The effects of trait redundancy and information content on hominin phylogenetic inference

Authors:

Levi Yoder Raskin[1,2,3 *], Maja Šešelj[3], Bárbara Domingues Bitarello[3 *]

* Corresponding authors: levi_raskin@berkeley.edu and bbitarello@brynmawr.edu.

1. University of California, Berkeley, Berkeley, CA
2. Haverford College, Haverford, PA
3. Bryn Mawr College, Bryn Mawr, PA

Author Contributions: Levi Yoder Raskin: conceptualization, methodology, formal analysis, writing - original draft, review, and editing, visualization. Maja Šešelj: resources, writing - review and editing, supervision. Bárbara Domingues Bitarello: methodology, validation, formal analysis, resources, writing - review and editing, visualization, supervision.

Competing Interest Statement: We are not aware of any competing interests.

Classification: Biological sciences, anthropology

Keywords: Human evolution, character analysis, morphological phylogenetics, fragmentary fossils, parsimony

This PDF file includes:

Main Text
Figures 1 to 5
Tables 1 to 4

Abstract

Paleoanthropological phylogenetic inference is based on characters assumed to be phylogenetically informative and mutually independent, but our understanding of which traits fulfill these requirements is limited. We assess the cladistic information content (CIC) of each trait in a widely-used hominin craniodental character matrix (107 traits) and quantify the degree of unique phylogenetic information contributed by each trait, groups of traits from a given anatomical unit (AU), and operational taxonomic unit (OTU) using the branch and bound (BAB) algorithm. We devise metrics based on permutation of character values along the taxa to quantify the phylogenetic information from a given trait, and removal of individual characters or of an AU to test for redundancy of each trait or an AU, respectively. We quantify the amount of phylogenetic information contributed by a given OTU by measuring the degree of topological change in the inferred tree when that OTU is removed. We use Robinson-Foulds distances to quantify the topological distance between these modified trees to a baseline tree inferred using all traits and taxa. Out of 107 traits, we only identify 23 traits (21.5%) that contribute unique phylogenetic information. Additionally, eight traits are informative but redundant, and thus can be removed from the character matrix without affecting the inferred topology. No single unit contains sufficient information for accurate phylogenetic inference. Because the challenge in identifying phylogenetically informative characters is present throughout paleontological clades, our approach and findings should be more broadly applicable to other paleontological taxa.

Significance Statement

Phylogenetic inference is dependent on the quality of the dataset, but little work has been done assessing hominin morphological character matrices for information content. Here, we present six new generalizable analyses to assess traits, anatomical units, and operational taxonomic units for information content under the parsimony framework. We find that only 21.5% of characters contribute unique information, isolated and fragmentary fossils contain insufficient information for robust phylogenetic inference, and hominin phylogenetic analyses are predicated on the validity of included dubious taxa. The methods we present are applicable to many other fossil clades and may allow for better selection of characters and operational taxonomic units for future systematic analyses.

Introduction

Paleontological phylogenetics endeavors to reconstruct evolutionary relationships in fossil taxa primarily using morphological data, as ancient DNA is not, generally speaking, available for most fossil taxa. The absence of genomic data imposes numerous challenges and introduces uncertainty in our ability to accurately reconstruct evolutionary relationships. Hominin phylogenetics is limited by the sparse nature of the fossil record, further complicated by differing rates of evolution across different anatomical structures — i.e., mosaic evolution (Parravicini and Pievani, 2019)— and homoplasy is potentially widespread (Collard and Wood, 2001; Wood and Constantino, 2007; Haile-Selassie et al., 2015; Ledogar et al., 2022; Gautney, 2023).

Given the sparse nature of the hominin fossil record, it seems reasonable to incorporate as much available morphological data (primarily dental and skeletal) as possible into the initial description and phylogenetic analyses of the material, though doing so biases analyses of the recovered material about the evolution of such traits and their phylogenetic signal — i.e., the tendency for closely-related organisms to look more similar (Blomberg et al., 2003). This circularity seems unavoidable in paleontological phylogenetics: by necessity, the traits of interest (i.e., those used to define taxa) are used in the very construction of the phylogenetic tree and, as such, understanding and interpreting their distributions across a phylogeny will inherently be biased (Strait, 2001; Foley, 2016; Ledogar et al., 2022). A way of circumventing this circularity in fossil-based phylogenetic inference is by relaxing the assumptions about what individual traits imply about hominin tree topology and instead looking at trait value distributions across a group of species using a tree-blind approach.

Several hominin morphological traits demonstrate large amounts of conflicting phylogenetic signals, whereby different traits in the same character matrix suggest different phylogenetic trees (Huelsenbeck and Bull, 1993). Some of this is likely due to homoplasy of masticatory characters (Gautney, 2023). For example, *Paranthropus* species (*P. robustus, P. boisei* and *P. aethiopicus*) possess masticatory structures that some have argued to be homoplasies, raising the hypothesis that this group may be paraphyletic (Collard and Wood, 2001; Wood and Constantino, 2007). While *Paranthropus* species do appear to exhibit shared, derived craniodental features related to increasing bite force (Wood and Constantino, 2007; Strait et al., 2013), those characteristics may have arisen independently in two geographic regions in response to selection for increased bite force (Ledogar et al., 2022). Our understanding of the relationship between the East and South African *Paranthropus (P.*

boisei and *P. aethiopicus* vs. *P. robustus*, respectively*)* is further complicated by the morphology of *Australopithecus africanus*, also from South Africa. *P. robustus* may be more closely related to *Au. africanus* to the exclusion of the East African *Paranthropus* (Wood and Constantino, 2007), but this interpretation is not supported by phylogenetic analyses that favor a single, monophyletic robust australopith genus (Strait and Grine, 2004; Dembo et al., 2015; Dembo et al., 2016; Mongle et al., 2019; Mongle et al., 2023). Furthermore, homoplasy in hominins is not exclusive to masticatory anatomy, as the entirety of the hominin skull may exhibit high levels of homoplasy (Collard and Wood, 2001), as well as conflicting phylogenetic signals that may be partially due to homoplasy (Gautney, 2023).

The mosaic nature of morphological evolution, observed in numerous hominin species, including *H. sapiens*, presents another challenge regarding phylogenetic inference of fossil specimens. This issue is well illustrated by two fossil hominins from South Africa with a seemingly confounding array of ancestral and derived characteristics relative to *Homo*: *Au. sediba* (Berger et al., 2010) and *H. naledi* (Berger et al., 2015). Given that phylogenies built using parsimony rely on character similarities, *Au. sediba* and *H. naledi* are extraordinarily challenging to position in the hominin tree: the former a *Homo*-like *Australopithecus* and the latter an *Australopithecus*-like *Homo* (Berger et al., 2010; de Ruiter et al., 2013; Berger et al., 2015). Fundamentally, mosaic evolution implies a modular organization of the organism with varying rate and direction of change acting on those units (Parravicini and Pievani, 2019). Thus, understanding the influence of anatomical units and individual characters on the resulting phylogenetic hypotheses is critical to understanding mosaicism's effect on parsimony-based phylogenetic inference.

Here, we assess the relative contribution of each trait and morphological unit in a well-established hominin character matrix of 107 craniodental traits (Mongle et al., 2019) to the resulting tree topology and focus on how individual traits, units, and taxa affect phylogenetic placement. We use multiple approaches involving removal and permutation of individual traits and anatomical units from the phylogenetic inference to assess the uniqueness and phylogenetic information contained within these traits. We find that 23 of the 107 traits are phylogenetically informative and not redundant (an additional eight are redundant). We also show that reliable phylogenetic reconstructions limited to isolated anatomical units cannot be inferred using traits from this commonly used dataset.

Materials and Methods

Dataset and Phylogenetic Trees

The dataset we used here (Mongle et al., 2019) contains 107 discrete craniodental characters covering the entirety of the cranium across 20 operational taxonomic units (OTUs), of which 13 are extinct and seven extant (SOM Table S2): *Colobus* (used to root the tree), *Papio*, *Hylobates*, *Pongo pygmaeus*, *Gorilla gorilla*, *Pan troglodytes*, *Sahelanthropus tchadensis*, *Ardipithecus ramidus*, *Au. anamensis*, *Au. afarensis*, *Au. africanus*, *Au. garhi*, *P. aethiopicus*, *P. robustus*, *P. boisei*, *Kenyanthropus platyops*, *Homo rudolfensis*, *H. habilis*, *H. ergaster*, and recent *H. sapiens*. We use the terms "trait" and "character" interchangeably.

All phylogenetic analyses were done in R v4.4.1 (Development Core Team R 2009) using the packages ape (Paradis and Schliep, 2019), TreeTools (Smith, 2019), and phangorn (Schliep, 2011). First, we constructed a phylogenetic tree from Mongle et al. (2019)'s dataset (characters listed in SOM Table S1), referred hereafter as the "baseline tree". We also inferred all test trees, and all tree topologies using the phangorn package's implementation of the branch and bound algorithm (BAB, Hendy and Penny, 1982) — which guarantees returning the most parsimonious (MP) tree(s) and maximum clade credibility to construct the consensus tree. We inferred all trees with the full 20 OTUs and, while all trees are depicted in figures as rooted with *Colobus*, the ingroup topology is the same whether *Colobu*s or *Papio* are used for rooting. All tree distance calculations were performed on unrooted trees and only on the hominin + *Pan troglodytes* subtree. All trees are plotted using the ggtree package (Yu et al., 2017).

