, it can also establish a precise chronological dating of relevant evolutionary changes. Moreover, the progressive optimization of techniques devoted to the analysis of the so-called fossil DNA has opened the way to the possibility of directly measuring the molecular evolution of such genes, yielding a much more accurate temporal perspective of the nature, pace, and magnitude of changes [45, 46]. This extraordinary methodological turn has made it possible for the first fragments of nuclear genes from Homo neanderthalensis to be cloned and sequenced [8, 47–49].

At this molecular level, FOXP2, generally regarded as the “gene of language” [50–54], has been the focus of much attention. This is the case due to two crucial facts: (i) the occurrence of two nonsynonymous changes in the sequence of the protein encoded by it [55], and (ii) the almost certainty that the corresponding Neanderthal gene also contains both substitutions. This suggests that the two modifications in point were not selected in correspondence with the emergence of Homo sapiens (around 200,000 years before present [55]), but in fact much earlier: within a common ancestor of our species and Homo neanderthalensis (c. 500,000 years before present [8]). The latter scenario has led different authors to speculate about the presence of a fully human Faculty of Language—or rather “language” in some general sense—in the second species [9, 56].

Interest in such comparative analyses has extended to other genes whose mutations appear to cause specific language impairments (these are conditions disjoint from neurological dysfunctions, mental retardation, broad cognitive deficits, a hearing impairment, or an inadequate exposition to linguistic stimuli during development). For instance, different substitutions in both DYXIC1 and ROBO1, two genes associated to dyslexia [57–59], have been positively selected in the human lineage: in the first case, after the separation of the evolutionary lines leading, respectively, to humans and the rest of higher primates [57] and, in the second case, between 12 and 16 million years ago [58]. As a complete genetic characterization of the Faculty of Language cannot be made with the exclusion of genes that, when mutated, impair other cognitive capacities besides language, the catalogue of genes of interest is expected to increase in the near future [60, 61].

All these findings point to new and exciting avenues of research concerning the evolution of the Faculty of Language. However, conclusions on this area are undermined by a very common, but unfortunately untenable, assumption that the existence of full-fledged linguistic abilities in other hominids can be automatically inferred from the presence of the human variant of any of these genes.

To begin with, even if many genes have presently been cloned from people affected by specific language disorders [60, 62, 63], paradoxical situations routinely arise. (i) Sometimes relevant genes are also expressed in brain regions not related to language processing, and even in tissues outside the nervous system. (ii) Sometimes such genes are mutated in people affected by other cognitive (i.e., non specifically linguistic) disorders or are simultaneously linked to diverse language impairments. (iii) In some individuals affected by a particular language disorder, the sequence of such “language genes” is normal (phenocopy), while (iv) the linguistic competence of some of the individuals endowed with an anomalous variant of one of these genes is not impaired at all (null-penetrance) or is just mildly impaired (reduced penetrance). Moreover, (v) the identity of such genes differs (to a certain extent) from one population to another or depending on the subtype of the disorder (for a review of different cases, see [60, 63]).

A second point of concern is how genes actually contribute to the regulation of the development and functioning of the neural substrate of the Faculty of Language. Several considerations are worth bearing in mind. (i) Genes do not directly determine language; they just synthesize biochemical products, which will be engaged in particular physiological functions. (ii) Ordinarily, the same gene plays different roles (i.e., contribute to different physiological functions) in diverse moments and body tissues during ontogeny (pleiotropy). Simultaneously, (iii) many genes usually contribute (each to a different extent) to the same biological process (polygenism). Finally, (iv) the extent to which a particular gene product contributes to such a biological process heavily depends on the precise balance it keeps, in a particular moment and place, with the biochemical products encoded by the remaining genes involved.

Other parameters besides genes themselves also contribute to the initial “wiring” of the neural substrate of the Faculty of Language. These include maternal factors (in essence, protein gradients inherited via the egg cytoplasm) and regulatory elements belonging to all levels of biological complexity between genes (and their products) and brain areas [64] (concretely, the metabolome, different subcellular
organelles, the diverse brain cells, the synaptic activities, and diverse specific brain circuits). Furthermore, information relating to the structural features and functional properties of the neural substrate of language could plausibly be generated as a consequence of the developmental process itself [65, 66]. Plus they could depend on general laws that apparently regulate the self-organization of biological systems [67, 68]. All these additional nongenetic factors, robustly appearing and acting at certain developmental stages, can be plausibly regarded as innate. Consequently, what can be deemed “innate” clearly transcends what can be regarded as “genetic” [69].

A third concern relates to the fact that complex regulatory mechanisms probably determine just the basic interconnection patterns among the diverse types of differentiated neurons involved (and, thereafter, the basic histological organization of the main anatomic macrostructures which conform the neural substrate of language). In itself, however, this need not produce fully operative computational devices, understood in more or less customary cognitive terms [70]. Frankly, no one has a clear picture as to how such systems may arise in nature in general, let alone minds/brains [71]. Another way of stating this important issue, emphasizing the distance between what happens at measurable brain levels and what is understood in more abstract mind terms, is that any brain prewiring must be compulsorily implemented by the feedback exerted by neural activity during language processing. Only in such a way is the ultimate cytoarchitecture of the neural substrate of the Faculty of Language achieved, with fully operative neural structures somehow resulting.

A fourth and crucial point to consider is that an increasing body of evidence suggests that most of the molecular changes occurred along our speciation have affected the transcriptome rather than the genome (and, consequently, the relevant protein sequences) [72–76]. These changes would essentially have carried the following: (i) modifications in the expression levels of different genes (and generally in the corresponding protein stocks) [72] and (ii) modifications in the spatiotemporal expression profiles of others, with the subsequent creation of new combined expression patterns; these are probably the basis for the appearance of new structural and functional compartmentalizations at the brain level and eventually of new cognitive capacities [77].

At the same time, both kinds of changes would have fundamentally been the result of the modification of the following: (i) the cis regulatory regions of gene expression (i.e., noncoding regions located proximal to coding sequences on the same DNA strand), as witnessed by the high number of positive selection signals observed in noncoding regions of genes associated with brain development and function [78, 79], (ii) the levels and/or modulating properties of certain transcription factors, components of signal transduction pathways, and noncoding RNAs (ncRNAs) [80] and (iii) the splicing patterns of mRNA of certain genes [81, 82], which could have affected on average around 6 to 8 percent of human exons, as compared with the chimpanzee [82]. Indeed, alternative splicing (i.e., the synthesis of different functional RNAs from the same primary transcript) is more frequent in the brain tissue than in any other [83]. Finally, it is also worth bearing in mind that other innate information storage systems, including mitochondrial DNA and epigenetic mechanisms, appear to exhibit a number of significant differences between humans and higher primates [84].

The case of FOXP2 leads to some particularly illustrative conclusions. The introduction of the human variant in mice produces interesting phenotypic alterations [85], while the human protein brings about in vitro and in vivo a variation in the transcriptional regulation pattern of the FOXP2 factor, as compared to what happens in the chimpanzee [86]. Nevertheless, we mostly lack information about the sequence (and evolutionary history) of the regulatory regions of the gene. Very probably, significant mutations in these regions have occurred throughout the evolution of our species, with effects on the biological activity of the FOXP2 protein probably surpassing those caused by the mutations accumulated in the coding sequence of the gene. Moreover, we also lack enough information about the sequences targeted by the gene [87] in other hominids (including Neanderthals) which have also been subject to positive selection during our recent evolutionary history. In other words, presently we are in the dark even about the existence of alternative expression patterns relatively to that of our species.

Advances in the molecular understanding of language are certainly fast and promising—but still inconclusive and relatively informative (connections between molecular and linguistic data are generally indirect and rather unclear. E.g., the protocadherin 11 gene pair (PCDH11X/PCDH11Y) has been suggested as a putative determinant of language via its role in cerebral asymmetry [88]. Cerebral asymmetry and right-handedness have recently been attributed to Neanderthals and pointed out as an additional clue in favor of the existence of complex linguistic capabilities in this species [89]. However, the relation between cerebral lateralization and language is not an implicational one [90–92]. Furthermore, the relation between lateralization and the peculiarities of the Faculty of Language qua computational system is unclear, if it exists at all [93]). There is, however, one thing we can be sure of: we cannot simply infer the presence of the Faculty of Language from just the existence of the human variant of a group of interesting genes, given all other relevant uncertainties.

2.2. On Speech and Hearing. Since Lieberman and Crelin’s analysis of the Neanderthal vocal tract [14], the debate on the speech capabilities of extinct hominid species has thrived, generally based on anatomical studies [15–19, 94]. This line of research has recently been complemented through the important finding, by the team working at Sierra de Atapuerca, of the ear ossicles of a Homo heidelbergensis. These have made possible to determine relevant features of this and other species’ hearing capabilities [20–22]. In this section, we would like to review these matters, although not attempting to be exhaustive—for example, we set aside studies focusing on anatomical features like the neural canal. Right from the onset we want to say that, in our view, it is unfortunate that this interesting debate should have been twisted into one
on the presence of full-fledged linguistic abilities. Given our discussion above, an inference from a modern speech and hearing apparatus to the Faculty of Language, as presently understood by linguists, is unwarranted. In what follows we unpack our arguments as explicitly as this context permits.

First, it is a contingent fact that some humans externalize their “linguistic thoughts” through a Vocal-Auditory interface; other humans achieve the same result through a Gestural-Visual interface. Thus the presence of a modern anatomy of the vocal tract and of the middle ear, *per se*, does not sanction an inference of the presence of the Faculty of Language (as they are soft organs, the outer and inner ear do not fossilize, so we can only extrapolate their structure in extinct species from comparative studies with other, closely related, extant species. Needless to say, the inner ear plays a determining role in the way the auditory signal is processed [95, 96]). On similar logical grounds, absence of the modern anatomy of the vocal tract would not allow us to infer absence of the Faculty of Language. For perspective, we would not conclude, if examining a human being who is incapable of gestural-visual abilities, a corresponding absence of linguistic abilities—on the basis of the fact that in other individuals the relevant interface does exist. Patently, correlations of this sort are weak. A being with a thoroughly modern Vocal-Auditory interface but no Faculty of Language is as easy to imagine as a being with archaic Vocal-Auditory interface but Faculty of Language “plugged” into some different externalization system (or none whatsoever). There is no a priori reason to dismiss these reasonable theoretical possibilities (see [97–99], for some evolutionary implications).

Secondly, one must carefully separate the perception of sounds as such from the categorization of such stimuli in linguistic terms. A given organism may be able to produce/perceive sounds like, say, [u] or [i], without this entailing that these sounds are interpreted as vowels by the organism in point (these are the sounds that, for a human subject, correspond to the vowels appearing in the English words “bood” and “bead”, resp. Note that English [u] and [i] are always long vowels and thus with a slightly extended duration than the corresponding short vowels found in other languages like Spanish. However the linguistic quality of these sounds is the same in both languages), without this entailing that these sounds are interpreted as vowels by the organism in point. In essence, vowels are linguistic units that, together with consonants, constitute the building blocks of syllables. The latter, in turn, are the basic constituents in the phonological structure of words and phrases, the basis for their characteristic rhythmic structure (rhythmic feet being still higher-order units) and so on. Thus a linguistic symbol is not just a sound, but an element in an intricate system of values that satisfy the famous slogan (attributed both to Saussure and to Meillet) that “chaque fait linguistique fait partie d’un ensemble où tout se tient.” The distinction is not just pedantic. Human babies in prelinguistic stages are able to perceive categorically vocalic sounds like [u] or [i] [100], but so are other mammals, primates included [101–105]. However, no one should attribute the category “vowel” to these nonlinguistic beings—unless one is ready to argue that, for these creatures too, relevant such percepts arrange themselves into a system of interrelated values.

From a biolinguistic perspective, the state of affairs just described means that human languages have their phonetic/phonological structure adjusted to production/perception capabilities of the species. However, perceptual capabilities as such are probably quite ancient. In all likelihood, they are associated to the evolution of the mammalian inner ear and its ability to perform the spectral analysis of complex waveforms in order to individuate their most intense harmonics. Moreover, the data on the perceptual capabilities of mammals are a clear indication that a perfect adjustment between production and perception is not something to be expected in all cases. It is observed in chimpanzees, whose vocalizations appear to contain sounds like human [a], [o], and [u], but nothing comparable to [i] or [e], which, given the sensitivity of their middle ear, are hard to discriminate for them [104]. That case contrasts with that of chinchillas, who perceive but do not produce a variety of the categories that enter the linguistic repertoire in some form [106]. More to the point of our concerns here, we simply do not know what may have happened in earlier hominid species and whether they could or could not discriminate sounds that they were not able to produce accurately [22, 61].

Lieberman [107] argued that if we were to find any evidence for the Faculty of Language in Neanderthals, and they externalized Faculty-of-Language expressions through the Vocal-Auditory interface, we would be able to predict that their phonetic inventory was smaller than ours. Lieberman furthermore suggested that this would be a phonetically less efficient system than ours. However, in point of fact anatomical evidence cannot tell us much about the Faculty of Language. All it can tell us is that human hearing capabilities antedate the apparition of *Homo sapiens*, whereas the modern configuration of the vocal tract seems to be a novelty of this species (note, moreover, that some of the features considered to be critical for the evolution of speech are not as uniquely human as has often been assumed. Thus, a descended larynx may be a human novelty as compared to other primates—but it is also observed in other mammals [108, 109]). Until we have additional, solid, evidence that this change might have had something to do with the emergence of a fully articulated language with a Faculty of Language interfacing a Vocal-Auditory system, the rest is sheer speculation (see [110, 111], for some interesting proposals about how research in this area might proceed).

2.3. An Archeological Approach to Language. Archaeologists studying the Paleolithic tend to agree that the transition to the Upper Paleolithic is one of the most complex, often elusive, research topics in this field [112]. Chronologically located at c. 40,000 years before present, it is a process that includes the demise of Neanderthals and not only the appearance of anatomically modern humans in Europe but their survival and their expansion into Eurasia—including areas that had never been inhabited before by older hominin species. Despite over forty years of studies and a few unquestionable advances, a clear and precise idea of how
this phenomenon took place still escapes us. Predictably, language and other abilities considered modern are central to the debate of the demise of Neanderthals and the endurance of anatomically modern humans (who are assumed to have had the same type of cognitive development level and faculties as present-day humans [6]). Far too often, it has been presupposed that these qualities are what ultimately made the difference between the fates of the two species (see [113], among others). Until a decade ago, a hypothetical revolution that would have taken place with the arrival of anatomically modern humans in Europe was seen as the spark lighting the intense changes the archaeological record of this period reveals [114]. A thorough revision of the African record [115] put an end to speculation, as it showed that every single “revolutionary characteristic” had been independently developed in that continent before being brought into Europe as part of a new-to-the-area “toolkit”—for over 100,000 years. With that result in mind, any analysis that attempts to shed light on the origins of language simply cannot be done on European data alone but must be based on much older remains, unearthed in African soil, where relevant traits first appeared. A simple linguistic argument for this view stems from the fact that the basic structure of natural languages (e.g., in their logical form) is roughly the same in all inhabited continents [116]. All existing evidence points to the direction that an underlying Faculty of Language, no matter how abstract, emerged in a focal African point and was subsequently carried to the confines of the planet.

Needless to say, the origins of the Faculty of Language cannot be directly studied by Archaeology as commonly understood, because of the lack of fossil evidence in this regard. At the same time, Archaeology aims at shedding light on how ancient humans lived and how they adapted to the environment and survived. Some of the ways in which this must have happened seem so intrinsically connected with the use of language that, no matter how invisible their fossil remains may have been, they ought to be taken into account when reconstructing the past. This is why a variety of archaeological projects have focused on this topic. We will concentrate now on the aforementioned set of innovations that left marks or remains in the record: a collection of traits also studied for other periods, involving economy, subsistence, technology, and so forth.

A word of clarification is relevant at this point. One traditional aspect in which archaeology may relate to linguistics, and vice versa, involves the sociocultural aspect of language, especially as encoded in lexical structures. What linguists call a lexicon—which can be defined as a repository of linguistic idiosyncrasies atomized into words—is undeniable a cultural artifact, aside from a fundamental cognitive component that the Faculty of Language interfaces with. Plainly, the way in which humans live affects how they record their living history through their words. In the process skills, traditions, instructions and other forms of “know how” repository may have been coded. It takes little reflection to realize that such explicit or implicit instructions would have such effects as cutting production time for tools or allowing for elaborate living dwellings, let alone artistic or metaphorical creations. In this regard, more or less sudden innovation in any of the latter—some of which do leave fossil records—can be seen as an indirect argument for a lexicon. Inasmuch as the lexicon presupposes a Faculty of Language, this then also constitutes indirect evidence for such a mental organ (see [117], where the argument is presented in a less meticulous fashion).

In the sections below we look into the traits that have been identified as “modernity indicators,” all part of the archaeological record. We will examine their relation to the Faculty of Language, as well as their implications for anatomically modern humans, their expansion out of Africa, and their survival.

2.3.1. Technology. Lithic industries constitute by far the largest corpus of remains in the archaeological record of the Paleolithic period. Accordingly, there is a large number of remains that can be grouped under the label of “technological markers”. We will focus on three types that appeared at different moments during the Middle Stone Age. Among these are microliths (c. 70 ka), points (c. 250 ka), and blades (c. 280 ka) [115].

Middle Stone Age lithic industries represent a radical change from previous industries, both morphologically and technologically: relevant assemblages include smaller tools and new types, such as blades and microliths. Both were thought to have been part of composite tools, formed by a nonlithic section into which several lithic pieces would be inserted. Points were vastly represented across the African Continent during the Middle Stone Age; during that time they constitute a clear case of regional artifact style, a modernity indicator. These were also hafted to shafts, to be used as projectiles [118].

Composite tools imply aspects that cannot be related to older types of tools. These include forward planning (standardized microliths as replacement of similar older pieces broken during use) and the preparation and complementation of different types of materials that had to be worked in different ways and in separate stages, very likely well before needed. These traits are typically attributed to a stage of cognitive development that can perhaps be related to the Faculty of Language.

Complex bone technology appears in the African record at around 110 ka, during the Middle Stone Age. It contains impressive pieces like the Katanda harpoons and points from D. R. Congo [119]. These materials are dated well before the appearance of the split base points that signal the presence of the earliest Aurignacian in Western Europe. McBrearty and Brooks [115] consider that the African bone-working tradition has its origins much earlier in the Pleistocene. The Middle Stone Age levels show that its development was widespread in that continent.

Bone tools and artifacts manufactured using ivory, antler, and shell supports (organic technology) are considered modernity markers. Complex organic technology is one of the innovations that signal the onset of the Upper Paleolithic in Europe. It is at this point that we have the first evidence that relevant materials are worked using techniques that are different from those used to produce lithic tools. Previous
attempts to use bone exist, but relevant artifacts are crudely worked [120, 121]. Of course, the innovation highlighted here relates not so much to the type of materials employed but to the way in which the materials were worked to make tools. Several techniques (polishing, sawing, and abrasion) started at that time to manufacture organic tools. Split-base bone points are an example of artifacts manufactured using such methods.

The abovementioned types of tools have implications within economic parameters. While some saw Middle Stone Age populations as mostly scavengers [122], it is now clear that they hunted. Moreover, their use of points indicates that they had no need to get close to game, which boosted survival rates at the same time that it improved productivity. Chase identifies a specific type of hunting practice as exclusive to 

and early Homo sapiens were present in Africa [115]. Within the extensive list of archaeological traits that start appearing at that time, the following are those classified as symbolic:

(i) regional artifact styles;

(ii) self-adornment objects (ornaments are defined by Mellars as “small objects for which [there is] no obvious functional explanation” [127]. Note that Chase [117] cautions against this direct relationship and offers an ethnographical list of perforated, grooved and serrated artefacts with practical functions), like beads (from c. 82 ka) and other perforated pieces;

(iii) use of pigment—processed (from c. 280 ka onwards);

(iv) notched and incised objects—organic and inorganic materials (from c. 105 ka);

(v) image and representation—also called “naturalistic art” (from c. 45 ka in Africa);

(vi) burials with grave goods, ochre, and ritual objects [128].

A cautionary note is due here. Although McBrearty and Brooks label the list above “modern human behaviors,” so far very few anatomically human moderns remains have been found in association with symbolic artifacts. The same is true about any components of assemblages to which these remains are commonly assigned (Middle Stone Age in Africa and early Upper Paleolithic/Aurignacian in Europe) (physically, the makers of the early Aurignacian are poorly known [6], but fossils from Moravia and the Czech Republic link those to anatomically modern humans rather than to other, older, populations). Finally, there are cases of Neanderthal remains associated to Chatelperronian objects—for example, Saint-Césaire, Poitou-Charentes (France) [129, 130], some of which would qualify as symbolic artifacts according to some definitions.
It is also noteworthy that, in some instances where allegedly symbolic material has been found in relation with Neanderthal remains, bona-fide symbolism has been ruled out as a falsifiable explanation. For instance, while anatomically modern human burials can certainly be considered ritualistic, Neanderthal burials are often described as merely hygienic [131–133]. In a different instance, ochre was found in the latest Mousterian level and the earliest Upper Paleolithic level at l’Arbreda Cave in Serinyà, Banyoles (Spain): to its excavators for the last few decades, this ochre is a sign of symbolic behavior during the earliest Upper Paleolithic at the site. However, these researchers do not understand what the same type of remains mean in the Mousterian layer below [134]. Thus, curiously, the very same ochre found in that layer is not related to symbolic activities [135].

One more interesting debate concerns what happened at around 40 ka, when the aforementioned modern behaviors enter the European scene [136]. Then Neanderthals, who for over 200 ka had developed assemblages entirely devoid of symbolic artifacts, appear to start producing objects of the relevant kind. Did Neanderthals suddenly get the Faculty of Language, or some such symbolic engine? A large number of researchers have preferred to see this situation as a case of acculturation of Neanderthals at the hands of innovative and versatile anatomically modern humans [135].

To sum up, considerable disagreement exists among experts, and a lot of work still needs to take place before the fossil evidence yields more information about the Faculty of Language [117]. But a more serious cautionary note should be added from the perspective of linguistics, concerning the validity of taking the vestiges of symbolic behavior or “symbolic culture” as unquestionable evidence for the presence of the kinds of complex abilities commonly associated to the Faculty of Language.

Factually, a collection of cultural practices correlate with the presence of anatomically modern humans, and a number of these fall under the rubric of symbolic behavior or culture. Given this correlation, the following is often assumed as a valid inference:

1. Symbolic Culture $\rightarrow$ the Faculty of Language.

This inference is often supplemented by its converse:

2. the Faculty of Language $\rightarrow$ Symbolic Culture.

In essence, this presupposes that a Symbolic Culture is only possible with language and that a Symbolic Culture is a necessary consequence of language. Thus,

3. the Faculty of Language $\rightarrow$ Symbolic Culture.

However, these inferences are invalid, based as they are on a false premise that the use of linguistic symbols is a special case of symbolic behavior. We need to clarify this.

We do not question the idea that a Symbolic Culture may indeed be characterized as an instance of bona-fide symbolic behavior. This basically means that relevant practices partake of signification systems established between different entities, through the relations they stand in with respect to other entities within the system (the most explicit formulation of this idea is to be found in Renfrew [137] and Noble and Davidson [7]. It has never been challenged neither by evolutionary anthropologists nor by archaeologists). The problem is that the Faculty of Language does not respond to this characterization for two different reasons: first, the Faculty of Language is not a behavior, symbolic or otherwise, but a natural system of computation. Second, the semantics of natural languages does not seem to be the product of the kinds of relations that make cultural symbols meaningful.

As pointed out by Eco [138], cultures can only be understood as complex and opaque systems of significations. They are complex because the meaning of each particular component depends on the relations it establishes with the other components of the system. They are opaque because we will hardly be able to know the meaning of a particular symbol unless we know how it is used (this is, e.g., one of the arguments for what Renfrew [137] calls the cognitive-processual approach to archaeology and against what he calls the interpretive approach). So in order to properly understand the meaning of a particular element of the symbolic culture of a group of, say, early humans we should know how it was used in its context (how its use related to that of other elements of the same cultural set). It is highly doubtful that the same general conditions extend to natural languages, pace Wittgenstein [139] and Ryle [140] (although Wittgenstein’s work is perhaps the main reference for Anthropology and Archaeology [141]).

The linguistic point is simple. Once we know the meaning of given words (DOG, UNICORN, BROWN, GREEN, etc.) we automatically gain access to the meanings of combinations thereof (BROWN DOG, BROWN UNICORN, GREEN DOG, etc.). This is so even without previous familiarity with the situations in which these symbols might be appropriate. In short, the semantics of natural languages possesses two well-established properties that no cultural system of symbols exhibits: compositionality and productivity. The only known explanation for these linguistic conditions is through the action of a computational system capable of dealing with hierarchical structures.

The contrast is thus clear. While the meanings of the elements making up a Symbolic Culture are opaque until we enter in contact with that Symbolic Culture (to participate in/observe/be informed of the practices in which these elements become meaningful), nothing of this sort applies to the meanings of linguistic complex expressions. These we naturally grasp as we hear them, even with no prior exposition and in the absence of corresponding entities or situations. Nothing of this comes as a surprise once we accept that Symbolic Culture and the Faculty of Language are very disparate entities: Symbolic Cultures are systems of complex and intricate culturally acquired behaviors, while the Faculty of Language is a natural component of the mind/brain of certain organisms (for detailed presentations of this argument see Fodor [30, 142–144]. Wittgenstein was aware of the consequences of the argument and, therefore, tried to show that the semantics of utterances is in fact not compositional [145]. Such a view has not been very influential in semantic studies [30, 146, 147]).
From all this it follows that “meaning” in a cultural system of symbols (or “meaningSC”) is probably quite different from “meaning” in a natural language (or “meaningFL”). A crucial consequence of this is that the capacity of dealing with meaningSC does not presuppose or entail the capacity of dealing with meaningFL. The putative connection between the two in humans is a contingent fact on which we cannot base reliable generalizations. The logic is corroborated by the fact that nonhuman apes appear to be able to acquire symbolic systems, at least under experimental conditions [148]—and perhaps even in the wild [149]. No primate, however, has ever been able to acquire/develop a full-fledged “language,” or even rudimentary versions thereof involving some serious combinatorial syntax.

3. A Biolinguistic View on the Neanderthal/Modern Divide

3.1. Linguistic Complexity. Once again, the Faculty of Language is a natural computational system, capable of constructing complex expressions with a hierarchical structure and nuanced dependencies holding at arbitrarily long distances. The hierarchical structure of linguistic expressions is illustrated by the simple fact that the sentence “The boy says that he likes apples” contains the sentence “He likes apples”, and could be contained within the sentence “Everybody knows that […]”. This means that linguistic utterances are not mere concatenations or linear arrangements of symbols but are instead hierarchically organized sets of units, as the following bracketing representation shows:

(4) [Everybody knows [that the boy says [that he likes apples]]].

As for long-distance dependencies, note for example that (4) contains a subject pronoun (he) optionally coreferring with the subject noun phrase (the boy) in the middle sentence. This preferred (though not obligatory) reading is captured by coindexation:

(5) [Everybody knows [that the boy says [that he likes apples]]].

Further dependences exist in this sentence under the form of subject/verb agreement (“everybody knows”, “the boy says,” and “he likes”). Arbitrarily large amounts of linguistic material can be introduced in between the verb and its subjects, without this having effect in the robustness of the dependency. Thus observe that

(6) Everybody in this (large but still very [very… charming …]) house knows that the boy with the red hat (that was a present form good old Santa …) says that he uncontrollably (and in fact even morbidly …) likes apples.

In addition, linguistic expressions can contain more subtle forms of long-distance dependences, technically known as “displacement” relations. This point can be illustrated with Wh-interrogative sentences, in which question words show up far away from the position where they receive interpretation as verbal arguments, as in (7) (where “e” represents the “empty position” left behind by question word after “moving”):

(7) [What does everybody know [that the boy says [that he likes e]]].

