Reconstructing the ancestral phenotypes of great apes and humans (Homininae) using sub-species level phylogenies

Keaghan Yaxley[1] and Robert Foley[2]

[1]Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, UK

[2]Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, UK

Abstract

By their close affinity, the African great apes are of interest to the study of human evolution. While numerous researchers have described the ancestors we share with these species with reference to extant great apes, few have done so with phylogenetic comparative methods. One obstacle to the application of these techniques is the within-species phenotypic variation found in this group. Here we leverage this variation, modelling common ancestors using Ancestral State Reconstructions (ASRs) with reference to subspecies level trait data. A subspecies level phylogeny of the African great apes and humans was estimated from full-genome mtDNA sequences and used to implement ASRs for fifteen continuous traits known to vary between great ape subspecies. While including within-species phenotypic variation increased phylogenetic signal for our traits and improved the performance of our ASRs, whether this was done through the inclusion of subspecies phylogeny or through the use of existing methods made little difference. Our ASRs corroborate previous findings that the Last Common Ancestor (LCA) of humans, chimpanzees and bonobos was a chimp-like animal, but also suggest that the LCA of humans, chimpanzees, bonobos and gorillas was an animal unlike any extant African great ape.

Keywords

African Apes - Ancestral State Reconstruction - Apes - Bayesian - Comparative Biology - Homininae - Human Evolution - Last Common Ancestor - Phylogenetics - Phenotypic Variation

Introduction

Since the time of Darwin and Huxley, the relationship between humans and apes has been a source of both controversy and information about human evolution. Since the establishment of the techniques of molecular evolutionary biology, the phylogenetic relationships have been clarified. This has not led to the closing down of an area of research, but rather has opened up new opportunities. In particular, phylogenetic techniques allow the reconstruction of ancestral states – for example, the nature of the last common ancestor of humans and chimpanzees. However, these techniques are sensitive to the domains of data selected, the amount of within and between taxon variation, and the pattern of evolution across the branches. Here we reconstruct hominin ancestral states across several domains and explore the influence of sub-species patterning among the apes.

The Homininae

Humans, the four species of African great apes - chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*) and the eastern and western gorillas (*Gorilla gorilla* and *Gorilla beringei*) - and their extinct relatives comprise a monophyletic clade usually referred to as the subfamily Homininae. Many studies have compared these species in order to make inferences about the evolutionary history of humans, with a considerable effort being dedicated to describing the Last Common Ancestor of chimpanzees, bonobos and humans ($LCA_{H–P}$). Typically, this has involved referential modelling - treating an extant ape as analogous to the $LCA_{H–P}$ for at least some traits of interest. Chimpanzees tend to be the preferred candidate (Pilbeam & Lieberman, 2017), but bonobos have also been put forward (Parish, De Waal, & Haig, 2006), and others have argued that the fossil species like *Ardipithecus ramidus* point to a more generalised ape, unlike any modern species (Lovejoy, 2009; Lovejoy *et al.*, 2009a,b; White *et al.*, 2009, 2015). Some studies used a more formal phylogenetic method, but did not use Ancestral State Reconstructions (ASRs) (Wrangham, 1987; Foley & Lee, 1989).

ASRs are a class of phylogenetic comparative methods which use the trait data and phylogeny for a set for taxa to estimate the state of a trait in the last common ancestor of those taxa. ASRs are routinely applied to published phylogenies to test macroevolutionary hypotheses (for recent examples see Limeri & Morehouse, 2016; Pereira *et al.*, 2017; De Meester, Huyghe, & Van Damme, 2019; Audino, Serb, & Marian, 2019). Yet only a handful of studies have applied these to the Homininae (Duda & Zrzavy, 2013; Herlyn, 2016; Schroeder & von Cramon-Taubadel, 2017). For example, Herlyn (2016) through maximum parsimony inferred ancestral states for the primates, including the Homininae, based on published species-level, while Duda and Zrzavy (2016) undertook a species-level reconstruction of 65 discrete character traits for all extant apes, using both maximum likelihood and maximum parsimony techniques.

One obstacle to the use of ASRs with the Homininae is the considerable amount of within-species phenotypic variation. For example, between populations chimpanzees are known to vary in mean body size for both male and female individuals (Smith & Jungers, 1997; Grabowski, Hatala, & Jungers, 2018), group size (Furuichi, 2009) and tool use (McGrew, 2010a). Phenotypic variation is particularly well documented for cultural behaviour in the Homininae, with different research sites reporting different suites of putatively cultural tools and behaviours in chimpanzees (Whiten *et al.*, 1999; Sanz & Morgan, 2007; Langergraber *et al.*, 2011), bonobos (Hohmann & Fruth, 2003) and, most recently, gorillas (Robbins *et al.*, 2016). If we were simply to take the species mean these traits we would (1) sacrifice information about the variation in a population and (2) risk attributing a trait value to a species which is not actually represented in any living population.

However, there are ways that these challenges can be overcome. Firstly, ASR methods now exist that can incorporate phenotypic variation into their analysis, allowing researchers to assign multiple trait values to a single tip (Pagel *et al.*, 2004; Felsenstein, 2008; Bruggeman, Heringa, & Brandt, 2009; Goolsby, Bruggeman, & Ané, 2017). Secondly, much of the phenotypic variation reported among the Homininae is found between the genetically and geographically distinct subspecies of both gorillas and chimpanzees. Given this, it is plausible that the problem of phenotypic variation can be overcome by (1) using ASR techniques that explicitly account for it, and (2) treating subspecies rather than species as the operational taxonomic unit. In fact, a few such studies already exist (McGrew, 2010b; Schroeder & von Cramon-Taubadel, 2017). We also expect the inclusion of subspecies to improve the performance of most ASRs as increasing taxon sampling is known to reduce the uncertainty of ancestral state estimates (Alisbury & Kim, 2001).Thus, we propose that the application of phylogenetic comparative methods to the Homininae, may be improved by the inclusion of subspecies-level trait data and a subspecies phylogeny.

A topology for Homininae subspecies

While once a topic of debate, the topology of the Homininae is now well understood (Goodman, 1963; Sarich & Wilson, 1967; Wilson & Sarich, 1969; Sibley & Ahlquist, 1987; Lebedev *et al.*, 2000; Stone *et al.*, 2002; Salem *et al.*, 2003; Thalmann *et al.*, 2006; Becquet *et al.*, 2007; Caswell *et al.*, 2008; Gonder *et al.*, 2011). In particular, molecular studies have untangled the relationships among the four purported subspecies of chimpanzees - Western (*Pan troglodytes verus*), Central (*Pan troglodytes troglodytes*), Eastern (*Pan troglodytes schweinfurthii*) and the Nigeria-Cameroon chimpanzee (*Pan troglodytes ellioti*) (Goldberg & Ruvolo, 1997; Gonder, Disotell, & Oates, 2006; Gonder *et al.*, 2011; Oates, Groves, & Jenkins, 2009; Hey, 2010; Wegmann & Excoffier, 2010; Bjork *et al.*, 2011; Prado-Martinez *et al.*, 2013) - and the four subspecies of gorilla - the Western lowland and cross-river varieties (*Gorilla gorilla gorilla* and *Gorilla gorilla diehli*), and the Eastern lowland and mountain varieties (*Gorilla beringei graueri* and *Gorilla beringei beringei*) (Anthony et al., 2007; Prado-Martinez et al., 2013; Das et al., 2014). While split time estimates for the

Homininae phylogeny are still being debated (Hobolth et al., 2011; Prüfer et al., 2012; Langergraber et al., 2012; Scally & Durbin, 2012; Moorjani et al., 2016a; Moorjani, Gao, & Przeworski, 2016b; Besenbacher et al., 2019) - we now have a well resolved subspecies-level topology with which to implement phylogenetic comparative methods (Figure 1).

In this study we use subspecies-level trait data to implement an ASR of the Homininae. We do this with a view to (1) determining whether including subspecies helps to account for within-species phenotypic variation, and (2) describing the trait states of the Homininae ancestral species ($\text{LCA}_{\text{H-P}}$ and $\text{LCA}_{\text{G-HP}}$), using several continuous traits relating to morphology, life history, sociality, behaviour and ranging.

Methods -

Collating trait data -

This study uses subspecies level trait data to reconstruct the ancestral trait values of the Homininae LCAs. Data were collated for fourteen different continuous traits from 22 different studies published between 1999 and 2016. A description of each of these traits is provided in Table 1, while Table 2 summarises the observations collected for each trait indicating if data were missing for any taxa and the study from which the data were sourced.

Table 1 - Definitions of all of the traits reconstructed in this study. Where appropriate these definitions have been sourced from the relevant literature.

Cultural traits - behavioural traits which have been proposed as putatively cultural because they vary between research sites and meet other criteria (Whiten *et al.*, 1999).

Body size - the average mass of both male and female adults in kilograms (kg).

Effective population size (N_e) - the size of an idealised population that would give rise to the rates of inbreeding and changes in gene frequencies observed in the population of interest (Wang, Santiago, & Caballero, 2016).

Census population size - the estimated number of wild individuals.

Day journey distance - the average distance (km) that individuals travel per day.

Home range - the total area (km^2) in which a group move and live. Typically contains all necessary resources.

Infant mortality - the percentage of infants who die in their first year of life.

Interbirth interval - the average number of months between births for female individuals.

Gestation length - the average number of days between fertilisation and birth.

Age at weaning - the average age in days of juveniles that they cease to wean from their mothers.

Age at first reproduction - the average age in years at which individuals begin their first reproduction.

Community size - the average number of individuals in temporally stable groups, for chimpanzees, bonobos and humans these differ to party size (Lehmann & Boesch, 2004).

Party size - the average number of individuals in temporary subgroups, typically associated with fission-fusion societies such as those of chimpanzees (Lehmann & Boesch, 2004).

Table 2 - The trait observations for Homininae taxa, including the source references. The rows represent each of the traits included in the study, while the columns represent the eleven taxa included in the phylogenetic analysis: *H. sapiens* (Hs), *P. t. verus* (Ptv), *P. t. troglodytes* (Ptt), *P. t. ellioti* (Pte), *P. t. schweinfurthii* (Pts), *P. paniscus* (Pp), *G. g. gorilla* (Ggg), *G. g. diehli* (Ggd), *G. b. beringei* (Gbb), *G. b. graueri* (Gbg) and *Po. abelii* (Poa) . A + indicates that data wer available for the taxa while multiple + indicate the number of sites from which observations were taken, while a - indicates that no data were available for this taxa (continued next page)..

Trait	Hs	Ptv	Ptt	Pte	Pts	Pp	Ggg	Ggd	Gbb	Gbg	Poa
Cultural surveys	+[1*]	++++[2,3]	++[2,4,5]	-	++++++[2,3]	+[6]	+++[7]	-	++[7]	-	++[8]♦
Body size male (kg)	+[1]	+[9]	+[9]	-	+[9]	+[9]	+[9]	-	+[9]	+[9]	+[9]
Body size female (kg)	+[1]	+[9]	+[9]	-	+[9]	+[9]	+[9]	-	+[9]	+[9]	+[9]
Effective population size	+[10]	+[10]	+[10]	+[10]	+[10]	+[10]	+[10]	+[10]	-	+[10]	+[10]
Census population size	+[11]	+[12]	+[12]	+[12]	+[12]	+[12]	+[12]	+[12]	+[12]	+[12]	+[12]
Day journey distance (km^2)	+[11]	++[14]	-	-	++[14]	++[14]	+[14]	-	+[14]	-	+[14]
Home range (km^2)	+[11]	++[14]	-	-	++[14]	++[14]	+[14]	-	+[14]	-	+[14]
Infant mortality 1yr (%)	+[1]	+[13]	-	-	+[13]	+[13]	+[13]	-	+[13]	+[13]	+[13]
Interbirth Interval (months)	+[1]	++[15]	-	-	+++[15]	++[15,17]	+[7]	-	+[15]	+[4]	+[16]
Gestation length (days)	+[18]	+[13]	-	-	++[13]	+[13]	++[13]	-	++[13]	++[13]	++[13]†
Age at weaning (days)	+[1]	+[13]	-	-	+++[13]	+[13]	+[13]	-	+[13]	+[13]	+[13]†

Table 3.2- continued.