Phylogenetic information content and redundancy metrics

We used the cladistic information content (Steel and Penny, 2005) metric (CIC, as implemented in the TreeTools package) as a baseline quantifier of whether a trait has intrinsic phylogenetic information, regardless of other characters and as to the character's fit to any proposed tree. This metric establishes an expected amount of information that can be compared against the observed information content from the removal and permutation analyses by quantifying how many binary trees (i.e., rooted phylogenetic trees with exactly two descendant lineages stemming from each node) are excluded by a single trait by taking the negative natural logarithm of the proportion of binary trees for which the trait of interest is homoplasy-free (Thorley et al., 1998; Steel and Penny, 2005; Cotton and Wilkinson, 2008). For example, for a trait with either a unique or an identical state

for every tip, this proportion is 1 and $CIC = 0$ (i.e., no binary trees are excluded). In contrast, a trait with few states shared across several of the tips would likely have a relatively high CIC. Because CIC is a relative metric with no fixed upper bound, here we discuss a trait's CIC as its rank (CIC_{rank}) in relation to the value for all traits. We resolved ties by using the minimum position: e.g., if two traits have the fourth largest CIC value, they both have $CIC_{rank} = 0.97$ and the fifth largest CIC value has $CIC_{rank} = 0.95$. We also binned traits as defined "low CIC" ($CIC_{rank} < 0.25$), "mid CIC" ($0.25 \leq CIC_{rank} < 0.75$), and "high CIC" ($0.75 \leq CIC_{rank}$).

We devised six complementary approaches to assess the phylogenetic information content of effect of each character, anatomical unit (AU), and OTU on the resulting parsimony-inferred tree, i.e., its "phylogenetically-blind information content". Briefly, we either permuted individual trait values across OTUs, removed traits from the character matrix altogether, or removed taxa from the character matrix. We then inferred the test tree from the modified character matrix, and computed the topological distance between the test tree and the baseline tree according to Robinson and Foulds (1981) as implemented in the ape package. The Robinson-Foulds distance (hereafter referred to as RF) calculates the minimum number of steps from one tree topology to another by repeatedly partitioning a tree and counting the partitions in one tree that are not present in the other (Robinson and Foulds, 1981; Penny and Hendy, 1985).

In the permutation approach, we started with the baseline tree and, for each of the 107 characters (one at a time), permuted the character values across the hominin OTUs by sampling without replacement. We restricted permutations solely to the hominin clade and only measured RF within the hominin +*Pan* subtree. Phylogenetic relationships among extant primates are well-established from genomic data (e.g., Shao et al., 2023); therefore, including extant outgroups in permutations or comparisons may risk identifying traits that are informative on a higher taxonomic level (e.g., across catarrhines) and not necessarily within hominins. We then 1) inferred the MP tree(s) from the modified character matrix using the BAB algorithm and used maximum clade credibility to build the consensus test tree, and 2) computed RF between the baseline and each of the test trees. We repeated this process several times to generate a distribution. To find an optimal number of permutations, we repeated this process with 100, 250 and 500 permutations. The lowest p-value was for the comparison between 100 and 500 permutations for character 41 (MWU $p = 0.059$, SOM Table S4), indicative that 100 permutations may be too low for other, untested

characters, so we used 250 permutations. Permuting a character before inferring the tree allowed us to directly test how variability in that character affects the phylogenetic position of the tips. If a character was allowed to vary across the character matrix and those permutations created a distribution with many changes to the topology of the resulting tree, we concluded that the observed character distribution contains substantial information for the inferred clustering of tips. We call this new approach permutation Robinson-Foulds distance (PRF) and refer to the distribution of these values as the PRF distribution (Table 1). For comparisons across analyses, we use the median of this distribution ($\widetilde{PRF}$) and the range between the 25% and 75% percentiles (interquartile range, PRF_{IQR}) to characterize the spread of the PRF distribution.

For the "removal RF (RRF)" approach (Table 1), for each of the 107 characters, one at a time, we removed them from the character matrix and, following the steps above, we inferred a test tree from the modified character matrix. Here there is only one RRF value per trait rather than a distribution, and this value was compared across traits. We devised RRF to assess to what degree of uniqueness of the phylogenetic information contributed by a character to the baseline tree. We tested for correlation between the different metrics ($CIC,\ \widetilde{PRF},\ RRF$) for each trait using Kendall's rank correlation. as implemented in the *cor.test* function from the R package stats.

Using both PRF and RRF, we classified traits in terms of their phylogenetic information content (not to be confused with cladistic information content, defined below) Briefly, we considered traits with $\widetilde{PRF} = 0$ uninformative, traits with $\widetilde{PRF} \geq 2$ or $RRF \geq 2$ uniquely informative, and traits with $\widetilde{PRF} \geq 2$ and $RRF = 0$ informative, but redundant. $RF = 2$ represents one change in the number of partitions created from the internal branches— i.e., the most minor non-polytomic topological change (Robinson and Foulds, 1981; Penny and Hendy, 1985). For example, the topology of *Paranthropus* in our baseline tree (Figure 1) has *P. aethiopicus* as the outgroup to *P. boisei* and *P. robustus*. If shuffling a trait's values (and leaving all others intact) changed the character matrix, causing *P. boisei* to become the outgroup to *P. aethiopicus* and *P. robustus*, we would have $RF = 2$ and consider the trait as phylogenetically informative. But if that change resulted in a polytomy ($RF = 1$) or did not change the topology at all, we considered that trait as uninformative. Similarly, if the removal of a character did not affect the resulting tree ($RRF = 0$) even though its permutation did ($\widetilde{PRF} \geq 2$), we inferred that the phylogenetic information present

in that character is redundant, meaning the information was replicated in other traits and still present when said trait was removed from the inference.

In an analogous approach to RRF, "unit removal Robinson-Foulds distance" ($URRF$, Table 1), we carried out the same three steps described for RRF but for all traits within an anatomical unit (AU) removed simultaneously. Traits were grouped based on location in the broader cranial structure into one of ten possible AUs (SOM Table S5): frontal ($n = 9$), nasal ($n = 1$), maxilla ($n = 13$), mandible ($n = 15$), dentition ($n = 25$), zygomatic ($n = 7$), temporal ($n = 14$), parietal ($n = 5$), occipital ($n = 9$), and "complex"($n = 9$), which included traits involving more than one bone (e.g. the trait cranial capacity). This approach allowed us to assess the impact on the inferred phylogeny of removing a set of traits which are more likely to not be independent. In a complementary approach ("single unit Robinson-Foulds distance" [$SURF$], Table 1), we instead used solely the traits in a given AU to infer a test tree which was compared to the baseline tree. This approach allowed us to assess the relative weight of these AUs on the resulting hominin phylogeny and to test whether the information contained in isolated fossil finds consisting of a single bony element (e.g., the mandible) is sufficient to resolve any recognizable clades from the baseline tree.

To test whether an AU had an unexpected amount of phylogenetic information, we generated null distributions for SURF and URRF by randomly sampling 100 random sets of n traits from the character matrix, where n is the number of traits in a given AU (or 107 minus that number for URRF analyses) and contrasted our empirical values with their respective distributions. We report two-tailed p-values and assumed $\alpha = 0.05$ for significance testing. For significant cases, we also report the direction of the effect and Bonferroni adjusted p-values for multiple comparisons. We did not report SURF or URRF values for the nasal AU due to its consisting of a single character. We also used this null distribution framework to assess whether the RF between other test trees (the informative traits set and the uniquely informative set) and the baseline trees was unexpected.

Lastly, we used an analogous approach to RRF to measure the effect of the presence of an OTU on the inferred test tree topology ("taxon removal RF, TRRF, Table 1). This approach required extra care to ensure compatibility when comparing our test tree (inferred with 19 OTUs) and baseline tree (inferred with 20 OTUS). For *Colobus*, *Papio*, *Hylobates*, *Po. pygmaeus*, and *G. gorilla*, because we already limit all RF calculations to the 15 OTUs in the hominin + *Pan* subtree, we were able to directly compare the TRRF test tree (inferred with 19 OTUs, 15 included in the RF calculation) with the baseline tree (inferred with 20 OTUs, 15 included in the RF calculation). For

hominins and *Pan*, we dropped those taxa from the baseline tree and then calculated RF. Using *P. boisei* as an example: the TRRF test tree was inferred with 19 OTUs (missing only *P. boisei*), but only 14 of which are used in the RF calculation (*P. boisei* and the taxa that are always removed for the RF calculation: *Colobus*, *Papio*, *Hylobates*, *Po. pygmaeus*, and *G. gorilla*). This test tree was then compared against the baseline tree after dropping the non-*Pan* extant outgroups and the ingroup taxon being assayed here. In this case, the baseline tree is inferred with 20 OTUs, only 14 of which are used to calculate RF.