The ones just reviewed are well-established linguistic facts [28, 150]. They are, in a nutshell, the reasons behind the characterization of the Faculty of Language as a “Type 1” system in terms of its computational complexity, using as a reference point the Chomsky Hierarchy of formal grammar [40–43], which we review next.

3.2. Formal Complexity. The said hierarchy defines different classes of “formal languages” (or corresponding grammars), arranged in an increasing scale of complexity. In these mathematical constructs, a “formal language” is understood as a set of strings of symbols generated under certain general admissibility conditions. Crucially, for a system of these characteristics to work, a finite collection of rules, describing the admissible strings in the language, can be produced and result in a computation that halts at some point (it should be easy to see, given the characterization of a “formal language” just introduced, that this mathematical notion is not synonymous to the biolinguistic notion of language defined at the outset of this paper. The notions are however related in some abstract sense [151]).

This ensures that, given the rules of a language (technically, its grammar), some computational device exists capable of mechanically generating any of the strings of the language in question. However, since the arrangement of symbols in a string may be more or less intricate, in definable ways, some languages may require more or less complex devices to generate them. The complexity of a language-generating device (an automaton) is essentially defined in terms of the amount and sophistication of its memory resources. Simply put, more complex languages can only be generated by automata with the appropriate memory resources. This distinction is what underlies the traditional classification of languages, grammars, and corresponding automata from Type 3 (or “finite-state”, the simplest ones) to Type 0, the most complex.

The Chomsky Hierarchy therefore provides a useful frame of reference to determine the complexity of Turing-computable problems (the architecture of relevant automata was defined by Alan Turing by theoretically imagining a logical processor writing operational steps on an infinite tape, one step at a time. When one speaks of “memory” within this system, one is basically referring to the ability to designate sections of the writing tape not so much for the purposes of carrying the computation forward but rather with the purpose of storing instructions to be used at later computational times. Different memory regimes determine, in the end, the overall complexity of the “formal languages” so characterized). Any such problem, inasmuch as it is computationally tractable, may be expressed by way of a “formal language” in the Hierarchy—success in this task being just a matter of identifying the critical properties of the problem. Type 3 languages are so simple that they can be
described by an automaton with no memory (strings in this type of language are in fact like beads arranged in a linear fashion, with no further internal structuring).

Next up in the hierarchy, we find Type 2 or context-free languages. These need an automaton with enough memory to keep track of what structure is being built, while some embedded substructure is being further constructed. Context-free languages already provide a good measure of complexity for natural languages, since most of their structures fall within the computational capabilities of a “push-down” automaton. This is so called because it contains a simple memory “stack,” such that the last item stored into the “stack” is the first one to come out, the top of said “stack” always being involved when recalling items from memory.

Push-down automata may actually describe some long-distance dependencies: those that happen to be “nested” within one another, as in the English example in (8a) (part of the sentence “Peter says that John wants to let Mary read the book”). However, natural language dependencies are known to also be “crossed-serial.” One relevant example is the translation of (8a) into a language like Dutch:

\[
\text{That John wants to let Mary to read the book}
\]

\[
(8a)
\]

\[
\text{dat Jan Marie het book wil laten lezen}
\]

\[
(8b)
\]

Push-down automata are not equipped to deal with the sorts of dependencies in (8b). This is so because no simple “stack” regime can allow the computation to, at the same time, establish a dependency between two items in the computation and to continue holding an item in between those two in active memory—for subsequent computation. “Stacks” are too simple minded a memory: the entire set of stored items has to be active up to the very last item in storage, and this last item must be the first to be recalled.

A more powerful kind of automaton is needed to generate cross-serial dependencies, either with allowable manipulations within the “stack” (not just at the top) or a different sort of memory procedure. It is this property of natural languages that makes them characterizable as Type 1, or context-sensitive, within the scale of computational complexity defined by the Chomsky Hierarchy (see Table 1 for a summary). Indeed, most theoretical discussion in the last half century has ultimately centered around the issue of precisely how (and when) linguistic structures happen to be of this complex sort.

Table 1: The first three levels of complexity of the Chomsky Hierarchy, with a formal example of the kinds of structures each generates. A context-free grammar may keep track of the number of symbols in every subset so long as it deals with a maximum of two correlated subsets. If the same number of symbols in three (or more) correlated subsets is required within a given “formal language,” a context-sensitive grammar is required to describe it [152].

<table>
<thead>
<tr>
<th>Level of complexity</th>
<th>Language</th>
<th>Sample string</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 3, finite-state</td>
<td>a^b^c^</td>
<td>aabbbcc</td>
</tr>
<tr>
<td>Type 2, context-free</td>
<td>a^b^c^</td>
<td>aaaabbc</td>
</tr>
<tr>
<td>Type 1, context-sensitive</td>
<td>a^b^c^</td>
<td>aaabbbccc</td>
</tr>
</tbody>
</table>

3.3. Is the Complexity of Knots Relevant to the Archaeology of Language?

From our evolutionary perspective, an interesting question arises in light of the formal facts just reviewed. One way to determine whether a given hominid species had the Faculty of Language (as presently exhibited by us) would be to test their computational capabilities—in order to determine how high they were within the Chomsky Hierarchy. Obviously this is not doable in any direct fashion, since the fossil record does not contain direct linguistic evidence of the right sort. Nevertheless, there may well be an indirect manner to proceed that could take advantage of fossilized remains, by seeking traces of language in domains that, while not being directly defined as linguistic, may presuppose a “technical intelligence” that could well be, in some sense at least, parasitic on the Faculty of Language [153]. The prospect is realistic inasmuch as, as emphasized at the outset, this faculty interfaces with other cognitive systems and, through these, with general cognition and the mechanisms underlying behavior [33, 154].

From this perspective the key is to observe relevant aspects of the fossil record with a “grammatical lens,” thus asking what sort of algorithm would computationally describe a given rule-governed behavior. If such an algorithm happens to be low within the Chomsky Hierarchy, not much can be surmised from the exercise, since behaviors thus described are common in animal cognition. But if the opposite is the case, and a hypothesized algorithm to describe a given behavior happens to fall high within the Chomsky Hierarchy, the result would potentially be significant. This is so because it is very rare to find bona-fide complex computational behaviors in the natural world. When or if such behaviors are isolated and properly described, three possibilities emerge for them: (i) that they correspond to a mental capacity that is totally unrelated to the Faculty of Language, (ii) that they depend, instead, on some interface with the computational procedure that the Faculty of Language presupposes, or (iii) that the said behaviors obey the conditions of a faculty that actually underlies both the Faculty of Language and whatever is responsible for the inferred behavior.

The idea of connecting artifactual properties with the presence of language (in some sense) is not new. But such
exercises are rarely accompanied by rigorous criteria for evaluating the validity of underlying correspondences [155]. The foregoing discussion is aimed at addressing this methodological concern. A proposal originally made by Uriagereka et al. [156, 157], concerning the capacity to tie knots, is a model story of what we are arguing for. It involves a unique behavior within apes, whose computational description falls high within the Chomsky Hierarchy and whose results are inferable from the fossil record.

Particular knotting techniques (say, “clove hitch,” “Eskimo bowline,” and so on) are cultural practices. What interests us, instead, is the fact that, underlying these technical traditions, a certain natural capacity exists that, so it seems, is not accessible to other primates. One way to assess the complexity of knot tying is to resort to their mathematical characteristics. Knot-theory is the branch of topology that deals with the nature and properties of knots. From this perspective knots are conceptualized as elastic, closed, and tangled strings. The most basic knot (the unknot) is like a circle (i.e., a string joined by its two ends) lying on a single plane. More complex knots are constructed by crossings of the string, such that some parts thereof lie on more than one plane. The minimal knot is a string with three crossings (the so-called threefoil knot). An important area of knot-theory is to determine whether a complexly tangled string is the unknot, and if not what kind of knot it is (the “unknotting problem”). Formal details aside, the task of determining whether any given string is knotted is known to have a complexity comparable to the one needed to process an expression in a natural language (for given knots the computational complexity can be greater than that needed to process linguistic expressions, which moves us into obscure issues dealing with mathematical intuition in humans. Still, for our point to be relevant it is enough to think of simple knottings of the sort routinely used at work in human societies. For an introduction to knot-theory see [158], and [159] for complexity issues).

When actually making a knot, humans must, at a certain point in time, relate a portion in the knot with the background “figure”. Intuitively, this is an operation in which both grouping and long-distance-like relations are implied [156, 157, 160]. If so (un)tying knots (or determining whether a tangled string is knotted) seems to require an underlying computational system of Type 1 (or even a more powerful system). Once again, such a system is context-sensitive, which is to say capable of keeping track of the computational history until the overlapping(s) needed for knotting take(s) place. The process as a whole can be modeled by storing some (arbitrary) elements A, B, C, in that order, in some computational stack, to then proceed to relate element A at the bottom of the stack to some element D in the current state of the computation (the crossing). This could be represented as in (9a), which is to be compared to (9b):

(9a) \[ D_1 \ [ C \ [ B \ A_1 ]] \]
(9b) \[ \text{What, does everybody know [that the boy says [that he likes e]]} \].

Abstracting away semantically irrelevant symbols, the formal parallel with a long-distance dependency should be clear.

Again, the correspondence itself could signal the existence of a grammar for knots. However, these sorts of dependencies are exceedingly rare within the primate world. It seems more plausible to ascribe the parallelism in (9) either to the bona-fide Faculty of Language (extended in the appropriate cognitive direction [161]) or to a deeper cognitive system underlying both that faculty and some putative system specific to knots. If either of the latter conditions holds, finding structures of the sort in (9a) in the fossil record would argue for the Faculty of Language being in place by that time.

One last important clarification is in order: being able to learn a specific motor sequence to tie a knot is not sufficient evidence for inferring complex cognitive capacities. The scarce literature on knot-tying abilities in humans has only focused on how people learn to tie a knot either by instruction or by imitation [162, 163]. This may tell us something about how a cultural practice may have been transmitted, but it says nothing about the process of inventing new knots, which humans have been doing for millennia. Similar confusions often arise in the literature on animal behavior, for example when attempting to demonstrate whether birds can parse nested dependencies [164]. As has been noted by many [165, 166], training a bird to successfully identify a couple of such dependencies may tell us nothing about the ability involved in creating any new such expression. Full creativity has always been what is most puzzling about human language.

3.4. Setting the Record Straight. Knots are not directly attested in Anatomically Modern Humans until 27 ka B.P., by means of weaving, both in clothing and clothing representations [167]. However, they can be inferred long before that, from purposefully perforated ornaments (beads, teeth, shells, etc.) and small projectile technology (arrow heads, arches, and harpoons), the oldest evidence of which is about 90–75 ka ([168–170], and [171] for an even earlier date). Those dates, of course, are tantalizingly close to what is presumed to be in the range of the emergence of the Faculty of Language.

From this perspective, a very intriguing issue is whether Neanderthals (or, for that matter, other hominids) were capable of knotting behaviors. Perforated shells dated at 50 ka and older, found at the Aviones site, Spain, have been presented as an indication that Neanderthals shared some of these practices with Anatomically Modern Humans [13]. The case merits serious examination, but the fact that the perforations in point are not deliberate (that is, they are indisputably due to natural causes) makes one wonder to what extent the relevant ornament was used in a deliberate way as a carefully crafted and prominently worn piece of jewelry, possibly a mark of social status.

Possibilities to interpret the relevant data, even from a biolinguistic point of view, are multiple. However, the most promising seem to be two. The first is that the Faculty of Language is an anatomically modern human evolutionary novelty among primates. By and large, this possibility fits well with the strong contrast between the material culture of anatomically modern humans and Neanderthals.
The extremely diversified and dynamic character of the former could be a reflection of the open-ended productivity of the Faculty of Language, a consequence of its computational properties [177]. Note that a computational system of Type 2 is enough for genuine “recursion” (or systematic and unlimited nested embedding) [33, 178]. This is the formal property that explains why sentences have no upper limit of components and, therefore, that there is no upper number of possible utterances a language allows [179]. However, recursion is still insufficient to deal with the real complexity of human languages. The qualification is in order because it could even be the case that Neanderthals had achieved one level of complexity without reaching the other see [180–183] for perspective. The second hypothesis is that the Faculty of Language is a feature shared by both anatomically modern humans and Neanderthals, already present in their common ancestor (i.e., a sapiens synapomorphy).

The latter position would rationalize Krause et al.’s finding concerning the antiquity of the FOXP2 genetic variant [8]. It also would explain the abilities underlying the ornaments of Aviones site (see above; [13]) particularly if some independent proof is found of their use as beads tied together by some sort of thread. This hypothesis would be consistent with the possibility that there might exist other Type 1 computational systems in nature—beyond the human Faculty of Language. One case to examine seriously involves species of weaver birds that tie knots as a part of their nest construction techniques. Some of these knots are (near) equivalents of human knots [184, 185] (some apes in captivity have been reported to tie simple knots [186]). The jury is of course out on whether this means that the birds in point have the cognitive equivalent of a Faculty of Language, albeit with nest-building consequences instead of anything familiar to humans. Key to answering that question would be to determine the level of complexity the birds can attain in their knot-tying abilities. Patently, some of their relevant knots are more complex than others, and the question is whether any or all of them can be produced by the compilation of a motor skill (for example, the knots used in Michel and Harkins’ experiment [162] are simple (the sheepshank, the butterfly knot, and the “magic” slip-knot), and yet only 37% of their subjects were able to learn to tie the three of them by just attending demonstrations, i.e., observing the necessary motor sequence to tie them).

Clarifying either position should also help us understand the late cultural achievements of Neanderthals: whether they constitute an acculturation effect from contacts with anatomically modern humans [12, 187], an independent cultural development [11, 130], an intraspecific “last minute” evolutionary event, or even the result of interbreeding [49]. It is good to have new tools to address such open questions.

4. Conclusions

In presenting the biolinguistic approach to the study of language, this paper has developed some ideas on how that approach may contribute to the study of human history. For the brand of linguistics we represent, the communicative or even symbolic aspects of language are not as central as its formal properties. Only by focusing on these properties are we able to draw a clear picture of the computational resources necessary to generate a human language. The natural cognitive computational system capable of deploying these resources—the Faculty of Language, an organ of our minds/brains—is thus our main subject of inquiry. Given this basic assumption, we have explored how research on Evolutionary Anthropology can be complemented: by qualifying or reassessing the interpretation of the existing data. It is our hope that future research will benefit from the perspective offered here, by strengthening the interdisciplinary stance that the inquiry into the origins of our species necessarily requires. When we need to date remains, we turn to Physics and Chemistry; when studying hominid remains, we ask palaeobiologists and geneticists—and experts on nutrition are consulted when palaeodietists are the focus. Similarly, when aiming at progress in the study of language and its origins, linguistics should prove to be useful. We hope it has been.

Acknowledgments

This work has been carried out through the project Biolinguística: evolución, desarrollo y fósiles del lenguaje (FFI2010-14955) from the Ministerio de Ciencia e Innovación (Spain) and partially cofunded by FEDER funds (EU). It also received partial support from the Generalitat de Catalunya through Grant 2009SGR1079 to the Centre de Lingüística Teòrica of the UAB. The authors want to express their gratitude to two anonymous IJEB reviewers for their helpful comments.

References

Research Article


Matthew L. Sisk¹ and John J. Shea²

¹ Interdepartmental Doctoral Program in Anthropological Sciences, Stony Brook University, Stony Brook, NY 11794-4364, USA
² Department of Anthropology, Stony Brook University, Stony Brook, NY 11794-4364, USA

Correspondence should be addressed to Matthew L. Sisk, matthew.sisk@stonybrook.edu

Received 15 September 2010; Accepted 31 January 2011

Academic Editor: John Gowlett

Copyright © 2011 M. L. Sisk and J. J. Shea. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Despite a body of literature focusing on the functionality of modern and stylistically distinct projectile points, comparatively little attention has been paid to quantifying the functionality of the early stages of projectile use. Previous work identified a simple ballistics measure, the Tip Cross-Sectional Area, as a way of determining if a given class of stone points could have served as effective projectile armatures. Here we use this in combination with an alternate measure, the Tip Cross-Sectional Perimeter, a more accurate proxy of the force needed to penetrate a target to a lethal depth. The current study discusses this measure and uses it to analyze a collection of measurements from African Middle Stone Age pointed stone artifacts. Several point types that were rejected in previous studies are statistically indistinguishable from ethnographic projectile points using this new measure. The ramifications of this finding for a Middle Stone Age origin of complex projectile technology is discussed.

1. Introduction

Recent fossil discoveries and genetic analyses indicate that Homo sapiens evolved in Africa by at least 200 thousand years ago (ka) [1, 2]. From this point until the present there are two very different patterns in both geographic range and behavior. Until 50 thousand years ago, Homo sapiens remained endemic to Africa with only a brief expansion into the contiguous Levantine corridor [3, 4]. Around 50 ka, there is considerable evidence for a dispersal that expanded the range of Homo sapiens throughout the Old World [5–8]. Absolute dates for this dispersal vary considerably, though most fall in the range of 60 ka to 35 ka (e.g., [9, 10]). Most researchers agree, however, that following this dispersal, Homo sapiens fossils are associated with a material culture more closely resembling that of ethnographic hunter-gatherers than that of previous hominins [11–13]. Thus, this dispersal has long been thought to mark a behavioral “revolution” or a significant shift to more complex behaviors unique to Homo sapiens. These complex behaviors may include the production of labor-intensive stone, antler and bone tools, the transfer of raw materials across long distances, the creation of a wide variety of personal adornments and other symbolic objects, and the development of more complex subsistence behaviors (i.e., broad-spectrum foraging and specialized big game hunting). The mode and tempo of this dispersal are hotly debated topics in paleoanthropology (e.g., [4, 11, 14]), but it is generally assumed that these novel behavioral adaptations played an important role.

Traditionally, these post-50 ka behavioral changes have been viewed as part of a single process: the evolution of “modern” human behavior. However, there is evidence for the presence of some of these components earlier in the African Middle Stone Age (MSA) [11]. Furthermore, there is no compelling reason to assume synchronicity in all of these components [15]. An adaptive shift that resulted in specialized hunting would not necessarily also cause the development of personal adornments, although synergy between these, and all factors, remains a possibility. Thus, while some researchers still view this change holistically, as a single watershed event (e.g., [4]), many others have turned to investigating the sources and results of adaptive shifts analytically in terms of particular component behaviors (e.g., [15–17]). In particular, one of the most striking aspects of
Homo sapiens adaptation after 50 ka is the occupation of a wide variety of environments. Thus, in attempting to understand the evolutionarily significant behavioral changes that led to the success of Homo sapiens, special attention should be paid to those adaptations that increase ecological versatility among recent human populations.

Complex projectile technology is one such adaptation. Complex projectiles, like the bow/arrow or spearthrower/dart, are composite technologies that propel a high-velocity projectile by storing or enhancing energy in a non-projectile component [18, 19]. Simple projectile weapons, in contrast, are those that rely solely on human mechanical energy for propulsion, like hand-cast spears, javelins, and throwing sticks. Complex projectile technology is a universal part of contemporary human adaptations, found among societies ranging from hunter-gatherer bands to industrial states, indicating that these technologies confer a significant ecological advantage. Furthermore, living human groups use projectiles to hunt prey of dramatically differing sizes and in diverse habitats, ranging from arctic deserts to tropical rainforests [20, 21]. Thus, these technologies increase hunting versatility and allow the construction of a wider ecological niche.

The development and widespread use of complex projectile technologies undoubtedly marks an important transformation in the adaptation of Homo sapiens. Their origin is definitively Pleistocene [22], and they likely contributed to the success of the 50 ka expansion of Homo sapiens into western Eurasia [18, 23]. However, the origin and nature of early projectiles remains an enigma.

While simple projectile weapons, like javelins, have been found in Lower and Middle Paleolithic contexts [24, 25], the oldest definitive complex projectile weapons date to relatively late in the Paleolithic [26]. Most of the materials used in the manufacture of complex projectile weapons, such as wood, cordage and sinew, are highly perishable. It is generally assumed that these technologies predate their oldest known examples, but how widely and for how long remain points of dispute. Despite issues of preservation, the use of complex projectile weapons can still be inferred from more durable materials. In several cases, microwear and residue analyses have suggested at least simple projectile use (i.e., hand-cast spears with stone tips) in the Middle Paleolithic/Middle Stone Age (e.g., [27–30]). Yet, such microwear and residue traces preserve only rarely, and the time-consuming process of detecting them limits analysis to small judgmental samples. Rough similarities in the form of Paleolithic points and stone projectile points of ethnographic or recent historical groups can also give some indication of projectile use [31, 32] but care must be taken when forming an analogy between modern and ancient behaviors or technologies based on gross similarity. Recent projectile points, especially those in museum collections, are often heavily retouched into a specific form. These forms are often thought to serve as stylistic markers for the group [33, 34], although the chronological controls of these markers remains somewhat contentious [35–38]. Given that the interplay of stylistic and functional variation in recent and ethnographic point forms is not fully understood, it is clear that gross similarity in form cannot be directly used to compare points from time periods like the Paleolithic, where cultural associations are unknown.

Furthermore, experimental work has demonstrated that unretouched, minimally retouched [39, 40], and even wooden points [41, 42] function adequately as projectile armatures. This suggests a possibility that some strategies involving the use of complex projectile technologies may evade archaeological detection except in cases of extraordinary preservation (e.g., waterlogging, freezing, or dessication). While there may be some functional advantage in crafting the perfect projectile point, the advantages must have been weighed against the “costs” of time and effort. Thus, it is important for researchers to look not only at the particular forms of recent projectile points as guides to recognize ancient examples, but also at ballistically significant metrics derived from ethnographic and experimental data.

1.1. Approaches Using the Metrics of Stone Points. Thus, many researchers have begun to use such ballistically significant measurements in determining if a given class of Paleolithic points was functionally capable of serving as projectile armature (e.g., [23, 41, 43, 44]). Using measures like the weight, convergence angle, or cross-sectional area, these studies have demonstrated the potential systematic or occasional projectile use of certain tool types. It should be stressed that these studies, including the current one, only comment on potentiality. Metrics that resemble modern projectiles could be the result of a different technological adaptation for small, pointed pieces. No single measure can or should be used as a definitive test of projectile usage. Regardless, studies like these are necessary first steps in investigating Paleolithic projectile use. At the very least, metric comparisons of known projectile points to prehistoric tools of unknown function can help archaeologists prioritize samples of the latter artifacts for more diagnostic studies employing microwear, breakage, and residue analysis.

One factor that limits the utility of metric analyses is our current understanding of factors influencing projectile penetration in prehistoric technological contexts. Ballistics measures are typically applied to, and derived from, firearms or high-powered bows with metal tips. Thus, these may not be accurate when applied to technologies available in the Paleolithic. Experimental studies (e.g., [40, 45, 46]) that have tested the accuracy of these measurements generally show that while simple measures may serve as a proxy, they may also be biased towards certain forms. As experimental work continues to untangle the penetration efficiency of stone points propelled by different technologies and at various ranges, it is vital that models of Paleolithic projectile use are updated as our knowledge of these measurements grows.

1.2. Tip Cross-Sectional Area and Perimeter. Most important among these ballistic measurements is the tip cross-sectional area (TCSA) [23, 47, 48]. The TCSA should represent the force necessary to penetrate a target to a lethal depth. Thus, the larger the TCSA, the more force that is needed. As a mechanical rather than a stylistic constraint, Shea [23] argued that the TCSA of a given artifact type should
have a central tendency that resembles ethnographic or experimental projectile points in order to be considered a plausible projectile armature. TCSA also has the advantage of being an easy calculation, requiring only the maximum width and thickness of a stone point.

Previous work has shown a clear separation in the TCSA of different projectile types (bow/arrow, spearthrower/dart, throwing spears) resulting from effective force and the characteristics of the technology [40, 45, 47]. The TCSA of archaeological points can then be compared to collections of hafted ethnographic arrowheads and dart-tips to give an idea of the projectile technology for which they were designed [40, 45, 47]. Artifacts with TCSA values that resemble ethnographic projectile points may then be interpreted as potential effective projectile armatures. In particular, Shea [23] applied this measure to samples of points from Africa, Southwest Asia and Europe to show when the earliest plausible stone projectile points occur in these regions. In this study, points that were statistically indistinguishable from ethnographic projectiles were only found in the Upper Paleolithic and the Late Stone Age (younger than 40 ka). This finding does not support a hypothesis of earlier projectile use, although it is noted that nonsystematic or occasional use would likely go undetected [23]. A further complicating factor arises from archaeologists’ practice of grouping stone tools of widely varying sizes together into the same artifact type. This practice can result in plausible projectile points, which are usually relatively small, passing undetected among large numbers of very large points.

Initial experimental work supports the utility of TCSA as an acceptable proxy for potential projectile use [40, 45]. However, its utility may actually derive from the fact that it tracks another measure, the tip cross-sectional perimeter (TCSP). Measures from ballistics, like TCSA, predict a case where the projectile, such as a bullet, pulverizes the target. This makes the area an important predictor of the size of the hole, which is then used as a proxy for the energy needed to penetrate to a lethal depth. Stone points instead work by effectively slicing a hole in the target. In this case, the size of the hole is controlled by the point’s outer margin or its perimeter, rather than by the area [40].

Both TCSA and TCSP are composed of the same measurements, maximum thickness and width, so in all cases they should be very highly correlated. Therefore, previous trends isolated by comparing TCSA values are likely accurate, but they may result from tracking variation in TCSP rather than from the analytic power of TCSA.

There are some limitations to using TCSP analytically. In particular, the perimeter measure is more sensitive to shape differences [40, 47]. Many modern ethnographic projectile points are bifacially retouched, yielding a roughly rhomboid or biconvex cross-section. However, simple pointed flakes have a cross-section more accurately represented by a triangle or trapezoid (Figure 1). When modeling the cross-section of a point, one of these estimates must be chosen. For the area, the calculation is the same in either case, but for the perimeter, these two estimates result in different values (Figure 1). However, because using the triangular measure increases the perimeter, it actually reduces the likelihood that a given class will be grouped with bifacial ethnographic arrowheads and dart tips (Figure 2) and is thus, in some ways, a more restrictive measure.

2. Methods

This analysis uses TCSP on pointed artifacts from African MSA contexts. To directly compare the results of a TCSP measure to a known TCSA-based study, it utilizes the same dataset of lithic point measurements used in Shea’s [23] analysis of TCSA across the Paleolithic Old World. Because the current analysis is primarily concerned with the potential African origins of complex projectile technology, only the point classes from the African Middle Stone Age are included (n = 1863; see Table 1). It should be noted that, in a few cases, the total number of artifacts in each sample is slightly
less than that of Shea [23]. This is due to the removal of point measurements that, upon reexamination, appear to be duplicates or otherwise questionable.

These points were compared to the same ethnographic control samples as the previous analysis [23]. These include a collection of 118 arrowheads and 10 spearthrower dart tips from of the American Museum of Natural History measured by Thomas [49] and 30 dart tips from various other museums measured by Shott [50]. Summary data for all of these points are reported in Table 1. In this study, we do not include the comparative sample of experimental thrusting spear points used by Shea [23]. This sample was excluded both because of our focus on complex projectiles and because this experimental sample may not represent the actual variability in this technology. We instead assume that TCSA or TCSP values exceeding those of the ethnographic sample are outside the functional threshold of complex projectiles.

The MSA point dataset contains both bifacially and unifacially worked types, so both the triangular and rhomboid measures were used. The ethnographic arrow and dart comparison sample are all classified as bifacially worked and the perimeter estimated as a rhombus. This difference in the calculation of values for both the archaeological and comparative sample may introduce error into the analyses. Therefore, all analyses were conducted twice, once with the variable perimeter estimation and once with all perimeters estimated by the rhomboid measure.

3. Results

As discussed above, the triangular estimation results in a larger perimeter measure. This then reduces the likelihood of a triangular point type being classified with the bifacial ethnographic controls. Therefore, the analysis using the rhomboid estimation for all points will be discussed only in cases where the estimation changes the grouping. All data, including which estimation was used for the variable measure, are reported in Table 2.