Trait	Hs	Ptv	Ptt	Pte	Pts	Pp	Ggg	Ggd	Gbb	Gbg	Poa
Age at first reproduction (years)	+[1]	++[13]	-	-	+++[13]	+[13]	-	-	+[13]	-	+[13]
Party Size	+[1†]	++[19]	+[19]	-	+++++[19]	++[20]	+[21∇]	-	-	-	+[22∇]
Community Size	+[1‡]	++[19]	+[29]	-	++++[29]	++[20]	+[21∇]	-	-	-	+[22∇]

1 - Marlowe (2010)
2 -Whiten *et al.* (1999)
3 - Langergraber *et al.* (2011)
4 - Whiten *et al.* (2001)
5 - Sanz and Morgan (2007)
6 - Hohmann and Fruth (2003)
7 - Robbins *et al.* (2016)
8 - van Schaik *et al.* (2003)
9 - Smith and Jungers (1997)
10 - Prado-Martinez *et al.* (2013)
11 - Marlowe (2005)
12 - IUCN Red List Assessments (Robbins *et al.*, 2008; Bergl *et al.*, 2016; Fruth *et al.*, 2016; Humle *et al.*, 2016; Maisels *et al.*, 2016a,b; Oates *et al.*, 2016; Plumptre *et al.*, 2016; Singleton *et al.*, 2017)
13 - Reichard and Berelli (2008)
14 - Dunbar (2000)
15 - Boesch and Boesch-Acherman (2000)
16 - Shumaker *et al.* (2008)
17 - De Lathouwers and van Elsacker (2005)
18 - Jukic *et al.* (2013)
19 - Wrangham (2000)
20 - Furuichi (2009)
21 - Parnell (2002)
22 - van Schaik (1999)

* No cultural survey available for *H. sapiens.* Hadza tool kit was used as a proxy.
♦ Survey data were also collected for four *Po. pygmaeus* sites
† No data were available for *Po. abelii* so *Po. pygmaeus* was used instead.
† The mean size of female Hadza foraging parties was used (5.31).
∇ The distinction between communities and parties is not meaningful for gorillas and orangutans, therefore the same value was used in both traits.
‡ The mean Hadza camp size was used (30.4).

H. sapiens are a member of the Homininae, and thus it is important that this species is included in our analyses. However, collecting trait data for humans presents a particular challenge compared to other hominids. While gorillas, chimpanzees and bonobos all exhibit some phenotypic and cultural variation between populations, it pales in comparison to the variation found within and between contemporary human societies (Foley & Lahr, 2011). Rather than attempt to capture this variation, we opted to use a single small-scale human society to represent all traits. The Hadza of Tanzania were selected not because they represented the ancestral human state, but because they occur in areas close to those in which extinct hominins lived (Lahr & Foley, 2016). Given their lifestyle and locality, we might expect that the Hazda operate under similar environmental constraints (Marlowe, 2010), at least providing some limits to the scale differences between contemporary humans and the *H. sapiens* lineage the study attempts to reconstruct. Unless otherwise noted, these data were sourced from Marlowe (2005, 2010).

Collecting cultural trait data

Of the data collected, these included putatively cultural traits reported in chimpanzees (Whiten *et al.*, 1999, 2001; Sanz & Morgan, 2007), bonobos (Hohmann & Fruth, 2003), gorillas (Robbins *et al.*, 2016) and orangutans (van Schaik *et al.*, 2003). To define a behavioural trait as cultural, these studies all employed the criteria first developed by Whiten et al. (1999) - behaviours that were present in some communities, but absent from others without an obvious ecological explanation, were defined as cultural. This resulted in a data set of 98 putatively cultural behaviours taken from 24 separate study sites. In order to summarise these data as a continuous variable, we used two different metrics, calculating both for all 24 sites. The first metric, hereafter referred to as the Cultural Count, was simply the number of cultural behaviours expressed at each of the study sites. The second metric was a modification of Shannon's H (Shannon, 1948) designed to measure the cultural diversity of each research site. To make Shannon's H suitable for the cultural data we treated each research site as equivalent to a community, each trait as a species and the frequency scores of those traits (absent, present, habitual, customary) as their abundances. Because the different cultural traits are not comparable between all studies, Shannon's H was calculated across all sites within each genus using the R package Vegan v.2.5-2 (Oksanen *et al.*, 2018),. We also note that because these cultural surveys have not been applied to human populations, we cannot measure cultural diversity for *H. sapiens*, and thus data are missing for this taxon.

Estimating the subspecies level phylogeny

In order to implement a subspecies-level phylogenetic estimate for the Homininae, we took full genome mtDNA sequence data from the Great Ape Genome Project (Prado-Martinez *et al.*, 2013), randomly selecting one sequence for each Homininae species and subspecies, except *G. b. beringei* for which there was no sequence available. We also selected a full genome mtDNA sequence for the Sumatran Orangutan (*Po. abelii*) to act as the out-group in our phylogenetic analyses. Sequence identification and accession numbers can be viewed in the supplementary materials (S2).

Sequences were aligned using MUSCLE (Edgar, 2004) executed via the sequence management software Geneious (Kearse *et al.*, 2012). The final alignment was 15,495 bases long. To account for within-genome rate variation, the alignment was partitioned into non-coding and the first, second and third codon positions of coding regions. PartitionFinder2 (Guindon *et al.*, 2010; Lanfear *et al.*, 2012, 2016) was used to select substitution models for each of the four partitions (full details in supplementary materials, S2).

Phylogenetic inference was implemented under a Bayesian framework using BEAST v.2.4.8 (Bouckaert *et al.*, 2014). To deal with between lineage rate variation we used the lognormal uncorrelated relaxed clock model (Drummond *et al.*, 2006). The clock model was calibrated by the divergence date of the human and chimpanzee-bonobo lineages - the prior for the calibration took the form of a log-normal distribution with a lower hardbound of 6 Myr, a mean of 7 Myr and no upper bound. These parameters encompass the range of split time estimates reported in two recent studies (Moorjani *et al.*, 2016a; Besenbacher *et al.*, 2019). The Markov chain Monte Carlo was set to fifty million generations with trees sampled every fifty thousand generations. The first 20 per cent of these samples were discarded as burnin. We extracted the Maximum Clade Credibility (MCC) tree from our final posterior distribution of 801 trees (Figure 2) and attached a tip for *G. b. beringei* halfway along the terminal branch of its sister subspecies *G. b. graueri*.

Ancestral State Reconstructions

Before any ASRs were conducted, all trait data were transformed using the natural log so as to be expressed on a ratio scale, ensuring that the model would be reconstructing relative changes in trait values rather than absolute changes. ASRs were implemented within a Maximum Likelihood framework using the statistical language and environment R (R Development Core Team, 2008) and the package Phylopars (Bruggeman *et al.*, 2009; Goolsby *et al.*, 2017). Phylopars allows for ASRs even when data is missing for some tips, using stochastic mapping procedures to assign a value to tip based on its phylogenetic position and the overall distribution of trait values. This is preferable to simply pruning tips where data is missing and compounding the effects of incomplete taxon sampling (for discussion see Pybus & Harvey, 2000; Rosenberg & Kumar, 2001; Wiens & Tiu, 2012). Moreover, Phylopars can handle multiple observations of the same trait for a given tip by assuming autocorrelation between those observations, thus allowing for within-subspecies variation to be included in the analyses.

ASRs were implemented under three different evolutionary models - Brownian Motion (BM) Ornstein-Uhlenbeck (OU) and Early Burst (EB). The best fitting model was selected using the Akaike Information Criterion (AIC). From this best fitting model, we then extracted ancestral state values and 95 per cent confidence intervals at each node in the phylogeny and for each trait.

Phylogenetic signal and significance testing

As part of the ASR procedure for each trait we estimated Pagel's λ (1994) using Phylopars. Pagel's λ estimates phylogenetic signal - the tendency for related taxa to express similar traits - by comparing the correlation of traits among taxa to the correlation that expected under Brownian motion ($\lambda = 1$, correlation exactly as expected under a Brownian evolutionary model, and $\lambda = 0$, no correlation). To determine whether the estimated phylogenetic signal was statistically meaningful, we used a loglikelihood-ratio test, comparing the likelihood of the MCC tree to that of a null or star phylogeny (i.e. $\lambda = 0$).

We also took an alternative measure of phylogenetic signal, Blomberg's *K* (2003). Like Pagel's λ, Blomberg's *K* describes the phylogenetic signal of a trait compared to the signal expected under a Brownian motion model of character evolution. Unlike λ, *K* is capable of distinguishing cases where phylogenetic signal is greater than expected under Brownian motion ($K < 1$, signal lower than expected; $K = 1$, signal exactly as expected; $K > 1$, signal higher than expected). Unfortunately, Blomberg's *K* could not be calculated using Phylopars, and thus were there are multiple or missing observations for a tip we use the imputed mean trait values that were calculated by Phylopars to measure λ. The function phylosig from the R package phytools (Revell, 2012) was used to measure Blomberg's K for each of our traits and we tested whether K was significant using the randomisation test described by Blomberg *et al.* (2003), hereafter referred to as Blomberg's test.

We note that Münkemüller *et al.* (2012) demonstrated that both tests have limited power and are susceptible to false-positives when estimated for small trees (< 20 tips). Additionally, while λ and K both measure phylogenetic signal, they do so differently - λ using maximum-likelihood to find the parameter value that best explains the data and K comparing the amount of observed variance to that expected under Brownian motion (Kamilar & Cooper, 2013) - and as such often can produce inconsistent results (Münkemüller *et al.*, 2012). Therefore, as the phylogeny used in this study has eleven tips, we believe it is prudent to only treat traits where both significance tests returned p-values ≤ 0.05 as phylogenetically significant.

In addition to the two measures of phylogenetic signal, we also applied Pagel's δ transformation to each of the traits (Pagel, 1999b,a). Pagel's δ is sensitive to variation in the rate of evolution over time ($\delta < 1$, the overall rate of evolution has slowed towards the tips; $\delta = 1$, rate is constant across the tree; $\delta > 1$, the rate has accelerated towards the tips).

Comparing subspecies- and species-level analyses

To assess the effect of subspecies level trait data on phylogenetic signal, we first created a species level version of our MCC tree. We then calculated λ and K on this species level tree but only for the traits that were found to be phylogenetically significant. We also the applied the best fitting model, determined by the subspecies analysis, to the species tree and extracted the median variance in the ancestral state estimates for every node shared between the two trees.

To assign trait values for these tests, we used two different approaches: (1) for each species' tip we drew a trait value at random from its descendent subspecies - thus to assign male body size to *P. troglodytes*, we randomly selected a value from all of the body sizes reported for the four subspecies. We repeated this process one hundred times for each of the traits; (2) we simply treated all subspecies data as multiple observations for the relevant species. Thus, to assign male body sizes to *P. troglodytes,* we used all of the observations reported for the four subspecies of chimpanzee.

Results

Phylogenetic signal for species versus subspecies trees

For thirteen of the fifteen traits λ was greater than 0.9, suggesting that the MCC subspecies tree was a strong predictor of the traits' distributions. While λ was high in a majority of cases, the loglikelihood-ratio tests found that phylogeny was a significant predictor of trait distribution for only four of the reconstructed traits: male body size (χ^2= 10.75, df = 1, *p*-value = < 0.01) and female body size (χ^2 = 5.92, df = 1, *p*-value = < 0.01), community size (χ^2 = 5.94, df = 1, *p*-value = < 0.01) and gestation length (χ^2 = 4.64, df = 1, *p*-value = 0.03).