Table 1. Summary of analyses to quantify the phylogenetic information content and uniqueness of each character or anatomical unit. RF: Robinson-Foulds distance; AU: anatomical unit. [a], following (Cotton & Wilkinson, 2008); [b], higher and lower here are in a relative sense, i.e., compared to other traits (CIC, PRF, RRF, TRRF), units (URRF, SURF), or to a null distribution (URRF, SURF).

Name (acronym)	Metric	Interpretation[b]
Cladistic Information Content (CIC)	CIC: The negative logarithm of the proportion of binary trees for which the trait is homoplasy free.[a]	The *a priori* grouping potential of a trait based on the number of states and their distribution across taxa, regardless of a trait's fit to a tree. Higher CIC: the trait higher potential to contribute cladistic information to an inferred tree.
Permutation RF distribution (PRF)	$\widetilde{PRF}$: the median of the PRF distribution of RF values between the baseline and each permutation test tree.	The degree of phylogenetic information contributed by a trait to the inferred tree. Higher $\widetilde{PRF}$: the trait contributes more information[d]. Higher PRF_{IQR}: greater instability or variance, i.e., new additions to species hypodigms are more likely to change topology.
Trait removal RF (RRF)	RRF: The RF between a baseline and a test tree inferred from a character matrix with one trait removed.	How much the absence of a trait affects the resulting tree. I.e., the degree of unique information contributed by a trait to the inferred tree. Higher RRF: the trait contains more non-redundant (unique) phylogenetic information. Lower RRF: trait either does not contain information or the information it does contain is replicated elsewhere (other traits).
Unit removal RF (URRF)	$URRF$: The RF between a baseline and a test tree inferred from a character matrix with all traits in a given AU removed.	How much the absence of traits in an AU affects the resulting tree. I.e., the degree of unique information contributed by a set of traits to the inferred tree. Higher $URRF$: the unit contains more non-redundant (unique) phylogenetic information. Lower RRF: trait either does not contain information or the information it does contain is replicated elsewhere (other traits).
Single unit RF (SURF)	$SURF$: RF between a baseline and a test tree inferred solely from characters in a given AU.	The unique phylogenetic information content in an anatomical unit compared to an equally sized random set of traits. Higher SURF: AU alone leads to inferences increasingly farther from the baseline tree.
Taxon removal RF (TRRF)	$TRRF$: RF between a baseline tree and a test tree inferred missing each taxon one at a time.	The information contributed by each taxon to the resolved topology. Higher TRRF: the taxon contributes more information to the character matrix.

Results

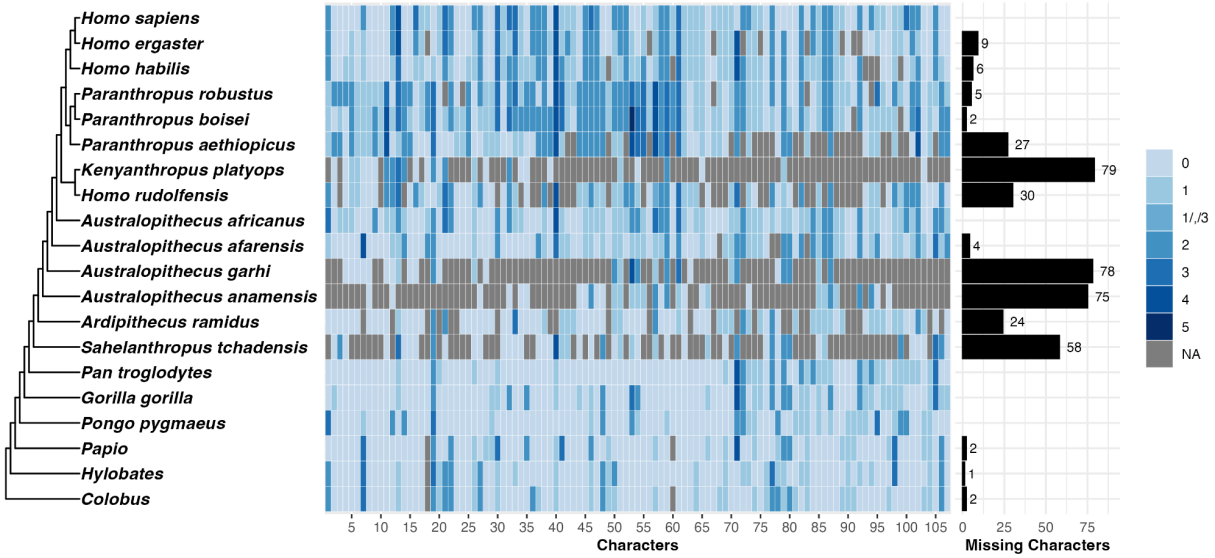
Inferring the baseline tree

The dataset compiled by Mongle et al. (2019) consists of 20 OTUs (SOM Tables S1-S3), including 13 hominin fossil taxa, modern humans, and six non-human extant catarrhine primates. This dataset is limited by the amount of missing data in the fossil record. *K. platyops* (79 characters missing), *Au. garhi* (78 characters missing), *Au. anamensis* (75 characters missing), and *S. tchandensis* (58 characters missing) are particularly affected (Figure 1, SOM Table S1). Only three of the 107 traits have data from all extinct hominin OTUs: reduced canines (trait 51), molar crown area (trait 54), and relative enamel thickness (trait 59) (Figure 1, SOM Table S3). Because this dataset — and the hominin record in general — are sparse, understanding which characters are redundant or altogether devoid of information may help maximize the amount of relevant phylogenetic information gained from a scant fossil record.

Using all 20 OTUs and 107 traits, we inferred 3 MP topologies (parsimony score 386) and their consensus tree, hereafter referred to as the baseline tree (Figure 1,). We compared the hominin + *Pan* baseline subtree topology to the parsimony-inferred tree from Mongle et al. (2019) and found that they are broadly similar ($RF = 6$). Our baseline tree grouped *H. rudolfensis* and *K. platyops* as sister taxa and positioned them as a sister clade to the *Homo-Paranthropus* clade, as opposed to Mongle et al. (2019)'s parsimony-derived tree (parsimony score 390), which places *K. platyops* as the outgroup to *Homo*, with *H. rudolfensis* ancestral to *H. habilis*, *H. ergaster*, and *H. sapiens*. Secondly, our baseline tree inferred *Au. garhi* as directly ancestral to *Au. afarensis*, whereas in Mongle et al. (2019) *Au. afarensis* is directly ancestral to *Au. garhi*.

Our goal here was not to infer the putatively "most correct" hominin tree. Rather, we explored how different hominin craniodental characters affect phylogenetic hypotheses (i.e., trees). Indeed, all our tests are conditioned on this baseline tree, and the baseline tree topology does affect the magnitude of RRF and $\widetilde{PRF}$. Using Mongle et al., 2019's parsimony tree topology as a baseline, values were different, as expected, but highly correlated ($p < 0.001$ in both cases, SOM Figure S1). Our baseline tree represents a reasonable hypothesis which, since all test trees were inferred identically, should enable relative comparison in information between characters.

Figure 1. Baseline tree for 20 OTUs. Left, a consensus tree topology (cladogram) inferred using the branch and bound algorithm with maximum clade credibility. Center, heatmap showing the coding for each taxon (rows) and character (columns). Grey indicates missing data. Characters are ordered sequentially (see SOM Table S1 for details). Right, bar plot showing the number of missing characters for each taxon ($n = 107$ characters). The tree is rooted with *Colobus*.



Cladistic information content (CIC)

One of our key goals was to characterize the degree of phylogenetic information content of individual traits and sets of traits. While the terms cladistic information content and phylogenetic content are occasionally used interchangeably in the literature — with both referring to the cladistic information content (Steel and Penny, 2005) — we make an explicit distinction between them for the purposes of our study. CIC is a function of the number of states and how many OTUs carry each state, but bears no relationship to how the trait values are distributed across the proposed tree (Steel and Penny, 2005). On the other hand, our RRF and PRF approaches were envisioned as ways to capture the degree of phylogenetic information a trait contributed to the proposed phylogeny, and as such are inherently dependent on the distribution of trait values across the OTUs. Throughout, we use CIC values as a measure of the "potential" of a trait to harbor phylogenetic information; thus, a trait with $CIC = 0$ should not be expected to have any phylogenetic information content (as measured by our approaches), whereas a trait with relatively high CIC has the potential to contain some degree of phylogenetic information content.

We find that all but one of the traits have at least some degree of variation as captured by CIC. The CIC values across the 107 traits are normally distributed ($p = 0.229$, Shapiro-Wilk's normality test, $\alpha = 0.05$) with mean of 12.673, standard deviation of 5.297 (Figure 2A), and ranging from 0 (character 83, "ossification of the petrous aspect", coded as 0 in *H. sapiens* and 1 or "?" [missing] in all other OTUs) to 25.841 (character 54, "molar crown area"). Because we are using CIC as a relative measure, hereafter we express CIC values in terms of their rank (CIC_{rank}) based on their CIC value divided by the number of traits. The resulting quartiles can be directly interpreted: "low CIC" ($CIC_{rank} < 0.25$), "mid CIC" ($0.25 \leq CIC_{rank} < 0.75$), and "high CIC" ($0.75 \leq CIC_{rank}$).