Figures 3 and 4 show box and whisker plots for the TCSA and TCSP estimations of the ethnographic arrowhead and dart tips and each of the MSA point types. The upper and lower quartiles of each ethnographic class are highlighted to illustrate any overlap between these and the MSA point types. The TCSA measure shows the dart tips’ upper and lower quartile area overlapping with those both types of points from Porc Epic and some of the subsamples of Aterian points. The TCSP shows overlap between the upper and lower quartiles of the dart tips and all MSA point samples except the three Klassies River Mouth samples and one of the Aterian subsamples.

Table 2 reports a series of independent sample t-tests comparing each of the artifact types to the ethnographic arrowhead and dart tip samples. All of the MSA types were larger in both TCSA and TCSP than ethnographic arrowheads at a high level of significance (p < .0001). As reported in Shea [23], all MSA TCSA values are also significantly greater than the ethnographic dart tips (p < .05). However, in several cases, the TCSP measures for the MSA points fell within the variation of the dart-tips. These include bifacially worked points from Porc Epic and two samples of Aterian points (from Aoulef and Azrag). Additionally, if the rhomboid estimation for TCSP is applied to the unifacial points from Porc Epic, these also fall within the variation of modern dart tips.

4. Discussion

As previously noted [40], TCSP appears to be an accurate proxy for projectile effectiveness. From this study of African Middle Stone Age points it is apparent that the TCSP follows a pattern similar to what previous studies found for the TCSA. However, where previous work identified trends in the TCSA, this study found several samples of MSA points that are statistically indistinguishable from ethnographic dart tips. While in some cases this could be the result of small sample size (e.g., the Aterian points from Aoulef), other types are more conclusively similar to dart tips (e.g., bifacial points from Porc Epic). A comparison of the box and whisker plots also qualitatively illustrates that there is more overlap in TCSP between the ethnographic and MSA archaeological samples. This indicates that at least some classes of Middle Stone Age points could have served as effective projectiles. This result also raises the possibility of ancient spearthrower use in Africa, expanding the geographic range of this weapon system significantly beyond its known ethnographic occurrence.

The types of MSA points that fall within ethnographic point variation also yield some interesting results. In particular, if typological classifications accurately reflect use, one would expect all Aterian points to yield a similar statistical pattern. Here, two subsets of the Aterian sample fell within the range of ethnographic dart-tips and two, and the lumped sample, were significantly larger. This would seem to indicate
Table 1: Summary statistics and information for the control and MSA samples.

<table>
<thead>
<tr>
<th>Industry</th>
<th>N</th>
<th>Width</th>
<th>Thickness</th>
<th>Cross-section estimate</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrowheads</td>
<td>118</td>
<td>15.09</td>
<td>3.93</td>
<td>1.29</td>
<td>Thomas 1978 [49]</td>
</tr>
<tr>
<td>Dart-tips</td>
<td>40</td>
<td>23.05</td>
<td>4.45</td>
<td>1</td>
<td>Thomas 1978 [49]; Shott 1997 [50]</td>
</tr>
<tr>
<td>KRM MSA I: Triangular flakes</td>
<td>71</td>
<td>33.61</td>
<td>6.04</td>
<td>2.17</td>
<td>S. Wurz</td>
</tr>
<tr>
<td>KRM MSA II Lower: Triangular flakes</td>
<td>528</td>
<td>35</td>
<td>7.85</td>
<td>3.87</td>
<td>S. Wurz</td>
</tr>
<tr>
<td>KRM MSA II Upper: Triangular flakes</td>
<td>298</td>
<td>31.77</td>
<td>7.17</td>
<td>2.95</td>
<td>S. Wurz</td>
</tr>
<tr>
<td>South African Stillbay points</td>
<td>203</td>
<td>27</td>
<td>7.8</td>
<td>3.83</td>
<td>T. Minichillo</td>
</tr>
<tr>
<td>Blombos Cave Stillbay points</td>
<td>239</td>
<td>26.56</td>
<td>9.77</td>
<td>4.46</td>
<td>M. Soressi</td>
</tr>
<tr>
<td>Porc Epic Bifacial Points</td>
<td>94</td>
<td>23.61</td>
<td>5.82</td>
<td>2.66</td>
<td>D. Pleurdeau</td>
</tr>
<tr>
<td>Porc Epic Unifacial Points</td>
<td>306</td>
<td>23.15</td>
<td>5.6</td>
<td>2.22</td>
<td>D. Pleurdeau</td>
</tr>
<tr>
<td>Aterian tanged points: Aoulef</td>
<td>4</td>
<td>26</td>
<td>3.56</td>
<td>1.32</td>
<td>T. Tillet; J. Shea</td>
</tr>
<tr>
<td>Aterian tanged points: Asriouel</td>
<td>46</td>
<td>26.26</td>
<td>5.33</td>
<td>1.86</td>
<td>T. Tillet; J. Shea</td>
</tr>
<tr>
<td>Aterian tanged points: Azrag</td>
<td>18</td>
<td>24.17</td>
<td>5.09</td>
<td>1.76</td>
<td>T. Tillet; J. Shea</td>
</tr>
<tr>
<td>Aterian tanged points: Izouzaden</td>
<td>12</td>
<td>31.92</td>
<td>5.84</td>
<td>1.62</td>
<td>T. Tillet; J. Shea</td>
</tr>
<tr>
<td>Aterian tanged points: Bir El Ater</td>
<td>41</td>
<td>29.63</td>
<td>8.19</td>
<td>1.7</td>
<td>Peabody Museum; J. Shea</td>
</tr>
<tr>
<td>All Aterian tanged points</td>
<td>124</td>
<td>27.65</td>
<td>6.78</td>
<td>2.01</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 2: African MSA samples compared to ethnographic arrowheads and dart tips by independent sample t-tests. * p < .05; ** p < .01; *** p < .001; n.s.: the two samples cannot be distinguished with 95% confidence; var: point estimated with the triangular measure, but using the rhomboid measure changes the result.

<table>
<thead>
<tr>
<th>Industry</th>
<th>N</th>
<th>Mean TCSA</th>
<th>Mean TCSP</th>
<th>Versus arrowheads</th>
<th>Versus dart tips</th>
</tr>
</thead>
<tbody>
<tr>
<td>KRM MSA I: triangular flakes</td>
<td>71</td>
<td>167.98</td>
<td>72.68</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>KRM MSA II lower: triangular flakes</td>
<td>528</td>
<td>216.77</td>
<td>77.66</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>KRM MSA II upper: triangular flakes</td>
<td>298</td>
<td>180.33</td>
<td>70.51</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>South African Stillbay points</td>
<td>203</td>
<td>144.64</td>
<td>57.73</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Blombos Cave Stillbay points</td>
<td>239</td>
<td>142.85</td>
<td>56.65</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Porc Epic bifacial points</td>
<td>94</td>
<td>103.29</td>
<td>50.25</td>
<td>***</td>
<td>n.s.</td>
</tr>
<tr>
<td>Porc Epic unifacial points</td>
<td>306</td>
<td>88.43</td>
<td>50.93</td>
<td>***</td>
<td>*** var</td>
</tr>
<tr>
<td>Aterian tanged points: Aoulef</td>
<td>4</td>
<td>81.13</td>
<td>54.76</td>
<td>***</td>
<td>* n.s.</td>
</tr>
<tr>
<td>Aterian tanged points: Asriouel</td>
<td>46</td>
<td>103.46</td>
<td>56.76</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Aterian tanged points: Azrag</td>
<td>18</td>
<td>85.28</td>
<td>52.02</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Aterian tanged points: Izouzaden</td>
<td>12</td>
<td>170.83</td>
<td>70.33</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Aterian tanged points: Bir El Ater</td>
<td>41</td>
<td>124.28</td>
<td>63.66</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>All Aterian tanged points</td>
<td>124</td>
<td>116.38</td>
<td>60.02</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

that Aterian points (which are typologically defined by a tang or shoulder) were not always used for the same purpose. These larger points may have been used differently, possibly as multipurpose tools (e.g., [51]) and/or as armatures for larger thrusting or hand-cast spears.

This is also true for the two samples of Stillbay points. Even with the largest examples removed, the TCSA of these two samples significantly exceeds that of the ethnographic complex projectile points [23]. When the full samples are analyzed with TCSP they remain significantly larger than ethnographic dart tips, but the box and whisker plot reveals large overlapping areas with the ethnographic dart tips. As has been noted previously, Stillbay points vary widely in size [52]. This may then be a case of one technological style being
used to create tools for a variety of uses. Thus, while there is no statistical support for the Klasies River Mouth or Blombos Stillbay points falling within the variation of ethnographic projectiles, the overlap of the smaller examples is intriguing and merits further study.

The triangular flakes from Klasies River Mouth are significantly larger than the comparative samples in both TCSA and TCSP. As noted by Shea [23] these triangular flakes are a broad typological category and likely represent a variety of different things. These triangular flakes are among the largest and most variable used in this or in Shea’s [23] study (Table 1). In this case, a lack of overlap could potentially represent a large, highly variable, type with a small sample of projectile points contained within. More likely, though, it simply reflects archaeologists’ longstanding habit of lumping together into the same tool type artifacts of widely-variable size.

The strongest case for plausible complex projectile use in the MSA comes from the two collections from Porc Epic. The overlap between the Porc Epic points, both unifacial and bifacial, and ethnographic dart tips shows that these points were created in response to particular morphological constraints. Currently, there is a better indication of plausible projectile use for the bifacially worked pieces, but sample size may play a role here (n = 94 for bifacial points versus n = 306 unifacial points). Both of these samples are larger than the ethnographic dart-tip sample. Thus it is possible that the ethnographic controls themselves do not fully capture the variability in complex projectile point morphology. The Porc Epic samples are conservatively dated to between 60 and 70 ka [53], only slightly earlier than the older estimated dates of permanent dispersal of *Homo sapiens* out of Africa.

These data provide compelling indications of complex projectile usage in the MSA, but it is important to reiterate that these relationships only show plausibility. The metric properties of stone points from Porc Epic, and the other samples discussed above, are not irrefutable evidence for the presence of complex projectiles in the MSA. Nevertheless, they do indicate that many of these tools could have been effective armatures for complex projectile weapons. Conclusive proof for the antiquity of complex projectile weapons in Africa will depend on confirmation or refutation from independent lines of evidence such as microwear, residue analysis, and zooarchaeological studies of MSA predation strategies.

### 5. Conclusions

Although it is a more complex measure relying on an estimation of the cross-sectional geometry, TCSP is a more accurate measure of plausible projectile usage than TCSA. By using TCSP, we were able to isolate significant relationships in the African MSA that were hidden in the TCSA data. In actuality, it is likely the interplay of TCSA and TCSP that predicts projectile effectiveness. A point that maximizes perimeter at the expense of area would be so thin and so fragile as to be useless. The inverse, a point that maximizes area at the expense of perimeter, may not function as an effective armature. Further models should analyze both these measures, as area remains an important predictor of durability [54], while perimeter controls actual penetration effectiveness.

Additionally, the methodology for TCSP used here only works for points. Unfortunately, some of the best candidates for early projectiles in Africa are backed pieces [55]. If these
objects were hafted as projectile armatures they would not resemble the ethnographic points used in this study. Alternative measurements and comparative samples are needed for backed pieces, and such investigations are currently underway by several research teams (e.g., [17, 46, 48, 56]).

The data presented in this paper demonstrate that plausible projectile armatures are present in several tool types of the Middle Stone Age and more are likely hidden in the variability encompassed by typological classifications. The current project demonstrates both the complexity of analyzing potential projectile use from a restricted sample and the possibility of refining these techniques. From these analyses, it seems apparent that the use of complex projectiles has its origins in the African MSA. Populations armed with complex projectiles are more ecologically versatile and can access a broader niche than those without them. Thus, the development of this technology was an important component of the adaptive change in Homo sapiens populations that allowed the successful colonization of the Old World at 50 ka. Complex projectiles are, however, not the only component of this behavioral transformation. It is only through continued collaborative analysis, experimentation, and interpretation that we can compile these data into a full picture of these technologies and how they contributed to the evolutionary success of Homo sapiens.

Acknowledgments

The authors are grateful to the researchers who kindly provided metric data to J. J. Shea (for this project including S. Wurz, T. Minichillo, M. Soressi, and T. Tillet). Ian Wallace and Jennifer Henecke provided invaluable comments on an earlier draft. Initial funding for the experimental collection was provided by a Wenner-Grenn Foundation for Anthropological Research grant to J. J. Shea.

References


Research Article

The Implications of the Working Memory Model for the Evolution of Modern Cognition

Thomas Wynn and Frederick L. Coolidge

Departments of Anthropology and Psychology, University of Colorado at Colorado Springs, Colorado Springs, CO 80918, USA

Correspondence should be addressed to Frederick L. Coolidge, fcoolidge@uccs.edu

Received 17 September 2010; Accepted 18 January 2011

Academic Editor: Parth Chauhan

Copyright © 2011 T. Wynn and F. L. Coolidge. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

What distinguishes the cognition of biologically modern humans from that of more archaic populations such as Neandertals? The norm in paleoanthropology has been to emphasize the role of language and symbolism. But the modern mind is more than just an archaic mind enhanced by symbol use. It also possesses an important problem solving and planning component. In cognitive neuroscience these advanced planning abilities have been extensively investigated through a formal model known as working memory. The working memory model is now well-enough established to provide a powerful lens through which paleoanthropologists can view the fossil and archaeological records. The challenge is methodological. The following essay reviews the controversial hypothesis that a recent enhancement of working memory capacity was the final piece in the evolution of modern cognition.

1. Introduction

Ever since the publication of *The Human Revolution* twenty years ago [1], many of the most exciting and contentious debates in paleoanthropology have revolved around the evolution of modern humans. The research and discussion have focused on two intertwined goals: (1) understanding the evolutionary transition from archaic to modern anatomy and behavior, and (2) tracing the specific evolutionary scenario by which the transition took place. In pursuing the first goal, paleoanthropologists strive to identify the derived characteristics that distinguish modern anatomy and behavior from those of more archaic humans, and propose selective hypotheses to account for them. This goal is in keeping with overall desire of paleoanthropologists to understand the entire sweep of human evolution. The second goal is the more difficult, and certainly the more contentious. Many paleoanthropologists would like to trace the actual sequence of events that led to modernity: who made the transition? Where did it happen? When did it happen? And, perhaps most contentious of all, what happened to archaic populations such as Neandertals? Archaeologists have been intimately involved in the pursuit of both goals. They have tried to identify derived features of modern behavior, as preserved in the archaeological record, and have proposed hypotheses to account for the emergence of modern behavior [2–5]. But they have also attempted to trace the specific evolutionary scenario through use of the same archaeological remains [6–9]. As with the fossil remains, the second goal has focused primarily on the fate of a single population from southern or eastern Africa, and the consequences that ensued for archaic groups, especially Neandertals.

Our interest is primarily in goal #1. In particular, we are interested in documenting the emergence of modern cognition, relying as much as possible on macroevolutionary evidence supplied by archaeology. Until recently, archaeological treatments of the cognitive components of modernity have been dominated by variously developed assertions concerning language and symbolic culture. This focus on symbolism resulted from two quirks in the history of palaeolithic archaeology. The first is an accident of archaeological discovery. The first palaeolithic culture attributed to early modern humans was the European Upper Palaeolithic (ca. 40,000–14,000 years ago), and the most
spectacular component of it was its art—cave paintings, figurines, and personal ornaments. The Neandertals who preceded the Upper Palaeolithic in Europe had little or none of this, and the contrast was stark. The presence of art implied the presence of meaning, and modern symbolic sensibilities, or so the reasoning went. Thus, in Europe modern anatomy (Cro-Magnon) and modern symbolism came to be linked, a scenario that was then applied with little criticism to other continents. The second quirk is linked to the history of archaeological method and theory. In the 1970s, palaeolithic archaeologists began to take social science seriously as a source of interpretive models. At that time social science had a strong structuralist orientation, and Chomsky's hypotheses about language were paramount. As a result, many archaeologists came to consider language to be the \textit{sine qua non} of humanness. Indeed, this opinion was so generally held that few ventured to question its appropriateness when the “modern question” came to dominate the attention of paleoanthropology in the 1990s. Several influential summaries attributed modern behavior to the emergence of modern language, on the assumption that enhanced communication ability would have had obvious evolutionary advantages \cite{1, 5}. A few archaeologists took a more critical stance. Davidson and Noble, for example, carefully parsed language into its grammatical and symbolic components, and, based on the work of Wittgenstein and James Gibson, argued that true symbolism was a late emerging faculty \cite{10, 11}.

Symbolism and language are certainly components of modern thinking, and documenting their evolution is important. However, alone they are insufficient to account for all of the features of the modern mind. The modern mind is not, we contend, simply an archaic mind augmented by symbolism and language. There are arguably several other components, including problem solving and long range planning abilities. A few archaeologists have appreciated this—most notably Mithen \cite{12}, Ambrose \cite{13}, and Davidson and Noble \cite{14}. The latter, for example, argued that the colonization of Australia required planning abilities only possible via modern language—so we are not alone in this contention. Our approach differs from theirs in its theoretical and methodological bases. For the last decade, we have taken a specific cognitive model, that of executive functions and working memory, and used it to interpret archaeological remains. The result has been a different and controversial picture of the emergence of one of the components of modern thinking \cite{15–18}.

2. Executive Functions and Working Memory

As the label implies, executive functions encompass the brain’s ability to plan and strategize. The term and concept were first developed by neuropsychologists working with brain damaged adults, initially Russo-Japanese War soldiers with head wounds. Neuropsychologists such as Luria \cite{19} described patients with damage to the frontal lobes who retained full language faculties but who were unable to carry out complex, purposive, goal-directed actions. They were unable to evaluate the success or failure of their own actions, and unable to alter their future behavior. Neural activity centered in the frontal lobes was clearly very important for higher level thinking. Executive functions of the frontal lobes were:

“...at the heart of all socially useful, personally enhancing, constructive, and creative abilities... Impairment or loss of these functions compromises a person's capacity to maintain an independent, constructively self-serving, and socially productive life no matter how well he can see, hear, walk and talk, and perform tests” \cite{20}, p. 281.

By the 1960s executive functions had become a well-established cognitive array in the field of neuropsychology, and neuropsychologists had devised a number of assessment tools that could be used to evaluate brain damaged individuals (e.g., the Tower of London test). Adults with executive function (EF) deficits have a difficult time functioning on their own and often have poor social interaction skills (e.g., interacting with sales people to purchase an item). The significant role of EFs in social life led Barkley \cite{21} to suggest that the evolution of EFs occurred via selection for effective social cognition.

Coolidge and Wynn \cite{15} initially proposed that EFs, as defined primarily by neuropsychologists, were a better candidate than language for the neural development that led to modern thinking. It soon became clear to them, however, that a well-developed model in cognitive psychology encompassed many of the same cognitive phenomena as EFs, but had the advantage of being more explicit in terms of specific abilities, and also being based on experimental research with normal, that is, not brain-damaged, individuals. This model is working memory.

3. The Working Memory Model

The concept of working memory (WM) was first developed by Baddeley and Hitch in 1974 \cite{22} as an elaboration of the older concept of short-term memory. At its most basic, WM is the mind’s ability to hold in attention, and process, task-relevant information in the face of interference. As an example we ask the reader to perform the following task: as you read this paragraph, remember the final word in each of the sentences and after completing the paragraph, recite these terminal words in order, from memory. This is a test of working memory capacity. The more words you can remember, the greater your WM capacity. Note that there are two components to this task. First, you must remember a sequence of words. This is a classic short-term memory test, and in itself is not too taxing. But in this context you must do it while reading, which interferes with remembering the words. The WM model has been perhaps the single most researched and successful cognitive model of the last forty years. It has integrated and synthesized research results from several allied fields such as psychology, neurology, and neuropsychology. Even more important, psychometric measures of WM capacity correlate with a wide variety of critical cognitive abilities, including reading comprehension,
vocabulary learning, language comprehension, language acquisition, second language learning, spelling, story telling, fluid intelligence, and general intelligence (e.g., [16]). The correlation with fluid intelligence is especially important because fluid intelligence is one’s ability to solve novel problems. It is less influenced by learning and culture than general intelligence (IQ) and tied directly to one’s problem solving ability. Interestingly, WM has even more recently been referred to as the new intelligence [23]. Thus, the WM model is a natural heuristic for enquiring into the evolution of modern thinking.

As currently understood, WM is not a single, simple, neural system but a set of interlinked abilities. The current WM model, as set out by Baddeley [24], consists of an attentional pan modal processor (the “central executive”), two subsystems (the “phonological loop” and the “visuospatial sketchpad”), and a temporary memory store (the “episodic buffer”). The phonological loop is dedicated to auditory phenomena, and maintains and rehearses auditory information either vocally or subvocally. The visuospatial sketchpad is a distinct subsystem that processes visual information (shapes and locations). The two subsystems can operate simultaneously, so that, for example, one can perform a visuospatial task and a speech task with minimal interference. The episodic buffer holds information provided by the subsystems in active attention where it can be processed by the resources of the central executive. The central executive of WM performs most of the processing including attention, active inhibition (or, e.g., suppress distracting stimuli or prepotent responses), decision making, planning, sequencing, temporal tagging, and the updating of information in the two subsystems. It also serves as the chief liaison to long-term memory and language comprehension. The central executive takes control when novel tasks are encountered, and one of its most important functions is to override pre-existing habits and inhibit prepotent responses.

Neurologically, WM is primarily a network of the prefrontal cortex (PFC) but also relies on extensive linkages to parietal and temporal lobes and also connections to subcortical regions. The dorsolateral PFC has long been associated with executive functions, but there appears to be no single neural structure that can be isolated. The central executive, for example, appears to emerge from the interplay of diverse cortical and subcortical systems (e.g., [25]). The phonological loop may be the most isolable neural network of the system. Aboitiz et al. [26] have argued that the phonological loop is a specialized auditory-vocal sensorimotor circuit connecting posterior temporal areas with the inferior parietal lobe and the ventrolateral prefrontal cortex. What is clear, however, is that WM is a complex neural network consisting of neural pathways that interlink much of the neocortex. As such, adult phenotypes are likely to be the result of structural and regulatory genes governing neural development, and also individual developmental context.

Working memory is a trait that varies in modern populations, and the variability correlates with performance on several measures of intelligence, including language comprehension and planning. Much of this variability appears to be under strong genetic control. Coolidge et al. [27], in an analysis of child and adolescent twins as rated by parents, found that a core of executive functions including planning, organizing, and goal attainment, was highly heritable (77%) and most likely due to an additive (polygenic) genetic influence. In a study specifically focused on general WM functions, Ando et al. [28] found a strong additive genetic influence (43–49%). And on phonological storage capacity, Rijsdijk et al. [29] found a 61% additive heritability. Friedman et al. [30] demonstrated that executive functions are correlated because they are controlled by a highly heritable (99%) common factor that could not be explained by simple intelligence or perceptual speed, and yet they can be separated because of other genetic influences that may be unique to particular executive functions. They concluded that the combination of general and specific genetic influences makes the executive functions among the most heritable psychological traits.

Today working memory is arguably the most researched and well-understood model in cognitive neuroscience. As of 2007, more than 15,000 articles had been published containing the term “working memory” [31]. This much research is bound to engender disagreements and controversy but is also likely to make significant progress. The general features of WM are now well delimited and understood; it is the details that drive most research. As such it is a powerful model for understanding the evolution of the modern mind.

4. Methodological Considerations

The challenge in cognitive archaeology is methodological. The archaeological record itself is impoverished compared to the experimental and ethological data sets of most cognitive science. However, archaeological data are the residue of activities that occurred in the past, and as such are the only direct evidence of past behavior and past minds. Tapping this data reservoir requires carefully constructed arguments that must be both cognitively valid and archeologically credible. In practice, a valid cognitive archaeological argument must have three components.

(1) The cognitive ability under investigation must be well defined by cognitive science. Common sense categories such as “abstract” or “complex” are just too vague to allow selection of valid attributes that could be applied to archaeological remains. Unfortunately, it is just such common-sense categories that underpin most archaeological arguments for modern cognition. It is noteworthy how rarely archaeologists have taken the trouble to inform themselves about cognitive science. Exceptions to this naivety include the work of Mithen [12], who based his analysis on concepts in developmental psychology, Davidson and Noble [10, 14], who drew on Wittgenstein and the ecological psychology of James Gibson, Ambrose [32], who has used the McDaniel and Einstein’s concept of prospective memory, and [33], who have relied on neurosemiotic theory [33]. Archaeologists must inform themselves about cognitive science if they wish to make substantive contributions to the study of the evolution of the human mind.
The cognitive science literature is immense and diverse, and much like evolutionary science, there are many factions and schools of thought. One cannot simply dip into it and pull out a useable model. One must understand the intellectual context in which it developed and in which it is used. The payoff is well worth this effort—experimentally or ethologically justified descriptions of cognitive abilities.

(2) The archaeologist must identify activities that would require the cognitive ability under investigation. This is the key methodological step. Unfortunately, the cognitive literature itself rarely addresses the kinds of activities that archaeologists can document. Such activities tend to be messy (metaphorically and actually) and difficult to operationalize in the laboratory. Some cognitive scientists do try to incorporate real world activities in their discussion, but for the most part archaeologists must themselves identify the appropriate activities, based either on their own experimental protocols (see, e.g., [34] or [35]), or on their understanding of the cognitive ability in question (see [36] for an example using spatial cognition). It is here that the value of explicit cognitive models becomes apparent. Because the WM model identifies response inhibition as an important component of the central executive, we can ask the tractable question “What activities require response inhibition?”, and generate a list of activities that would be visible archaeologically. A strict standard of parsimony must apply; the activities must require the cognitive ability. If an activity (say, specialized mammal hunting) could be performed using a less powerful form of cognition (e.g., procedural memory), then the less powerful form must be given precedence.

(3) The archaeologist must define attributes of the activities that would preserve in archaeological record and which can reliably stand for the activity. This is the essential archaeological piece to the argument, and is a step required in any archaeologically based reasoning. One of the major challenges in this step is equifinality. Often many activities can produce identical or very similar archaeological residues (e.g., hunting versus high-end scavenging leave similar butchery traces). Again a strict standard of parsimony must apply; one must be confident of the link between archaeological traces and the reconstructed activity. For example, some (e.g., [37]) argue that evidence for Neandertal burial in Middle Eastern sites is evidence for modern symbolic ability. However, the evidence is more parsimoniously explained as minimal corpse treatment by Neandertals with strong emotional attachment, and grief at the loss of the deceased.

Archaeological credibility is no different for cognitive archaeology than it is for any other archaeological interpretation. The evidence must have been acquired by sound field and analytical techniques, and it must be reliably situated in time and space. These requirements are easily stated but not easily met. Indeed, one could argue that the preponderance of time, energy, and resources in any archaeological research is devoted to these practical issues. But this does not in turn mean that archeological credibility is more important than cognitive validity in the structure of a cognitive interpretation. Both are equally necessary.

5. Archaeological Evidence for Modern Working Memory Capacity

We have already set out our case for the first component of our archeological argument for the evolution of modern cognition. Working memory is a well-defined, voluminously documented component of the modern mind. Moreover, it is an ability that varies in modern people, and it is an ability possessed by nonhuman primates at a comparatively reduced capacity. Modern WM capacity must have evolved over the course of human evolution. Hints at increasing capacity (beyond an ape range) can be identified as far back as Homo erectus [17]. But when did it achieve modern levels, something we have labeled “enhanced working memory” (EWM)?

The second step in the analysis is to identify activities that require not just WM, but EWM. This presents two related practical problems. First, WM capacity is typically measured in terms of numbers of discrete items (e.g., terminal words in a series of read sentences). We cannot apply such tests in prehistory, and thus a simple quantitative measure is unavailable. We must rely on behavioral correlates of WM capacity. Second, psychological tests of WM capacity rarely include activities that would leave an archaeological signature; it is necessary for the archaeologist to select the appropriate activities. In practice this requires that we, as researchers, make ordinal comparisons of everyday tasks. Because these judgments are ordinal (e.g., more versus less), they are by nature not fine-grained. For example, we will argue that planning months and years in advance is a feature of modern executive thinking enabled by modern WM. It is fairly easy to cite modern examples. But what would constitute archaic WM? We can argue that prehistoric groups who did not demonstrate appropriate activities did not have modern WM, but it is effectively impossible to assign a number. Moreover, there is always the danger of under assessing WM if we rely on only a few kinds of activities. We must therefore use a variety of different activities if we wish to have reasonable confidence in our assessment. Below we will identify technological activities, subsistence activities, and information processing activities that we suggest are reliable indicators of EWM.