For eight of the fifteen traits *K* was greater than one, again suggesting that there was high phylogenetic signal for several of the collated traits. Blomberg's test also found this signal to be significant for all traits except census and effective population size, and infant mortality (for *p*-values see Table S4.1). All of the traits that were found to be significant by the loglikelihood-ratio tests were also found to be significant for Bloomberg's test. Thus, we treat only these traits - male and female body size, community size and gestation length - as phylogenetically significant.

Neither cultural index was found to be phylogenetically significant although signal was high for cultural count ($\lambda = 0.99$; $K = 2$). To explore whether our decision to use the Hadza toolkit as a proxy for cultural count in *H. sapiens* was affecting our analysis (42 tools, twice as large than any other site included in the study), we reran the ASR omitting the Hadza toolkit. Phylogenetic signal ($\lambda = 0.63$; $K = 0.3$) was lower when the Hadza toolkit was omitted and remained insignificant ($\chi^2 = 0.09$; df = 1; *p*-value = 0.76). For a brief discussion of the phylogenetic signal and reconstruction of the cultural indices see the supplementary material, S3.

Model choice was relatively consistent between traits. For the vast majority of traits (twelve of fifteen), Brownian Motion (BM) was found to be the best fitting model with the lowest AIC value. Census size and effective population size were best explained by an OU model of evolution, while community size was the only trait in which an EB model was selected.

For three of the phylogenetically significant traits, δ was found to be greater than one (male body size, $\delta > 2.99$; female body size, $\delta > 2.99$; gestation length, $\delta = 1.4$), while community size was found to lower than one ($\delta = 0.14$). In fact, for male and female body size δ exceeded the maximum value of the parameter space (2.99). Figure 3 shows the δ transformed trees for each of the four significant traits, plotted against the MCC tree. For δ values for non-significant traits see Table S3.1.

Subspecies- compared to species-level analyses

The median value of the permuted species-level λ was considerably lower than the median estimates for the subspecies tree for all traits except community size, while the median value of the permuted species-level K was lower for everything other than female body size (Table

3). Similarly, the subspecies-level estimate of K was higher than the median value of the species-level estimates for all traits except female body size.).

The λ values for the species-level analysis, where all observations were included, were roughly equivalent to those reported in the subspecies analysis, except for gestation length which was considerably lower ($\lambda_{sub} > 0.99$; $\lambda_{species} < 0.01$). In contrast, the K values were much more variable between the two tests, finding similar levels of phylogenetic signal for male body size ($K_{sub} = 0.91$; $K_{species} = 0.95$), but higher signal on the species level tree for female body size ($K_{sub} = 0.41$; $K_{species} = 1.04$), and much lower signal for both gestation length ($K_{sub} = 1.8$; $K_{species} = 0.51$) and community size ($K_{sub} = 4.7$; $K_{species} = 1.34$).

Figure 4 compares the variance of the reconstructed ancestral estimates in both the subspecies- and species-level analyses, for each of the four phylogenetically significant traits. Here we can see that for the $LCA_{G\text{-}HP}$, LCA_{G}, $LCA_{H\text{-}P}$, LCA_{P} and the root of the phylogeny the variance of the subspecies was overall lower than the variance of the permutated species analyses. However, we also see that the variance of the subspecies-level analysis and species-level analysis where all observations were included, are roughly equivalent to one another.

Table 3 - Comparison of phylogenetic signal of subspecies and species level trait data for the traits where λ and K were found to be signficant. Table includes the λ and K when calculated using the subespecies tree (λ_{sub} and K_{sub}), the species tree with all observations (λ_{species} and K_{species}), the median λ and K of the permuted species level analyses (λ_{permute} and K_{permute}) and the 25th and 75th quantiles, and the number of runs where the *p*-value was found to be less than 0.05.

Trait	λ_{sub}	λ_{species}	λ_{permute} (25th %-75th %)	Sig. runs	K_{sub}	K_{species}	K_{permute} (25th %-75th %)	Sig. runs
Male body size	0.97	>0.99	0.17 (<0.01-0.79)	4	0.91	0.95	0.29 (0.18-0.49)	9
Female body size	0.91	>0.99	0.82 (<0.01->0.99)	9	0.41	1.04	0.47 (0.29-0.67)	17
Gestation Length	>0.99	<0.01	<0.01 (<0.01-0.74)	0	1.8	0.51	0.3 (0.24-0.5)	0
Community size	>0.99	>0.99	>0.99 (>0.99->0.99)	13	4.7	1.34	1.46 (1.24-1.68)	96

Ancestral State Estimates

The results of the analysis suggest that the body size of the $\mathrm{LCA_{G–HP}}$ for both males (82.13kgs; 95%CI = 43.03-156.78) and females (51.71kgs; 95%CI = 27.92-95.78) was roughly intermediate to those found in modern African great apes, and larger than any extant chimpanzee. Thereafter, the trends diverge (Figure 5). The male and female body sizes of the gorilla lineage increasing by 8.92kg/Myr and 2.72kg/Myr until the LCA of *Gorilla* ($\mathrm{LCA_{G}}$). In contrast, between the $\mathrm{LCA_{G–HP}}$ and the $\mathrm{LCA_{H–P}}$ the human-chimpanzee-bonobo lineage declined in both female and male body size by 5.44 kg/Myr and 1.69 kg/Myr, respectively. Body sizes estimated for the $\mathrm{LCA_{H–P}}$ were 63.07 (95%CI = 36.86-107.94) kg for male individuals and 45.79 (95%CI = 27.44-76.42) kg for female individuals. While the female value reported here falls at the upper end of the body sizes reported in *Pan* (33.2kg - 45.8kg), the ancestral male size is larger than all the values reported in *Pan* (42.7 kg - 59.7 kg) and even the value reported for *Homo* (53.03kg).

The reconstruction also found that the body sizes of bonobos and the chimpanzee subspecies are highly derived from those of the LCA of bonobos and chimpanzees ($\mathrm{LCA_{P}}$), which had a male body size of 49.33 kg (95%CI = 35.97-67.65) and a female body size of 38.40 kg (95%CI = 28.42-51.89). In particular, the sister subspecies *P. t. troglodytes* and *P. t. schweinfurthii* appear to have diverged rapidly in body sizes for both sexes since their split. This is consistent with the very high δ values reported for both body sizes (>2.99), which suggest that much of the evolution of this trait occurred on the shallowest branches of the phylogeny.

Gestation length is highly clustered by genus, with very little change arising among species and subspecies. The reconstruction suggests that the $\mathrm{LCA_{G–HP}}$ had a gestation length of 255.21 (95%CI = 235.73-276.3) days, intermediate to those found in modern hominids.

Thereafter the gestation length increased in the *Gorilla* lineage (1.66 days/Myr), while the *Homo* and *Pan* lineage declined until the $\mathrm{LCA}_{\mathrm{H\text{-}P}}$ (1.4 days/Myr). The *Pan* lineage continued to decline (2.75 days/Myr) while the *Homo* lineage increased (2.51 days/Myr), suggesting that the longer gestation periods of gorillas and humans evolved convergently.

According to the best fitting model, the majority of change in community size has occurred among the deeper branches of the phylogeny. This is also reflected in the comparatively low δ value calculated for this trait and the fact that this was the only trait where the EB model was preferred over BM. While the *Gorilla* lineage community size declined after the $\mathrm{LCA}_{\mathrm{G–HP}}$ (0.71 individuals/Myr), the *Homo* and *Pan* lineages generally increased towards the $\mathrm{LCA}_{\mathrm{H–P}}$ (2.11 individuals/Myr). This suggests that the large group sizes of *Pan* and *Homo* may have evolved convergently. While the group size recorded for *H. sapiens* is slightly less than those reported for *Pan*, it is worth noting that we use the average Hadza camp site size and that there is considerable variation in this trait for the Hadza, with some camps exceeding 150 individuals (Marlowe, 2010).

Discussion

Phylogenetic signal and taxonomic resolution

We have shown that while phylogenetic signal, measured both λ and K, may be high for many traits, it is only significant for body size, community size and gestation length. Our analyses demonstrate that incorporating subspecies-level trait data, or at least data on phenotypic variation, increased phylogenetic signal and reduced uncertainty in ancestral state estimates. However, it did not matter if this phenotypic variation was analysed using a subspecies level-phylogeny or a method that assumes autocorrelation between multiple within-species observations.

These results should be interpreted cautiously. Phylogenetic signal is a measure of patterning. It can only tell us if the distribution of traits among a group of related taxa conforms or departs from the distribution expected under Brownian motion. Simulation studies have shown that different evolutionary scenarios, with very different dynamics (eg. neutral versus selection) can produce similar measures of phylogenetic signal (Hansen, Pienaar, & Orzack, 2008; Revell *et al.*, 2008). We also note that just because a trait's phylogenetic signal was found to be non-significant by the criteria of this study, does not prove that this trait has no phylogenetic signal. The power of both tests are limited when trees are small (Münkemüller *et al.*, 2012), such that if our analyses were expanded to a larger sample of primates, we may find that other traits were phylogenetically significant. In fact, Kamilar and Cooper (2013) reported significant phylogenetic signal across primates for four of the traits that we report as non-significant in this study (age at weaning, home range, infant mortality and interbirth interval), as well as gestation length and body size. Nonetheless, phylogenetic signal is useful insofar as it allows us to limit our interpretation of our ASRs to those traits where phylogeny was shown to be a significant predictor of their current distribution.

Phylogenetic signal can also be used to interrogate the performance of our subspecies-level analysis compared to our species level analyses. While phylogenetic signal was generally higher on the subspecies level analysis compared to the permuted species level analyses (the exception being community size were estimates remained high on both) the results of the total data species analysis were essentially the same as the subspecies analysis. This was also true of the variance in ancestral state estimates for each of the phylogenetically significant traits.

This suggests that while the inclusion of subspecies-level traits data improved the performance of our ASRs, the inclusion of a subspecies phylogeny made little difference.

This is surprising, as we expected phylogenetic signal to be higher on the subspecies tree because (i) subspecies are similar and only separated by shallow branches, thus conforming to the expectations of Brownian motion, and (ii) increased taxon sampling has been shown to reduce the amount of variance in ancestral state estimates (Alisbury & Kim, 2001). Phylopars handles within-species phenotypic by modelling a layer of variability by assuming autocorrelation between observations reported for a single tip, in our case a single species (Felsenstein, 2008; Bruggeman *et al.*, 2009). If these estimates of autocorrelation correct for structured differences among subspecies, it may mean that functionally they are very similar to a subspecies level analysis. Regardless, our results speak to the importance of including and modelling within-species phenotypic variation in ASRs.

Ancestral state estimates

The reconstruction implemented in this study estimates that the body size of the $\mathrm{LCA}_{\mathrm{G\text{–}HP}}$ was broadly intermediate to those of modern African great apes, while the body size of the $\mathrm{LCA}_{\mathrm{H\text{–}P}}$ fell somewhere around the upper limit of those reported in modern *Pan* lineages. These results are seemingly contrary to the findings of several other studies, both theoretical and empirical, that suggest that the body sizes of the $\mathrm{LCA}_{\mathrm{G\text{–}HP}}$ and $\mathrm{LCA}_{\mathrm{H\text{–}P}}$ were equivalent to modern chimpanzees (Pilbeam, 1996; Richmond & Strait, 2000; Grabowski & Jungers, 2017; Pilbeam & Lieberman, 2017).

Intriguingly our estimates for the body size of the $\mathrm{LCA}_{\mathrm{H\text{–}P}}$ are similar to those estimated for some of the oldest known fossil hominin species. Grabowski (2018) used the scaling relationships between body mass and osteological traits in chimpanzees to estimate a body size average of 45 kg for the hominin genus *Orrorin* and a range of 41.9-59.3 kg for genus *Ardipithecus*. While this study's estimate for male body size is greater than either reported by Grabowski, the ASR also predicts that male body size declined rapidly over the *Homo* lineage, meaning that our estimate may in fact be on track to reach that of the two archaic Hominins. The LCA$_{H–P}$ female body size however, which falls comfortably within Grabowski's estimates, declines at a much slower rate of only 0.28 kg/Myr.