Phylogenetic information content

We proposed two approaches to quantify the degree of phylogenetic information content: removal (RRF) and permutation Robinson-Foulds (PRF) distances. A common method to assess the relationship between traits and trees and how effective maximum parsimony is at describing the evolutionary pattern consists of permuting the tip values across OTUs (Alroy, 1994; Laurin, 2004). This allows an assessment of whether the observed distribution of traits along a tree is likely to have

occurred randomly. Here, we elaborate on this idea by testing how arrangements that contain no phylogenetic signal change the inferred topology: assuming that each trait in the character matrix exhibits some amount of phylogenetic signal, does the disruption of that signal by removal or permutation result in changes to the inferred topology? We hypothesized that if a trait contributes phylogenetic information to the inferred topology, the permutation of its values would disrupt the information carried by the trait and result in an topological distance to the baseline tree, and that its removal entirely would either result in an increased distance (if the information carried by the trait is unique) or not (if the information is redundant with information carried by a different trait that was not removed).

The distributions of both RRF and $\widetilde{PRF}$ are right-skewed (Figure 2B-C), with same range (between 0 and 10), same medians (0 for both), and similar means (1.08 and 1.25 for RRF and $\widetilde{PRF}$ respectively). Across all traits, RRF and $\widetilde{PRF}$ are highly correlated with each other ($\tau = 0.863$, Kendall's rank correlation, $p < 2.2 \times 10^{-16}$ Figure 2F). This is sensible given that in one case (RRF) we remove a trait entirely, and in the other (PRF) we destroy the information it carries. We only inferred even-numbered $RRF$and $\widetilde{PRF}$ values (Figure 2B-C) — i.e., we never inferred a polytomy in the hominin ingroup. Both metrics are also modestly but significantly positively correlated with CIC ($CIC\ v.\ RRF$: $\tau = 0.26,\ p = 0.00072$; $CIC\ v.\ \widetilde{PRF}$, $\tau = 0.342,\ p = 7.1 \times 10^{-6}$; Kendall's rank correlation; Figure 2D-E). Based on our criteria for classifying traits as informative/uninformative and unique/redundant (Figure 3A), 76 traits (71%) were uninformative — i.e., permutation across OTUs has no effect on the baseline tree. The remaining 31 traits (29%) contain variable degrees of redundant or unique phylogenetic information: 23 (21.5 % of all traits) are non-redundant ("unique") in the phylogenetic information content they carry (i.e., $\widetilde{PRF} \geq 2$ and $RRF \geq 2$), and the remaining 8 traits, while potentially informative ($\widetilde{PRF} \geq 2$) carry variation that is captured by other traits — i.e, they are redundant ($RRF = 0$). Throughout the results and discussion, we often highlight traits that are commonly studied in paleoanthropology.

Figure 2. Distributions of different metrics and their correlations across 107 characters. A)-C) Distributions (bin size of 1) for A) cladistic information content, B) B) removal Robinson-Foulds distance, and C) median permutation Robinson-Foulds distance (250 permutations for each trait). D-F) Correlations (τ, Kendall) between D) $RRF\ v.\ CIC$, E) $\widetilde{PRF}\ v.\ CIC$, and F) $\widetilde{PRF}\ v.\ RRF$. The orange line shows the linear regression.

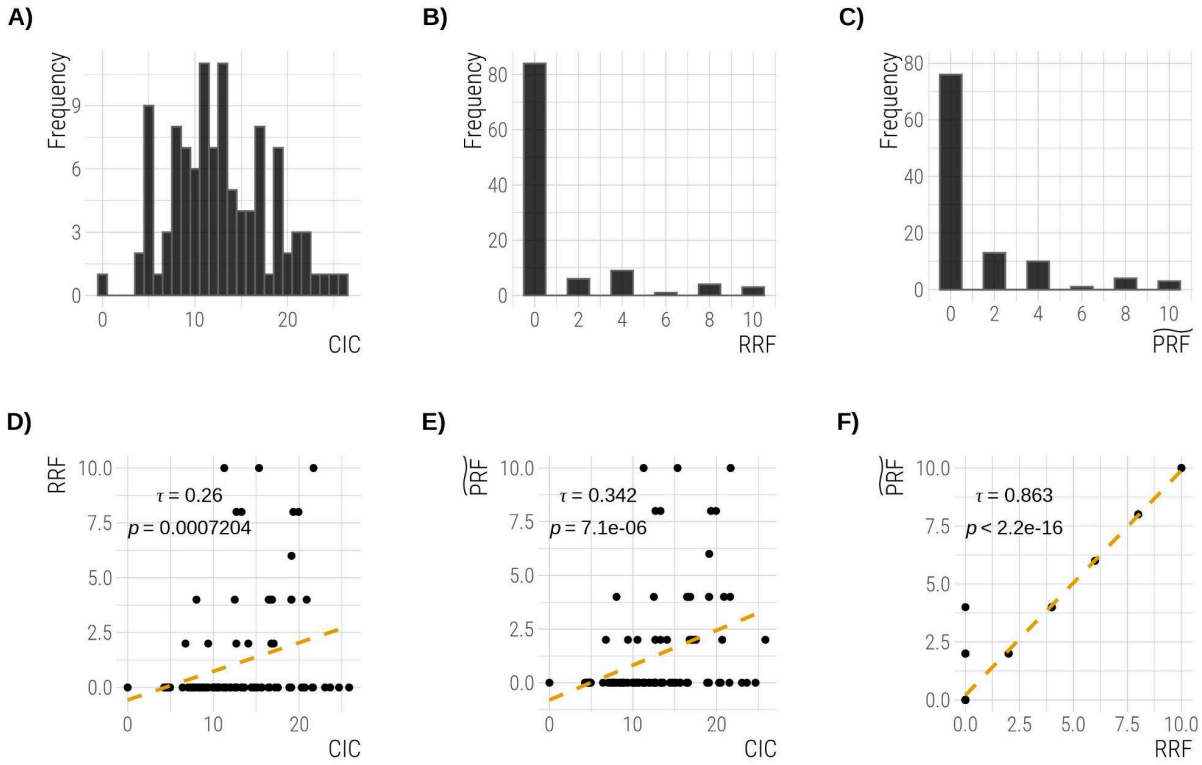
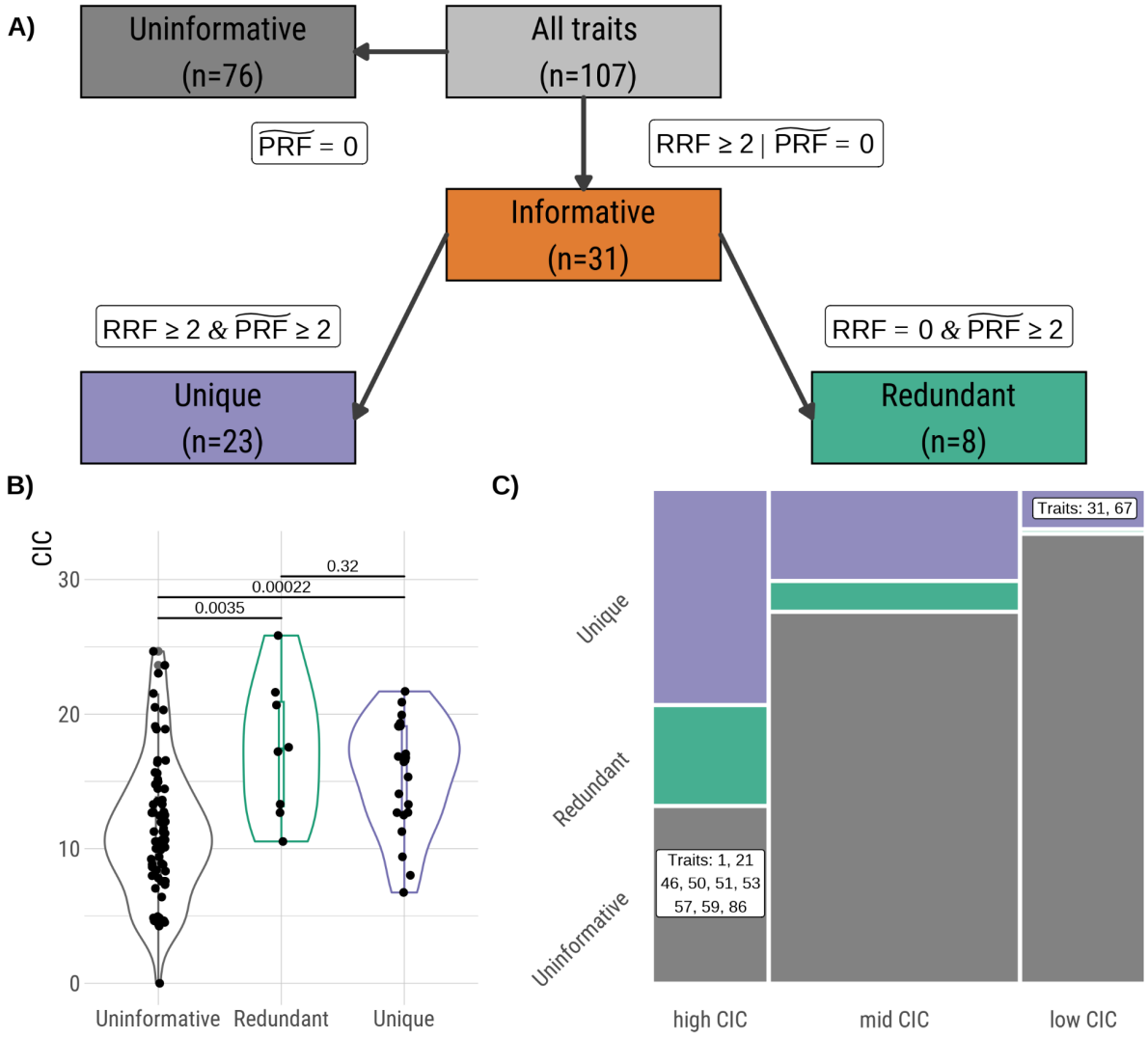