The final step in the analysis is to scour the archaeological record for the earliest credible evidence for the activity in question. There are several inherent pitfalls in this step. One, the problem of equifinality, we have already touched upon. A second is simple serendipity. Much of the evidence we seek requires good preservation—a rarity in palaeolithic sites—but we also need the good fortune to find such sites. Archaeologists have little control over these factors. However, we need not kowtow to the dictum “absence of evidence
is not evidence of absence.” Often absence of evidence is evidence of absence. Nevertheless, it is always dangerous to conclude that evidence from one site will always be the oldest, or to adhere to too strict a chronology. As a corollary, it is important to use as many different kinds of archaeological evidence as possible. If the archaeological evidence for many different activities all point to the same chronological conclusion, then confidence in the conclusion improves. Finally, some archaeological evidence is direct—archaeologists find physical remains of the activity. But some is indirect; the archaeological remains strongly imply the presence of the activity. For example, archaeologists have occasionally found actual traps made of wood and fiber. As you might suppose, these are very rare because the constituent materials rarely survive the ravages of time. The oldest such examples are only about 8,000 years old. So, was the first use of traps only 8,000 years old? Archaeologists think not, but the evidence is indirect, primarily in the form of animals that could not be effectively killed or captured without the use of traps. This indirect evidence pushes traps back to perhaps 70,000 years ago [38], a considerable difference.

6. Technical Evidence for Enhanced Working Memory

The irony for archaeologists is that technology is the most visible activity in the archaeological record, but one of the least likely to require the resources of EWM. Most tool making and tool use relies, often exclusively, on a style of thinking known as expertise or expert performance [39–41]. This kind of thinking relies on procedural cognition and long-term memory—motor action patterns learned over years of practice and/or apprenticeship. It is also largely nonverbal. Very little of the problem solving ability of EWM is ever devoted to tool use. Instead, flexibility in tool use comes from the large range of procedures and solutions learned over years. The millions of stone tools produced over human evolution tell us mostly about this other cognitive system, not WM. It is not that WM was never used, just that it is almost impossible to eliminate procedural cognition as a candidate for the cognition behind the tool use or in question (e.g., equifinality and parsimony). Nevertheless, there are technical systems that do require EWM, and which cannot be reduced to procedural cognition. Most of the good examples (e.g., alloyed metals and kiln fired ceramics) appeared so late in human evolution as to engender little controversy or interest (ca. 6,000 years ago). There are just a few that extend much further back.

6.1. Traps and Snares. “Facility” is a term for relatively permanent immobile constructions built onto or into the landscape [42]. Perhaps the most common facilities used by hunters and gatherers are traps and snares, which are facilities designed to capture or kill animals (including fish). Facilities, including traps and snares, are often multicomponent gadgets, occasionally very heavy, that are time-consuming to build, and which operate remotely, occasionally in the absence of direct human engagement. It is the remote action that implicates EFs and EWM. To make a trap one must project present action toward a future, uncertain result. This requires the long range planning in space and time of modern EFs, and relies significantly on the response inhibition of the central executive of WM (delayed gratification).

Direct archaeological evidence for traps and snares, as mentioned above, have a relatively shallow antiquity. Actual wooden fish traps date back 4,500 years in North America, and a few thousand years earlier in Europe, that is, not much earlier than the alloyed metals passed over above. The oldest direct evidence of a kind of trap appears to be the “desert kites” of the Middle East [43]. These are lines of piled stone cairns, often hundreds of meters long, converging on a stone enclosure. There were used to hunt gazelle, and the oldest are about 12,000 years old.

Indirect evidence pushes traps and snares back to about 35,000 and perhaps even 75,000 years ago. At Niah cave on Borneo, Barker et al. [44] have evidence of extensive remains of bush pigs by about 35,000 BP, an animal best hunted using nets or snares. Similarly, Wadley [38] has recently argued that extensive blue duiker remains at Sibudu are indirect evidence for using traps by 70,000-year-old Middle Stone Age people in South Africa. In sum, traps and snares supply direct evidence for modern WM back to 12,000 BP, and indirect evidence back to 75,000 BP.

6.2. Reliable Weapons. Twenty-five years ago, Peter Bleed introduced a distinction in technical systems that has important cognitive implications, that between “maintainable” and “reliable” weapons [45]. The former require comparatively less effort to produce but are easier to fix ("maintain") when necessary, for example, when damaged through use. Most stone tools, even from recent time periods, qualify as maintainable. Reliable weapons, on the other hand, are designed to assure function, that is, to reduce as far as possible the chances for failure. As such they tend to be overdesigned, complex in the sense of having several interrelated parts, hard to maintain, and often heavy. They often require long periods of “down time” for their construction and maintenance, and are most often intended to be deployed over short time spans of heavy use. Bleed developed this distinction as a way to understand the difference between simple stone tipped thrusting spears and the sophisticated projectile systems of North American Paleoindians, which included spear throwers, flexible aerodynamic shafts, replaceable foreshafts, and thin, fluted stone points. However, the distinction between maintainable and reliable applies generally to all technologies, not just weapons. The guiding principle behind reliable systems is that the investment of time and labor well in advance of need will maximize future success. More recently, Shea and Sisk [46] have taken a related but narrower focus and argued that the use of complex projectile weaponry (spear throwers and bows and arrows) is a good marker of modern technical prowess. “We use the term ‘complex projectile technology’ to refer to weapons systems that use energy stored exosomatically to propel relatively low mass projectiles at delivery speeds that are high enough to allow their user to inflict a lethal puncture wound on a target from a ‘safe’ distance” (p. 102). They consider this development significant enough to qualify as a
derived feature of modern behavior. Reliable weapons, and in particular complex projectile weapons, rely on the executive function ability to plan over long stretches of time, and especially the response inhibition of WM (i.e., do not hunt today, even if you are hungry, but instead invest your effort in producing tools more likely to succeed tomorrow), and contingency planning (if the foreshaft breaks, slip in a new one; it is quicker than making an entire spear).

The archaeological record in North America clearly places reliable weapons back to Paleoindian times, at least 11,500 BP (roughly the same age as the earliest desert kites in the Near East, which were also reliable technical systems). Earlier examples rest on our ability to judge time investment and effectiveness of technical systems. Following Pike-Tay and Bricker [47], we believe that one earlier type of Palaeolithic artifact qualifies as being a component of a reliable system, and certainly an element of complex projectile technology—the bone and antler projectile points (a.k.a. sagaies) of the European Upper Palaeolithic. To make these artifacts, artisans used stone tools to remove appropriately sized blanks from a piece of bone or antler, often after soaking the raw material, and then carved the blanks into specific shapes (split based, barbed, etc.). Most were spear points hafted directly onto shafts, but others were harpoon heads, designed to come off the shaft while attached to a line. There are many examples of reworked points, attesting to the time required to make one from scratch. The most spectacular examples of such projectile points, which include the harpoons, date from the Late Upper Palaeolithic, about 14,000–18,000 BP, with slightly simpler systems extending back to 30,000 years ago. In Africa, bone points date back even earlier, perhaps as early as 90,000 years ago in the Congolese site of Katanda [46, 48]. The European evidence is more compelling because of the contemporary evidence for managed foraging (see below), and evidence for spear throwers and harpoons, which imply systems of gear. As yet the early African evidence consists of just the bone points, but it is provocative nonetheless.

6.3. Hafting. Hafting—attaching a stone tool to a shaft—has itself often been touted as a technological and even cognitive watershed in human evolution [13, 32]. Hafted tools represent the first time Palaeolithic people united separate elements into a single tool. These compound tools consist of three distinct elements: the stone tool (usually a spear point), the shaft, and the haft itself. It was the haft that was the challenge because it had to withstand significant impact forces when the tool was used. Spears withhafted stone points represent a clear escalation in the human-prey arms race, and it is fair to emphasize their importance in technological history. But their cognitive significance is harder to assess. Much hinges on how the hafting was done. A simple haft using a naturally available glue has different implications than a haft requiring days of soaking animal tendons followed by controlled, heated drying of the lashings on the shaft. The former is straightforward, single-sitting task, while the latter is a multiday procedure. In a sense, the former leans towards maintainable, the latter toward reliable in the maintainable-reliable continuum. It is only the latter that carries clear implications for EWM capacity. Hafting also calls out for a discussion of invention, the conscious design of an innovative technology. Someone had to design the first haft; it could not have occurred by accident. And it would be very informative, from a cognitive perspective, to know just how that person came up with the idea. The frustrating answer is that we just do not know. We can speculate, but our speculations cannot then be used as data for a cognitive interpretation.

The earliest evidence for hafting extends back probably 200,000 years in Europe, the Middle East, and Africa [49–52], and includes examples by Neandertals and modern humans. Thus far, at least, these early hafts seem to be of the simpler, single-sitting task variety, though certainly collection of natural adhesives adds a component of complexity to the task, and Grünberg [53] and Koller et al. [54] have argued that the production of birch pitch required sophisticated knowledge of heating temperatures. It was only after 100,000 years ago that there is evidence for multiday hafting procedures. The best evidence comes from Sibudu in S. Africa (the same site as the indirect evidence hunting blue duikers with snares) at about 70,000 BP. Here hunters used a mixture of acacia gum, a little beeswax, and powdered ochre to produce an adhesive that had to be carefully dried using fire [34]. Although in theory such hafting could be accomplished by procedural cognition, the variety of constituents required for the adhesives, and the multiday procedure itself, imply the use of modern WM, particularly response inhibition and contingency planning.

To summarize, three lines of technical evidence are in broad agreement. Convincing archaeological evidence extends easily back to 18,000 years BP or so, but there are strong examples going back as far as 70,000 years BP in Africa. Earlier than that is only the single example of simple hafts, which cannot alone bear the weight of assigning modern WM. It is important to reiterate that technology is not a domain of activity that easily documents WM capacity. Procedural cognition can be effective and flexible, and can encompass almost all technical activity. Certainly hafting, or even complex projectile technology, could not alone stand as evidence for modern executive thinking (nor, we should emphasize, do [46] make such a claim). Of the examples we cite, the only one that might stand alone as an argument for modern cognition is the example of traps. Technical evidence works better when it supports or corroborates evidence from other domains.

7. Foraging Systems

Next to technology the domain of activity most visible in the archeological record is subsistence—acquiring and processing food. And like technology, archaeologists’ arguments for modern subsistence systems have been heavily distorted by the record of the European Upper Palaeolithic, especially its later phases, which included examples of specialized hunting of single species such as reindeer or mammoth. These were no doubt impressive subsistence systems but specialization per se does not actually require the planning resources of modern EFs and WM. It can easily be organized and executed
by expert procedural cognition. In fact this is arguably a more appropriate cognitive strategy because it consists of well-learned, automatic responses that can be selected and deployed quickly in dangerous situations. Neandertals were very good at this kind of thinking and, no surprise, we have extensive evidence for specialized hunting [55, 56]. Thus, it is necessary to eschew this war-horse of modernity and identify subsistence activities that actually do require modern EFs and WM.

Modern people manage their food supply. This is obvious in agricultural economies, where activities must be planned on a yearly scale (for nontropical systems). It clearly relies on the long-range planning of EFs and, more specifically, the response inhibition that is a key component of modern WM (e.g., retaining a portion of the harvest for replanting even in the presence of extreme want). But agriculture is not the only form of managed foraging. Most of the hunting and gathering systems archaeologists have recognized as “complex” also qualify [57]. Good recent ethnographic examples include foragers of the Northwest Coast of America, the Arctic, and Australia. In Northern and Western Australia, hunter-gatherers systematically burn tracts of land in order to encourage a second green-up of grass, which attracts herbivores. They rotate the tract to be burned every year, and do not return to a tract for at least a decade [58]. This is a managed system, with planning over long periods, and response inhibition. Another component of modern hunting and gathering systems is a marked division of labor by age and sex [59]. It requires coordination of separate labor pools, which weakly implicates WM and its executive functions (organization, delegation, disputation, etc.), but more importantly is manifested in the tropics by increased reliance on small, seasonal resources (plants and small animals) that require scheduled harvesting, typically by women and children.

Archaeological evidence for agriculture extends back to 10,000 years BP on several continents, and evidence for managed forms of hunting and gathering back another several thousand years in the guise of Archaic, Mesolithic, and Epipaleolithic cultures all over the world. An especially good example is that of the Epipaleolithic site of Abu Hureyra in Syria [43]. Here a group of hunters and gatherers established a sedentary community based on hunting gazelle and gathering a wide variety of local plants. When the local conditions became much drier 11,000–10,000 years ago these people did not simply shift the focus of their hunting and gathering; they changed its very basis by beginning to cultivate rye. The interesting point is not so much the broad spectrum hunting and gathering but the inventive response to changing conditions. These people were clearly using the planning abilities enabled by EWM.

Finding evidence for managed foraging that is earlier than the end of the Pleistocene is fraught with problems, mostly linked to preservation, but also to mobility patterns of earlier hunter-gatherers who rarely settled in permanent sites like Abu Hureyra. The amount of refuse is much less, and harder to characterize. Nevertheless, there are several provocative earlier examples. A well-known example is that of late Upper Palaeolithic reindeer hunters [60] of southwestern Europe (ca. 18,000 years BP). Here it is not the specialization that is telling (see above), but the evidence for a tightly scheduled hunting system in which herds were intercepted and slaughtered at specific locations during migrations, but at other times of the year were hunted individually using a different set of tactics. Though other resources were used, reindeer were the clear focus year-round, using a seasonally adjusted strategy that included periods of down-time during which the hunters made and maintained their complex technical gear (see above). At about the same time, hunters on the Russian Plain used a system in which they killed large numbers of animals during late summer/early fall and then cached large quantities of meat in underground storage pits for freezing and future consumption [61]. Storage and delayed consumption are strong evidence for modern WM.

Earlier evidence is largely indirect. At Niah Cave on the island of Borneo [44], archaeologists have recovered large quantities of pollen from plants that flourish on recently burned areas. The local tropical conditions are quite wet, and the pollen far exceeds what one would normally expect to find, suggesting extensive human-induced burning. This evidence dates to sometime between 42,000 and 28,000 years BP. Earlier still is the evidence for hunting blue duiker in South Africa using snares or traps (70,000 years BP, see above). Of similar antiquity is evidence from other South African sites for extensive use of corms (fleshy, semisubterranean stems), which are features of plants that flourish on burned landscape, suggesting as at Niah human use of fire as an ecology altering tool [62]. Kuhn and Stiner [59] argue that this broadening of the subsistence base in South Africa is an indication of division of labor by age and sex. In sum, the archaeological evidence for managed foraging parallels the evidence of technology. There is a strong signature going back 18,000 years or so, and a weaker, but still provocative, set of isolated examples going back 70,000 years.

8. Information Processing

Thus far in our discussion we have focused primarily on the long-range planning and response inhibition components of modern EFs and EWM and have traced them archaeologically through the technological and subsistence records. We now shift focus to problem solving, another of the executive functions enhanced through an increase in WM capacity. Working memory is the active problem solving “space” of the modern mind. We use WM to construct analogies, perform thought experiments, make contingency plans, and even make metaphors. It is how and where we bring things together in thought; however, even modern WM has a limited capacity, because the episodic buffer is a limited capacity store. If the capacity of this store is depleted by holding raw information, little comparison and processing can also occur (try multiplying two four-digit numbers in your head). One solution to the problem that modern humans regularly use is externalization of some of this information, that is, holding the information outside of the mind itself. This is an aspect of extended cognition, which has recently received significant attention in cognitive science [63], and even archaeology [64, 65]. Our interest is on the implications that extended
cognition holds for WM, and the primary effect is to extend WM capacity by relieving the necessity to hold information in the episodic buffer, thereby freeing capacity for the processing components of the central executive. Examples of such externalized storage systems abound in the modern world—writing, numbers, calculators, and so on. External systems need not be artifactual—one can, for example, count on one's fingers—but they often are artifactual, which gives us an avenue to follow into the prehistoric past.

It is uncontroversial to assert that early writing and accounting systems, which date back at least 5,000 years, were external information storage. Systems of clay tokens, used for accounting purposes, extend the record back several thousand years into the early Neolithic [65, 66]. We pick up the trail about 12,000 years ago at the site of Grotte du Tai in western France [67].

The plaque in Figure 1 above appears to have been a record keeping device. Someone engraved a series of long lines crossed by groups of slashes on a piece of flat bone. Marshack, and later d’Errico [68], examined the markings microscopically and determined that they were produced by different tools, probably in different episodes. Marshack famously argued that it was a lunar calendar, but d’Errico concluded simply that it was an external memory device. We do not know what the engraver was tracking, but it was clearly something, attesting to a desire to externalize information, thereby freeing up WM capacity for processing. Similar objects, not quite as elaborate, date back in Europe to about 28,000 years BP [2].

Earlier still, and equally provocative are therianthropic figurines from Germany, the most famous of which is the Hohlenstein-Stadel figurine (see Figure 2). This is an image of lion-headed person (or human-bodied lion) carved in elephant ivory, roughly 28 cm high. It is about 32,000 years old [69]. It is certainly an evocative piece, and has inspired much discussion about symbolism and Upper Palaeolithic religious thinking. It also has a number of important implications for cognition [70], one of which concerns WM. The figurine is an externalized abstraction. Such a creature does not exist in the real world, and it must have been metaphorically glued together, initially at least, in the WM of some Upper Palaeolithic person. The problems people need to solve are not always practical issues in day-to-day life. They are also social, and even metaphysical. The Hohlenstein-Stadel figurine is the externalization of such a metaphysical problem, and its externalized presence frees up WM of the artisan, and also observers, to ponder other related existential issues.

So far, our discussion of externalized information processing has not yielded any surprises. Suggesting that Upper Palaeolithic people in Europe 32,000 years ago exercised modern cognition is neither a novel nor a controversial conclusion. But what about earlier? Can we push externalized information back as far as the early evidence of traps, or managed foraging? The answer is yes, but it requires a slightly different take on a famous set of artifacts—the Blombos beads.

Blombos Cave is a site on the coast of South Africa whose Middle Stone Age levels date back at least 77,000 years. These MSA levels have famously yielded engraved bones, shaped and engraved pieces of ochre, bone awls, and marine shell beads [3, 33, 71]. These are among the earliest putatively modern artifacts yet found, and make a strong case for extending many of the components of modern behavior and cognition back to this early period. But the initial enthusiasm has recently been tempered by more sober critiques. d’Errico and Henshilwood [71], for example, argued that the presence of decorative beads indicates that the inhabitants had fully syntactical language. This conclusion was then elegantly challenged by Botha [72], who pointed out that d’Errico and Henshilwood had not made explicit and convincing bridging arguments linking beads to language. Henshilwood and Dubreuil [33] have replied, providing part of the linkage (a very nice example of a productive scholarly exchange), but the implications of the beads are still not entirely clear. We suggest that an alternative approach is to look at these artifacts not in their possible symbolic role, but as externalized information storage. Henshilwood and Botha agree that the shells with punched holes were beads, and that the beads were worn as ornaments. But why does one wear beads? One answer is that one wears beads to send information about oneself to another person. This could be an explicit message about social status (“I am an adult”, “I am wealthy”, and so on), or an implicit message (“I am a good mate prospect”), but by changing how others view the wearer, the wearer is
externalizing information about him or herself. Curiously, there is an alternative function for these beads that neither Henshilwood nor Botha have considered. They might have been tally devices, used to keep track of (remember) some sequential phenomena (much like rosary beads). The social implications of this option are different from the decorative bead interpretation, but the information implications are similar: beads were an externalized store of information, freeing WM to devote space in the episodic buffer for processing information, rather than just holding it in attention.

This evidence of externalized information storage is provocative. We live in a modern world where externalized information has come to dominate, perhaps even overwhelm, our daily lives, and the thought that it had its roots far back in the stone age is certainly provocative. However, for our topic at hand—working memory—external storage of information is actually an ambiguous signature. Externalization of information would release storage space for episodic buffers of whatever capacity, not just modern capacity. There need only be enough capacity to hold the external device as a token of some kind in the episodic buffer as one performs processes upon it. Because of this ambiguity, we suggest that evidence for the use of external storage devices cannot, on its own, provide compelling evidence for modern WM. It can, however, stand as corroborating evidence for assessments established by other means. As such, it supports the picture painted using technology and subsistence, that is, strong evidence in this case going back to perhaps 30,000 years, and weaker evidence extending further to 77,000 BP.

9. Conclusion and Discussion

At the outset of this essay we noted that archaeologists who study the problem of modern human origins typically address two rather different subgoals—the emergence of modern culture/cognition, and/or the specific evolutionary scenario by which one or more archaic populations made the transition to modernity. Our documentation of the final enhancement of working memory is primarily a contribution to the first of these subgoals. Evidence from neuroscience clearly identifies a planning and problem solving ability that is isolable neurologically and behaviorally from symbolic and language abilities. This component of modern thinking is working memory. Archaeological evidence indicates that human WM capacity underwent an enhancement to the modern range in the relatively recent past. Given the serendipitous nature of archaeological preservation and recovery, assigning a precise date for this development is not yet possible. Modern WM capacity was certainly in place by 30,000 years ago, but there is scattered evidence that it may be as old as 77,000 years ago. Despite this range of dates, it is clear that an enhancement of WM capacity was one of the final developments in human cognition that produced the modern mind.

Perhaps not unexpectedly, most of the criticism of our hypothesis has been aimed at its implications for the second sub-goal pursued by archaeologists—the narrative scenario of just who became modern [73, 74]. We admit to fueling this fire by directly addressing the issue of Neandertal cognition [41, 75, 76]. The archaeological signature of Neandertals is well known, but does not provide evidence for the enhancement of WM that can be found elsewhere. But beyond this fairly direct contrast, the archaeological record for EWM fits several alternative scenarios for just who become modern and when they did.

(1) Alleles for enhanced WM accompanied the parietal hypertrophy that distinguishes the brains of Homo sapiens sapiens from those of archaic Homo sapiens such as Neandertals [77]. The parietal hypertrophy is not itself evidence for an increase in WM capacity because WM is primarily a frontal lobe function, with significant neural links to the parietal and temporal lobes, and the basal ganglia. But clearly something did evolve in the brains of Homo sapiens sapiens by about 200,000 years ago [50], and perhaps WM capacity accompanied this development. If true, it leaves us with a chronological gap. Archaeological evidence for enhanced WM does not appear until at least 130,000 years later. There are two ways to account for this lacuna.

(a) The alleles for enhanced WM at first yielded only a very modest reproductive advantage, and it was not until a significant proportion of the population (presumably African) expressed the enhanced phenotype that group planning and problem solving began to provide a more marked advantage, which powered a subsequent rapid expansion after 70,000 years ago. We do not know whether or not such a sequence could even be modeled in microevolutionary terms, and at this point it remains conjecture.

(b) The alleles for enhanced WM yielded an immediate phenotypic advantage, which resulted in modern problem solving ability, but because its expression played out through learned cultural mechanisms, the ratchet effect initially allowed for only slow, almost imperceptible change. Essentially, enhanced WM had little to work with until cultural knowledge had accrued more and more components. If the archaeological record was more complete, we would see an accelerating rate of culture change over the course of 150,000 years. Some would argue that this is precisely what we do see, but given the limited number of data points such a conclusion is unwarranted, at least for now.

We find the 100,000–150,000 year gap between the first anatomically modern humans, and evidence for EWM to be troubling, and are unconvinced, thus far, by either solutions (a) or (b).

(2) Alleles for enhanced WM occurred by mutation in anatomically modern African population after 200,000 years ago, and probably after 100,000 years ago. Here the serendipity of archaeological discovery clouds the chronology. Some evidence—for example,
the indirect evidence for use of snares at Sibudu [38]—is as old as 75,000 BP. But abundant evidence for enhanced WM did not appear until about 30,000 year ago. As such, the archaeological record does fit an “Out of Africa” scenario in which a local southern or eastern population of modern humans expanded rapidly out of Africa following the demographic crash that occurred about 70,000 years ago [6]. It may well turn out, when the archaeological record for southern and eastern Africa are more thoroughly documented, that this version will prevail. But this is not the only feasible scenario. It is also possible that the final enhancement of WM occurred closer to 30,000 years BP, and spread rapidly via gene flow to populations all over the world. The archaeological evidence for enhanced WM just does not have the resolution to resolve this specific evolutionary puzzle.

A final observation concerning human cognitive evolution in general is appropriate. Evidence for the evolution of working memory fits nicely into a mosaic account of human cognitive evolution. Some components of modern cognition evolved long ago. Spatial cognition, for example, was modern by 500,000 years ago [78, 79], and evolved in circumstances very different from those of the last 100,000 years. Procedural cognition also has considerable antiquity, with archaic humans such as Homo heidelbergensis and Neandertals demonstrating modern procedural abilities [41, 80]. Symbolism, though a poorly defined cognitive ability, has roots stretching back perhaps 300,000 years (pigment use at Twin Rivers in Zambia [81]). In a very real sense, search for the evolution of modern cognition is a fool’s game. The components of modern cognition, like the components of modern anatomy, evolved at different times for different reasons. True, the final package did not come together until after 100,000 years ago, with working memory perhaps the final piece. But this was preceded by many other developments equally important to the modern mind.

References


Research Article
The Role of Genetic Drift in Shaping Modern Human Cranial Evolution: A Test Using Microevolutionary Modeling

Heather F. Smith¹, ²

¹ Department of Anatomy, Arizona College of Osteopathic Medicine, Midwestern University, 19555 North 59th Avenue, Glendale, AZ 85308, USA
² School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287-2402, USA

Correspondence should be addressed to Heather F. Smith, hsmith@midwestern.edu

Received 16 September 2010; Revised 3 January 2011; Accepted 14 January 2011

Academic Editor: Bing Su

Copyright © 2011 Heather F. Smith. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The means by which various microevolutionary processes have acted in the past to produce patterns of cranial variation that characterize modern humans is not thoroughly understood. Applying a microevolutionary framework, within- and among-population variance/covariance (V/CV) structure was compared for several functional and developmental modules of the skull across a worldwide sample of modern humans. V/CV patterns in the basicranium, temporal bone, and face are proportional within and among groups, which is consistent with a hypothesis of neutral evolution; however, mandibular morphology deviated from this pattern. Degree of intergroup similarity in facial, temporal bone, and mandibular morphology is significantly correlated with geographic distance; however, much of the variance remains unexplained. These findings provide insight into the evolutionary history of modern human cranial variation by identifying signatures of genetic drift, gene flow, and migration and set the stage for inferences regarding selective pressures that early humans encountered since their initial migrations around the world.

1. Introduction

Recent studies have revealed that human cranial morphology, whether quantified using absolute linear dimensions or relative geometric morphometric techniques, largely reflects population history among humans [1–9]. Certain functional and developmental modules (FDMs) of the cranium have been identified as reflecting relationships based on neutral molecular markers [4–9]. In particular, patterns of variation in the 3D morphology of the temporal bone, upper face, basicranium, individual bones of the cranial vault [5–9], and to a lesser degree the mandible [10] have been demonstrated to reflect genetic distances among human populations. The assumption behind these findings is that the morphology of these FDMs is evolving largely neutrally, in a similar manner to neutral molecular loci, and can consequently be utilized in a similar manner to estimate genetic relationships among human specimens or samples in the absence of direct molecular data. The relationship between cranial morphology and genetic distances in humans indirectly suggests that human cranial morphology is evolving primarily neutrally; however, the influence of selection has not been explicitly tested and therefore cannot be definitively discounted. Alternative methods exist for directly testing the impact of microevolutionary processes on morphology, through the use of microevolutionary modeling.