We have to be cautious when interpreting these ancestral state estimates, as they represent a hypothetical value derived from a necessarily simplistic model of trait evolution. A BM model given enough evolutionary time, will tend to produce ancestral values that are the intermediate to those found in the sampled taxa. If the large size of male gorillas was driven by strong positive selection, this rate may not be captured in a reconstruction, leading to overestimated ancestral values. Nonetheless, what these results do show is that even under a deliberately simplistic model, we predict some evolutionary change across the branch connecting the LCA$_{G-HP}$ to the LCA$_{H-P}$.

We should also look to deeper evolutionary history to interpret these findings. While the body size estimates of the LCA$_{G–HP}$ are similar to those reported in extant Ponginae (Smith & Jungers, 1997), like *H. sapiens* the Sumatran, Bornean (*Po. pygmaeus*) and Tapanuli (*Po. tapanuliensis*) orangutans are relics of a once diverse lineage. Ponginae fossil species show considerable body size variation and include the largest known hominoid *Gigantopithecus blacki* (Zhang, Harrison, & Yingqi Zhang, 2017). Additionally, one of the most frequently commented features of the hominid fossil record is the striking variation in the sizes of these fossil species (Pilbeam & Gould, 1974; Jungers & Susman, 1984; Jungers *et al.*, 2016; Pilbeam & Lieberman, 2017; Grabowski *et al.*, 2018) and body size is known to be labile in primates more generally (Smith & Jungers, 1997; Grabowski *et al.*, 2018). If changes in body size are prolific among the Ponginae and the hominins, we might expect that they should also be frequent in the deeper history of the Homininae. Future ASRs of body size in the Homininae could be improved by including data for fossil hominins (Finarelli & Flynn, 2006), however this would require reliable estimates of both the species' body sizes (Grabowski *et al.*, 2015) and their place on the phylogeny (Collard & Wood, 2000).

The results of this analysis suggest that both the *Gorilla* and *Homo* lineages have experienced a convergent increase in the length of gestation, while *Pan* has steadily declined. This pattern is broadly consistent with those found in female body size, and thus some of the change in gestation length may be a consequence of allometry. Kamilar and Cooper (2013), who found strong phylogenetic signal for gestation length across 213 primates species, also suggested that this was a likely consequence of a correlation between life history and body mass. To explicitly measure the evolutionary correlation among these traits, an additional ASR would need to be implemented that modelled the evolution of both traits simultaneously. In *Homo,* selection for larger infant brain size may have also played a role in lengthening gestation (Cunnane & Crawford, 2003)

It is worth noting that while gestation periods are clearly delineated between genera in our dataset, other studies have reported considerable variation in this trait among some Homininae species. For example, Jukic *et al.* (2013) reported that the gestation length range in healthy human births is 37 days (247-284 days), encompassing most of the range of our dataset, and Roof et al. (2005) reported that the mean gestation period of captive chimpanzees was 217.3 days (n = 272 female chimpanzees), approximately eleven days shorter than those reported in our data set. If all species of Homininae showed high phenotypic plasticity for gestation period, then it would be difficult to determine how much of the interspecific differences in this trait are plastic responses to these species living in different environments.

Community size was particularly interesting as it was the only phylogenetically significant traits for which BM was not the best fitting model. Instead, the EB (Harmon *et al.*, 2010), was favoured. EB described an evolutionary scenario where a period of rapid trait evolution is followed by a slowdown or stasis. The ASR itself estimates that the community size of the $LCA_{G\text{-}HP}$ was approximately intermediate to that of extant African great apes. Thereafter, the trait diverges, with the gorilla lineage rapidly evolving smaller community sizes, while the community size of the chimpanzee-bonobo-human lineage increases. This trend for larger communities continues convergently for both the chimpanzee-bonobo and human lineage after they diverge at the $LCA_{H–P.}$

As with body size, it is worthwhile considering an alternative scenario which could produce similar ancestral estimates to those reported here. In this scenario the community size of the $LCA_{G–HP}$ is that of gorillas, or near enough, and after divergence the community size of the gorilla lineage remains relatively unchanged, while the chimpanzee-bonobo-human lineage is subject to strong directional selection for larger communities. The EB model, which is an extended BM model where the magnitude of dispersion decreases through time (Harmon *et al.*, 2010), would not distinguish this scenario from one where the $LCA_{G–HP}$ is intermediate to the extant African great apes.

Conclusion

The aims of this study were to reconstruct the ancestral states of the Homininae for a variety of continuous traits known to vary among subspecies, and to assess the performance of subspecies-level ASRs. To do this, we collated published data on traits from different domains that were known to vary among subspecies, and reconstructed the hominoid evolutionary history using Bayesian phylogenetic inference techniques. The inclusion of phenotypic variation at the levels of species and subspecies, lead to generally higher phylogenetic signal and lower uncertainty for our ancestral state estimates. However, it is not clear whether including a subspecies phylogeny is preferable to using techniques that model autocorrelation among within-species observations. Our best models estimate that for the phylogenetically significant traits the $\mathrm{LCA}_{\mathrm{H\text{-}P}}$ was broadly similar to a chimpanzee, while the $\mathrm{LCA}_{\mathrm{G\text{-}HP}}$ exhibited some important differences, including larger body sizes, a longer gestation period and smaller communities. Future research should focus on including fossil evidence in ASRs as this adds additional evidence regarding the evolution of morphology even behavioural traits (Lister, 2014), and can reduce uncertainty (Finarelli & Flynn, 2006). However, this would require reliable phylogenetic hypotheses regarding the relationship of extant species to extinct ones. Phylogenetic techniques that integrate both molecular and morphological data to reconstruct the evolutionary history - so called ‘Total Evidence Phylogenies’ (Ronquist *et al.*, 2012; Ronquist, Lartillot, & Phillips, 2016; Wood *et al.*, 2013) - represent a promising avenue.

Acknowledgments

We would like to thank Dr Erik Gjesfjeld, Dr Enrico Crema and Dr Robert Attenborough for their advice and guidance. Additionally, we would like to thank the three anonymous reviewers who’s feedback substantially improved this manuscript.

References

Alisbury BAS & Kim J. **2001**. Ancestral State Estimation and Taxon Sampling Density. *Syst. Biol* **50**: 557–564.

Anthony NM, Johnson-Bawe M, Jeffery K, Clifford SL, Abernethy KA, Tutin CE, Lahm SA, White LJT, Utley JF, Wickings EJ & Bruford MW. **2007**. The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 20432–6.

Audino JA, Serb JM & Marian JEAR. **2019**. Ark clams and relatives (Bivalvia: Arcida) show convergent morphological evolution associated with lifestyle transitions in the marine benthos. *Biological Journal of the Linnean Society* **126**: 866–884.

Becquet C, Patterson N, Stone AC, Przeworski M & Reich D. **2007**. Genetic Structure of Chimpanzee Populations. *PLoS Genetics* **3**: e66.

Bergl RA, Dunn A, Fowler A, Imong I, Ndeloh D, Nicholas A & Oates JF. **2016**. <i>Gorilla gorilla ssp.<i> diehli. *The IUCN Red List of Threatened Species 2016*.

Besenbacher S, Hvilsom C, Marques-Bonet T, Mailund T & Schierup MH. **2019**. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nature Ecology & Evolution* **3**: 286–292.

Bjork A, Liu W, Wertheim JO, Hahn BH & Worobey M. **2011**. Evolutionary History of Chimpanzees Inferred from Complete Mitochondrial Genomes. *Molecular Biology and Evolution* **28**: 615–623.

Blomberg SP, Garland T, Jr. & Ives AR. **2003**. Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution* **57**: 717–745.

Boesch C & Boesch-Achermann H. **2000**. *The Chimpanzees of the Taï Forest*. Oxford: Oxford University Press.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A & Drummond AJ. **2014**. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis (A Prlic, Ed.). *PLoS Computational Biology* **10**: e1003537.

Bruggeman J, Heringa J & Brandt BW. **2009**. PhyloPars: estimation of missing parameter values using phylogeny. *Nucleic Acids Research* **37**: W179–W184.

Caswell JL, Mallick S, Richter DJ, Neubauer J, Schirmer C, Gnerre S & Reich D. **2008**. Analysis of Chimpanzee History Based on Genome Sequence Alignments (G McVean, Ed.). *PLoS Genetics* **4**: e1000057.

Collard M & Wood B. **2000**. How reliable are human phylogenetic hypotheses? *Proceedings of the National Academy of Sciences of the United States of America* **97**: 5003–6.

Cunnane SC & Crawford MA. **2003**. Survival of the fattest: fat babies were the key to evolution of the large human brain. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **136**: 17–26.

Das R, Hergenrother SD, Soto-Calderón ID, Dew JL, Anthony NM & Jensen-Seaman MI. **2014**. Complete Mitochondrial Genome Sequence of the Eastern Gorilla (*Gorilla beringei*) and Implications for African Ape Biogeography. *Journal of Heredity* **105**: 846–855.

Drummond AJ, Ho SYW, Phillips MJ & Rambaut A. **2006**. Relaxed Phylogenetics and Dating with Confidence (D Penny, Ed.). *PLoS Biology* **4**: e88.

Duda P & Zrzavy J. **2013**. Evolution of life history and behavior in Hominidae: Towards phylogenetic reconstruction of the chimpanzee–human last common ancestor. *Journal of Human Evolution* **65**: 424–446.

Dunbar R. **2000**. Male mating strategies: a modelling approach. In: Kappeler PM, ed. *Primate Males: Causes and Consequences of Variation in Group Composition*. Cambridge: Cambridge University Press, 259–268.

Edgar RC. **2004**. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**: 1792–7.

Felsenstein J. **2008**. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *The American naturalist* **171**: 713–25.

Finarelli JA & Flynn JJ. **2006**. Ancestral State Reconstruction of Body Size in the Caniformia (Carnivora, Mammalia): The Effects of Incorporating Data from the Fossil Record (T Oakley, Ed.). *Systematic Biology* **55**: 301–313.

Foley RA & Lahr MM. **2011**. The evolution of the diversity of cultures. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**: 1080–1089.

Foley RA & Lee PC. **1989**. Finite Social Space, Evolutionary Pathways, and Reconstructing Hominid Behavior. *Source: Science, New Series* **243**: 901–906.

Fruth B, Hickey JR, André C, Furuichi T, Hart J, Hart T, Kuehl H, Maisels F, Nackoney J, Reinartz G, Sop T, Thompson J & Williamson EA. **2016**. *Pan paniscus* . *The IUCN Red List of Threatened Species 2016*.

Furuichi T. **2009**. Factors underlying party size differences between chimpanzees and bonobos: a review and hypotheses for future study. *Primates* **50**: 197–209.

Goldberg TL & Ruvolo M. **1997**. Molecular phylogenetics and historical biogeography of east African chimpanzees. *Biological Journal of the Linnean Society* **61**: 301–324.

Gonder MK, Locatelli S, Ghobrial L, Mitchell MW, Kujawski JT, Lankester FJ, Stewart CB & Tishkoff SA. **2011**. Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 4766–71.

Gonder MK, Disotell TR & Oates JF. **2006**. New Genetic Evidence on the Evolution of Chimpanzee Populations and Implications for Taxonomy. *International Journal of Primatology* **27**: 1103–1127.

Goodman M. **1963**. Man’s Place in the Phylogeny of the Primates as Reflected in Serum Proteins. In: Washburn SL, ed. *Classification and human evolution*. London: Methuen & Co. Ltd, 204–334.

Goolsby EW, Bruggeman J & Ané C. **2017**. Rphylopars : fast multivariate phylogenetic comparative methods for missing data and within-species variation (R Fitzjohn, Ed.). *Methods in Ecology and Evolution* **8**: 22–27.

Grabowski M, Hatala KG, Jungers WL & Richmond BG. **2015**. Body mass estimates of hominin fossils and the evolution of human body size. *Journal of Human Evolution* **85**: 75–93.

Grabowski M, Hatala KG & Jungers WL. **2018**. Body mass estimates of the earliest possible hominins and implications for the last common ancestor. *Journal of Human Evolution* **122**: 84–92.