Figure 3. Classification of traits in terms of phylogenetic information content. A) A flowchart explaining the categorization of the 107 traits based on RRF and $\widetilde{PRF}$ values. B) Distribution of CIC values across uninformative and informative (redundant or unique). Boxplots are defined by the 25% and 75% quantiles, the lower whisker as the lowest observation greater or equal than the lower whisker $-\ 1.5\ \times\ IQR$ and the upper whisker as the highest observation lower than the upper whisker $+\ 1.5\ \times\ IQR$, where IQR is the interquartile range. Horizontal bars show comparisons and p-values (Mann Whitney U test). C) Proportion of unique, redundant, and uninformative traits within bins of CIC: "low CIC" ($CIC_{rank} < 0.25$), "mid CIC" ($0.25 \leq CIC_{rank} < 0.75$), and "high CIC" ($0.75 \leq CIC_{rank}$). Traits annotated in the figure are those that fall within the "uninformative" and "high CIC" intersection (bottom right) and "unique" and "low CIC" (top right). Trait names and definitions are listed in SOM Table S1.



Confirming our intuition that CIC values are a measure of the "potential" of a trait to harbor phylogenetic information, we find that the only trait with $CIC = 0$ (character 83) also has $RRF = 0$ and $\widetilde{PRF} = 0$ — not surprising, that this is one of the traits with the highest degree of missing data (11 OTUs, Figure 1, SOM Table S3) and has only 2 coded states, with only *H. sapiens* coded as 1— and can be safely assumed to be devoid of phylogenetic information (SOM Table S6). In fact, as expected CIC is strongly negatively correlated with the number of OTUs with missing data ($\tau =- 0.412, p = 3.494 \times 10^{-9}$), whereas RRF and $\widetilde{PRF}$ are similarly but less intensely affected (RRF: $\tau =- 0.174, p = 0.031$; $\widetilde{PRF}$: $\tau =- 0.239, p = 0.0029$). Interestingly, the negative correlation persists for CIC when looking only at the 31 informative traits ($\tau =- 0.334$, $p = 0.014$), whereas for RRF and $\widetilde{PRF}$ it disappears (RRF: $\tau = 0.122, p = 0.405$; $\widetilde{PRF}$: $\tau = 0.076, p = 0.614$).

Lastly, CIC is generally higher for informative traits than for uninformative ones (MWU, uninformative *vs.* redundant, $p = 0.0035$;. uninformative *vs.* redundant: $p = 0.00022$, Figure 3B). Nevertheless, no significant difference was detected between CIC values for unique *vs.* redundant traits (MWU, $p = 0.32$, Figure 3B), indicating that CIC alone would not have allowed this level of insight — i.e., uniquely v. redundantly informative — about the traits.

Uniquely informative traits

The 23 uniquely informative traits (SOM Table S7) are those whose removal and permutation both affect the baseline tree. Importantly, the fact that they are informative does not mean they are intrinsically "better". In fact, given our criteria, it is entirely possible for a trait to be uniquely informative but also that it informs an incorrect phylogeny (we do not assume the baseline topology is correct), or that it informs particular placements that are otherwise (in its absence) not robust. For this set of traits, RRF and $\widetilde{PRF}$ are perfectly correlated, so we mention only their $\widetilde{PRF}$ values. Their CIC_{rank} values span from 0.131 (character 67) to 0.963 (character 33, Figures 2A, 3B, SOM Table S5), once again indicating that CIC alone would not have allowed the detection of these traits as a singular set.

The three most informative traits (SOM Table S7) are the size of the postglenoid process (character 33, $CIC_{rank} = 0.963$, $\widetilde{PRF} = 10$, $PRF_{IQR} = 6$), the orientation of the mandibular premolar row arcade shape (character 62, $CIC_{rank} = 0.701$, $\widetilde{PRF} = 10$, $PRF_{IQR} = 2$), and the prominence of the lingual ridge of the mandibular canine (character 52, $CIC_{rank} = 0.430$, $\widetilde{PRF}$ =10, $PRF_{IQR} = 0$). Removal or permutation of any of these traits results in major changes to the baseline tree, and in particular affecting the position of the *Paranthropus* clade (SOM Figure S3-S5). For example, for character 33 both the removal and one example permutation tree (SOM Figure S3), in contrast to the baseline, *Homo* is monophyletic and the *Paranthropus* clade is inferred as differentiating much earlier, with *Paranthropus* as an outgroup to the clade containing *Au. garhi*, *Au. afarensis*, *K. platyops*, *Au. africanus*, and *Homo*. The removal and most common PRF test tree were identical to one another (SOM Figure S3).

To ensure that these 23 uniquely informative traits were truly not redundant to one another and not all informative to a few, tenuously supported relationships in the baseline tree (as exemplified by traits 33, 52, and 62 and their effect on the *Paranthropus* clade), we compared all RRF test trees from the 23 uniquely informative traits to each other. Their pairwise topological distances ranged from $RF = 0$ to $RF = 12$, with most trees being quite distant from one another with ($\overline{RF} = 7.22, \widetilde{RF} = 8$) (SOM Figure S6).

To gain further insight into what CIC and and our RRF and $\widetilde{PRF}$ metrics can capture, we argue that if a trait is uniquely informative but falls within the low CIC bin, it provides more phylogenetic information than expected based on CIC alone. We find two traits in this category (Figure 3C, top right): the horizontal distance between the temporomandibular joint and the M^2/M^3 (character 31, $CIC_{rank} = 0.206$, $\widetilde{PRF} = 4$) and the configuration of the superior orbital fissure (character 67, CIC_{rank} 0.131, $\widetilde{PRF} = 2$). Character 67's low CIC can be explained by its 8 missing OTUs and only 2 coded states (Figure 1). Nevertheless, it is highly informative to the baseline tree: removal of this trait places *K. platyops* as a sister group to *Homo* + *Paranthropus*, in contrast to the baseline tree, where *K. platyops* is grouped with *Homo* + *Paranthropus*. Conversely, if a trait is uninformative and falls within the high CIC bin we argue that the trait provides less information than expected based on CIC.

Redundant traits

Among the 31 informative traits, we found eight traits which, though informative ($\widetilde{PRF} \geq 2$), contributed redundant information ($RRF = 0$). Among these, five have high CIC: facial prognathism (character 13, $CIC_{rank} = 0.925$), the depth of the mandibular fossa (character 32, $CIC_{rank} = 0.822$), the width of the mandibular extramolar sulcus (character 48, $CIC_{rank} = 0.953$), the shape of the supraorbital contour (character 80, $CIC_{rank} = 0.813$), and molar crown area (character 54, $CIC_{rank} = 1$). Despite having the highest cladistic information content, the latter contributes no unique phylogenetic information ($RRF = 0,\ \widetilde{PRF} = 2,$ $PRF_{IQR} = 4$), i.e, it is redundant with at least one other character in the matrix.

To test the ability of these uniquely informative and informative traits to resolve the baseline hominin topology (Figure 1) as a set, we inferred two test trees: one from the 31 informative traits and one from the 23 uniquely informative traits. The one inferred from all 31 informative traits differed from the baseline tree by $RF = 12$ and is not significantly different from the null expectation for this number of traits ($p = 0.58$), but the test tree inferred from just the 23 uniquely informative traits differed from the baseline tree by only $RF = 4$ and is significantly less distant than expected ($p = 0.02$; SOM Figure S7).