The patterns of phenotypic diversity within a species are central to inferring its modes of evolutionary diversification. Lande’s quantitative approach to evolutionary theory can be applied to assess the relative effects of genetic drift and selection in a sample [11]. Lande’s model is based upon the formula: $B_t = G(t/N_e)$, in which $B_t$ is the between-population V/CV matrix during generation $t$, $G$ is the additive V/CV matrix, and $N_e$ is the effective population size [11]. For phenotypic data from contemporaneous groups, $G$ is typically proportional to the within-group morphological V/CV matrix ($W$) [12–15]. Therefore, $W$ can be substituted for $G$ such that $B \propto W(t/N_e)$. Within a contemporaneous comparison, $t$ and $N_e$ will be constants, so the equation can be simplified further to $B \propto W$. In other words, if populations have diversified primarily through neutral evolutionary processes, then their within-group variance/covariance
(V/CV) matrices should be proportional to their between-group V/CV matrices [11, 16, 17]. Significant deviations from this pattern indicate that selection has impacted the morphology of a particular region, and the null hypothesis of neutrality is rejected [11, 16, 17]. This approach has been taken to evaluate the role of genetic drift in the evolution of cranial shape in New World Monkeys [18–21] and fossil hominins [22, 23]. Weaver and colleagues applied statistical tests based on Lande’s model to human and Neandertal linear cranial dimensions and concluded that morphological divergence between the two species was consistent with a null hypothesis of drift [23]. However, to date, Lande’s model has not been applied to the comparison of landmark-based cranial data among individual human populations with the purpose of assessing the impacts of selection and drift.

Several genetic models, including the Isolation by Distance (IBD) model [24–26] and the serial founder effect model [27, 28], predict that genetic divergence among demes increases with geographic distance between them. Therefore, one means of investigating the extent to which morphological diversity has been influenced by neutral or selective forces is to compare human craniofacial variation to geographic distance. Human cranial morphology, as captured by linear dimensions, generally has an association with genetic distance on a global scale [2, 4]. However, Relethford [2] notes that the accordance of data to an IBD model does not definitively clarify the evolutionary history of the phenotype. Environmental conditions also grade geographically, and morphological clines can be adaptive. Thus, adherence of a phenotype to a cline does not conclusively indicate that neutral factors are dictating its variation.

In population genetics, adaptation is often identified by first accounting for variation that falls within the potential range of neutral evolution (e.g., [29–31]). This tenet derives from the Neutral Theory of Molecular Evolution, which argues that the majority of genetic variation is likely to be selectively neutral [29–31]. According to the Neutral Theory, the rate of divergence among groups will be determined by the mutation rate (\(\mu\)), which is neutral [29]. Mutation generates new variation, and genetic drift acts upon it. There are two primary models in population genetics that describe neutral divergence patterns among groups [32–34]. According to the Constant Heritability (CH) model, the degree of heritability \((h^2)\) of a character dictates the rate of neutral divergence among groups [32, 33]. Under the Mutation-drift Equilibrium (MDE) model, the expected genetic variance is calculated as \(2N_e\sigma_m^2\), in which \(N_e\) is the effective population size and \(\sigma_m\) is the morphological variance. A population is considered to be in mutation-drift equilibrium when the rate of new mutations (causing new variation) is balanced by genetic drift (causing homozygosity). Therefore, morphological characters will reflect population history if populations approximate an equilibrium between mutation and genetic drift (MDE model), and/or new mutations impact the V/CV structure minimally (CH model). According to both models, the divergence rate among groups is determined by the strength of genetic drift, which is in turn dictated by \(N_e\) and \(h^2\).

Approaching the evolution of modern human cranial morphology from a population genetics framework, such as through the application of molecular-based models and microevolutionary modeling, provides a basis for understanding the patterns and variation that characterize humans today. Using an assumption of neutrality as the primary mode of microevolution unless demonstrated otherwise allows cranial form to be evaluated objectively, and deviations from the pattern expected under a neutral model can be investigated further and adaptive explanations sought out. This research expands upon previous studies investigating patterns of cranial morphology in modern humans, interpreting them in the context of geographic dispersion and migration, and revealing the microevolutionary processes which produced the variation in cranial shape observed among our species today.

2. Methods

2.1. Data Collection. Fifty-two landmarks capturing the shape of the basicranium and splanchnocranium were digitized in samples of sixteen modern human populations (Table 1). The three-dimensional coordinates of each landmark were recorded using a Microscribe G2 point digitizer (Immersion Corp.) and sent to a laptop computer. These landmarks are standard craniometric landmarks [35–37] and were selected for the current study following findings by Smith et al. [6, 7] that the morphology depicted by each set of landmarks overlaying is significantly correlated with genetic distances in humans (Table 2). The landmarks were divided into subsets representing four different FDMs of the skull: the basicranium, upper face, mandible, and temporal bone. FDMs are portions of a morphological structure, in this case the skull, that share a common biomechanical function and/or developmental pathway.

The analyses of the basicranium, upper face, and temporal bone consisted of all 16 populations. However, due to the fragmentary nature of museum collections, many specimens were found to be missing mandibles. Consequently, the number of populations with adequate sample sizes to be included in the mandibular morphology analysis was reduced compared to the other analyses. Ten of the sixteen populations contained a sufficient number of mandibular specimens (Table 1); however, this reduced number is equivalent to those used in several other similar studies [4, 5, 7]. Sex determination was conducted by the author for each specimen, using traditional sexually dimorphic cranial traits [35], and approximately equal sex samples were then utilized for each population. Each population sample consisted exclusively of adult individuals, as indicated by the eruption of the third molars and/or fusion of the sphenoccipital synchondrosis. An approximation of the average geographic coordinates for each population was obtained using Google Earth.

2.2. Geometric Morphometric Analyses. Morphological coordinate data for each FDM were superimposed separately using Generalized Procrustes Analyses [38–41] in MorphoJ 1.02 [42]. During this step, all specimens were scaled,
Table 1: Human population samples included in the present study, with sample sizes and museum locations. AMNH, American Museum of Natural History; NMNH, National Museum of Natural History; BNHM, British Natural History Museum. *Subset of populations included in the analysis of mandibular morphology.

<table>
<thead>
<tr>
<th>Population</th>
<th>Region</th>
<th>n</th>
<th>Museum locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cameroon*</td>
<td>Africa</td>
<td>44</td>
<td>AMNH</td>
</tr>
<tr>
<td>Khoisan*</td>
<td>Africa</td>
<td>43</td>
<td>AMNH, BNHM</td>
</tr>
<tr>
<td>Pare</td>
<td>Africa</td>
<td>27</td>
<td>AMNH</td>
</tr>
<tr>
<td>French</td>
<td>Europe</td>
<td>50</td>
<td>NMNH, BNHM</td>
</tr>
<tr>
<td>Russians</td>
<td>Europe</td>
<td>36</td>
<td>AMNH, NMNH</td>
</tr>
<tr>
<td>Han Chinese*</td>
<td>East Asia</td>
<td>50</td>
<td>AMNH, NMNH</td>
</tr>
<tr>
<td>Malay</td>
<td>East Asia</td>
<td>50</td>
<td>AMNH, NMNH</td>
</tr>
<tr>
<td>Japanese*</td>
<td>East Asia</td>
<td>40</td>
<td>BNHM, NMNH</td>
</tr>
<tr>
<td>Mongolian*</td>
<td>East Asia</td>
<td>44</td>
<td>AMNH</td>
</tr>
<tr>
<td>Siberian natives*</td>
<td>East Asia</td>
<td>51</td>
<td>AMNH, NMNH</td>
</tr>
<tr>
<td>Southern Indians*</td>
<td>South Asia</td>
<td>50</td>
<td>AMNH</td>
</tr>
<tr>
<td>Australian aborigines*</td>
<td>Oceania</td>
<td>48</td>
<td>AMNH, NMNH</td>
</tr>
<tr>
<td>Papua New Guineas*</td>
<td>Oceania</td>
<td>34</td>
<td>AMNH</td>
</tr>
<tr>
<td>Solomon Islanders</td>
<td>Oceania</td>
<td>31</td>
<td>AMNH, NMNH</td>
</tr>
<tr>
<td>Greenland Inuit*</td>
<td>North America</td>
<td>43</td>
<td>AMNH, NMNH</td>
</tr>
<tr>
<td>Mexican Indians</td>
<td>North America</td>
<td>44</td>
<td>NMNH</td>
</tr>
</tbody>
</table>

Table 1: Human population samples included in the present study, with sample sizes and museum locations. AMNH, American Museum of Natural History; NMNH, National Museum of Natural History; BNHM, British Natural History Museum. *Subset of populations included in the analysis of mandibular morphology.

rotated, and translated, such that any remaining variation among them could be attributed directly to shape differences. Principal Components Analyses (PCA) were subsequently conducted, and Principal Component (PC) scores obtained. Mahalanobis D2 distances between pairs of populations were calculated based on the morphology of each FDM. In the context of geometric morphometric analyses of cranial morphology, D2 distances have been argued to be especially appropriate for estimating distances among groups due to their shared properties with the molecular distances with which they should covary if the morphology is evolving primarily neutrally (e.g., [43]). Patterns of morphological distances were also visualized using Multidimensional Scaling (MDS) plots, in which the relationships indicated by a distance matrix are forced onto three dimensions.

2.3. Geographic Comparisons. Geographic distances between each pair of populations were calculated from their approximate average geographic coordinates using great circle distances, a haversine formula in which the distance between two points \( D = 2R \arctan(\sqrt{\text{hav} \theta / (1 - \text{hav} \theta)}) \), in which \( \theta \) is the central angle between the two points and \( R \) is the radius of the Earth. Additionally, following Ramachandran et al. [44], distance calculations included the consideration of migratory waypoints, geographic points through which early migrating humans would have likely passed in order to travel between continents. In accordance with Ramachandran et al. [44], the following waypoints were used: Anadyr, Russia (64N, 117E), Cairo, Egypt (30N, 31E), Istanbul, Turkey (41N, 28E), Phnom Penh, Cambodia (11N, 104E), and Prince Rupert, Canada (54N, 130W). This adjustment in geographic distance calculation for intercontinental distances generates a more realistic estimate of the actual amount of migratory distance between two human populations, taking potentially insurmountable geographic barriers such as large bodies of water into consideration.

In order to compare the statistical association between geographic distance and morphological distance, a Regression Analysis was conducted separately between the great circle distances for each pair of populations and the Mahalanobis distances based on the 3D morphology of each of the FDMs. The alpha level was set at \( \alpha = 0.05 \) for significance, and the slope (\( \beta \)) was used to infer the degree of continuity between the variables. The \( R^2 \) value was interpreted as the degree of morphological variation explained by geographic patterning. Geographic distances were also visualized using a Multidimensional Scaling (MDS) plot in order to compare the distribution of populations to those based on morphology.

2.4. Lande’s Method. The likelihood that the patterns of cranial variation observed in humans today can be explained by genetic drift can be assessed by comparing among- and within-population V/VC matrices [16]. The V/VC matrices for each FDM were calculated in Statistica 9.1 using a Multiple Analysis of Variance (MANOVA) in which the morphological traits were treated as dependent variables and the populations as independent variables. Each V/VC matrix was subsequently simplified into its Principal Components for comparative purposes. Following Ackermann and Cheverud [19], PC scores were calculated for each population mean by multiplying population means by the standardized eigenvectors: \( Y = XE \), in which \( Y \) is the PC score of a population mean, \( X \) is the matrix of population means, and \( E \) is the matrix of standardized eigenvectors. For each PC, the among-population variance was calculated as the variance among the population mean PC scores.

If the morphology of an FDM has diversified primarily through genetic drift, then Lande’s model predicts that the between-group variation will be proportional to the within-group variation \( (B \propto W) \). Here, this model is applied by comparing the within-population variances to the among-population variances. A Regression Analysis was conducted between the within- and among-population variances. For each FDM, a slope of \( \beta = 1.0 \) is consistent with the hypothesis of drift, while any significant deviations from 1.0 are suggestive of nonneutral patterns of microevolution.

As an additional test of whether the covariance structure was similar among and within populations, a Mantel test [45, 46] of these covariance matrices for each FDM was conducted in PopTools, an add-on for Microsoft Excel. FDMs, found to have significant correlations \( (P < .05) \), were interpreted as having similar covariances among and within populations. It should be noted, however, that this test considers only the lower diagonal of a matrix, and therefore the variances within each variable are not incorporated nor tested. As a result, it should not be considered a direct test of
Table 2: Cranial landmarks included in each functional and developmental module.

<table>
<thead>
<tr>
<th>Landmark</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basicranium</strong></td>
<td></td>
</tr>
<tr>
<td>Basion</td>
<td>Midline point on the anterior margin of the foramen magnum</td>
</tr>
<tr>
<td>Condylar foramen</td>
<td>The posterior point on the margin of the condylar foramen</td>
</tr>
<tr>
<td>Condyle anterior</td>
<td>Most anterior points on the occipital condyles</td>
</tr>
<tr>
<td>Condyle posterior</td>
<td>Most posterior points on the occipital condyles</td>
</tr>
<tr>
<td>Inferior nuchal</td>
<td>Midline point on the inferior nuchal line</td>
</tr>
<tr>
<td>Inion</td>
<td>Most posterior point on the external occipital protuberance</td>
</tr>
<tr>
<td>Mastoidale</td>
<td>Most inferior point on the mastoid process</td>
</tr>
<tr>
<td>Opisthion</td>
<td>Midline point at the posterior margin of the foramen magnum</td>
</tr>
<tr>
<td><strong>Temporal Bone</strong></td>
<td></td>
</tr>
<tr>
<td>Anterior articular</td>
<td>Most anterior point on the articular surface of the articular eminence</td>
</tr>
<tr>
<td>Auriculare</td>
<td>A point on the lateral aspect of the root of the zygomatic process at the deepest incurvature</td>
</tr>
<tr>
<td>Entoglenoid</td>
<td>Most inferior point on the entoglenoid process</td>
</tr>
<tr>
<td>Jugular</td>
<td>Most lateral point of the jugular fossa</td>
</tr>
<tr>
<td>Lateral eminence</td>
<td>Point on the center of the lateral margin of the articular surface of the articular eminence</td>
</tr>
<tr>
<td>Lateral ovale</td>
<td>Most lateral point on the margin of the foramen ovale</td>
</tr>
<tr>
<td>Mandibular fossa</td>
<td>Deepest point within the mandibular fossa</td>
</tr>
<tr>
<td>Mastoidale</td>
<td>Center of the inferior tip of the mastoid process</td>
</tr>
<tr>
<td>Medial articular</td>
<td>Most inferior point on medial margin of articular surface of the articular eminence</td>
</tr>
<tr>
<td>Petrous apex</td>
<td>Apex of petrous part of the temporal bone</td>
</tr>
<tr>
<td>Porion</td>
<td>Most superior point of the external auditory meatus</td>
</tr>
<tr>
<td>Postglenoid</td>
<td>Most inferior point on the postglenoid process</td>
</tr>
<tr>
<td>Supracondyloid gutter</td>
<td>Point of inflection, where the braincase curves laterally into the supraglenoid gutter, in the coronal plane of the mandibular fossa</td>
</tr>
<tr>
<td>Tympanic</td>
<td>Most inferolateral point on the tympanic element of the temporal</td>
</tr>
<tr>
<td>Zygion</td>
<td>Most lateral point on the zygomatic arch</td>
</tr>
<tr>
<td><strong>Face</strong></td>
<td></td>
</tr>
<tr>
<td>Dacryon</td>
<td>Point on the medial orbit at which the frontal, lacrimal, and maxilla intersect</td>
</tr>
<tr>
<td>Ectoconchonion</td>
<td>The intersection of the most anterior surface of the lateral border of the orbit and a line bisecting the orbit along its long axis</td>
</tr>
<tr>
<td>Frontomalarae</td>
<td>Most laterally positioned point on the frontozygomatic suture</td>
</tr>
<tr>
<td>Temporale</td>
<td>Most anterior midline point on the frontal bone</td>
</tr>
<tr>
<td>Glabella</td>
<td>Point of intersection between the frontonasal suture and midsagittal plane</td>
</tr>
<tr>
<td>Nasion</td>
<td>The lowest point on the margin of the orbit</td>
</tr>
<tr>
<td><strong>Mandible</strong></td>
<td></td>
</tr>
<tr>
<td>Condylion laterale</td>
<td>Most lateral point on the mandibular condyle</td>
</tr>
<tr>
<td>Coronoid process</td>
<td>Most superior point on the coronoid process of the mandible</td>
</tr>
<tr>
<td>Gnathion</td>
<td>Most inferior midline point on the mandible</td>
</tr>
<tr>
<td>Gonion</td>
<td>A point along the rounded posteroinferior corner of the mandible between the ramus and the body</td>
</tr>
<tr>
<td>Infradentale</td>
<td>Midline point at superior tip of the septum between the mandibular central incisors</td>
</tr>
<tr>
<td>M1-M2 contact</td>
<td>Projected (laterally) onto the alveolar margin</td>
</tr>
<tr>
<td>Mandibular notch</td>
<td>Most inferior point in the mandibular notch</td>
</tr>
<tr>
<td>Mesial P3</td>
<td>Most mesial point on mandibular P3 alveolus, projected onto the alveolar margin</td>
</tr>
<tr>
<td>Pogonion</td>
<td>Most anterior point on the mental eminence</td>
</tr>
</tbody>
</table>
Lande’s method but instead a means of further investigating covariance structure.

Alternate methods for investigating the proportionality of between- and among-group variance/covariance structure exist. Most notably, Flury [47] proposed a hierarchical method of testing for common principal components (CPCs) among matrices, which has been employed as a test for Lande’s model [48, 49]. However, it has been noted that this approach may generate spurious results in analyses involving large samples, since even small deviations in biological variation can appear statistically significant and cause the null hypothesis of neutrality to be rejected [18, 20]. Therefore, that approach was not employed here.

3. Results

3.1. Morphological Variation among Populations and Geographic Distance Comparisons. The matrices and multidimensional scaling plots of Mahalanobis D2 distances among populations indicated variation in the degree of similarity among groups in the morphology of the various FDMs (Figure 1). In general, most populations grouped roughly according to continent (Figure 1); however, the various continental groups exhibited differing degrees of dispersion and overlap with other groups.

The three populations from Oceania (Australians, Papuans, and Solomon Islanders) were highly divergent in the morphology of the temporal bone, and basicranium as a whole. In fact, the D2 distances among the Oceanic populations were among the highest recovered for these FDMs. Despite some geographic patterning, the MDS plots revealed overlap between continental groups in basicranial and temporal bone morphology, and the African and Oceanic populations did not cluster together (Figures 1(a) and 1(b)).

The D2 distances based on facial morphology suggested some shared structure between the populations from Africa and those from Oceania (Figure 1(c)). The East Asian populations grouped together, as did the European samples. The New World groups were similar along two of the three dimensions, but the Inuit showed the greatest morphological affinity to the native Siberians.

With regard to mandibular morphology, the African populations clustered together, as did the East Asian populations (Figure 1(d)). However, the Oceanic populations were again widely divergent both from each other and from their Pacific neighbors in East Asia. The Inuit fell near the East Asian samples in dimensions 1 and 2 but exhibited a negative score along dimension 3.

The great circle distances including waypoints among populations indicated the migratory distances required for each group to migrate to the center of the other’s average geographic location (Table 3, Figure 2). The Regression Analyses between great circle distances and morphological distances revealed variation in the relationship between geography and shape for each FDM (Table 4). Morphological D2 distances based on the basicranium and mandible had moderate ($R = 0.322$ and $0.358$, resp.) and significant ($P = .0003$ and .0156, resp.) correlations with great circle geographic distances.

The morphological patterns of temporal bone shape were significant ($P = .0367$) but low ($R = 0.191$). Basicranial D2 distances, on the other hand, were nonsignificant ($R = 0.110$, $P = .152$). The $R^2$ values of all comparisons were low (0.012–0.128).

3.2. Lande’s Model. The Regression Analysis of the PCs representing within- and among-population V/CVs indicated differences in the associations between these factors for the various FDMs (Table 5). For three of the four FDMs evaluated—the basicranium, face, and temporal bone—the slope was not significantly different than 1.0 (Table 5). The $R^2$ values for these three FDMs indicated that a substantial proportion of variation was shared (86.5–95.6%) between the within- and among-population variances.

For mandibular morphology, however, a different pattern emerged. The slope of the regression equation was $\beta = 0.746$, although the correlation between the two variables was nonsignificant ($P = .254$) and the 95% confidence interval ranged widely from $\beta = 0.275$–1.217. The $R^2$ value was 0.335 (Table 5).

The Mantel tests of within- and among-population covariances revealed highly significant correlations ($P < .001$) between the patterns for all four FDMs (Table 6).

4. Discussion

FDMs of the skull known to reflect population history in humans are generally assumed to be evolving primarily through neutral microevolutionary processes, such as genetic drift, gene flow, and mutation. As such, their patterns of variation behave very much like neutral molecular loci in that variation should accumulate at a relatively constant rate and grade geographically along clines. However, cranial morphology is affected by a number of complex and varied influences, including functional constraints and pressures of the masticatory apparatus, remodeling of osseous tissue, and indirectly through climate and diet, in ways that neutral genetic markers are not. As such, it is perhaps unreasonable to expect any functional aspect of morphology to behave according to a strict molecular model; however, the application of this framework can provide a starting point for identifying microevolutionary signatures that can be subsequently explored further.

The application of Lande’s model [16] to subsets of cranial data indicated that the null hypothesis of predominantly neutral evolution cannot be rejected for the basicranium, temporal bone, and upper face, cranial regions which have been found previously to reflect population history [5–8]. The 3D shape of the temporal bone and face also track geography, with populations located in greatest proximity to each other sharing the most morphological affinities.

However, despite the significance of these correlations, their $R^2$ values are low, suggesting that much of the variance is unexplained by this factor alone.

Basicranial morphology was found to have a nonsignificant relationship with great circle distances. One possible explanation for this result is that, while this FDM is generally evolving primarily neutrally across humans as a species, a few
<table>
<thead>
<tr>
<th>Country</th>
<th>Australia</th>
<th>Cameroon</th>
<th>China</th>
<th>France</th>
<th>India</th>
<th>Inuit</th>
<th>Japan</th>
<th>Malay</th>
<th>Mexico</th>
<th>Mongolia</th>
<th>Pare</th>
<th>PPNG</th>
<th>Russia</th>
<th>San</th>
<th>Siberia</th>
<th>Solomon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0</td>
<td>17337</td>
<td>8962</td>
<td>16434</td>
<td>9121</td>
<td>21859</td>
<td>13208</td>
<td>5276</td>
<td>20621</td>
<td>10226</td>
<td>17913</td>
<td>3384</td>
<td>13494</td>
<td>19758</td>
<td>12571</td>
<td>3644</td>
</tr>
<tr>
<td>Cameroon</td>
<td>17337</td>
<td>0</td>
<td>10026</td>
<td>6740</td>
<td>8462</td>
<td>20012</td>
<td>12767</td>
<td>12053</td>
<td>18774</td>
<td>9705</td>
<td>3110</td>
<td>16563</td>
<td>6367</td>
<td>3058</td>
<td>10732</td>
<td>17540</td>
</tr>
<tr>
<td>China</td>
<td>8962</td>
<td>10026</td>
<td>0</td>
<td>8611</td>
<td>3646</td>
<td>13233</td>
<td>3053</td>
<td>3678</td>
<td>11995</td>
<td>1264</td>
<td>10602</td>
<td>8188</td>
<td>5380</td>
<td>12447</td>
<td>3708</td>
<td>9165</td>
</tr>
<tr>
<td>France</td>
<td>16434</td>
<td>6740</td>
<td>8611</td>
<td>0</td>
<td>7871</td>
<td>17722</td>
<td>11034</td>
<td>11150</td>
<td>16484</td>
<td>8028</td>
<td>7316</td>
<td>15660</td>
<td>2691</td>
<td>9161</td>
<td>8609</td>
<td>16637</td>
</tr>
<tr>
<td>India</td>
<td>9121</td>
<td>8462</td>
<td>3646</td>
<td>7871</td>
<td>0</td>
<td>16703</td>
<td>6517</td>
<td>3837</td>
<td>15465</td>
<td>4496</td>
<td>9038</td>
<td>8347</td>
<td>5856</td>
<td>10883</td>
<td>7000</td>
<td>9324</td>
</tr>
<tr>
<td>Inuit</td>
<td>21859</td>
<td>20012</td>
<td>13233</td>
<td>17722</td>
<td>16703</td>
<td>0</td>
<td>11595</td>
<td>16575</td>
<td>6304</td>
<td>12211</td>
<td>20588</td>
<td>21085</td>
<td>13789</td>
<td>22433</td>
<td>9742</td>
<td>22062</td>
</tr>
<tr>
<td>Japan</td>
<td>13208</td>
<td>12767</td>
<td>3053</td>
<td>11034</td>
<td>6517</td>
<td>11595</td>
<td>0</td>
<td>7924</td>
<td>10357</td>
<td>3062</td>
<td>13343</td>
<td>12434</td>
<td>7355</td>
<td>15188</td>
<td>3279</td>
<td>13411</td>
</tr>
<tr>
<td>Malay</td>
<td>5276</td>
<td>12053</td>
<td>3678</td>
<td>11150</td>
<td>3837</td>
<td>16575</td>
<td>7924</td>
<td>0</td>
<td>15337</td>
<td>4942</td>
<td>12629</td>
<td>4662</td>
<td>8210</td>
<td>14474</td>
<td>7287</td>
<td>5650</td>
</tr>
<tr>
<td>Mexico</td>
<td>20621</td>
<td>18774</td>
<td>1195</td>
<td>16484</td>
<td>15465</td>
<td>6304</td>
<td>10357</td>
<td>15337</td>
<td>0</td>
<td>10973</td>
<td>19350</td>
<td>19847</td>
<td>12551</td>
<td>21195</td>
<td>8504</td>
<td>20824</td>
</tr>
<tr>
<td>Mongolia</td>
<td>10226</td>
<td>9705</td>
<td>1264</td>
<td>8028</td>
<td>4496</td>
<td>12211</td>
<td>3062</td>
<td>4942</td>
<td>10973</td>
<td>0</td>
<td>10281</td>
<td>9452</td>
<td>4519</td>
<td>12126</td>
<td>2554</td>
<td>10429</td>
</tr>
<tr>
<td>Pare</td>
<td>17913</td>
<td>3110</td>
<td>10602</td>
<td>7316</td>
<td>9038</td>
<td>20588</td>
<td>13343</td>
<td>12629</td>
<td>19350</td>
<td>10281</td>
<td>0</td>
<td>17139</td>
<td>6943</td>
<td>2699</td>
<td>11308</td>
<td>18116</td>
</tr>
<tr>
<td>PPNG</td>
<td>11702</td>
<td>16563</td>
<td>8188</td>
<td>15660</td>
<td>8347</td>
<td>21085</td>
<td>12434</td>
<td>4662</td>
<td>19847</td>
<td>9452</td>
<td>17139</td>
<td>0</td>
<td>12720</td>
<td>18984</td>
<td>11797</td>
<td>989</td>
</tr>
<tr>
<td>Russia</td>
<td>16061</td>
<td>6367</td>
<td>5380</td>
<td>2691</td>
<td>5856</td>
<td>13789</td>
<td>7355</td>
<td>8210</td>
<td>12551</td>
<td>4519</td>
<td>6943</td>
<td>12270</td>
<td>8788</td>
<td>4692</td>
<td>13697</td>
<td>13697</td>
</tr>
<tr>
<td>San</td>
<td>19758</td>
<td>3058</td>
<td>12447</td>
<td>9161</td>
<td>10883</td>
<td>22433</td>
<td>15188</td>
<td>14474</td>
<td>21195</td>
<td>12126</td>
<td>2699</td>
<td>18984</td>
<td>15561</td>
<td>0</td>
<td>13153</td>
<td>19961</td>
</tr>
<tr>
<td>Siberia</td>
<td>12571</td>
<td>10732</td>
<td>3708</td>
<td>8609</td>
<td>7000</td>
<td>9742</td>
<td>3279</td>
<td>7287</td>
<td>8504</td>
<td>2554</td>
<td>11308</td>
<td>11797</td>
<td>4692</td>
<td>13153</td>
<td>0</td>
<td>12774</td>
</tr>
<tr>
<td>Solomon</td>
<td>3644</td>
<td>17540</td>
<td>9165</td>
<td>16626</td>
<td>9324</td>
<td>22062</td>
<td>13411</td>
<td>5650</td>
<td>20824</td>
<td>10345</td>
<td>18116</td>
<td>989</td>
<td>13697</td>
<td>20274</td>
<td>12774</td>
<td>0</td>
</tr>
</tbody>
</table>
populations may still be differentially affected by selection. In fact, this FDM contains some potentially adaptive aspects of morphology, in particular with respect to climate. Three samples in this study could be characterized as “cold-adapted,” the Inuit, Siberian natives, and, to a lesser degree, the Mongolians. These three groups cluster together on the MDS plot (Figure 1(a)), indicating affinities in basicranial morphology, as captured by the current set of landmarks. It has been noted that extremely cold climates may have a disproportionate effect on the basicranial morphology of populations exposed to them, causing them to differ from expected patterns of morphology [6, 46]. The functional pressures of a diet of tough and partially frozen items and paramasticatory behaviors may result in adaptive and plastic responses in the morphology of the masticatory apparatus, including the temporomandibular joint (TMJ), as these factors result in extensive stress on the masticatory apparatus [48–52]. Ethnographic literature has documented the Inuit practice of using the anterior dentition as a “third hand” [53], contributing to additional stresses to be accommodated by the skull. The skeletal structure of the Inuit skull has been modeled as a series of adaptations.