Grabowski M & Jungers WL. **2017**. Evidence of a chimpanzee-sized ancestor of humans but a gibbon-sized ancestor of apes. *Nature Communications* **8**: 1–10.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W & Gascuel O. **2010**. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**: 307–321.

Hansen TF, Pienaar J & Orzack SH. **2008**. A Comparative Method for Studying Adaptation to a Randomly Evolving Environment All use subject to JSTOR Terms and

Conditions. *Evolution* **62**: 1965–1977.

Harmon LJ, Losos JB, Jonathan Davies T, Gillespie RG, Gittleman JL, Bryan Jennings W, Kozak KH, McPeek MA, Moreno-Roark F, Near TJ, Purvis A, Ricklefs RE, Schluter D, Schulte II JA, Seehausen O, Sidlauskas BL, Torres-Carvajal O, Weir JT & Mooers AØ. **2010**. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* **64**: 2385–2396.

Herlyn H. **2016**. The phylogenetic system of primates—character evolution in the light of a consolidated tree. *Organisms Diversity and Evolution* **16**: 689–713.

Hey J. **2010**. The Divergence of Chimpanzee Species and Subspecies as Revealed in Multipopulation Isolation-with-Migration Analyses. *Molecular Biology and Evolution* **27**: 921–933.

Hobolth A, Dutheil JY, Hawks J, Schierup MH & Mailund T. **2011**. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome research* **21**: 349–56.

Hohmann G & Fruth B. **2003**. Culture in Bonobos? Between-Species and Within-Species Variation in Behavior. *Current Anthropology* **44**: 563–571.

Humle T, Boesch C, Campbell G, Junker J, Koops K, Kuehl H & Sop T. **2016**. *Pan troglodytes* ssp. *verus*. *The IUCN Red List of Threatened Species 2016*.

Jukic AM, Baird DD, Weinberg CR, McConnaughey DR & Wilcox AJ. **2013**. Length of human pregnancy and contributors to its natural variation. *Human reproduction (Oxford, England)* **28**: 2848–55.

Jungers WL, Grabowski M, Hatala KG & Richmond BG. **2016**. The evolution of body size and shape in the human career. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **371**: 20150247.

Jungers WL & Susman RL. **1984**. Body Size and Skeletal Allometry in African Apes. In: *The Pygmy Chimpanzee*. Boston, MA: Springer US, 131–177.

Kamilar JM & Cooper N. **2013**. Phylogenetic signal in primate behaviour, ecology and life history. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P & Drummond A. **2012**. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence dat. *Bioinformatics* **28**: 1647–1649.

Lahr MM & Foley RA. **2016**. Human Evolution in Late Quaternary Eastern Africa. In: *Africa from MIS 6-2: Population Dynamics and Paleoenvironments, Vertebrate*.

Lanfear R, Calcott B, Ho SYW & Guindon S. **2012**. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular Biology and Evolution* **29**: 1695–1701.

Lanfear R, Frandsen PB, Wright AM, Senfeld T & Calcott B. **2016**. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular Biology and Evolution* **34**: 772–773.

Langergraber KE, Boesch C, Inoue E, Inoue-Murayama M, Mitani JC, Nishida T, Pusey A, Reynolds V, Schubert G, Wrangham RW, Wroblewski E & Vigilant L. **2011**. Genetic and ‘cultural’ similarity in wild chimpanzees. *Proceedings. Biological sciences* **278**: 408–16.

Langergraber KE, Prüfer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, Robbins MM, Schubert G, Stoinski TS, Viola B, Watts D, Wittig RM, Wrangham RW, Zuberbühler K, Pääbo S & Vigilant L. **2012**. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 15716–15721.

De Lathouwers M & Elsacker L Van. **2005**. Reproductive Parameters of Female Pan paniscus and P. troglodytes: Quality versus Quantity. *International Journal of Primatology* **26**.

Lebedev YB, Belonovitch OS, Zybrova N V., Khil PP, Kurdyukov SG, Vinogradova T V., Hunsmann G & Sverdlov ED. **2000**. Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. *Gene* **247**: 265–277.

Lehmann J & Boesch C. **2004**. To Fission or to Fusion: Effects of Community Size on Wild Chimpanzee (*Pan troglodytes verus*) Social Organisation. *Source: Behavioral Ecology and Sociobiology* **56**: 207–216.

Limeri LB & Morehouse NI. **2016**. The evolutionary history of the 'alba' polymorphism in the butterfly subfamily Coliadinae (Lepidoptera: Pieridae). *Biological Journal of the Linnean Society* **117**: 716–724.

Lister AM. **2014**. Behavioural leads in evolution: evidence from the fossil record. *Biological Journal of the Linnean Society* **112**: 315–331.

Lovejoy CO, Latimer B, Suwa G, Asfaw B & White TD. **2009a**. Combining Prehension and Propulsion: The Foot of *Ardipithecus ramidus*. *Science* **326**: 72–72.

Lovejoy CO, Simpson SW, White TD, Asfaw B & Suwa G. **2009b**. Careful climbing in the Miocene: the forelimbs of *Ardipithecus ramidus* and humans are primitive. *Science (New York, N.Y.)* **326**: 1–8.

Lovejoy CO. **2009**. Reexamining Human Origins in Light of *Ardipithecus ramidus*. *Science* **326**: 74–74, 74e1-74e8.

Maisels F, Strindberg S, Breuer T, Greer D, Jeffery K & Stokes E. **2016a**. *Gorilla gorilla* ssp. *gorilla*. *The IUCN Red List of Threatened Species 2016*.

Maisels F, Strindberg S, Greer D, Jeffery K, Morgan DL & Sanz C. **2016b**. *Pan troglodytes* ssp. *troglodytes*. *The IUCN Red List of Threatened Species 2016*.

Marlowe FW. **2005**. Hunter-gatherers and human evolution. *Evolutionary Anthropology: Issues, News, and Reviews* **14**: 54–67.

Marlowe FW. **2010**. *The Hadza: Hunter-Gatherers of Tanzania* (U of C Press, Ed.). Berkley: Univeersity of California Press.

McGrew WC. **2010a**. Chimpanzee technology. *Science (New York, N.Y.)* **328**: 579–580.

McGrew WC. **2010b**. In search of the last common ancestor: new findings on wild chimpanzees. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **365**: 3267–76.

De Meester G, Huyghe K & Van Damme R. **2019**. Brain size, ecology and sociality: a reptilian perspective. *Biological Journal of the Linnean Society* **126**: 381–391.

Moorjani P, Amorim CEG, Arndt PF & Przeworski M. **2016a**. Variation in the molecular clock of primates. *Proceedings of the National Academy of Sciences* **113**: 10607–10612.

Moorjani P, Gao Z & Przeworski M. **2016b**. Human Germline Mutation and the Erratic Evolutionary Clock. *PLOS Biology* **14**: e2000744.

Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, Schiffers K & Thuiller W. **2012**. How to measure and test phylogenetic signal. *Methods in Ecology and Evolution* **3**: 743–756.

Oates JF, Doumbe O, Dunn A, Gonder MK, Ikemeh R, Imong I, Morgan BJ, Ogunjemite B & Sommer V. **2016**. *Pan troglodytes* ssp. *ellioti*. *The IUCN Red List of Threatened Species 2016*.

Oates JF, Groves CP & Jenkins PD. **2009**. The type locality of *Pan troglodytes vellerosus* (Gray, 1862), and implications for the nomenclature of West African chimpanzees. *Primates* **50**: 78–80.

Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens HMH, Szoecs E & Wagner H. **2018**. vegan: Community Ecology Package.

Pagel M. **1994**. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proceedings of the Royal Society B: Biological Sciences* **255**: 37–45.

Pagel M. **1999a**. The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. *Syst. Biol* **48**: 612–622.

Pagel M. **1999b**. Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884.

Pagel M, Meade A, Barker D & Thorne J. **2004**. Bayesian Estimation of Ancestral

Character States on Phylogenies. *Systematic Biology* **53**: 673–684.

Parish AR, De Waal FBM & Haig D. **2006**. The Other “Closest Living Relative”: How Bonobos (*Pan paniscus*) Challenge Traditional Assumptions about Females, Dominance, Intra- and Intersexual Interactions, and Hominid Evolution. *Annals of the New York Academy of Sciences* **907**: 97–113.

Parnell RJ. **2002**. Group Size and Structure in Western Lowland Gorillas (*Gorilla gorilla gorilla*) at Mbeli Bai, Republic of Congo. *American Journal of Primatology J. Primatol* **56**.

Pereira EB, Pinto-Ledezma JN, de Freitas CG, Villalobos F, Collevatti RG & Maciel NM. **2017**. Evolution of the anuran foam nest: trait conservatism and lineage diversification. *Biological Journal of the Linnean Society* **122**: 814–823.

Pilbeam D. **1996**. Genetic and Morphological Records of the Hominoidea and Hominid Origins: A Synthesis. *Molecular Phylogenetics and Evolution* **5**: 155–168.

Pilbeam D & Gould SJ. **1974**. Size and scaling in human evolution. *Science (New York, N.Y.)* **186**: 892–901.

Pilbeam DR & Lieberman DE. **2017**. Reconstructing the Last Common Ancestor of Chimpanzees and Humans. In: Muller MN, Wrangham RW, Pilbeam DR, eds. *Chimpanzees and Human Evolution*. London: The Belknap Press of Harvard University Press, 22–141.

Plumptre A, Nixon S, Caillaud D, Hall JS, Hart JA, Nishuli R & Williamson EA. **2016**. *Gorilla beringei ssp. graueri*.

**Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O’Connor TD, Santpere G, Cagan A, Theunert C, Casals F, Laayouni H, Munch K, Hobolth A, Halager AE, Malig M, Hernandez-Rodriguez J,

Hernando-Herraez I, Prüfer K, Pybus M, Johnstone L, Lachmann M, Alkan C, Twigg D, Petit N, Baker C, Hormozdiari F, Fernandez-Callejo M, Dabad M, Wilson ML, Stevison L, Camprub C, Carvalho T, Ruiz-Herrera A, Vives L, Mele M, Abello T, Kondova I, Bontrop RE, Pusey A, Lankester F, Kiyang JA, Bergl RA, Lonsdorf E, Myers S, Ventura M, Gagneux P, Comas D, Siegismund H, Blanc J, Agueda-Calpena L, Gut M, Fulton L, Tishkoff SA, Mullikin JC, Wilson RK, Gut IG, Gonder MK, Ryder OA, Hahn BH, Navarro A, Akey JM, Bertranpetit J, Reich D, Mailund T, Schierup MH, Hvilsom C, Andrés AM, Wall JD, Bustamante CD, Hammer MF, Eichler EE & Marques-Bonet T. **2013**. Great ape genetic diversity and population history. *Nature* **499**: 471–475.

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

Pybus OG & Harvey PH. **2000**. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings. Biological sciences* **267**: 2267–72.

R Development Core Team. **2008**. R: A language and environment for statistical computing.

Reichard UH & Barelli C. **2008**. Life History and Reproductive Strategies of Khao Yai Hylobates lar: Implications for Social Evolution in Apes. *International Journal of Primatology* **29**: 823–844.

Revell LJ, Harmon LJ, Collar DC & Oakley T. **2008**. Phylogenetic Signal, Evolutionary Process, and Rate (T Oakley, Ed.). *Systematic Biology* **57**: 591–601.

Revell LJ. **2012**. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

Richmond BG & Strait DS. **2000**. Evidence that humans evolved from a knuckle-walking ancestor. *Nature* **404**: 382–385.

Robbins MM, Gray M, Kümpel N, Lanjouw A, Maisels F, Mugisha A, Spelman L & Williamson L. **2008**. *Gorilla beringei* ssp. *beringei*. *The IUCN Red List of Threatened Species 2016*.