Uninformative traits

More than 70% of the traits assessed here are uninformative (Figure 3A). Though the criterion to group them as such was that $\widetilde{PRF} = 0$, we note that all of them also have $RRF = 0$. Among the 76 uninformative traits, nine are also "high CIC" traits (Figure 3C, bottom left), once again showing that CIC alone is not a good predictor of whether a trait will actually inform an inferred tree topology, but rather of a trait's potential to have such an effect. These include traits that have been often studied in paleoanthropology: sagittal cresting (character 20, $CIC_{rank} = 0.589,\ \widetilde{PRF} = 0$), anterior pillars (character 4, $CIC_{rank} = 0.112, \widetilde{PRF} = 0$), reduction of incisors (character 50 , $CIC_{rank} = 0.972, \widetilde{PRF} = 0$), reduction of canines (character

51, $CIC_{rank} = 0.981$ $\widetilde{PRF} = 0$), and relative enamel thickness (character 59, $CIC_{rank} = 0.944$, $\widetilde{PRF} = 0$).

Furthermore, 57 uninformative traits also had $PRF_{IQR} = 0$. Thus, for those 57 traits, a majority of trees built from random permutations of their values did not topologically differ from the baseline tree. Strikingly, for 22 of these 57 (SOM Table S6), $RF = 0$ for all permutation test trees — i.e., none of the 250 permutations resulted in a test tree topologically different from the baseline. Included among them are the inclination of the nuchal plane (character 40), the presence of an extensive mesial groove on the upper canine (character 70), distinctiveness of the angular process of the mandible (character 74), glabellar prominence relative to the sella (character 78), robusticity of canines (character 85), and several other characters related to dental morphology (SOM Table S5, SOM Table S6). Unexpectedly, *K. platyops* is not coded for any of these traits ($p = 0.00045$, from 100,000 resampled units of size 22) and *Au. garhi* is only coded for 3 ($p = 0.103$).

To test that these 22 "highly uninformative" traits truly had no impact on the baseline tree inference, we inferred a test tree excluding these 22 traits from the character matrix and found that it was identical to the baseline tree ($RF = 0$), confirming that these traits, as a block, represent the strongest candidates for exclusion from future hominin character matrices (SOM Table S6).

Effects of anatomical units on the inferred topology

We assigned each of the 107 craniodental traits to one of 10 anatomical units (AUs), where nine represent individual bones and the tenth (the "complex" AU) encompasses traits that include more than one bone (Table 4). The number of traits per unit ranges from $n = 1$ (nasal) to $n = 25$ (dentition). To gain further insight on the amount of information contained within individual AUs relative to the baseline tree, we implemented two approaches. In one approach (URRF) we looked at distances of test trees obtained by removing an entire AU to the baseline tree, and in the other (SURF) by only including traits in those AUs and excluding all other AUs (Table 1). Because the AUs contain different numbers of traits, we contextualize the SURF and URRF values to their respective null distributions, obtained by sampling "pseudo AUs" of the same size as actual AUs (see Methods).

Overall, what drives URRF values for most of the AUs is the number of traits it contains rather than any innate information contained in the units as a block. URRF is barely significantly negatively correlated with the number of traits on each unit ($\tau = -0.554, p = 0.0496$, Kendall's

rank correlation). Thus, while in an absolute sense we find different URRF values associated with the different AUs, with dentition containing the most phylogenetic information ($URRF = 14$) followed by the maxilla ($URRF = 12$), mandible ($URRF = 10$), and zygomatic ($URRF = 10$, Table 4), only the maxilla has a significantly higher URRF than expected given the number of traits in the AU ($p = 0.03, p - adj = 0.27$, Table 4, SOM Figure S8). However, the phylogenetic tree inferred solely from maxillary traits is very distant from the baseline tree and not unexpected ($SURF = 18,\ p = 1.00$). In general, SURF values across all AUs fall well within the expected range given their respective null distributions. However, SURF values are marginally correlated with the number of traits in each AU ($\tau =- 0.632,\ p = 0.0572$, Kendall's rank correlation).

Though the phylogenetic trees inferred from the mandibular and dentition AUs are well within expectation given the number of traits they contain (SOM Figure S8), we nevertheless consider them in greater detail given the prevalence of isolated mandibles and teeth in the hominin fossil record. Both differ substantially from the baseline tree. Among other differences, in the mandibular-traits-only tree ($SURF = 12$, Figure 4A), *Ar. ramidus, Au. anamensis*, and *S. tchadensis* are not inferred in the hominin clade: *Ar. ramidus* is inferred as a sister taxon to *Pan troglodytes*, and *Au. anamensis* and *S. tchadensis* are inferred in a clade with *G. gorilla* and *Papio*. Similarly, in the dentition-only tree (SURF=20, Figure 4B), all hominins are inferred in a monophyletic hominin clade, but *S. tchadensis* is inferred as more derived than *Ar. ramidus* and *Au. anamensis*, while *H. ergaster* is inferred as a sister group to a clade containing *H. sapiens*, *Au. garhi*, and *Paranthropus*.

Moreover, while all reported topologies here are the consensus most parsimonious trees with the highest clade credibility (see Methods), the BAB algorithm returned an astounding number of MP trees for all SURF analyses — 72,520 and 1,278 equally parsimonious trees for the mandibular and dentition-only trees, respectively (Table 4). In contrast, we found 3 MP baseline trees from the unmodified hominin tree. The number of most parsimonious trees is not significantly correlated with the number of traits contained in the unit ($\tau =- 0.390,\ p = 0.224$, Kendall's rank correlation).

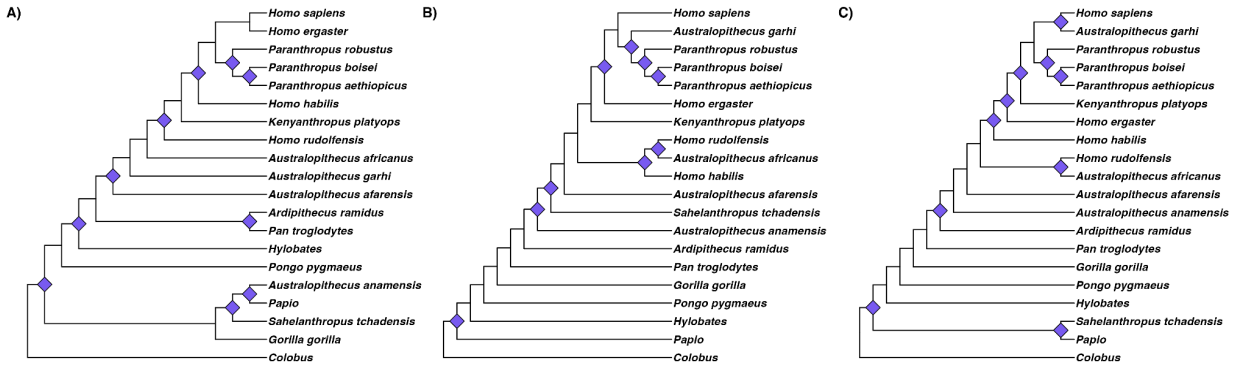
In brief, under the framework of maximum parsimony with the BAB algorithm, and the dataset and AU definitions used here, we were unable to discriminate across much of the tree space when using only traits from single AUs. Because mandibular and dental traits are so often used to place fossil hominin specimens, we also assessed how far from the baseline tree is a test tree inferred jointly from mandibular and dental traits ($n = 40$, Figure 4C). While using mandibular and dental

traits jointly more effectively specifies tree space (only one MP tree is returned by BAB), the MP tree is substantially distant from the baseline tree ($RF = 16$) — though this is not unexpected ($p = 0.53$) and shows that those units in concert are not sufficient to closely resolve the baseline topology. Moreover, using them jointly leads to a tree that is more distant from the baseline than the dentition-only trees (Figure 4).

Table 4. Robinson-Foulds distances (RF) of each region of the skull. AU, anatomical unit. URRF, unit removal RF; SURF, single unit RF. For the URRF and SURF columns, the p-value based on sampling 100 pseudo-AUs of same length as the AUs are shown. Units are ordered by $SURF$ value. Nasal is omitted due to only containing one trait (SOM Table S5). [a], the number of MP trees returned by the branch and bound algorithm. [b] The SURF analysis was unable to complete due to computational limitations. Null distributions and values are shown in SOM Figure S8. Trait names and definitions are listed in SOM Table S1.