Figure 1: Multidimensional scaling (MDS) plot of Mahalanobis D2 distances among populations based on three-dimensional morphology of the: (a) basicranium, (b) temporal bone, (c) face, and (d) mandible. All plots use the following color scheme: Africa = red; Asia = orange; North America = green; Europe = yellow; Oceania = blue.
study is difficult to tease apart definitively from geographic distance or population history, however, because the current cold-adapted populations are located comparatively closely to each other and share a fairly recent common ancestry [44]. Future studies incorporating cold-adapted samples from more distantly related groups, such as those from northern Europe, could shed additional light on this topic by separating out the effects of extreme climate in high latitudes from genetic relatedness.

One apparent deviation from geographic patterning in morphology is the widely divergent shape of the basicranium, temporal bone, and face among the three Oceanic populations. While these groups are located within a reasonably circumscribed geographic area (Australia, Papua New Guinea, and Solomon Islands), they are quite morphologically distinct. However, this is not surprising given that they are separated from each other by large bodies of water, which could have hindered the degree of possible gene flow among these groups by necessitating watercraft travel between islands. Additionally, several studies have suggested that Melanesia was likely colonized multiple times [54–58] and that speakers of the Papuan and Oceanic language groups have maintained a substantial degree of differentiation [54–56, 58]. Geographic distance is frequently used as a proxy for genetic distance, but, in cases of multiple colonization events of a region by genetically differentiated groups, this assumption may be unfounded. Friedlaender and colleagues found that molecular distances for populations within Melanesia are substantially higher than a worldwide average and show essentially no relationship with geographic distances [58]. Similarly, molecular studies have revealed substantial genetic delineation between Australian indigenous peoples and Papuans, which has been interpreted as separate colonization events from different ancestral stock [54–56]. In the case of basicranial and temporal bone microevolution, it is clear that simple geographic distances cannot explain the patterns of morphology among Oceanic groups, just as it does not explain their genetic relationships.

Genetic drift is rejected as the predominant mechanism influencing mandibular shape in *Homo sapiens*. This finding is consistent with previous studies that have attempted to approach this question in an alternate manner [7, 10]. Interpretations as to the utility of mandibular morphology for reconstructing population history have been varied and depend upon which populations are evaluated and the type of molecular data used for comparison. Smith [7] found that the shape of the mandible in nine human populations

Table 6: Results of Mantel test comparing among- and within-population covariances. All correlations were significant.

<table>
<thead>
<tr>
<th>FDM</th>
<th>$R$</th>
<th>$R^2$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basicranium</td>
<td>0.998</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Face</td>
<td>0.999</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mandible</td>
<td>0.868</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Temporal bone</td>
<td>0.999</td>
<td></td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Figure 2: Multidimensional scaling (MDS) plot of geographic great circle distances among populations, incorporating waypoints. Geographic regions are depicted as follows: Africa = red; Asia = orange; North America = green; Europe = yellow; Oceania = blue.

Table 4: Results of Regression Analysis of geographic great circle distances and Mahalanobis D2 distances based on the morphology of each functional and developmental module (FDM). Significant correlations are indicated in bold.

<table>
<thead>
<tr>
<th>FDM</th>
<th>$R$</th>
<th>$R^2$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basicranium versus geography</td>
<td>0.110</td>
<td>0.012</td>
<td>.1516</td>
</tr>
<tr>
<td>Temporal bone versus geography</td>
<td>0.191</td>
<td>0.037</td>
<td>.0367</td>
</tr>
<tr>
<td>Face versus geography</td>
<td>0.322</td>
<td>0.103</td>
<td>.0003</td>
</tr>
<tr>
<td>Mandible versus geography</td>
<td>0.358</td>
<td>0.128</td>
<td>.0156</td>
</tr>
</tbody>
</table>

Table 5: Results of Regression Analysis of within- and among-population variance/covariance (V/CV) matrices. Significant correlations are indicated in bold.

<table>
<thead>
<tr>
<th>FDM</th>
<th>$\beta$</th>
<th>95% confidence</th>
<th>$R^2$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basicranium</td>
<td>0.960</td>
<td>0.798–1.122</td>
<td>0.895</td>
<td>.0096</td>
</tr>
<tr>
<td>Face</td>
<td>0.995</td>
<td>0.926–1.065</td>
<td>0.956</td>
<td>.0048</td>
</tr>
<tr>
<td>Mandible</td>
<td>0.746</td>
<td>0.275–1.217</td>
<td>0.335</td>
<td>.2539</td>
</tr>
<tr>
<td>Temporal bone</td>
<td>0.940</td>
<td>0.801–1.079</td>
<td>0.865</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

to generate and dissipate powerful occlusal forces, one of which involves a thickened tympanic plate [48, 49]. These factors have also been documented to affect some aspects of facial morphology; however, the facial landmarks utilized here specifically avoid the oral cavity and its associated alveolar bone. The extent of cold adaption in the basicranial
did not significantly correspond with their genetic distances based on short tandem repeat (STR) data; however, in a comparison to molecular distances based on Alu insertion polymorphisms and using a smaller number of populations, the association became statistically significant. Additionally, Nicholson and Harvati [10] found that mandibular shape tracks loosely with geography, and that functional demands such as diet and climate also play a role. These studies highlight the complexity of the patterns of mandibular morphology currently present in our species and the variety of microevolutionary processes that likely contributed to its current distribution.

Overall, the shape of the human skull, whether quantified using linear measurements or three-dimensional landmarks, reflects population history to a large degree [1–9]. Much of the microevolutionary history of this region, in particular the temporal bone and upper face, has involved largely neutral mechanisms and therefore reflects population history. However, simple geographic distance models [24–28] cannot explain the extent of the variation observed in individual FDMs, and specific functional aspects of the skull, such as the masticatory system, require selective explanations and interpretations. A population genetics framework provides a starting point for inferring the multicausal mechanisms of human cranial evolution, understanding these processes and their impacts, and retracing human evolution.

The subtext of many studies investigating human cranial evolution goes beyond understanding how the morphological variation has evolved. In the absence of molecular data in the paleoanthropological record, many researchers wish to identify phylogenetically informative aspects of morphology that can be used as a proxy for genetic data to address questions about hominin phylogenetic relationships and relative genetic distances among individual hominin specimens. Research into the relationship between cranial morphology and genetic relationships in humans is a crucial step in this process; however, the assumption that the patterns characterizing Homo sapiens will also apply to nonhuman species has not been tested. Future studies investigating similar factors in other catarhine primates can elucidate this question by using phylogenetic bracketing. Cranial FDMs found to be evolving predominantly neutrally and reflecting genetic relationships both within humans and among other Old World primates can be evaluated in the hominin fossil record to confidently infer phylogeny.

She thanks Ian Tattersall, Gisselle Garcia, and Gary Sawyer of the AMNH, Dave Hunt of the NMNH, and Robert Kruszynski of the NHM for permission to measure physical anthropology collections in their care. An earlier version of this manuscript was improved by the comments of Bing Su and two anonymous reviewers.

Acknowledgments

This project was supported by the National Science Foundation (BCS-0622570) and The Wenner-Gren Foundation (Grant no. 7499). The author would like to thank Chris Stojanowski for insightful discussions about microevolutionary modeling, Charles Roseman for guidance regarding the application of population genetics principles to the study of human cranial morphology, the late Charlie Lockwood for direction on geometric morphometrics and statistical methods, Verne Simons for help with the calculation of great circle distances, and Brent Adrian for assistance with references and discussions of inverse OWL matrices.

**References**


Research Article

A 150-Year Conundrum: Cranial Robusticity and Its Bearing on the Origin of Aboriginal Australians

Darren Curnoe

School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, NSW 2052, Australia

Correspondence should be addressed to Darren Curnoe, d.curnoe@unsw.edu.au

Received 15 October 2010; Accepted 16 December 2010

Academic Editor: Bing Su

Copyright © 2011 Darren Curnoe. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The origin of Aboriginal Australians has been a central question of palaeoanthropology since its inception during the 19th Century. Moreover, the idea that Australians could trace their ancestry to a non-modern Pleistocene population such as Homo erectus in Southeast Asia have existed for more than 100 years, being explicitly linked to cranial robusticity. It is argued here that in order to resolve this issue a new program of research should be embraced, one aiming to test the full range of alternative explanations for robust morphology. Recent developments in the morphological sciences, especially relating to the ontogeny of the cranium indicate that character atomisation, an approach underpinning phylogenetic reconstruction, is fraught with difficulties. This leads to the conclusion that phylogenetic-based explanations for robusticity should be reconsidered and a more parsimonious approach to explaining Aboriginal Australian origins taken. One that takes proper account of the complex processes involved in the growth of the human cranium rather than just assuming natural selection to explain every subtle variation seen in past populations. In doing so, the null hypothesis that robusticity might result from phenotypic plasticity alone cannot be rejected, a position at odds with both reticulate and deep-time continuity models of Australian origins.

1. Introduction

The origin of modern humans remains a core topic of palaeoanthropology. Although four major models are presently being debated, only the out-of-Africa suite has received strong support from interpretations of both the fossil record and DNA [1–6]. Yet, the evolutionary origin of Aboriginal Australians remains controversial owing to the presence of considerable variability in cranial morphology during the Pleistocene and interpretations of its possible phylogenetic importance [1, 2, 7–18]. Moreover, there is sharp disagreement about the possible alternative causes of this variation and its significance to a global understanding of the evolutionary history of modern Homo sapiens [1, 2, 11, 12, 17, 18].

Long before a human fossil record was known for Australia, various speculative evolutionary sequences were devised linking nonmodern hominins to Aboriginal Australians. In particular, it was proposed that Pithecanthropus (Homo erectus) or the late-surviving Ngandong population of this species (sometimes referred to as Homo soloensis [14]) played a role in their origins. This idea has been an enduring theme of palaeoanthropology for more than 100 years: from Klaatsch [19] through to Westaway and Groves [18]. A recent major review of the question of modern human origins [1] identified three major issues for Australian palaeoanthropology to be resolved: (1) the relationship of the first Australians to later inhabitants of the continent, (2) whether late Pleistocene morphological diversity may have been accentuated by the severity of the last glacial maximum, leading to isolation and the forcing of morphological change in some Australian populations, and (3) if archaic populations such as those known from Ngandong did survive into the late Pleistocene, an analogous situation to that in Europe might have existed, raising the possibility of gene flow with dispersing H. sapiens. All three points are clearly interrelated. Regarding the first, the very earliest fossil remains from Australia (e.g., Willandra Lakes Human 3) do seem to fit metrically and morphologically within the range of living Aborigines [8]. Stringer’s [1] points 2 and 3 relate to the possible cause(s) of cranial robusticity in some Pleistocene/early Holocene Australians. That is, whether such features arose as a result of natural selection...
acting on populations within Australia or were brought here by people who evolved from, or hybridised/admixed with, a nonmodern population in Southeast Asia (i.e., the Solo/Ngandong hominins).

Australian palaeoanthropological theory and method continues to be dominated by adaptationist accounts [20] of robusticity and population history. The assumption that the cranium is optimised part by part and that atomising its form into traits assumed to be heritable units, functionally discrete, to have been shaped by natural selection and, therefore, positively associated with reproductive success, remains the core proposition of the field. In the present contribution, it is argued that the failure to fully consider alternative (nonadaptationist) approaches is a major reason why the interrelated issues of firstly, the cause(s) of cranial robusticity and, secondly, its relevance to reconstructing the origins of Aboriginal Australians remain unresolved. This paper commences with a review of the history of ideas regarding cranial robusticity and the origins of Aboriginal Australians. Then, an alternative to atomisation, herein called the ontogenetic approach, is described and some key concepts underpinning it are introduced. Finally, this approach is applied to some characters used to support the Ngandong ancestry model for Aboriginal Australians.

2. History of an Idea

Ever since Blumenbach’s De Generis Humani Varietate Nativae of 1795, European scientists have been attempting to understand the affinities and hence origins of Aboriginal Australians. Blumenbach had at hand the skulls of several Australians provided to him by Joseph Banks [21]. However, only with the wider exploration of the continent by Europeans and the beginnings of a global trade in human cranial remains during the 19th Century was a systematic effort made to understand their skeletal morphology [22]. During this period, the deep antiquity of the earth was beginning to be established, including an ancient origin for humankind as presented in Lyell’s Antiquity of Man [23]. In Lyell’s volume, Huxley’s (1863) first comparisons of Aboriginal skulls to Neanderthal remains were noted, indicating various morphological resemblances (not exclusive though, or implying ancestry). As the first anthropologist to study human origins from comparative anatomical and fossil sources, and placing his ideas within the Darwinian evolutionary framework, Huxley can reasonably be considered the founder of the discipline of palaeoanthropology. Moreover, as Aboriginal Australians were central to his ideas, as well as to the early development of this scientific field, Huxley ensured their place at the centre of debate surrounding human origins, a position they have held for close to 150 years.

During the early 20th century, many researchers continued to focus on documenting similarities in cranial form between Australians and the Neanderthals [24–26], inspired by Huxley’s earlier and highly influential work (see Figure 1). However, following the discovery of fossil human-like remains in Indonesia by Dubois [27], the view eventually emerged that Australians may actually have descended from a local population within Southeast Asia, either Pithecanthropus [19] or Homo soloensis [28, 29]. Klaatsch [19] found it “impossible to believe that the Australian natives are descended from European palaeolithic man” (page 162) as proposed by Huxley (see other criticisms in [30, 31]). However, the idea of a regional evolutionary sequence between Aboriginal Australians and Pithecanthropus in Southeast Asia seems not to have been in the minds of these workers. Instead, they saw Aboriginal Australians as representatives of the most “primitive type” of living H. sapiens and had in mind a global evolutionary sequence in which “Solo Man” (Ngandong) and “Rhodesian Man” (Kabwe) were examples of “proto-Australians,” belonging together to living humans in H. sapiens [29]. Also, Dubois [28, 29] thought that the Wadjak remains he recovered from Indonesia were “Australoid” although they now seem to be terminal Pleistocene in age and are probably not related to Aboriginal Australians [32].

With the ideas of Weidenreich [33, 34] and Coon [35], the “Pithecanthropoid-Australoid” lineage hypothesis become a major feature of palaeoanthropological theorising. Both these workers believed that modern humans around the world had evolved from regional populations of Pithecanthropus (today H. erectus). Thus, the regional lineage from this taxon through to Aboriginal Australians was part of a global process, which Howells [36] dubbed the “Candelabra” hypothesis. Parenthetically, it should be noted that this idea was not universally supported; for example, Hrdlička [37] preferred a Neanderthal stage in human evolution and considered Aborigines to belong to an “old race” of “whites,” borrowing heavily from Huxley, while Birdsell’s [38] views are strikingly similar to the contemporary out-of-Africa theory.
The increase in the number of fossilised human remains recovered during controlled excavations from the 1960s onwards [39–43], and demonstration of a Pleistocene occupation of the Australian continent [44, 45] encouraged renewed interest in the Pithecanthropoid-Australoid lineage hypothesis [39–43]. Again, it must be stressed that not all researchers accepted this hypothesis, Macintosh in particular reversing his earlier endorsement [46].

During the 1980s and 1990s, this idea was reformulated, initially as the “Regional Continuity” hypothesis [47] and later as the global “Multiregional” model of modern human origins [48, 49]. At its core was the notion of a deep-time Southeast Asian-Australian clade supported by evidence of morphological continuity in the skull and dentition between lower and upper Pleistocene nonmodern hominins in Indonesia through to recent Aboriginal Australians. As Thorne and Wolpoff [47] stated, “in no other region can a specimen (i.e., Sangiran 17) be found that combines so many features that seem unique or at least of high frequency in Pleistocene Australians” (page 345; words in parenthesis added). It is important to note that many proponents of the multiregional model have since the early 1990s regarded H. erectus to be a junior synonym of H. sapiens [50]; thus, the emergence of modern humans is seen by them as a process occurring within a single, long lasting, and widely distributed evolutionary lineage, or species.

Moreover, Thorne [51–53] proposed that two populations had colonised Australia at different times during the Pleistocene in his “dihybrid” model, the merging of the two giving rise to modern Aborigines. The first population was the “robust” group descended from H. erectus and the second a later arriving “gracile” population originating in Pleistocene China and, like modern East Asians, evolving from East Asian H. erectus. Although the order of arrivals had to be revised (reversed) once it was established that gracile cranial traits like WLH3 were actually geologically much older than any robust remains recovered from various localities [15, 53].

Three recent variants of the regional continuity model have been published. The first emerged from a test of multiregionalism in Southeast Asia-Australia using the WLH50 calvaria (see Figure 2) [54, 55]. It was found that six of the seven Ngandong calvaria examined were phenetically closer to WLH50 than to any other specimens considered, including early modern humans from Skhul and Qafzeh. In this work, it was concluded that the results implied a “dual ancestry” for Aboriginal Australians because “there is no evidence suggesting WLH-50 can be grouped with either Late Pleistocene Africans or Levantines to the exclusion of the Ngandong sample” [55, page 296]. That is, it was argued that Australians are descended from both an ancient regional population (e.g., Solo/Ngandong) and recent modern humans from Africa through a process of reticulation (admixture/hybridisation). The present author has made similar conclusions employing different fossils and an alternative, multivariate, methodology [2]. However, his findings were argued to be consistent with the “assimilation” model of modern human origins [56, 57], a hypothesis receiving stronger support from genetic studies [2]. Weidenreich in a personal communication to Birdsell [58] changed his mind late in life and also thought Australians had a dual ancestry. Oppenheimer [59] has also suggested that interbreeding might account for Australian robusticity. The present author has, however, changed his views and no longer considers interbreeding to provide a parsimonious explanation for Aboriginal Australian morphology or the origins of these people [17].

The final recent variant was proposed by Webb [14] and borrows heavily from Birdsell’s “trihybrid” hypothesis [60, 61], as have Thorne’s later ideas [62]. Webb [14] speculated that the first population to colonise Australia was the species H. soloensis, an upper Pleistocene descendent of Javan H. erectus. He argued that it migrated to the continent as early as 130–150 ka and ultimately adapted to local conditions, founding the robust Australian Pleistocene/early Holocene population. Further, he contended that modern humans entered Australia sometime between 50 ka and 75 ka, tracing their origins back to Africa. These people are argued to have been “Negrito or Negrito-like,” being of small stature, a feature developed “external to Australia” [14, page 239]. The two species (H. soloensis and H. sapiens) are argued to have gradually formed a single population through “genetic mixing,” but the process was dominated by the modern group, “which had the larger population and constantly, albeit slowly, receives fresh genes through a series of migrations arriving throughout the glacial maximum and after, into the early Holocene” [14, page 269].

3. Some Objections to Continuity

Many objections to Southeast Asian-Australian regional continuity and the related dihybrid origins model have been offered over the last 30 years or so. Some opponents have argued that perceived similarities between robust individuals and H. erectus may simply represent the retention of plesiomorphies of later Homo by Aboriginal Australians.
Indeed, Larnach [64] found that Aborigines are closest in cranial morphology among living people to H. erectus but stressed “their resemblance is never close” (page 159). He also noted that the Solo/Ngandong population exhibits a number of autapomorphies not found among Aboriginal Australians (see also [65, 66]), and “if Australians are descended from Solo Man, then sometime during that descent these unique traits were lost and their sites reverted to a similar state to that obtaining in Homo erectus” [64, page 170]. Thus, if regional continuity is correct, Pithecanthropus (H. erectus) must have given rise to the robust Pleistocene Australian group in a separate event, the Ngandong population (H. soloensis) being a late surviving and autapomorphic descendent, and unrelated to Australians. Durband [67] has pointed out that the available data for Javan H. erectus especially from the facial skeleton (i.e., n = 1 or Sangiran 17) is too inadequate to make reliable phylogenetic inferences.

Moreover, the dihybrid model and by implication regional continuity has been rejected by most Australian workers on the basis of its failure to take adequate account of spatiotemporal variability in cranial and dental morphology, including size, within a single continental population [8–10, 12, 18, 68–75], or at the least its failure to reject this as a null hypothesis. Moreover, some individual features underpinning regional continuity such as vault thickness have been proposed as selected for in response to violence and cranial trauma [8] (see below for a discussion of epigenetic explanations for some robust traits). Finally, the conceptual underpinnings of the dihybrid/migrationist model have been deemed to be founded in a “palaeontological” (i.e., typological) rather than population-based approach [10], as would be demanded by the synthetic evolutionary paradigm.

4. An Alternative Program

As noted above, most studies of Australian robusticity have assumed that the morphological characters under investigation are genetic (or heritable). This concept is used in its broader sense to mean that the phenotype of an individual is a good predictor of the genotype [76]. However, it has become a truism of developmental biology that “the old and compelling idea that there exist specific genes for virtually every structural detail throughout the craniofacial complex is simply not true” [77, page 230]. The more that is learned about the genes and developmental processes forming the cranium, the less tenable the atomisation of complex organs like the cranium has become [78]. Some concepts described in this section indicate powerfully why this is the case and underpin the need for an alternative paradigm in Australian palaeoanthropology.

Estimates of the heritability for many commonly employed cranial measurements and traits are available (e.g., [79–82]). Although there are some uncertainties surrounding the application of sample-specific heritability estimates to other populations, they nonetheless provide insight into the potential reliability of particular traits/variables as well as a broader understanding of the heritability and integrated nature of cranial ontogeny. Heritabilities vary widely for standard craniometric variables: \( h^2 = 0.000 \pm 0.000 \) to \( 0.867 \pm 0.156 \) [81]. Moreover, many variables have been found to exhibit heritability estimates that do not differ from zero (i.e., \( h^2 = 0 \)) [79, 81]. In a recent investigation, only two commonly employed measurements have been found to exhibit high heritability values: palate breadth (MAB) and nasal height (NLH) [81]. Moreover, major length (GOL), breadth (XCB), and height (BBH) dimensions exhibited low to moderate, but statistically significant heritability estimates [81]. However, many important regional measurements such as frontal chord (FRC), bifrontal breadth (FMB), biastionic breadth (ASB), bizygomatic breadth (ZYB), and nasal breadth (NLB) are characterised by low and nonsignificant heritability estimates [81]. Additionally, these estimates show a spatial pattern whereby lower estimates tend to be associated with the face. In many cases, they are associated with measurements for areas involving attachment sites for the muscles of mastication [81] suggesting an important role for epigenesis in their ontogeny.

Endocranial dimensions show moderate to high heritability estimates including important measures of human basicranial size and angulation [80]. These results highlight the reasonably high heritability and considerable ontogenic integration (correlation among dimensions) of human endocranial form, but contrast with low-moderate heritability for measurements on the ectocranial surface [81].

In a recent study of the heritability of cranial dimensions employing a 3D approach, Martínez-Abadias et al. [82] found a broadly similar pattern to previous (2D) studies [79–81]. Overall, dimensions showed low-moderate heritability with about 72% being significant (i.e., \( h^2 > 0 \)). Interestingly, they found the face to be the region with the highest number of significantly heritable traits and highest mean heritability, followed by the cranial base and the neurocranium [82]. Within regions they found the orbit, nasal part, neurocranial vault, and basicranium to be characterised by low to moderate heritability, while the masticatory apparatus exhibited low heritability [82], a finding consistent with Carson’s [81] results. Between cranial regions, they also found low-moderate heritability confirming the concept of ontogenetic integration of the cranium. The findings of Martínez-Abadias et al. [82] show that the cranial base, neurocranium, and face are characterised by similar levels of heritability and also strongly point to an important role for epigenesis in ontogeny (see also [80, 81, 83–86]).

Epigenesis is defined as the developmental interactions among cells, tissues, and their environments [84]. It can translate localised developmental alterations into integrated and widespread morphological changes and provide a fundamental mechanism for introducing flexibility into developmental programs including phenotypic plasticity [83–86]. Lieberman et al. [87] have discussed three types of epigenetic interactions during ontogeny: (1) primary interactions, occurring at the cellular level, (2) secondary interactions, involving those between adjacent tissues during growth, and (3) tertiary interactions, in which interactions occur throughout ontogeny between cells within a unit (e.g., via hormones) and the rest of the organism as well as the environment.

One important and widely discussed concept underpinning epigenesis in the ontogeny of the cranium is the
functional matrix hypothesis (FMH) of Moss [88]. He described the FMH in the following way: “The developmental origin of all cranial skeletal elements (e.g., skeletal units) and all their subsequent changes in size and shape (e.g., form) and location, as well as their maintenance in being, are always, without exception, secondary, compensatory, and mechanically obligatory responses to the temporally and operationally prior demands of their related cephalic nonskeletal cells, tissues, organs, and operational volumes (e.g., the functional matrices)” [88, page 9]. The FMH explains many aspects of cranial development and is widely regarded to be an important concept surrounding secondary epigenetic interactions during cranial ontogeny (e.g., [77, 89]).

An important example of tertiary epigenetic interactions is the widely discussed “Wolff’s” law of bone transformation [90]. This “law” is actually a poorly defined and frequently criticised term usually applied as a “catch-all” concept to refer to the adaptation of bone to mechanical stimuli [77, 78, 91, 92]. Put simply, Wolff’s law is, “an extension of the old and trusted idea that form is interrelated with and inseparable from function... that bone grows and develops in such a manner that the composite of physiologic forces exerted on it are accommodated by the bone’s developmental processes” [77, page 233]. Historically, one problem with this concept’s use has been that it was frequently invoked to argue for uniform responses such as bone deposition in the face of biomechanical stress particular from muscle, when very often the opposite response, resorption, had occurred [77, 91]. Thus, while in its general form this law remains valid and useful, it is now widely acknowledged that just how bone responds to mechanical stress is complex and varies according to its location and whether stress is direct or is mediated by other tissue. The reader is referred to some recent and more detailed reviews of this concept [77, 78, 92].

The external environment through tertiary epigenetic interactions has long been argued to be an important determinant of cranial form. For example, the shapes of the neurocranium and nose have long been linked to climatic adaptation [93–100] and facial form to diet or masticatory practices (see [101–104]; see below for a more detailed discussion of this idea). A 3D study of cranial morphology compared to neutral genetic population distances [105] found that the human cranium does preserve a “signal” of population history. However, historical signatures are not equal across cranial regions and seem to be largely concentrated in the temporal bone and neurocranium [105].