Robbins MM, Ando C, Fawcett KA, Grueter CC, Hedwig D, Iwata Y, Lodwick JL, Masi S, Salmi R, Stoinski TS, Todd A, Vercellio V & Yamagiwa J. **2016**. Behavioral Variation in Gorillas: Evidence of Potential Cultural Traits (R Stanyon, Ed.). *PLOS ONE* **11**: e0160483.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL & Rasnitsyn AP. **2012**. A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera. *Systematic Biology* **61**: 973–999.

Ronquist F, Lartillot N & Phillips MJ. **2016**. Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**: 20150136.

Roof KA, Hopkins WD, Izard MK, Hook M & Schapiro SJ. **2005**. Maternal age, parity, and reproductive outcome in captive chimpanzees (Pan troglodytes). *American journal of primatology* **67**: 199–207.

Rosenberg MS & Kumar S. **2001**. Incomplete taxon sampling is not a problem for phylogenetic inference. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 10751–10756.

Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB & Batzer MA. **2003**. Alu elements and hominid phylogenetics. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 12787–91.

Sanz CM & Morgan DB. **2007**. Chimpanzee tool technology in the Goualougo Triangle, Republic of Congo. *Journal of Human Evolution* **52**: 420–433.

Sarich VM & Wilson AC. **1967**. Immunological time scale for hominid evolution. *Science (New York, N.Y.)* **158**: 1200–1203.

Scally A & Durbin R. **2012**. Revising the human mutation rate: implications for understanding human evolution. *Nature Reviews Genetics* **13**: 745–753.

van Schaik CP. **1999**. The Socioecology of Fission-fusion Sociality in Orangutans. *Primates* **40**: 69–86.

van Schaik CP, Ancrenaz M, Borgen G, Galdikas B, Knott CD, Singleton I, Suzuki A, Utami SS & Merrill M. **2003**. Orangutan cultures and the evolution of material culture. *Science (New York, N.Y.)* **299**: 102–5.

Schroeder L & von Cramon-Taubadel N. **2017**. The evolution of hominoid cranial diversity: A quantitative genetic approach. *Evolution* **71**: 2634–2649.

Shannon CE. **1948**. A Mathematical Theory of Communication. *The Bell System Technical Journal* **27**: 379–423.

Shumaker RW, Wich SA & Perkins L. **2008**. Reproductive Life History Traits of Female Orangutans (*Pongo* spp.). In: *Primate Reproductive Aging*. Basel: KARGER, 147–161.

Sibley CG & Ahlquist JE. **1987**. DNA hybridization evidence of hominoid phylogeny: Results from an expanded data set. *Journal of Molecular Evolution* **26**: 99–121.

Singleton I, Wich SA, Nowak M, Usher G & Utami-Atmono SS. **2017**. *Pongo abelii*.

Smith RJ & Jungers WL. **1997**. Body mass in comparative primatology. *Journal of Human Evolution* **32**: 523–559.

Stone AC, Griffiths RC, Zegura SL & Hammer MF. **2002**. High levels of Y-chromosome nucleotide diversity in the genus Pan. *Proceedings of the National Academy of Sciences* **99**: 43–48.

Thalmann O, Fischer A, Lankester F, Paabo S & Vigilant L. **2006**. The Complex Evolutionary History of Gorillas: Insights from Genomic Data. *Molecular Biology and Evolution* **24**: 146–158.

Wang J, Santiago E & Caballero A. **2016**. Prediction and estimation of effective population size. *Heredity* **117**: 193–206.

Wegmann D & Excoffier L. **2010**. Bayesian Inference of the Demographic History of Chimpanzees. *Molecular Biology and Evolution* **27**: 1425–1435.

White TD, Asfaw B, Beyene Y, Haile-Selassie Y, Lovejoy CO, Suwa G & WoldeGabriel G. **2009**. *Ardipithecus ramidus* and the Paleobiology of Early Hominids. *Science* **326**: 64–64, 75–86.

White TD, Lovejoy CO, Asfaw B, Carlson JP & Suwa G. **2015**. Neither chimpanzee nor human, *Ardipithecus* reveals the surprising ancestry of both. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 4877–4884.

Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW & Boesch C. **1999**. Cultures in chimpanzees. *Nature* **399**: 682–685.

Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Wrangham RW & Boesch C. **2001**. Charting cultural variation in chimpanzees. *Behaviour* **138**: 1481–1516.

Wiens JJ & Tiu J. **2012**. Highly Incomplete Taxa Can Rescue Phylogenetic Analyses from the Negative Impacts of Limited Taxon Sampling (CS Moreau, Ed.). *PLoS ONE* **7**: e42925.

Wilson AC & Sarich VM. **1969**. A molecular time scale for human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **63**: 1088–1093.

Wood HM, Matzke NJ, Gillespie RG & Griswold CE. **2013**. Treating Fossils as Terminal Taxa in Divergence Time Estimation Reveals Ancient Vicariance Patterns in the Palpimanoid Spiders. *Syst. Biol* **62**: 264–284.

Wrangham RW. **1987**. The Significance of African Apes for Reconstructing Human Social Evolution. In: Kinzey W, ed. *The Evolution of Human Behaviour: Primate Models*. Albany: State University of New York, 51–71.

Wrangham RW. **2000**. Why are male chimpanzees more gregarious than mothers? A scramble competition hypothesis. In: Kappeler PM, ed. *Primate Males: Causes and Consequences of Variation in Group Composition*. Cambridge: Cambridge University Press, 248–258.

Zhang Y, Harrison T & Yingqi Zhang C. **2017**. Gigantopithecus blacki: a giant ape from the Pleistocene of Asia revisited. *Am J Phys Anthropol* **162**: 153–177.

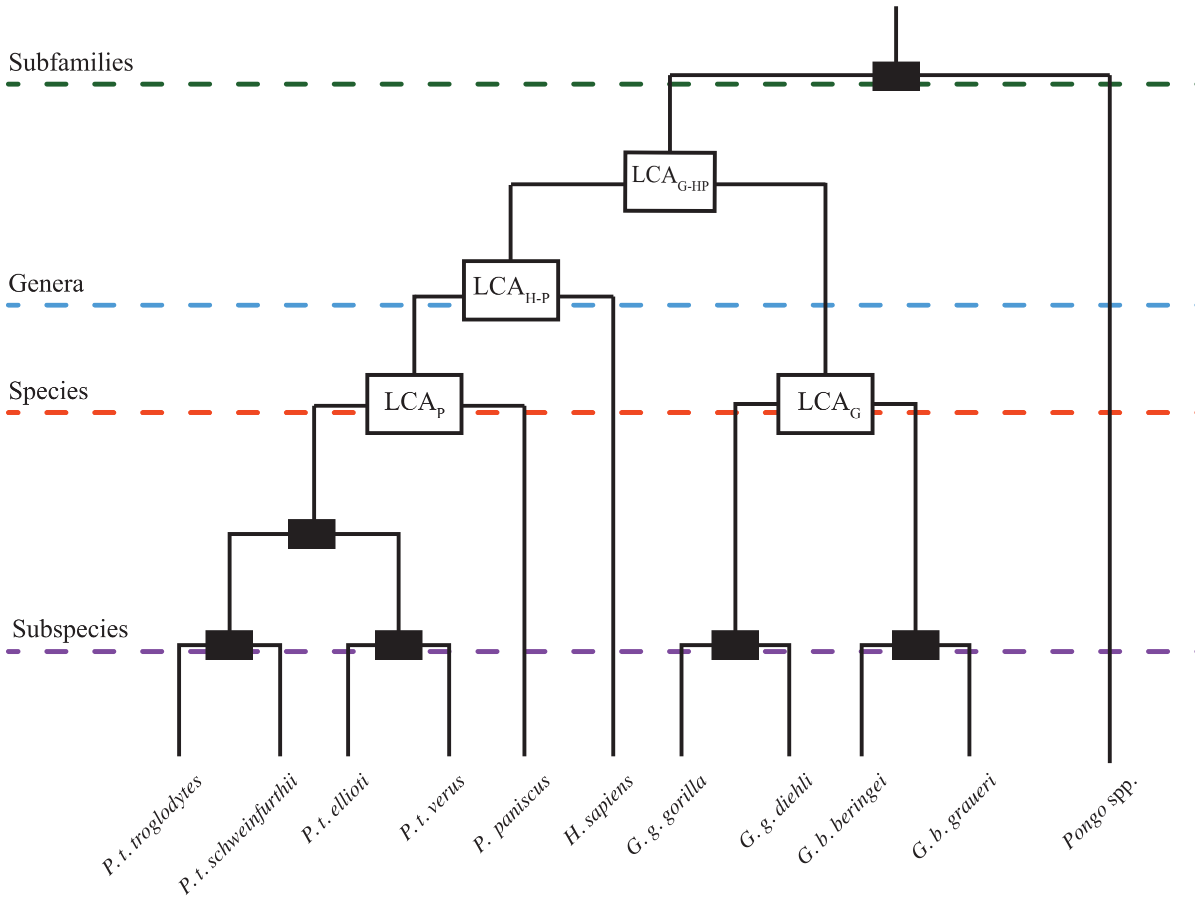

Figure 1. A subspecies level topology of the Hominidae. Each of the LCAs is identified by a white box placed at the relevant node. These include the last common ancestor of the Homininae (LCAG-HP), of humans, chimpanzees and bonobos (LCAH-P), of chimpanzees and bonobos (LCAP) and of eastern and western gorillas (LCAG). The smaller black boxes indicate the locations of common ancestors that were not explicitly discussed in this chapter. The coloured dashed lines define the points at which all lineages of a specified taxonomic level have been established. Thus, by the green line, all subfamilies are established as independent lineages, by the blue line all genera, by the red line all species and by purple line all subspecies. The lengths of the branches are for illustrative purposes and do not reflect any measure of evolutionary distance.