AU (number of traits)	Traits	URRF	SURF	MP Trees[a]
Zygomatic (n=7)	9, 10, 11, 77, 81, 102, 107	10 ($p = 0.26$)	22 ($p = 0.19$)	2,350,575
Dentition (n=25)	49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 70, 84, 85, 86, 87, 88, 89, 90, 91, 92, 97, 99	14 ($p = 0.23$)	20 ($p = 0.36$)	1,278
Frontal (n=9)	19, 26, 67, 71, 75, 78, 79, 80, 105	4 ($p = 1.00$)	20 ($p = 0.74$)	356,440
Occipital (n=9)	18, 21, 30, 39, 40, 42, 68, 96, 106	8 ($p = 0.60$)	20 ($p = 0.74$)	46,167
Maxilla (n=13)	2, 3, 4, 5, 6, 7, 8, 12, 13, 15, 76, 82, 101	12 ($p = 0.03$)	18 ($p = 1.00$)	37,014
Temporal (n=14)	24, 25, 27, 29, 33, 34, 35, 36, 37, 38, 65, 69, 72, 83	8 ($p = 0.93$)	14 ($p = 0.31$)	9,278,325
Mandible (n=15)	32, 43, 44, 45, 46, 47, 48, 62, 66, 73, 74, 93, 94, 95, 98	10 ($p = 0.45$)	12 ($p = 0.15$)	72,520
Complex (n=9)	14, 16, 17, 22, 28, 31, 41, 103, 104	8 ($p = 0.60$)	—[b]	
Parietal (n=5)	20, 23, 63, 64, 100	0 ($p = 0.60$)	—[b]	

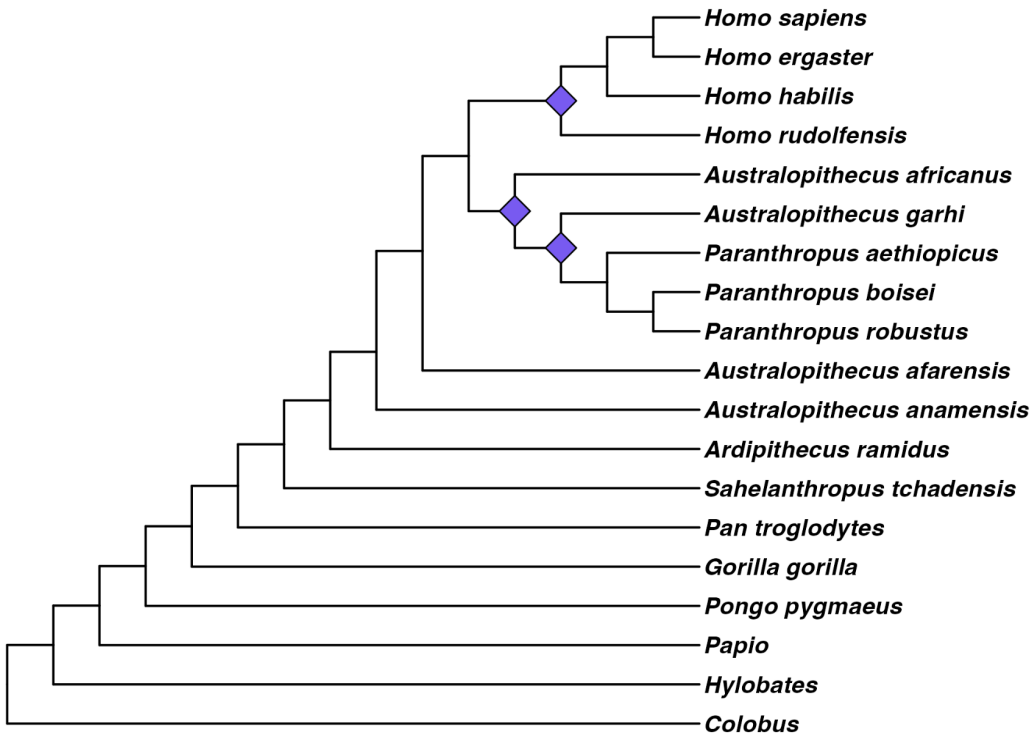
Figure 4. Hominin phylogenetic trees inferred from sets of characters from different anatomical units. Consensus tree topologies inferred using the branch and bound algorithm with maximum clade credibility using solely A) solely mandibular characters ($n = 15$, $SURF = 12$), B), solely dental characters ($n = 25$, $SURF = 20$), C) both dental and mandibular characters ($n = 40$, $RF = 16$). All trees are rooted with *Colobus*. Purple diamonds indicate all changes in relation to the baseline topology, but only the differences in the hominin + *Pan* subtree are considered in the RF calculation.



Effects of each taxon on the inferred phylogenetic tree

Removing each of the 20 taxa individually from the character matrix (TRRF, Table 1) allowed us to show the relative contribution of each taxon to the inferred baseline tree. These varied from taxa that do not affect the inferred hominin + *Pan* topology, ($TRRF = 0$, *P. aethiopicus*, *Au. garhi*, *Ar. ramidus*, and *Po. pygmaeus*) to various degrees of effect, ranging from moderate ($TRRF = 4$, *S. tchadensis, P. boisei, Papio, H. ergaster*, *Hylobates*, *G.gorilla,* and *Colobus*; $TRRF = 6$, *H. habilis, Au. africanus,* and *Au. afarensis,* SOM Table S8, SOM Figures S14-S23) to high ($TRRF = 8$, *P. robustus, H. rudolfensis, H. sapiens, K. platyops,* and *Au. anamensis*; $TRRF = 10$, *Pan troglodyte*s, SOM Table S8, SOM Figures S9-S13). We found no correlation between $TRRF$ and the number of missing traits in a given OTU ($\tau =- 0.037,\ p = 0.838$, Kendall's rank correlation), exemplified by *Au. garhi* ($TRRF = 0$; 78 missing traits) and *K. platyops* ($TRRF = 8$; 79 missing traits). The consensus MP tree inferred from the character matrix missing *K. platyops* is depicted in Figure 5. We highlight this particular case due to the long-standing questions surrounding the validity of this taxon. In contrast to the baseline topology, this tree positions *Homo* as monophyletic and *Au. garhi* and *Au. africanus* as sister taxa to the *Paranthropus* clade.

Figure 5. Hominin phylogenetic tree inferred without *K. platyops*. Consensus tree topology inferred from the character matrix excluding *K. platyops* using the branch and bound algorithm with maximum clade credibility ($RF = 8$ between this and the baseline tree minus *K. platyops*). The tree is rooted with *Colobus*. Purple diamonds show changed topology relative to the baseline tree.



Discussion

We applied several new analytical approaches to an existing dataset of 107 hominin craniodental characters and 20 OTUs (Mongle et al., 2019) to evaluate to what degree those characters contribute useful information for the phylogenetic placement of fossil hominins. Using the branch and bound algorithm within the maximum parsimony framework (still widely used when dealing with morphological datasets), we compared all test trees against a baseline tree, which we defined as the best-supported MP tree inferred from the full character matrix.

In a limited-evidence context such as the fossil record, the tree and the trait are not independent because the traits(s) of interest are often used to infer the tree. Thankfully, our approaches do not require that the most accurate (i.e., genomic) phylogeny be known, and we make no phylogenetic claims about what the correct hominin species tree may be. Rather, we report that the character matrix used here and widely across hominin phylogenetic studies (e.g., Strait et al., 1997; Strait and Grine, 2004; Dembo et al., 2015; Dembo et al., 2016; Mongle et al., 2019; Mongle et al., 2023; and Gautney, 2023) includes traits and taxa that, at the present time, may not be meaningful and instead, confounding, to the understanding of hominin evolution.

Below, we discuss our proposed metrics for phylogenetic information content and their relationship with the widely used cladistic information content, while highlighting traits found to be unexpectedly informative or uninformative. Next, we discuss the relative contribution of individual anatomical units and taxa to the inference of the baseline tree. Throughout, we contextualize the relevance of our findings for paleoanthropology.

Uniquely informative traits

We focused on the two aspects of morphological traits: phylogenetic information content (i.e., traits distributed across the OTUs such that the distribution informs the inferred topology) and the uniqueness (i.e., non-redundancy) of that information. Using our combination of trait removal and permutation, we identified 31 informative characters, 23 of which are non-redundant. While our test tree with the 23 uniquely informative traits was very similar to the baseline tree ($RF = 4$), the tree inferred from the 31 informative traits was not. The difference between these two sets of traits is solely due to the additional eight informative-but-redundant traits, and clearly exemplifies the dependence of inferred topologies on relatively small proportions of the overall character matrix.

The difference between the two sets is expected: the 23 uniquely informative traits are, by definition, not redundant, so the addition of eight redundantly informative traits adds unique information, presumably from the traits they are redundant with that are not included among the 23 uniquely informative traits. Given the similarity between the tree inferred from the 23 uniquely informative traits and the baseline tree, our approach is therefore well-suited to identifying the majority of informative traits in this specific character matrix and, likely, other datasets. However, we were not able to perfectly identify all uniquely informative traits in the character matrix, as otherwise the test tree inferred from this set of traits would be identical to the baseline tree. This indicates our approach is conservative in its detection of phylogenetically informative traits to the baseline tree considered here.

The most phylogenetically informative traits in our analyses were traits 33 (the size of the postglenoid process), 52 (the prominence of the lingual ridge of the mandibular canine), and 62 (the orientation of the mandibular premolar row dental arcade). It should be noted, however, that their observed information comes from, when removed or permuted, an inferred basal split of *Paranthropus,* a position that is inconsistent with our baseline tree and most of the literature (Strait et al., 1997; Strait & Grine, 2001; Lockwood & Tobias, 2002; Villmoare & Kimbel, 2011; Strait et al., 2013; Dembo et al., 2015; Dembo et al., 2016; Mongle et al., 2019; Braga et al., 2023; Mongle et al., 2023), with the notable exception of Parins-Fukuchi et al. (2019), though they used a combination of temporal and morphological data. If three traits, when allowed to individually vary across the tree, are enough to change the position of these taxa within the hominin tree —all of which occur at a crucial time for understanding the origin of *Homo*— , then the confidence we have in those relationships should be reevaluated. Our findings thus suggest that the derived position of the *Paranthropus* split seen in the literature and our baseline tree is tenuously supported.