The final concepts to be briefly considered here are those of modularity and integration. Modularity is a general property of biological systems, from molecular to ecosystemic levels of interaction [86]. In ontogeny, it represents the observation that morphological features do not vary independently but are integrated with each other, reflecting coordination in development, function, and evolution. They might be thought of as forming modules or “complexes that are highly integrated internally but are relatively independent of each other” [86, page 628]. The morphological characters within modules are characterised by three major properties [106]: (1) they collectively serve a common functional role, (2) they are tightly integrated by pleiotropic effects of genetic variation, and (3) they are relatively independent of other modules. This independence among structures (modules) allows unrelated components to vary and evolve separately, but the integration within the units maintains functionally necessary relationships among traits [85–87, 106–113]. Morphological integration assumes that developmentally/functionally related traits are coinherited and will produce coordinated responses in evolution [85–87, 106–113].

The cranium is divided along ontogenetic lines into three regions: the cranial base, neurocranium, and face. As the cranial base is in an evolutionary sense the oldest structure, it is phylogenetically highly conserved [112] and believed to be subject to stronger genetic influence in ontogeny than the neurocranium and face [77, 113–116]. Moreover, it widely is assumed that the face is the most sensitive region to epigenetic factors as it stops growing later in ontogeny than either the cranial base or neurocranium [77, 114, 115]. Thus, it is subject to greater influence from mechanical loading during mastication and from various environmental factors during ontogeny [116–119]. There is, however, considerable evidence that the face and neurocranium are characterised by both integration between regions and considerable modularity, or region-specific integration [77, 80–83, 85, 87, 115, 120–123]. Moreover, this pattern seems to characterise humans, chimpanzees, and gorillas [106, 120, 124, 125] and is known to be phylogenetically conserved in mammals [107, 126]. Cranial shape patterns appear early in ontogeny and remain from early childhood until adulthood [127], including in depository and resorptive cranial growth fields [115, 128]. Although this is somewhat simplistic as endocranial shape continues to change in humans through adolescence, well after brain growth it has ceased [129].

A range of studies aiming to test and develop Enlow’s ideas about facial and cranial growth [77, 115, 130] have been published over the past two decades and employed a range of techniques including 3D morphometrics [119, 122, 123, 130–144]. Many studies have focussed on interactions between the cranial base, neurocranium, and facial form, and some major findings include that larger relative brain size is associated with a larger basiocranial angle, while large faces may produce the opposite situation, although the influence of facial size seems to be weaker than for brain size [140]. Lieberman et al. [122, 123] also found that cranial proportions (neurocranial shape, or degree of brachycephaly versus dolicocephaly) depend on interactions between cranial base width and brain volume. This has been confirmed by 3D studies showing that changes in cranial base width play an important role in cranial shape variation, including facial width [127]. Moreover, studies of the heritability of cranial regions confirm strong covariance between the breadth measures of major developmental regions of the skull [82]. Additionally, independent of age and size, an important proportion of cranial shape variability seems to be traceable to differences in the position and orientation of the face and masticatory system relative to the braincase [127].
Experimental research using mice suggests that integration in the mammalian skull is highly structured following a hierarchical scheme dominated by covariation between the widths of the neurocranium and the basicranium and, to a lesser extent, also the face [85]. Thus, the cranial base can be thought of as the “skull’s central integrator” [122, 123]. It strongly influences overall cranial shape, constraining facial breadth, height, and length, and neurocranial breadth and length, helping prevent the “different regions from evolving independently and would preserve the functional and architectural requirements of the skull” [82, page 29].

5. Nonadaptationist Explanations and Australian Robusticity

The above review of some ontogenetic concepts and recent research findings provide the context for a reconsideration of the possible causes of cranial robusticity among Australians. There have been very few synthetic attempts to explain it using nonphylogenetic or nonadaptationist approaches. One important, but overlooked, example is that of Howells [145] who proposed that robusticity may have been “a phenotypic plastic response to some regional and transient environmental stimulus. That is, the phenomenon would be, not adaptation and selection, resulting in genetic change in the post-Mungo population, but a reversible phenotypic shift, on an unchanged genetic basis, toward larger size and related allometric effects on the face and mandible particularly, producing in some relatively smaller-brained individuals the special flattened and narrow frontal which is so striking in Cohuna and some of the Kow Swamp skulls” (pages 646-647). This explanation is broadly consistent with the ontogenetic approach outlined above, being in large part a putative example of tertiary epigenetic interactions.

The cultural practice of artificial deformation, another example of tertiary interactions, has been argued to have produced or exaggerated the acute angle of the frontal squama, angulation of the parietais, and aspects of occipital squama morphology in a small number of robust Australian crania such as Kow Swamp 5, Cohuna, some Coobool Creek crania and Nacurrie 1 [9, 146–151]. Other epigenetic factors have been proposed to explain vault thickness, including, increased levels of growth hormone [8], nutritional phosphorous deficiency [8], and inherited anaemia [13, 14, 152, 153]. The latter (pathology) hypothesis has been examined to some extent by Stuart-Macadam [154], Curnoe and Thorne [16], Westaway [155], and Curnoe [16] but requires further scrutiny (see below).

Lahr and Wright [156] using multivariate statistics considered spatial (or architectural) aspects of cranial form and proposed that the “superstructures” (see below) characteristic of cranial robusticity may be integrated and covary allometrically. For Australian Aborigines, they concluded that like other human populations the expression of robust characters was likely a response to some functional complex, possibly mastication. They also suggested that the “distinctive anatomical combination present in Australian crania of a very narrow vault and pronounced robusticity, does not represent a plesiomorphic state” inherited from either early modern humans or the Solo/Ngandong population [156, page 184]. Moreover, they proposed that allometry in combination with this regional morphology might explain extreme robusticity seen in the Pleistocene of Australia. However, in attempting to explain the origins of this cranial morphology, they turned again to an explanation involving natural selection during the settlement of Australia.

In recent research by the present author [17], the ontogenetic framework was adopted in an attempt to explore nonphylogenetic explanations for cranial robusticity among Pleistocene/early Holocene Australians. It was concluded that robusticity among these individuals may have been the result of the complex developmental and functional interplay between (1) a large neurocranium, (2) narrow cranial base, (3) large viscerocranium with considerable high midfacial projection, and (4) large dentition, especially the posterior teeth, with their resultant large jaws (mandible and maxilla) and high volume of masticatory muscles. While it was suggested that these features were probably heritable to some extent, other factors such as body size, advanced physiological age, environmental effects from the physical demands of a hunter-gatherer lifestyle in an arid zone, dietary factors including food abrasiveness and limited preparation of food, and the use of teeth as tools may all have been factors affecting (exaggerating) the expression of robusticity in the Australian context.

5.1. Cranial Thickening as Pathology in WLH50. As noted above, pathological processes are suggested to have greatly enlarged the already thick vault bones of WLH50 (Figure 2). Webb [13, 14, 152, 153] has listed three indicators of pathology in this individual: (1) uniform vault thickness, (2) identification of the “hair-on-end” sign on lateral radiographs of the calvaria, and (3) a thin vault cortex.

Figure 3 compares vault thickness at eight locations in WLH50 with the median and range for the remaining Willandra Lakes series, as measured by Webb [13]. The large absolute thickness of the WLH50 vault is striking...
although thickness at one landmark (inion) does lie well within the range of the remaining Willandra Lakes sample. It is also very clear that vault thickness in WLH50 is far from uniform, as it varies between locations in a similar magnitude to the medians for the Willandra Lakes series (Figure 3). The magnitude of change (difference) between each of eight locations was calculated for WLH50 and the rest of the Willandra Lakes series, and the median value is actually larger in this individual (2 mm versus 1 mm). Moreover, a nonparametric Mann-Whitney U-test showed that the sample medians are statistical indistinguishable ($n = 7/54$, $U = 187.5$, $P = .98$; Monte Carlo $P = .98$). Thus, WLH50 does not possess a vault of uniform thickness when compared to the rest of the Willandra Lakes Human sample.

Figure 4 is a lateral radiograph of WLH50 taken in 2006 by the author and A. Thorne. After careful inspection, no indication of the hair-on-end sign can be seen. Figure 5 is a lateral slice of a CT-scan of WLH50 also taken by the author and A. Thorne in 2006. Again, in this and every slice inspected by the present author (245 slices in three planes, or 735 images), no evidence of the hair-on-end sign can be found. There is, however, evidence for abrasion on the external cortex around bregma (Figure 5), which may have been mistaken by Webb for the vertical spiculations sometimes associated with porotic hyperostosis.

It is also clear in Figures 4 and 5 that the external table of WLH50 is far from thin. To assess this quantitatively, the ratio of diploe to external table thickness was calculated from thickness measurements made on CT-scans just lateral to the median sagittal plane by the author. External table thickness was chosen because along with the diploe this vault component is altered in cases of porotic hyperostosis [157]. The median of this ratio for WLH50 measured at nine locations shows the diploe to be 1.4 times as thick as the external table, while in WLH3 the median shows it to be 0.9 times as thick. While WLH50 certainly does exhibit relatively thinner external table, it is far from thin. Finally, the results of a Mann-Whitney U-test showed the median difference between WLH3 and WLH50 to be statistically nonsignificant ($n = 9/23$, $U = 74$, $P = .22$; Monte Carlo $P = .23$).

5.2. Ontogenetic Examination of Some Phylogenetic Characters. Table 1 lists 16 morphological traits for WLH50 used to support a role for the Solo/Ngandong population in the evolution of Aboriginal Australians (i.e., a reticulate or dual ancestry model) [54, 55]. These “nonmetric” traits were originally selected in order to avoid “duplicating features that seemed to reflect the consequences of the same anatomical variation” [55, page 294]. While somewhat ambiguous, this statement is understood by the present author to imply that traits were employed that were believed to be developmentally and functionally nonintegrated and, therefore, to be characterised by weak or absent covariation. This assumption is crucial to an evaluation of the phyletic valency of cranial robusticity in the Australian context and the implicit assumption that characters shared by Australian and Solo/Ngandong crania are homologous. (The reader is referred also to Bräuer et al. [158] who provided a critical evaluation of the character coding and statistical methods employed by Hawks et al. [54] and Wolpoff et al. [55].)

All of the characters shown in Table 1 belong to a single cranial (developmental) unit, namely, the neurocranium. A small part of the lateral cranial base is preserved while the facial skeleton is missing save a fragment of zygomatic bone. Moreover, these traits represent two subunits of the neurocranium most of them belong to the cranial vault (14), with the remaining (2) belonging to the orbit subunit (Table 1). Given the now well-established finding of ontogenetic modularity and integration of the neurocranium (e.g., [77, 80–83, 85–89, 106–144]), there are good reasons to be suspicious about the use of the atomisation in this instance.

According to the functional matrix hypothesis [88], the cranium comprises various functional cranial components (FCC), which are ontogenetically integrated. One FCC contains the brain, cerebellum, and ocular globe, sharing a common embryological origin in the neural tube [77]. This FCC
develops in an integrated way being subjected to common heritable and epigenetic factors. The neurocranium largely develops as a result of the passive displacement of the vault bones, occurring within a connective tissue stroma, which is enlarged by the growing brain [77]. As a result, all three parts of this FCC follow a common growth trajectory, reaching 90–96% of their adult size by about age 6-7 [77, 89, 115, 129]. However, the neurocranium also forms an integrated unit with the cranial base, the “neurobasicranial complex,” and the developing face also epigenetically “finetunes” basicranial morphology, indirectly altering the form of the cranial vault [122, 123, 131, 132, 136–144].

Additionally, all 16 characters are cranial “superstructures” as originally defined by Weidenreich [159]. He suggested that superstructures develop in response to tension (“pull”) from muscles, a hypothesis subsequently verified experimentally with respect to some cranial crests [77, 78, 91, 92, 115, 160, 161], implying that they “belong to, or sit only upon, the outer bony table” [91, page 208]. However, this situation is not universal for superstructures and their development is complex.

As Enlow has shown [77, 91, 115, 161], the external surface of a bone is frequently shifted from an endosteal position. This means that developmental changes within the endocranium as well as epigenetic factors acting on the external table determine the morphology of the vault. Moreover, different processes may be involved in remodelling the neurocranium in spatially adjacent regions, and different regions of the cranium may exert influence over the growth of a single structure at different ontogenetic stages. For example, about midway up the frontal squama is a reversal line from resorptive (inferiorly) to depositional (superiorly) remodelling [77]. Once the frontal lobes have largely completed their enlargement, the internal table stops growing, while the outer table continues to remodel, progressively separating the inner and outer tables, and replacing through resorption cancellous bone to form the frontal sinus. However, this continued growth now results directly from the anterior remodelling of the upper face or nasomaxillary complex [77]. Further, the orbit (subunit) is contained within this region, and the inner table of its roof (endocranial side) is resorptive while the external (orbital plate) is depository [77]. As the orbital roof continues to remodel inferiorly along with the growth of the frontal lobes of the brain, the orbit must grow anteriorly to provide sufficient room for the growing eye; it is displaced outwards with other bones of the orbit which form part of the face according to the V principle [77].

In the case of the supraorbital torus, also considered to be a superstructure by Weidenreich [159], there is now abundant evidence that this feature develops epigenetically as a result of the spatial relationships among the face, neurocranium, and cranial base rather than biomechanical stresses during mastication [77, 115, 162–167], contra [168]. For example, Figure 6 is a Holocene San skull indicating stress gradients from masticatory stress as analysed with the finite element method [166]. The supraorbital region is unaffected by strain in this homogenous 3D model (but see similar results using a heterogeneous model [165]).

Thus, while 14 characters might be regarded as superstructures of the cranial vault, developing to provide muscle attachments and remodelling in response to mechanical stress, research into the supraorbital torus makes clear that the two orbital roof features lack epigenetic influences from the muscles of mastication. The supraorbital develops as

---

Table 1: Cranial traits used in phenetic studies of WLH50 (characters from [55]).

<table>
<thead>
<tr>
<th>Character</th>
<th>Developmental subunit*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Angular torus</td>
<td>Vault</td>
</tr>
<tr>
<td>(2) Coronal keel</td>
<td>Vault</td>
</tr>
<tr>
<td>(3) Sagittal keel on frontal</td>
<td>Vault</td>
</tr>
<tr>
<td>(4) Lateral frontal trigone</td>
<td>Vault</td>
</tr>
<tr>
<td>(5) Linea obliquus strongly developed</td>
<td>Vault</td>
</tr>
<tr>
<td>(6) Mastoid crest</td>
<td>Vault</td>
</tr>
<tr>
<td>(7) Sagittal keel on parietal</td>
<td>Vault</td>
</tr>
<tr>
<td>(8) Postlambdoidal eminence</td>
<td>Vault</td>
</tr>
<tr>
<td>(9) Prebregmatic eminence</td>
<td>Vault</td>
</tr>
<tr>
<td>(10) Projecting inion</td>
<td>Vault</td>
</tr>
<tr>
<td>(11) Sulcus dividing the medial and lateral elements of the supraorbital torus or superciliary arches</td>
<td>Orbit</td>
</tr>
<tr>
<td>(12) Superior margin of the orbit blunt (as opposed to sharp)</td>
<td>Orbit</td>
</tr>
<tr>
<td>(13) Suprainiac fossa</td>
<td>Vault</td>
</tr>
<tr>
<td>(14) Supramastoid crest</td>
<td>Vault</td>
</tr>
<tr>
<td>(15) Temporal line forms a ridge</td>
<td>Vault</td>
</tr>
<tr>
<td>(16) Transversely extensive nuchal torus</td>
<td>Vault</td>
</tr>
</tbody>
</table>

* After Ackermann and Cheverud [120].

---

Figure 6: Stress (von Mises) distributions in a high-resolution three-dimensional computer simulation of cranial mechanics in a Holocene San skull (modified from [166]).
a result of a complex interplay between the shape of the neurocranium and breadth of the basicranium, growth of the frontal lobes of the brain, development of the frontal sinus, remodelling of the orbit to accommodate the growing eye, and growth of the nasal region, which develops in response to lung enlargement and in particular affects the superior orbital rim, and the anteroposterior length of the face seems also to be a determinant of its position relative to the frontal squama and, therefore, the degree of frontal recession seen [77, 115, 122, 123, 137, 162–171]. Moreover, crania with high midfacial projection exhibit a long anterior sphenoid resulting in a viscerocranium positioned anterior to the neurocranium [133, 134, 170, 171] and in a prominent supraorbital. Finally, the development of the supraorbital region may even be epigenetically influenced by exercise through the effects of hormones seemingly independently from (but commonly with) other structures such as thickened vault bones (see [172, 173]; see also [8]).

As noted above, Mitteroecker and Bookstein [106] have found four common cranial factors or trajectories across a range of primates including humans, which are integrated through genetic effects such as pleiotropy and linkage, as well as epigenetic processes.

**Factor 1.** An enlarged and prognathic maxilla, relatively small cranial capsule, cranial crests, and enlarged zygomatic arches are features epigenetically associated in primates mainly as a result of masticatory muscle action. These are all characteristics of cranial robusticity and while they can be observed as a package in better-preserved crania (e.g., Cohuna, Kow Swamp 1), some of them are observable in incomplete remains such as WLH50 (Table 1, Figures 2 and 7) [2, 8–14, 16–18, 41–43, 46–52, 54, 55, 64, 72, 151].

**Factor 2.** This pattern contrasts broad and short crania with narrow and long crania (brachycephalic versus dolichocephalic crania), including both the face and the neurocranium, and also involves changes in the overall size of the face relative to the neurocranium. All robust Australians (e.g., Cohuna, KS1, WLH50) exhibit narrow and long crania (Figures 2 and 7) [2, 8–14, 16–18, 41–43, 46–52, 64, 72, 151]. Overall neurocranial length, breadth, and height are characterised by moderate but significant levels of heritability, suggesting a large amount of variation in these features determined by epigenetic forces [79, 81, 82]. Moreover, while their faces are absolutely long and consistent with this factor (e.g., KS1, KS5, Cohuna), they differ in exhibiting absolutely and relatively broad faces [2, 8, 9, 11–14, 17, 18, 41–43, 47–49, 51, 52, 151]. However, facial shape has been found to exhibit weaker correlations with neurocranial and basicranial shape than these regions have with each other, perhaps explaining why the face is the most variable part of the skull in humans [85]. Additionally, the dimensions of facial length, breadth, and height are characterised by moderate heritability indicating that additive genetic variation accounts for approximately 30% of the phenotypic variation [79, 81, 82]. Further, facial breadth is well known to be under strong ontogenetic influence from the muscles of mastication [171, 174–178]. A positive correlation between total masticatory muscle size (cross-sectional area) and body size (stature and weight) has also been found [171, 174–178]. The body size of the WLH50 individual, like other Pleistocene Aboriginal Australians, would have been large by later Holocene and contemporary standards [8, 17]. In fact, metrical dimensions of size and endocranial volume reveal it to be one of the largest Pleistocene modern humans found anywhere in the world and comparable in size to early African crania such as from Herto, and larger than many nonmodern individuals (compare with data in White et al. [179]), including H. erectus calvarii [17].

**Factor 3.** Encompassing relative size of the midface and neurocranial globularity, two characteristics that are tightly associated during postnatal ontogeny. While sample size disallows a proper comparison, robust Australian individuals such as KS1 are certainly characterised by a relatively narrow midface (ZMB/NPH = c66%) compared with the gracile Keilor cranium (c76%) (data from [151]). Moreover, these and other crania are strongly contrasted in their degree of neurocranial globularity; robust crania exhibiting highly angulated vaults, with receding frontal squamae (Table 1, Figure 2) [2, 8, 9, 11–14, 17, 18, 41–43, 47–49, 51, 52, 151]. In contrast, individuals such as WLH1, WLH3, and Keilor possess more globular crania (see Figure 7) [2, 9, 13, 14, 17, 41, 42, 51–53, 71, 72, 151].

**Factor 4.** A roundish and relatively short and high neurocranium to elongated, ellipsoidal crania. Exaggeration of the latter generates an occipital bun and lambdoid flattening along with large browridges. Robust Australian crania all exhibit ellipsoidal crania (superior outline); many have an occipital bun and lambdoid flattening, and all exhibit prominent browridges by modern human standards (see Figures 2 and 7) [2, 8, 9, 11–14, 17, 18, 41–43, 47–49, 51, 52,
Although the WLH50 calvaria is actually relatively tall in addition to being very long, this probably results from the combination of large endocranial volume and very narrow cranial base (see [17, 156]; see also [180, 181]). A range of other factors interacting in a complex manner are also known to influence the development of the supraorbital part, cranial shape being one of them (see above).

It is interesting also to note that Mitteroecker and Bookstein [106] found that their Factors 1 and 2 were the most highly conserved among humans, chimpanzees, and gorillas, with all three sharing a single ontogenetic trajectory owing to the presence of tight integration during development. Many of the traits listed in Table 1 can be explained by variation within these two highly constrained modules. Trajectories for Factors 3 and 4 were found to deviate considerably in humans compared to chimpanzees and gorillas suggesting that certain human cranial features have decoupled during evolution [106]. They opined that this difference was best explained by local developmental factors unique to humans such as relative brain size, posture, and locomotion (Factor 3), and facial shape, mainly involving the browridge (Factor 4). In the end, this research indicates that selection does not act on individual features, but instead cranial traits form part of an integrated and heritable set of factors or trajectories. Australian cranial robusticity lies on the trajectory for all four factors but is never as extreme in its expression as seen among nonmodern hominins.

Figure 7 presents 3D models (scaled to the same length) constructed by the author from CT-scans of WLH3 and WLH50, a Holocene African hunter-gatherer and various Pleistocene (nonmodern) hominins. Compared to WLH5 and the Holocene San individual, WLH50 does exhibit a receding frontal squama, prominent browridge, and some projection of the occipital. However, compared to all nonmodern, middle Pleistocene calvarii, WLH50 is tall, has a more upright frontal squama, which bulges anteriorly (prebregmatic eminence), and has relatively rounded parietal and occipital profiles and mild lambdoidal flattening. In fact, it closely resembles WLH3 in all these respects. Moreover, it lacks the strongly projecting browridge and a posttoral shelf so evident in all middle Pleistocene remains, whether belonging to H. erectus, H. heidelbergensis/rhodesiensis, or archaic H. sapiens/H. helmei. Its profile including the frontal squama is quite rounded (globular) compared to nonmodern crania. Thus, the purported similarities between WLH50 and H. erectus seem to have been greatly exaggerated, particularly when a comparison involving Middle Pleistocene African crania, the putative ancestors of all modern humans [1], is made. Instead, WLH50 simply presents as a more rugged version of the WLH3 morphology, the two resembling each other in their angle of the posterior part of the frontal squama and profile of the parietals and occipital; when compared with the San cranium, and indicating regional-specific features for Australian crania. Moreover, the comparatively modest differences in shape between WLH3 and WLH50 seem to be explicable in terms of the ontogenetic trajectories described by Mitteroecker and Bookstein [106].

Another important recent study of cranial robusticity with strong bearing on the morphology of WLH50 and other Pleistocene/early Holocene Australians is that of Baab et al. [182]. They tested for integration/correlation among various robusticity traits in a global sample of humans including recent Aboriginal Australians. Overall, they found that crania with more prognathic faces, expanded glabellar and occipital regions, and longer skulls were more robust than those with rounder vaults and more orthognathic faces. This supports the findings of Lahr and Wright [156] and, more broadly, the results of Mitteroecker and Bookstein [106].

Baab et al. [182] also found evidence for significant positive but weak coexpression among all robust traits (frontal trigone, sagittal keel, infraglabellar notch, supraorbital torus, zygomatic torus and prebregmatic eminence) with the exception of the occipital torus. They did, however, find that the supraorbital torus is strongly and significantly correlated with an infraglabellar notch and moderately and significantly correlated with a zygomatic torus and a zygomatic trigone. Moreover, moderate and significant correlations were found between the zygomatic torus and bregmatic eminence, inferolateral rounding of the orbit, infraglabellar notch, and zygomatic trigone; between a bregmatic eminence and inferolateral rounding of the orbit; infraglabellar notch and zygomatic trigone. They concluded that epigenesis through masticatory function was the most likely cause of the development of these traits.

One feature studied by Baab et al. [182] failed to show correlation with the other robust traits they examined, namely, the occipital torus. Again this seems to be explicable in terms of epigenetic forces during ontogeny. For example, Perez et al. [183] have described a positive association between occipital muscle attachment areas and head shape (dolichocephaly) among males in a sample of South American hunter-gatherer crania. Thus, it may simply be an epigenetic consequence of an elongated cranium, in-line with other changes such as a protruding occipital and formation of an occipital bun [106, 122, 123, 180, 181]. A related character is the suprainiac fossa in WLH50 (Table 1). Functionally, this feature is suggested to provide attachment for nuchal ligament [184] or to represent an area of resorptive bone developing in response to bending forces along the nuchal torus [185]. Irrespective of its cause, this feature is actually found in both WLH3 and WLH50 (personal observation) and cannot, therefore, be considered a defining trait of Australian robusticity nor evidence bearing on their phylogenetic history.

The issue of vault ridges (or keeling) has not been explicitly addressed in most ontogenetic studies of the human cranium. Although Baab et al. [182] did find statistically significant moderate correlations between a sagittal keel and bregmatic eminence, both features are associated with cranial vault sutures. Their development may, therefore, be part of the integrated process of ontogeny and result from epigenetic factors relating to cranial shape and mechanical forces.

The research of Baab et al. [182] also highlights that while integration and modularity are fundamental to understanding cranial ontogeny, certain features sometimes show lower than expected correlation and covariance than expected. Similar findings have been made with respect to the endocranium [186–188].
A number of researchers have specifically addressed the issue of the mechanical effects of mastication on cranial sutures experimentally. In vivo studies [188] found that strain (tension or compression) along sutures was not the result of torsion but rather the localised effects of masticatory actions. Strain was found to be significant along all of the major vault sutures, and its effects varied depending upon the muscle involved. For example, while masseter contraction tensed the coronal suture, the temporal caused compression at this region. Moreover, an earlier study [189] found that even a relatively small bite force was sufficient to cause separation of the sagittal suture in juvenile monkeys, from which they concluded that “such separation might act as a major factor in the local control of osteogenesis of the sutures” [189, page 907]. Additionally, Sun et al. [190] confirmed earlier findings of significant masticatory strain along vault sutures and further found that synostosis (normal fusion) resulted in increased suture strain and enhanced bone growth on the ectocranial surface leading to thicker bones in adults. Finally, the sutures have also been suggested to play a role in dissipating masticatory strain [191], while the diploë may also play a similar role [165]; although these hypotheses require further testing.

Clinical interest in cranial ridges along the major vault sutures stems from understanding the causes and treatment of abnormal head shape as result of congenital or acquired conditions (such as birth injuries) or pathophysiological processes leading to clinical craniosynostosis (premature suture closure). Midline ridges are found within the clinical spectrum of such cranial deformities [192]. However, in the absence of changes to vault or facial shape, they are considered normal variants [193]. The metopic ridge can arise in an infant (before two years of age) from premature closure of the metopic suture, while a midline bony ridge over vertex from front to back may develop following premature fusion of the sagittal suture in adults prior to 40 years of age [192]. Coronal bony ridges can also form in cases of premature craniosynostosis, again arising any time before normal closure of the coronal suture (typically around 40 years of age [192]).

Premature synostosis of one or more sutures is accompanied by compensatory growth, both in other sutures, and by remodelling (appositional growth) of other parts of the skull [194]. Premature closure of sutures prevents separation of the bones and affects skull growth in a direction perpendicular to the affected suture, leading to skull deformations [194, 195]. Much later in life (i.e., around 40 years of age), premature craniosynostosis of the sagittal and coronal sutures can occur and leads in mild cases to the development of sagittal and coronal ridges. Among robust Australian crania, the parietals are relatively long and the lower occipital scale (nuchal plane) is unusually relatively short [151]. Moreover, premature craniosynostosis of the sagittal suture at around 40 years of age can result in dolichocephaly, accompanied by a narrow head with bitemporal widening, sometimes frontal and occipital bossing, and a midline bony ridge over the vertex from front to back [192].