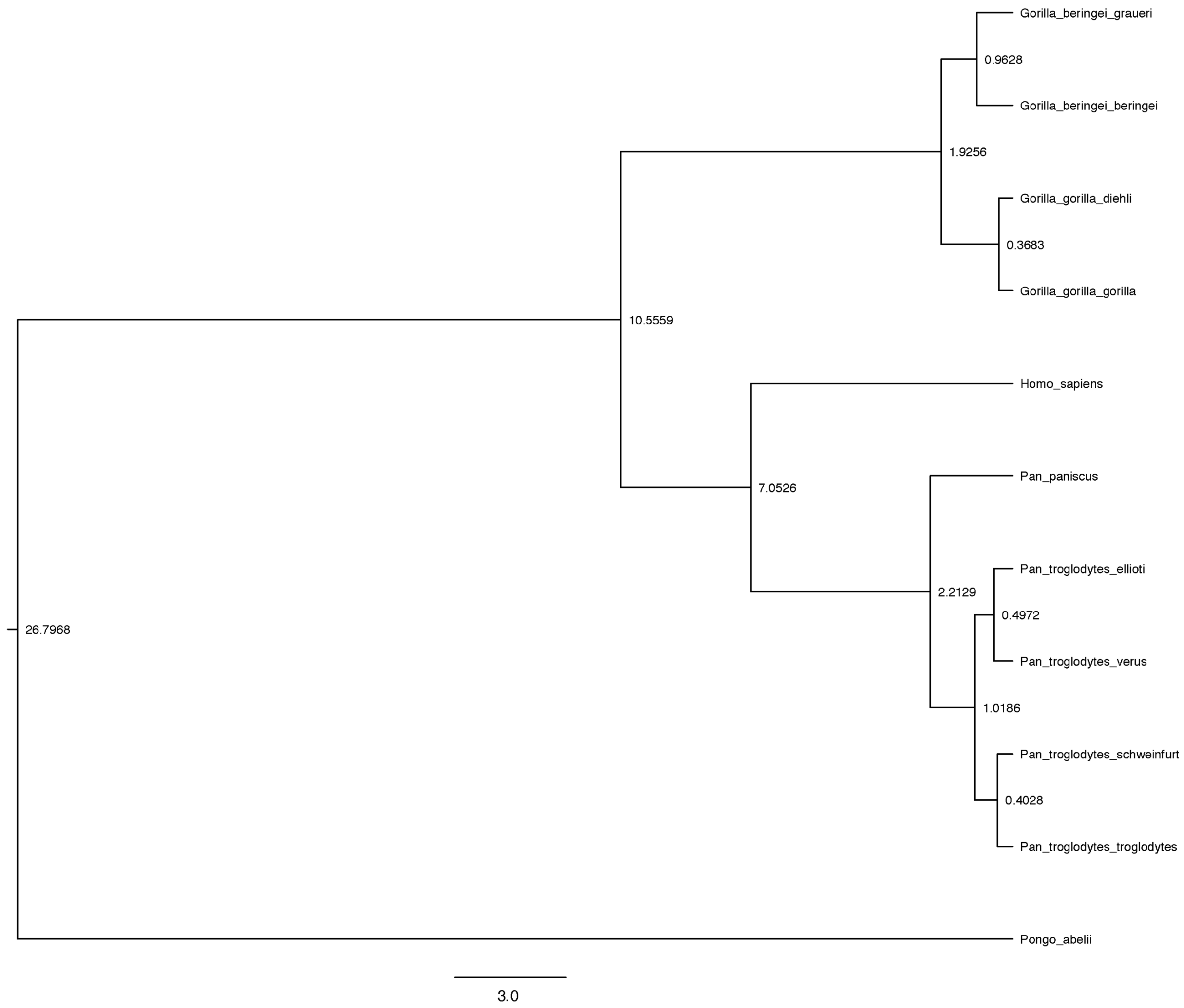

Figure 2. MCC tree extracted from the posterior sample of 801 trees. Estimated divergence times are given for each of the nodes in millions of years ago. A full mtDNA genome was not available for G. b. beringei so its tip was fixed half way along the G. b. graueri tip, thus the divergence date given for the *G. beringei* has not been derived using the molecular clock model and should be interpreted with caution.

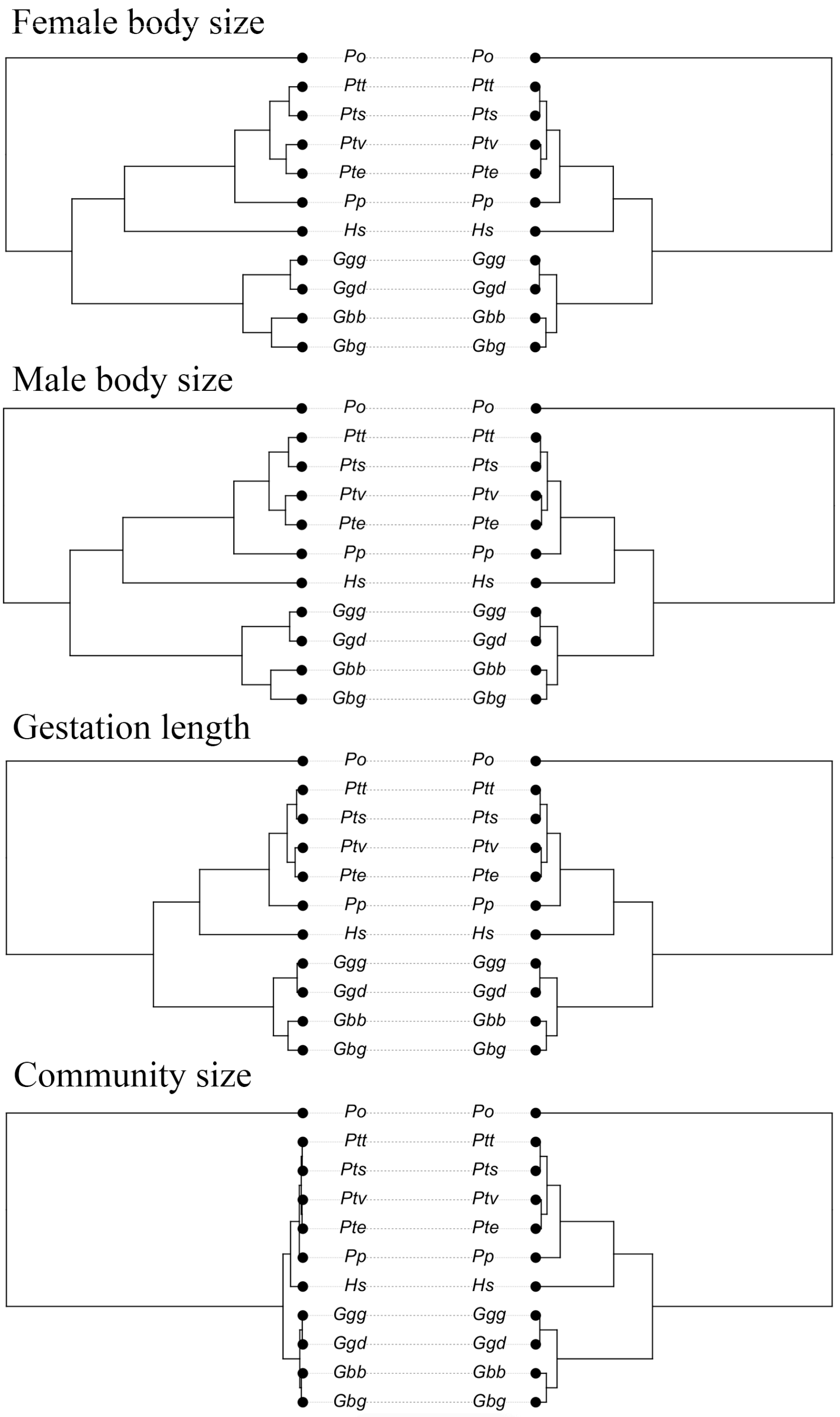

Figure 3. The MCC tree transformed by the δ of each trait (left) compared to the original MCC tree (right). Branch lengths on the transformed trees are now relative to the amount of evolutionary change that occurred along each branch. Here we can see that for body size and gestation length much of the evolution in the traits has occurred on the shallower branches of the tree. In contrast, community size shows that the deeper branches are most evolutionarily significant, as the average community size of the outgroup *Po. abelii* is considerably smaller than any value reported for the Homininae.

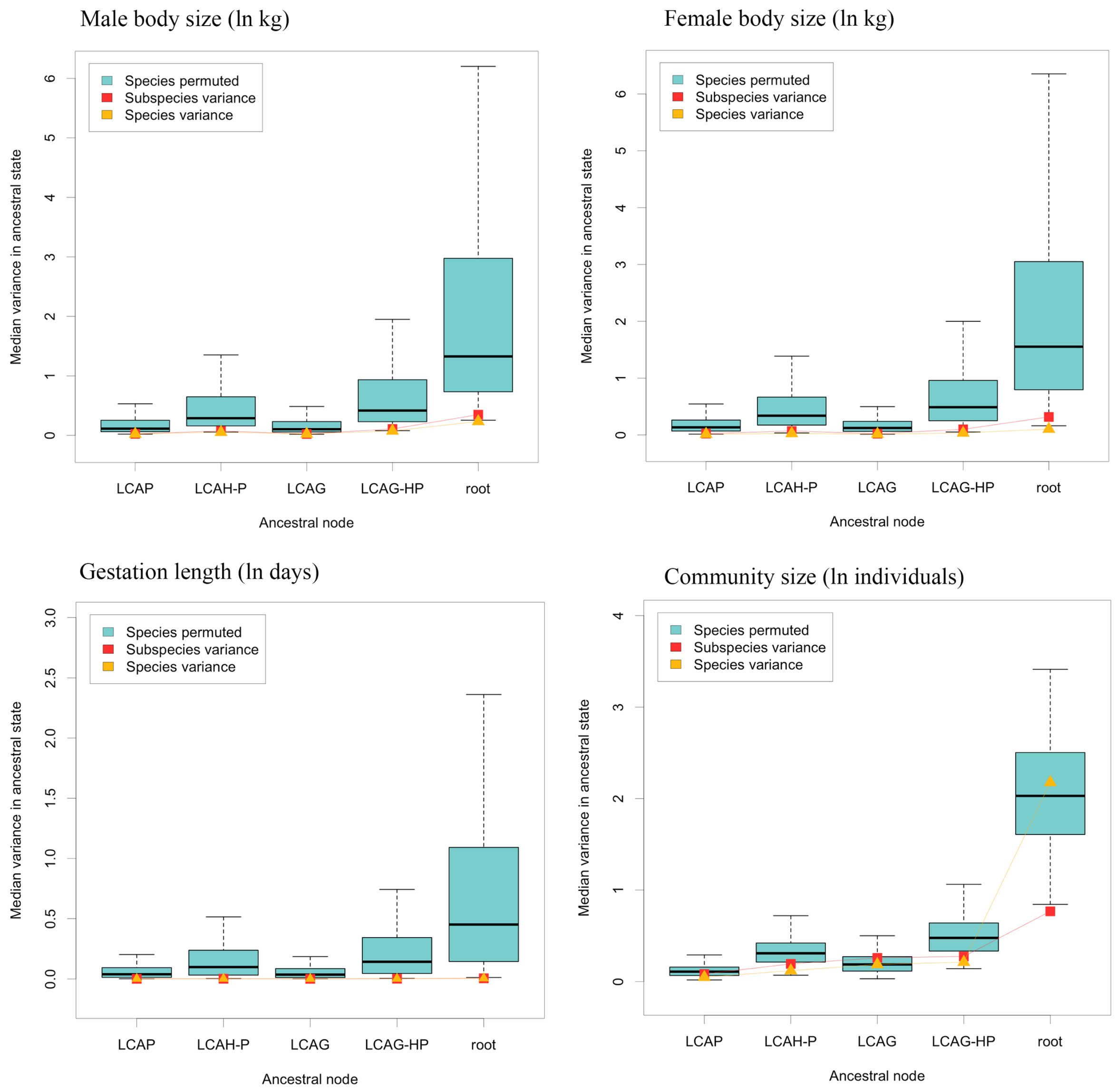

Figure 4. Variance for the permuted species level analysis (blue boxplots), the total data species level analysis (yellow lines and points), and the subspecies analysis (red lines and points). Here we can see that while both the total data species and the subspecies analysis generally outperform the permuted analyses, neither appears to be consistently better than the other.