The current clinical incidence of synostosis is 1 in 2,000–3,000 newborns although syndromic craniofacial malformations involving more than one suture occur in about 15% of cases [192]. Sagittal craniosynostosis is rarer, with an incidence of about 1 in 5,000 adults [192]. The aetiology of craniosynostosis is heterogeneous: hereditary, mechanical, teratogenic, and idiopathic [195]. Multiple genetic and environmental causes have been identified. Among the latter are rickets caused by vitamin D deficiency or resistance, chronic renal failure, hypothyroidism and hypophosphatasia, and multiple causes of abnormal fetal positioning in utero leading to constraint of the fetal skull [192]. Rickets is documented among human remains from the Willandra Lakes [153]. Moreover, synostosis may have been more common in the past among some populations. For example, White [196] found an incidence of 3% in a sample of normal Mayan skulls, rising to 31% among those from the precontact period; and in 41% of them premature synostosis was apparently explained by artificial deformation. This lends support to suggestions that artificial deformation may be an exaggerating influence in the robusticity of Pleistocene/Holocene Australians [9, 146–150].

6. Conclusions

The idea that Aboriginal Australians posses an unusual and phylogenetically informative pattern of cranial robusticity has been important in palaeoanthropology for about 150 years. There has, however, been a wide range of views about specific aspects of their evolutionary history. Particularly, whether or not Pleistocene nonmodern populations in Southeast Asia (i.e., *Pithecanthropus, H. soloensis*, or *H. erectus*) played a role in their evolutionary history. Irrespective of whether the multiregional, assimilation, or Out-of-Africa model has been supported, most investigations have taken the human cranium to be optimised part by part, and atomised its morphology into traits assumed to be heritable units, functionally discrete, shaped by natural selection and, therefore, to be positively associated with reproductive success in order to make phylogenetic inferences. Yet, a large body of research from the morphological sciences, including anatomy and embryology, pathology, genetics, and evolutionary biology makes it clear that these assumptions may be unrealistic and that atomisation should be used with great care in phylogenetic reconstruction.

The human (vertebrate) cranium is characterised by ontogenetic and functional modularity, with much integration/covariation among the elements within each module, as well as between them. Moreover, epigenesis is a powerful, pervasive, and very complex force, its influences being felt in all levels of organisation during ontogeny, from cellular interactions to those among adjacent tissues to interactions with the external environment of an organism. Additionally, while many aspects of cranial size and shape are heritable, epigenesis may determine 70% or more of variation seen within populations with well-established pedigrees [79–82]. Finally, many of the traits designated “robust” are also characterised by some ontogenetic integration and are likely to derive from a set of common epigenetic factors, which includes the spatial effects/constraints of size and shape, for
example, central integrating features such as the cranial base and mechanical effects from the muscles of mastication.

While there is clearly a strong message in this research for palaeoanthropology in general, its implications for studies of Australian Aboriginal robusticity are particularly acute. Phylogenetic reconstruction across all organisms and employing all methods involves a set of important assumptions about the traits used. For example, Lieberman [197] lists seven such assumptions, the first four being relatively uncontroversial, with the remaining three particularly relevant here although hardest to satisfy; (5) characters should reflect independent units of information about phylogenetic relationships, (6) they must be heritable if they are to provide reliable information on ancestry and descent, and (7) they should be equivalent units or homologous structures. The ontogenetic approach outlined here suggests that these assumptions probably cannot be met in the case of phylogenetic reconstructions involving robust Australian crania when using atomised traits. Most characters are unlikely to be independent units, being integrated in their development and influenced by a complicated set of epigenetic interactions [77, 83–89, 106–144, 160–167, 169–178, 180–183, 188–197]. They also show mostly low to moderate heritability [79–82] and are in some instances clearly not homologous in modern humans and nonmodern hominins.

For example, a character used widely in studies of the evolutionary origins of Aboriginal Australians is the possession of a flat frontal squama among robust individuals [2, 8, 9, 11–14, 16–18, 39–43, 47–49, 51, 52, 63, 72, 151]. Putting aside epigenetic explanations involving pseudopathology [17, 146–150], this trait fails to satisfy all three assumptions of phylogenetic reconstruction. The size and shape of the growing brain in accordance with the functional matrix hypothesis [77, 88, 114, 115] and its epigenetic interactions with the developing cranial base largely determine the size and shape of the vault. Moreover, there are also influences from the growing face although, they are more subtle than those of the basicranium [122, 123, 131, 132, 136, 138–144]. The anteroposterior length of the face seems also to be a determinant of its position relative to the frontal squama and, therefore, the degree of frontal recession present; crania with high midfacial projection exhibit a long anterior sphenoid, resulting in a viscerocranium positioned anterior to the neurocranium [122, 123, 131–134].

In nonmodern hominins like H. erectus, a flat frontal squama results from a complex combination of a small frontal brain lobe, broad cranial base, combined high midfacial projection and long anterior sphenoid. However, in anatomically modern humans with our large brains, including relatively broad and steep frontal lobe [198], angulation of the frontal squama results from the brain being epigenetically forced to grow excessively posteriorly by a narrow cranial base resulting in a long and tall cranium [17, 156, 180, 181]. Moreover, a flat frontal squama is associated with a browridge, the orbital roof being forced to grow forward to accommodate the growing eye, and an occipital bun, the result of the posterior growth of the brain within a narrow vault. Additionally, because the brain and its frontal lobe are relatively broader and more globular (taller) in modern humans and our face characterised by significant retraction [137, 170], the brow is never as prominent in modern humans as it is in nonmodern hominins including in WLH50 (see Figure 7) [122, 123, 181, 199]. As noted above, WLH50 is among the largest modern humans found in the fossil record globally and is comparable in size to the earliest modern remains such as from Herto (see data in [179]), and larger than many nonmodern crania including H. erectus. While large cranial size and a narrow cranial base, as seen in WLH50, are certainly plesiomorphic traits, they are not shared with H. erectus or other nonmodern hominins but are plesiomorphies of anatomically modern H. sapiens [17, 122, 123]. Moreover, they seem to have been important influences on the form of this calvaria and other Australian remains.

To conclude, there is now ample evidence that the atomisation of cranial robusticity has provided a misleading picture of the evolutionary history of Aboriginal Australians. Moreover, models suggesting that living Australians as demonstrated by their robust cranial morphology can trace their origins to the nonmodern hominins of Pleistocene Southeast Asia should be reconsidered in light of major developments in the morphological sciences. In 1976, Howells [145] issued a challenge to Australian palaeoanthropology to reject a null hypothesis that differences between gracile and robust crania can be explained by phenotypic plasticity alone. So far, this challenge has been ignored. It is time to reconsider the adaptationist program and to take a more parsimonious approach to explaining Aboriginal Australian origins, one that takes account of the complex processes involved in the ontogeny of the human cranium rather than just assuming that natural selection explains every subtle variation seen in past populations.

Acknowledgments

The author wishes to express his deepest gratitude to Alan Thorne for allowing him access to fossilised human remains from the Willandra Lakes in his care and for encouraging critical thought over a decade of friendship and collaboration. He would also like to acknowledge the traditional owners of the Willandra Lakes region who have continued to support scientific research about their history and on the remains of their ancestors. Thanks is extended to Rob Lindemann, Ella Onikul, and the staff of the Radiology Section of the Children’s Hospital Westmead for undertaking X-rays and CT-scans of WLH3 and WLH50 during 2006. James Brink is acknowledged for allowing the author to study fossils in his care and them to be CT-scanned. CT-scans of Florisbad and Holocene San remains were undertaken in 2008 at the Bloemfontein MediClinic; thanks to Leanne Carver, other radiographers, and the nursing staff at Drs Van Dyk and Partners. Fred Spoor is thanked for providing CT-scans of the Sangiran crania. CT-scans for other fossil hominins were purchased from the European Virtual Anthropology Network. The comments of two referees (including Emiliano Bruner) improved the paper. This research was funded by the Australian Research Council under nos. DP0877603 and DP0987985 and by the University of New South Wales.
References


[27] E. Dubois, “Pithecanthropus erectus, eine menschenaehnliche Uebergangs form aus Java,” Batavia, NY, USA, 1894.


S. G. Webb, S. Webb, “Cranial thickening in an Australian Hominid as a
D. Curnoe and A. Thorne, “Human origins in Australia: the
P. Stuart-Macadam, “Cranial thickening and anemia: reply
P. Brown, “Artificial cranial deformation: a component in
M. M. Lahr and R. V. S. Wright, “The question of robusticity
P. Stuart-Macadam, “A radiographic study of porotic
M. C. Westaway, “The Pleistocene human remains collection
M. C. Ravosa, “Ontogenetic perspective on mechanical
F. Weidenreich, “The brain and its role in the phylogenetic
S. L. Washburn, “The relation of the temporal muscle to
S. Wroe, K. Moreno, P. Clausen, C. Mchenry, and D. Curnoe,
[147] P. Brown, “Artificial cranial deformation: a component in
[148] P. Brown, “Nacurrie 1: mark of ancient Java, or a caring
[150] J. L. Clark, S. D. Dobson, S. C. Antón, J. Hawks, K. L. Hunley,
[151] M. C. Raadsheer, C. A. Dobson, R. H. Crompton et al., “Mas-
[152] S. E. Churchill, “Cold adaptation, heterochrony, and
[153] W. A. Weijs and B. Hillen, “Correlations between the cross-
[154] B. Endo, “Experimental studies on the mechanical
[156] M. C. Raadsheer, S. Kiliaridis, T. M. G. J. Van Eijden, F. C. Van


Review Article

Molecular Adaptation of Modern Human Populations

Hong Shi and Bing Su

State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology and Kunming Primate Research Centre, Chinese Academy of Sciences, Kunming 650223, China

Correspondence should be addressed to Hong Shi, shihong@mail.kiz.ac.cn and Bing Su, sub@mail.kiz.ac.cn

Received 11 October 2010; Accepted 14 December 2010

Academic Editor: Darren Curnoe

Copyright © 2011 H. Shi and B. Su. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Modern humans have gone through varied processes of genetic adaptations when their ancestors left Africa about 100,000 years ago. The environmental stresses and the social transitions (e.g., emergence of the Neolithic culture) have been acting as the major selective forces reshaping the genetic make-up of human populations. Genetic adaptations have occurred in many aspects of human life, including the adaptation to cold climate and high-altitude hypoxia, the improved ability of defending infectious diseases, and the polished strategy of utilizing new diet with the advent of agriculture. At the same time, the adaptations once developed during evolution may sometimes generate deleterious effects (e.g., susceptibility to diseases) when facing new environmental and social changes. The molecular (especially the genome-wide screening of genetic variations) studies in recent years have detected many genetic variants that show signals of Darwinian positive selection in modern human populations, which will not only provide a better understanding of human evolutionary history, but also help dissecting the genetic basis of human complex diseases.

1. Origins of Modern Humans

Two schools holding different opinions about where and when modern humans originated have been debating for more than two decades. The “Out-of-Africa” hypothesis of modern human origin was initially proposed in the mid-1980s [1, 2] by studying the genetic variations of mitochondrial DNA (mtDNA), and it was suggested that modern humans first originated in Africa about 200 kya (thousand years ago) then expanded to the other parts of the world and completely replaced the local archaic populations [3]. On the contrary, the alternative hypothesis of “Multiregional Origin”, suggested that modern humans were evolved not only in Africa but also in other continents, for example, East Asia and Europe independently [4], which explains the abundant hominid fossils unearthed outside of Africa and the evolutionary continuity not only in morphological traits but also in the spatial and temporal distributions [5, 6].

The “Multiregional Origin” hypothesis has been supported mostly by archeological data, while the “Out-of-Africa” hypothesis has been supported by both extensive genetic and archaeological evidence [1, 7–15]. By molecular dating, the most recent common ancestor of modern humans emerged only about 100–200 kya ago [8, 10, 13, 16–19], which is consistent with the fossil and culture records of anatomical modern humans firstly appearing in Africa about 150–200 kya ago [17, 20].

The Old World (including Africa, Europe, and Asia) is the area where the multiregionalists claimed independent origins of modern humans [21]. However, besides the numerous genetic studies on current human populations worldwide, genetic data from human fossil samples also supports the “Out-of-Africa” hypothesis. The ancient DNA analysis implies that Neanderthals is not the ancestor of modern humans [22–24], although a recent study suggested potential (but rather limited) gene flow between Neanderthals and modern humans [25].

East Asia is one of the most important regions for studying evolution and genetic diversity of human populations [26]. A great deal of efforts has been put on the origin of East Asian using different genetic markers. The autosomal [7, 9, 15, 27], mtDNA (mitochondrial DNA) [13, 28–33] and Y chromosome [11, 34–38] lines of evidence all support
an African ancestor of modern humans in East Asia. In addition, the genetic data also revealed the prehistoric migratory routes from Africa to East Asia. They first reached the southern part of East Asia (mainland Southeast Asia) [38, 39] then migrated northward to mainland China, Siberia, and eventually to America. There was also an early southward migration from mainland Southeast Asia to Australia and Papua New Guinea [37, 39, 40].

2. Prehistoric Migrations of Modern Humans and Natural Selection

According to the genetic data as well as the fossil records, the exodus of modern humans from Africa only started less than 100 Kya, initially along the coastal line, and then gradually expanded northward to high-altitude areas in the northern hemisphere and southward to Australia, and relatively recently to South America, where they had been facing different environments from their homeland in Africa.

Although the origin and expansion of modern humans is relatively recent, there are extensive variations in modern human populations in view of phenotypic traits (e.g., skin color) as well as genetic make-up (e.g., the diverse Y chromosome haplotypes). At DNA sequence level, most of the variations are considered neutral, that is, doing neither good nor bad, but genetic studies in the past ten years have revealed many examples of molecular adaptations during the prehistoric migrations of modern human populations when they encountered new environments and developed adaptive genetic features through natural selection.

Modern humans have occupied almost all possible environments globally since exiting Africa about 100 Kya. Both behavioral and biological adaptations have contributed to their success in surviving the rigors of climatic extremes including cold, strong UV radiation, and high-altitude, and so forth. Throughout their evolutionary history, human populations have been exposed to myriad pressures from natural selection. The confronted stresses (selective pressures) are the important factors influencing the fate of the variations in the human genetic background. The genetic changes arising from variable metabolic efficiency, resistance to pathogens, and even the preferences of potential mates, and natural selection produce the adaptable phenotypic patterns [41].

The process of natural selection induces an increase or decrease of the frequency of mutations that have an effect on fitness, and the favored heritable traits become dominant in successive generations. The advantageous mutations can rapidly increase in frequency coupled with linked variants (the hitchhiking effect), due to Darwinian positive selection [42–45]. Given enough time, an advantageous mutation eventually would accumulate from a single event to fixation in human populations. When selection is still ongoing, the genetic locus under selection is usually polymorphic, that is, on the way to fixation. In contrast, the deleterious mutations generally keep low frequencies because of their adverse effect on fitness. At the same time, both natural selection and demographic history can cause the departures from a neutral equilibrium. A rapidly expanding population can result in frequency increase of rare polymorphisms, but a population bottleneck would cause the loss of rare variants. Hence, when detecting the molecular signatures of natural selection, we need to consider these confounding factors.

Additionally, analyses on the genetic variation among populations will not only help understand the molecular mechanism of adaptation, but will also aid in identifying the differential susceptibility to diseases, differential responses to pharmacological agents, human evolutionary history, and the complex interactions of genetic variations with environmental factors.

3. Genetic Adaptations of Modern Humans

During the expansion of modern humans from Africa to the other parts of the world, they should have confronted different environmental changes, and the external stresses triggered the selective events during the course of migrations. For example, the heat stress had led to the emergence of hairless skin and more efficient sweating of modern humans, which was suggested as an adaptive feature for efficient hunting during the early history of modern humans in Africa [46].

At the same time, the environmental UV and the low temperature (when modern humans migrated to high-latitude areas) would have imposed additional selection on the hairless skin [47]. One reported example is the latitude-dependent distribution of a sequence polymorphism of the p53 gene. The p53 codon 72 has a common variant from proline to arginine (p53 Pro72Arg), and p53 Arg72 can partially escape the negative regulation by iASPP and lead to a prolonged half life of p53. Hence, p53 Arg72 has more activity than p53 Pro72 [48, 49]. When our ancestors left Africa and started to explore northward to high-latitude regions, the p53 Arg72 was selected by winter temperature and enriched in the populations living in the north. The more active p53 is more efficient in regulating processes such as metabolism (e.g., heat generating) and embryo implantation, which would result in higher fitness levels. Interestingly, this is only one side of the story. p53 has a major negative regulator, the MDM2 gene, counteracting with p53 in a “Yin and Yang” fashion. Mdm2 has a common polymorphism, MDM2 SNP309 (rs2279744), which is located in the MDM2 promoter sequence. A single-nucleotide change from T to G creates a binding site for the transcription factor SP1. As a result, the homozygotes for the G allele express more Mdm2 mRNA and protein compared to the homozygotes for the T allele [50]. Population data in both Europe and East Asia showed that the MDM2 SNP309G/G is selected by low UV radiation exposure. Migration of modern humans northwards to lower UV regions requires less p53 in order to prevent any adverse effects of p53 hyperactivity such as embryonic death. The higher level of Mdm2 is selected to counteract the high p53 levels resulting from the positive selection of the p53Arg72 allele, thereby conferring a selective advantage. This would prevent an overactive p53 pathway, while still enabling p53 to function as a tumor suppressor. The selection ensures the low p53 levels when UV poses no risk to health and survival and also provides a selective advantage, adapting to new environments by keeping...
the balance between p53 and MDM2. The signature of positive selection indicates that selection on p53Arg72 and MDM2 SNP309G/G is strong, leading to the rapid formation of an allele frequency gradient across different latitudes. A well-regulated p53 pathway has been shown to be crucial in many organisms, not only for tumor suppression but also for proper embryonic development and inflammatory responses [47].

Besides the climatic factors modern humans had to cope with during their expansion, there are other factors, such as food, life style, and pathogens, which could also provoke selection. Recently, the genome-wide genetic analyses in East Asian populations have indicated that a considerable number of genes have changed significantly among clusters that are involved in cardiotoxicity, hepatotoxicity, and nephrotoxicity [51]. This suggests that there had been selective forces driving the genes to convey metabolic detoxication to fight with those toxic substances absorbed from new food.

Food and life styles have influenced our genetic makeup, especially during the transition from hunting-gathering to farming societies during Neolithic time (starting about 10 Kya). Several reported good examples include alcohol drinking, starch-food diet, and milk consumption. The studies on human ADH (alcohol dehydrogenase) gene have suggested Darwinian positive selection on the genetic variation of ADH1B His47 [52–54]. Alcohol consumption is a recent agriculture-related life style, and the adaptive selection has begun about 10 Kya [55]. One reported study showed the strikingly high frequency of ADH1BHis47 in southeastern China where the earliest rice domestication originated. Rice has been used as the material to produce fermented food and beverages for a long time in southern China since early Neolithic time. The fermentation helps to preserve and enhance the nutritional value of foods and beverages. However, alcohol can lead to addiction and cause damages to human bodies, including nervous system dysfunction, tumor genesis, innate immune system modulation, and fetal alcohol syndrome. Therefore, genes involved in the ethanol metabolic pathway have become the target of selection when the ethanol-containing food and beverages had been routinely consumed by Neolithic populations in southern China. The systemic population analysis across China indicated that the ADH1B Arg47His sequence polymorphism is under the selection of alcohol consumption. The change of amino acid from Arg to His causes enzymatic activity alteration. The derived allele, ADH1B*47His, changes the pKa of the enzyme from 8.5 to 10.0 which is associated with 40- to 100-fold increase in Km and Vmax of alcohol metabolism. The individuals carrying the ADH1B*47His allele can rapidly metabolize the administered alcohol, therefore, resulting in quick eradication of ethanol, and therefore lower local exposure. On the other hand, the rapid eradication of alcohol also leads to the quick accumulation of aldehyde, the substance which causes flushing when people drink, preventing from overconsumption and possibly addiction. The ADH1B*47His is selected by the ethanol-containing food routinely consumed, and it provides a mechanism to utilize the benefit of fermented food and at the same time eliminating the adverse effects of alcohol [55].

The other two well-studied examples are the copy number variation of amylase gene for starchy food and the regulatory sequence variations of lactase for milk [56–58]. Starch consumption is a characteristic of agricultural societies. This adaptation raises the possibility of selective pressures on amylase, which is responsible for starch hydrolysis. The higher copy number of the salivary amylase gene (AMY1) is identified with positive selection by high-starch diets [58]. Higher AMY1 copy numbers and protein levels would improve the digestion of starchy foods and may buffer against the fitness-reducing effects of intestinal disease.

Lactase persistence (LP) is the ability to digest milk and other dairy products at adulthood, which is the acquired genetic adaptation, independently emerged in northern Europe and Africa. LP is most common in the nomadic tribes that practice pastoralism and is rare in the agricultural populations [59]. The studies indicated that LP is strongly associated with the T allele at a C/T SNP which is located at 13910 bp upstream from the lactase gene (LCT) [60, 61]. The T–13910 haplotype in the LCT promoter can enhance gene transcription [56, 57]. The LCT gene showed the strongest signal of positive selection in northern Europeans [44, 62].

The SNP T–13910 (rs4988235) is only present in a few African populations [63–65], suggesting that the T–13910 allele may not contribute to the lactase persistence in most Africans. A study on a large African population identified three SNPs, which are located within 100 bp of the European LP-associated SNP T–13910. They are significantly associated with the LP trait in African populations [66]. SNP C–14010 is common in Tanzanian and Kenyan pastoralist populations, whereas the SNP G–13915 and SNP G–13907 are common in northern Sudanese and Kenyans. The SNP C–14010, SNP G–13915, and SNP G–13907 in the LAC promoter can enhance gene transcription as the SNP T–13910 does in Europeans [66]. Genotyping across a 3-Mb region in these populations indicated that the African LP-associated variants are different from those in Europeans. This is an example of convergent and adaptive evolution due to the same selective pressure of the cattle domestication and adult milk consumption in Europeans and Africans.

4. Adaptive Evolution and Pathogen Resistance in Modern Humans

An important human adaptation is the ability to taste bitter compounds. The selective advantage of bitter taste can avoid ingesting toxic substances in plants. The systematic analysis of nucleotide and haplotype variations of bitter taste genes identified that these genes had been under natural selection [67]. For example, genetic analyses have detected signatures of balancing selection at the TAS2R38 locus in continental populations [68]. It is hypothesized that individuals with a low sensitivity to the bitter taste of cyanogenic compounds may have a survival advantage against malarial infection through a higher intake of these bitter compounds [69, 70], and the cyanide can inhibit the normal development of the Plasmodium falciparum parasite in vitro [71].

In addition, host genetic variation plays a key role in infectious disease susceptibility of humans. Through
recurrent exposure to pathogens, the genetic adaptations evolve to provide resistance on infection, and the immune systems could be subject to different pressures from bacteria, virus, and other pathogens.

Malaria is a major killer worldwide and its infection is the strongest selective force in recent human evolution. Most genetic adaptation to resistant malarial infection has impact on red blood cells or on immune response. The adaptations include the hemoglobinopathies gene: α-globin (α∗ thalassemia), β-globin variants (HbC and HbS), a number of HLA alleles, and the glucose 6-phosphate dehydrogenase (G6PD) A allele [72–77]. Common hemoglobinopathies, sickle cell trait and α∗ thalassemia, confer partial resistance to the fatal forms of malaria, which presents particularly high frequencies in the populations of the malaria areas in Africa. The disorders of hemoglobin structure, α∗ thalassemia, can protect against malaria infection from the transmission intensity and suggest the selection of the α∗ allele by malaria [73]. The β-globin variants, hemoglobin C (HbC; beta6Glu>Val) or hemoglobin S (HbS; beta6Glu>Val), can stimulate a higher immune response against the malaria antigens, suggesting that both HbC and HbS affect the early development of naturally acquired immunity against malaria [76].

The human leukocyte antigen (HLA) system is highly polymorphic among human populations, and the different components of the disease drive the evolutionary selection on HLA antigens. The selections on a few HLA alleles associated with resistance to malaria infection have been found in different populations, which have the unique HLA antigens prevalent in different malaria endemic areas. The HLA allele B*53 and DRB1*1302 are resistant to malaria in Gambia, Africa [78]. The resistant alleles are prevalent in multiple geographic areas, including the alleles of B*53, and A*2 in Maharashtra, DQB1, 0203, B*35, and B*27 in Mumbai, and A*0211 in New Delhi, India [79–81]. Additionally, the alleles B*56, DRB1*1001 and DPB1*0501 have been found in Thailand, HLA-DQ1*0502 in Vietnam and B*1513 in Malaysia [82–85].

The common polymorphism of G6PD A allele causes G6PD deficiency, which confers resistance to malaria [86–90]. G6PD A allele is associated with about 50% reduced risk of severe malaria in female heterozygotes and in male hemizygotes [91]. The geographical distribution of G6PD deficiency is consistent with evolutionary selection by malaria [87], and analysis of haplotypic structure at the G6PD locus supports the recent positive selection of the A allele [89, 90].

Another well-known example is the recent enrichment (less than 500 years) of mutations in several HIV-infection related genes. For example, the Δ32 mutation at the chemokine receptor 5 (CCR5) gene can defend against HIV virus infection [92].

5. Adaptive Evolution and Human Complex Diseases

Metabolic syndrome (including obesity, Type 2 diabetes, and hypertension) is one of the major human complex diseases, and current prevalence of metabolic syndrome attributes to the evolutionary history of modern humans. The most well-known theory explaining the prevalence of metabolic syndrome in modern societies is the “thrift gene” hypothesis [93]. Before agriculture emerged about 10,000 years ago, modern humans had lived in a lifestyle of hunting and gathering, which would easily lead to frequent cycles of feast and famine. Consequently, the genetic variants had been selected to promote the efficient absorption, storage, and utilization of nutrients in this environment, that are now maladaptive in modern civilization, increasing risk for metabolic diseases [94, 95]. For example, obesity is such a serious disease in modern society, in which agriculture provide exceedingly enough food to people. The adaptation of the efficient absorption, storage and utilization of nutrients would now produce side effect, and it becomes a severe problem to current human populations.

Salt retention is a adaptive character to ancient humans living in hot, humid areas, who consumed low levels of dietary salt [96]. The genetic polymorphisms that promote salt retention were selected in hot and humid environments, but now became a risk factor for hypertension. Genetic adaptations to a low-salt environment in the ancestral populations have turned into a risk for hypertension in current populations living in a high-salt environment [96].

6. Adaptive Evolutions under Extreme Environments

In the recent 100,000 years, modern humans have occupied almost all possible environments, and the adaptations have contributed to their success in surviving the environmental rigors. Among these environmental stresses, high-altitude hypoxia is the only condition in which traditional technology is incapable of mediating its effects. The high-altitude natives including Tibetans and Andeans have developed routes to high-altitude adaptation [97], Tibetans are likely the most adaptive population for high-altitude hypoxia, which is reflected by the elevated resting ventilation and low hemoglobin concentration compared to acclimatized lowlanders. In contrast, Andeans have developed a different strategy by increasing their hemoglobin concentrations and oxygen saturation [97]. The molecular mechanisms of these two high-altitude adaptation routes seem to be different [97]. Recent genome-wide genetic studies suggested that the genes involved in response for hypoxia are under strong natural selection. EPAS1 and EGLN1, the two hypoxia-related genes, were shown under strong natural selection in Tibetans, and there are several candidate sequence variations showing significant association with hemoglobin levels [98–100]. The genetic data suggested that during the long-term occupation of high-altitude areas, the functional sequence variations for acquiring biological adaptation to high-altitude hypoxia have been enriched in Tibetan populations though the detailed molecular mechanism is yet to be revealed.
7. Perspectives

Modern human populations have experienced varied environmental and social transitions since they left Africa about 100 Kya. Current proceedings in genetic studies have revealed that the observed molecular adaptations of modern human are usually regional events, reflecting the history of encountering different environments during the prehistoric migrations outside of Africa. At the same time, the genetic adaptations once developed during evolution may become deleterious at a later stage with the change of environments and life styles. With the rapid progress of genome-sequencing technology, there will be a large quantity of genome-wide variation data across the populations worldwide. More genetic loci will be identified contributing to varied adaptive features of modern human populations, which will not only provide a better understanding of human evolutionary history, but will also aid in dissecting the genetic basis of human complex diseases.

References