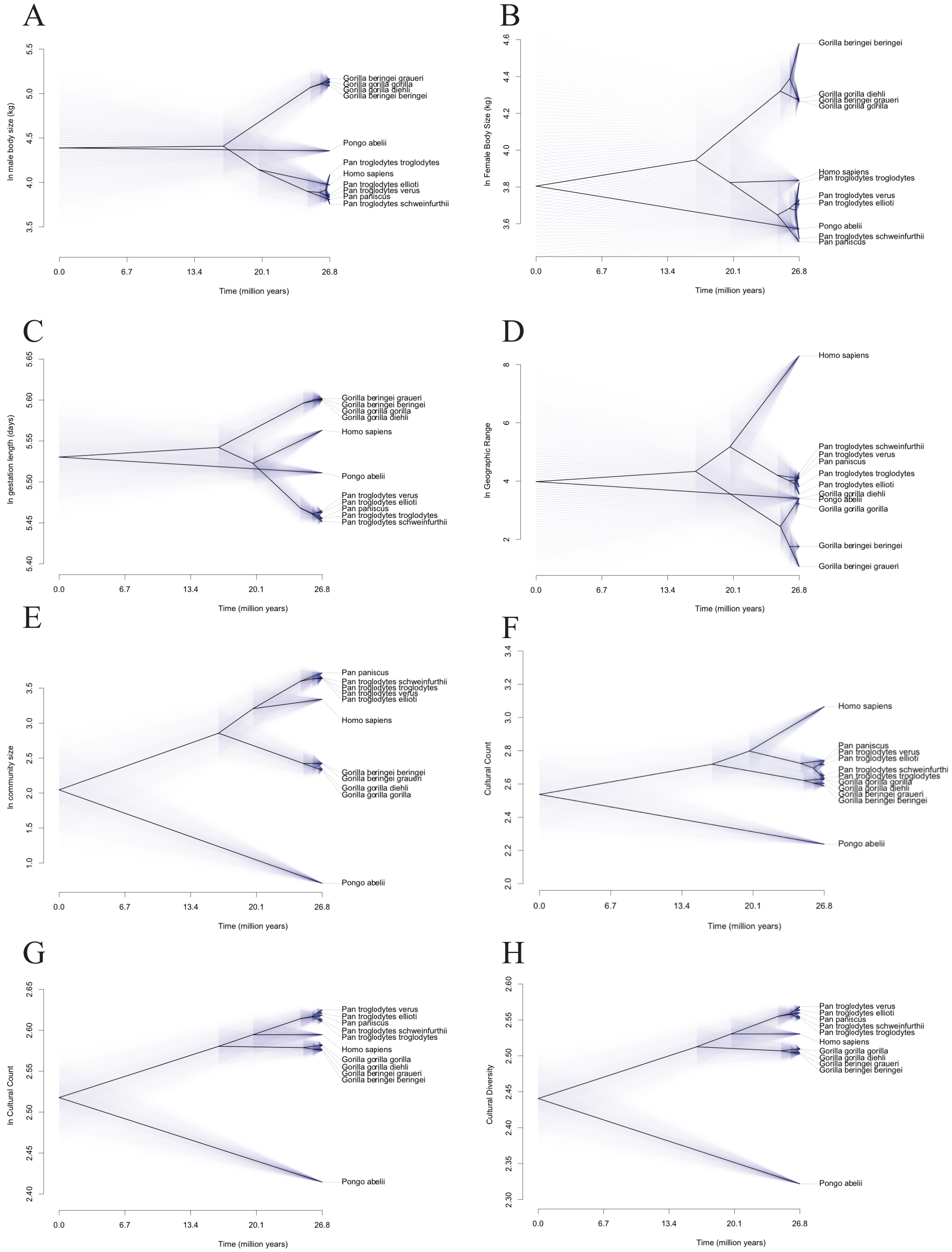

Figure 5. Density phenograms for traits including (A) male body size, (B) female body size, (C) gestation length, (D) geographic range, (E) community size, (F) cultural count, (G) cultural count without Hadza an

Appendices

S1- Phylogenetic reconstruction using BEAST

S1.1 - Files relevant to phylogenetic estimates

The phylogenetic reconstructions involved the use of several programs. To ensure that these procedures are reproducible a number of files have also been made available in the GitHub repository. These include:

- The nexus alignment produced with MUSCLE of the ten mtDNA full genomes sequences (alignment.nex).
- The xml files used to implement the BEAST2 analysis (trimmed_samples_alignment.xml)
- The log file for the BEAST analysis (trimmed_samples_alignment.log)
- All 1001 trees in nexus format (trimmed_samples_tree.trees)

The mtDNA sequences that were originally published in Prado-Martinez et al.[1] were accessed via the website of the Great Ape Genome Program (http://biologiaevolutiva.org/greatape). Table S1.1 includes the identification numbers and individual names for each of the mtDNA sequences used in this study.

Table S1.1 includes the taxa name, sample name and identification number for each of these sequences.

Table S1.1 - Individual names and identification numbers for each of the mtDNA sequences sourced from the Great Ape Genome Project.

Taxa	Individual's Name	Identification Number
G. b. graueri	Mkubwa	9732
G. g. diehli	Nyango	B646
G. g. gorilla	Kowali	9749
P. paniscus	Kumbuka	A928
P. t. troglodytes	Valliant	A957
P. t. verus	Bosco	9668
P. t. ellioti	Akawya-Jean	LWC2
P. t. schweinfurthii	Vincent	10037
Po. abelii	Elsie	A947
H. sapiens	Unidentified (Dai)*	HGDPO1307

* The human sample is only identified by ethnicity.

The raw sequence data are also available through the Sequence Read Archive (SRA) (SRP018689 and PRJNA189439).

S1.2 - Partitioning process

In order to account for genome level substitution rate variation, the mtDNA sequences used in this study were partitioned by the non-coding regions and the 1st, 2nd and 3rd codon positions of the coding regions. Using the alignment management software Geneious v.11.1.2 [2], we assigned these partitions manually, comparing an annotated full human mitochondrial genome (supplied through Geneious, accession number EU54545435) to the alignment of the ten hominid sequences. This allowed us to identify each coding and non-coding region in the mtDNA sequences and annotate them accordingly. We then further partitioned the coding regions by codon position. Here, we assume that patterns of molecular evolution have been similar within each of these partitions.

To select the appropriate substitution models for each of the four partitions we employed the model fitting programme PartitionFinder2 [3–5]. PartitionFinder2 estimates independent models of molecular evolution for each user defined partition. We used PartitionFinder to find the best fitting substitution model for each of the four partitions using the Akaike's Information Criteria (AIC). We only fit substitution models that could be implemented in the phylogenetic reconstruction software BEAST[6]. The executable file for used to parameterises PartitionFinder can be found on the GitHub repository (Partition_finder.cfg).

The best fitting substitution model at all partitions was the Generalized Time Reversible model (GTR) [7]. GTR is one of the most flexible and parameter-rich substitution models, allowing for rates to differ for all transitions and transversions and to change over time. The parameters for each of the partitions' GTR models are summarised in Table S1.2. The output file of PartitionFinder, which describes the best fitting substitution models for each of the partitions is available on the GitHub repository (best_scheme.txt).

Table S1.2 - Best fitting partition scheme according to PartitionFinder

Partition	Best Model	Proportion Invariant	Gama	Starting rates
Non-Coding	GTR	+	+	Estimated
1st Coding Position	GTR	-	+	Estimated
2nd Coding Position	GTR	+	+	Estimated
3rd Coding Position	GTR	+	+	Estimated

1.3 - Parameterising the BEAST analysis

Here we present all parametrisation choices I made for the BEAST. While substitution models were implemented separately for each partition under the models described by PartitionFinder, the tree and molecular clock models were linked for each partition. This forces the BEAST to find the best consensus for both topology and branch lengths based on the information provided by the separate substitutions models. This ensures that BEAST provides one phylogenetic estimate that best reflects the species tree, rather than a separate tree for each of the four partitions.

For each partition, we implemented the substitution model suggested by PartitionFinder. Where suggested, gamma categories were set to four, allowing BEAST to draw from four separate substitution rates for every branch on the phylogeny. The shape of this distribution was estimated by the programme itself. Where suggested, the proportion of invariable sites was automatically estimated for each partition: however, a starting value of 0.2 was given for each. Finally, the rates of specific transversions and transitions (i.e. the substitution matrix) were all estimated by BEAST rather than being derived empirically from the frequency of each nucleotide within the alignment.

We selected the log normal relaxed clock model [6] to make divergence-time estimates. The model was parameterised so that a different rate of molecular evolution could be estimated for every branch on the phylogeny. This was done to account for rate variation among lineages. In order to obtain absolute time scale estimates we calibrated the divergence date of the human, chimpanzee and bonobo lineages. To do so, we assigned a lognormal distribution as a prior from which the calibration could be drawn. We gave this distribution a median value of seven million years, a lower soft-bound of six million years and an upper soft-bound of ten million years, reflecting the spread of dates suggested by previous studies. This meant that for each run of the BEAST analysis a calibration date was drawn from this distribution.

The Monte Carlo Markov Chain (MCMC) was set to fifty million generations. Trees and chain logs were sampled every fifty thousand generations, creating a posterior distribution of 1001 trees. The first two thousand trees sampled were discarded as burnin.

S2 - Analyses scripts

The analyses were implemented in R. The scripts for executing each of these analyses as well as the original data can be accessed via a GitHub repository at: https://github.com/keaghanjames/Reconstructing_the_Homininae

S3 - λ, K and δ values and test statistics

Table S3.1 - For each of the reconstructed traits we present the λ value, the χ^2 and *p*-values of the loglikelihood-ratio test, the *K* value, the *p*-value of the Blomberg's test, and the δ value.

Trait	λ	χ^2	λ *p-value*	*K*	*K p*-value	δ
Age at first reproduction	> 0.99	1.43	0.23	1.44	< 0.01	1.20
Age at weaning	> 0.99	3.44	0.06	1.21	< 0.01	2.55
Female body size	0.91	5.92	0.01	0.41	< 0.01	> 2.99
Male body size	0.97	10.75	< 0.01	0.91	< 0.01	> 2.99
Census population size	< 0.01	< -0.01	1	0.05	0.86	> 2.99
Community size	> 0.99	5.94	0.01	4.78	< 0.01	0.14
Day journey length	> 0.99	2.31	0.13	0.86	< 0.01	1.89
Effective population size	< 0.01	< -0.01	1	0.05	0.77	> 2.99
Gestation length	> 0.99	4.64	0.03	1.82	< 0.01	1.40
Home range	> 0.99	0.69	0.41	0.85	< 0.01	0.85
Infant mortality	0.47	< 0.01	1	0.25	0.04	> 2.99
Interbirth intervale	>0.99	3.81	0.05	1.37	< 0.01	0.50
Party size	>0.99	2.70	0.10	4.47	< 0.01	0.11
Culture count	0.99	0.05	0.81	2.03	< 0.01	0.90
Culture count*	0.63	0.09	0.76	0.34	0.06	> 2.99
Culture diversity	0.86	0.20	0.66	1.40	< 0.01	1.17

* Culture count data without observations for *H. sapiens*

S4 - Reconstruction of cultural traits

Cultural count data were found to have a strong phylogenetic signal ($\lambda = 0.99$; $K = 2$) the λ of cultural diversity was low (0.32), while the K was relatively high (1.41). Moreover, neither trait's λ was found be phylogenetically significant when compared to the null star-phylogeny. While both measures increased in the Homininae after their divergence with Ponginae, the number of cultural traits has declined since the $\text{LCA}_{\text{G–HP}}$ for all lineages except that of *H. sapiens*. From this we might infer that the $\text{LCA}_{\text{G–HP}}$ was a more culturally proficient animal than chimpanzees, bonobos or gorillas. However, this may reflect that the BM model is inadequate when the evolution of a trait is under strong and consistent directional selection. If the *H. sapiens* lineage experienced a high degree of selection for cultural traits, then the rate of evolution along that branch may in fact exceed what can be explained by a BM model. This would cause the BM model to overestimate the rates on all other branches leading to the $\text{LCA}_{\text{G–HP}}$ and give ancestral state estimates which are essentially the intermediate of *H. sapiens* and all other tips on the tree.

This seems a likely explanation as we used the Hadza tool kit as a proxy for the cultural trait count in humans. The size of the kit is 42 tools, larger than any other site included in the study, the next highest being the 26 chimpanzee traits recorded at the northern and southern Taï Forests sites. To investigate the influence of this effect, we reran the analysis, this time omitting the Hadza toolkit. Thus, the *H. sapiens* tip was treated as missing data and assigned a value using the stochastic mapping procedure. This can be thought of as modelling an evolutionary scenario in which the human cultural capacity is unremarkable - they are just another great ape in this respect. Under this model there is a far more consistent trend towards increased cultural counts across the Homininae evolutionary history. However, phylogenetic signal ($\lambda = 0.63$; $K = 0.3$) was weaker compared to the version which included the Hadza toolkit and remained insignificant compared to the star phylogeny ($\chi^2 = 0.09$; df = 1; p-value = 0.76). We note that the trends in cultural diversity and cultural count without the Hadza toolkit are nearly identical. This may reflect that the overall driver in cultural diversity is actually the number of traits, as the simplest way to make a cultural repertoire more diverse is innovation.

References

1. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499,** 471–475 (2013).

2. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence dat. *Bioinformatics* **28,** 1647–1649 (2012).

3. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol. Biol. Evol.* **34,** 772–773 (2016).

4. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Mol. Biol. Evol.* **29,** 1695–1701 (2012).

5. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).

6. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol.* **4,** e88 (2006).

7. Tavaré, S. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lect. Math. Life Sci.* **17,** 57–86 (1986).