Uninformative and redundantly informative traits

In this study, we distinguish between a trait that carries no phylogenetic information towards the baseline tree and a trait that would otherwise be informative but is redundant with other traits. While one may be tempted to include as many traits as possible in morphological phylogenetic inference including fossils, doing so can bias inference: redundancy in variation captured by multiple characters artificially inflates the weighting of those characters in a parsimony analysis (Prendini and Wheeler, 2005; Mongle et al., 2019), which our combination of RRF and PRF metrics can detect. We found that 76 (71%) of the characters in the matrix are uninformative to the inference of the

baseline tree, and eight (26% of the 31 phylogenetically informative traits) are redundantly informative. Any traits that are either uninformative or redundant should be treated with caution when utilized in systematic arguments. For example, the presence or absence of anterior pillars (character 4) is a key trait in discussions regarding monophyly or paraphyly in *Paranthropus* (Lockwood & Tobias, 2002; Villmoare & Kimbel, 2011; Strait et al., 2013; Braga et al., 2023), as is the presence of a sagittal crest (character 20) in presumptive male *Paranthropus*, but both are uninformative according to our methods. That these traits contain no information may, perhaps counter-intuitively, further support the monophyly of *Paranthropus*: even when these two critical traits are excluded from the character matrix, the monophyly of the *Paranthropus* clade is still inferred.

Our approaches were able to detect informative and uninformative traits where a cladistic information content metric alone would have failed. Interestingly, among the 14 high CIC traits that are either redundantly informative or uninformative, 50% are dental traits. Included among those traits are relative enamel thickness (character 59, uninformative), reduction of the incisors (character 50), reduction of the canines (character 51), premolar crown area (character 53), and molar crown area (character 54, redundant). That these traits had the potential to be highly informative but were not suggests that their use in hominin phylogenetics should be done cautiously and only after a better understanding is reached as to why they did not contribute information. At the same time, their lack of information to the baseline topology does not necessarily mean they lack utility in all contexts. Rather, our findings suggest that context-specific assessments, such as those we present here, are necessary to determine whether these traits are informative in a given context.

The relative unimportance of the foramen magnum position (character 41, uninformative) is somewhat surprising given its intermediate CIC and that it is continually implicated in bipedal transitions (Brunet et al., 2002; Zollikofer et al., 2005; Russo and Kirk, 2013), though its relationship with positional behavior has been questioned (Neaux et al., 2017). Interestingly, when looking at all OTUs, the CIC value for foramen magnum position is in the third quartile of values (suggestive of its potential to be informative), but when we recalculated CIC values solely for *Pan* + hominin, the foramen magnum position's CIC is in the lowest quartile. This reinforces the importance of critically assessing which taxonomic levels are being used to estimate phylogenetic signal in a particular group, and at which taxonomic levels are said estimates of phylogenetic signal being applied. In brief, based on our findings, the taxonomic placement of a hypothetical hominin taxon with no preservation of the foramen magnum would likely be accurate, given that this trait does not seem to significantly affect topology.

A possible explanation for the relatively low contribution of the position of the foramen magnum to the inferred baseline tree might be related to issues with how continuous traits are discretized. While the hominin clade is defined by a progressively increasing capacity for bipedalism (Almécija et al., 2013), the inevitable information loss from reducing continuous variation into discrete character states may result in the discretized trait no longer carrying enough information to impact hominin phylogenetic relationships. If this is true, then our finding would indicate not a lack of relevance of the trait itself, but rather a lack of sufficient information resulting from discretization of a continuous trait. Traits associated with bipedalism, like the foramen magnum position, may thus represent a strong argument for the incorporation of continuous traits into phylogenetics, as others have suggested (e.g., Parins-Fukuchi, 2018).

Effects of anatomical units

A noteworthy finding in the individual AU analyses was the high number of resolved MP trees (i.e., those identified as equally parsimonious), indicating that single units alone are unable to distinguish between a large proportion of tree space, at least within a maximum parsimony framework. This finding holds special relevance for the commonly found hominin remains: dentition, mandibles, and maxillae.

No AU except the maxilla affected the tree more than expected. Two units, the frontal ($URRF = 4, n = 9$) and the parietal ($URRF = 0, n = 5$), elicited minor or no changes from the baseline tree when removed. The frontal bone's lack of information is not particularly surprising, given that this AU consists of relatively few traits, and much of the variation in traits found on the frontal that is potentially relevant in phylogenetics occurs in the more closely related Middle Pleistocene hominins not included in this study (and even in those taxa it does not appear to be phylogenetically informative, Athreya, 2009; Liu et al., 2022). In contrast, given the consistent increase in encephalization over hominin evolution (with the exception of *H. naledi* and *H. floresiensis*, though neither was included here) and the correlation between parietal morphology and encephalization (Püschel et al., 2021) noted in the literature, the observed lack of information in the parietal may be surprising. Increased encephalization in hominin evolution likely reflects cranial reorganization (especially around the face) with increasing bipedalism, potentially reflecting structural changes in the brain (Bruner et al., 2017; Alatorre et al., 2019). It is likely, however, that any variation we identified as informative in the parietal was also captured to a greater extent by traits such as cranial capacity and cerebral morphology, which fall in the "complex" category, thus

rendering parietal morphology redundant and supporting the theory that the traits present in the parietal are redundant to traits located outside the unit. Additionally, there are only five traits on the parietal in this dataset, and all five were themselves uninformative. So potentially, our observed lack of information in this AU may simply be due to the composition of the traits in this character matrix.

In addition, individual anatomical units (including the maxilla, mandible, and dentition) by themselves do not provide enough information to retrieve the baseline tree. Isolated mandibular remains are relatively common in the hominin fossil record and much research has relied on mandibular morphology and characters to make systematic arguments in the genus *Homo* (e.g., Schrenk et al., 1993; Bermúdez de Castro et al., 2007; Mounier et al., 2009; de Ruiter et al., 2013; Villmoare et al., 2015; Bermúdez de Castro et al., 2016). However, our findings are consistent with past research warning against using solely mandibular traits (Schmittbuhl et al., 2007; von Cramon-Taubadel and Lycett, 2014) when inferring hominin phylogenetic relationships. The tree we generated from mandibular characters only (not including the mandibular dentition traits, assigned to the dentition AU) was remarkably distant from the baseline tree, and contributed an amount of phylogenetic information expected for random sets of traits. Many of the notable differences in topology come from *Au. anamensis*, *S. tchadensis*, and *Ar. ramidus*, as well as *H. habilis* and *H. rudolfensis*. Our findings caution against hominin phylogenetic inference from isolated mandibular remains (at least based on the traits in this dataset and using the parsimony framework) because of its inability, on its own, to closely recover the baseline tree topology and the large number of equally MP trees returned. Additionally, we find that combining dental and mandibular characters does not lead to more robust phylogenetic inference using the traits in the Mongle et al. dataset. Firstly, dental characters alone are insufficient to differentiate early *Homo* from *Australopithecus*, evidenced by *H. rudolfensis* and *H. habilis* both being inferred as closer to *Au. africanus* than to any other *Homo* OTU. Secondly, when dental and mandibular traits are jointly used to infer a tree, *H. rudolfensis* is inferred as a sister taxon to *Au. africanus* and *H. sapiens* as a sister taxon to *Au. garhi*.

In sum, we urge caution and skepticism when attempting to phylogenetically place fossil remains based on isolated AUs alone, at least on the basis of the characters included in the Mongle et al. (2019) dataset and the maximum parsimony framework. Lastly, here the AUs were defined solely based on the bones of the skull, but in future analyses may include a prior grouping based on known functional correlation across those bones, such as functional complexes as described in Strait

(2001). Such an approach would allow researchers to shed further light on the findings described here.

Effects of individual taxa

In addition to evaluating the effects of individual traits and anatomical units, we also assessed the effects of individual taxa (be it outgroup or ingroup) on hominin phylogenetics, for the following reasons. Firstly, this character matrix incorporates taxa whose validity has been questioned; secondly, because outgroup choices can affect ingroup topology; and lastly, because the phenotypic variation in the fossil record in general has been binned into OTUs that cannot be independently confirmed (e.g., with genomic methods), and thus may bias the inferred topology in unintended ways.

Given the concerns of Kimbel (2013) and Carter et al. (2014) that only using *Pan* as outgroup for hominin systematics may increase the error and bias stemming from *Pan*-specific characters, Mongle et al. (2019) included multiple extant outgroups in their analyses. Our taxon-removal results support these concerns. The removal of any of the extant non-hominin primates — *Colobus*, *Papio*, *Hylobates*, *Pan troglodytes* , and *G. gorilla* — all result in different resolved hominin topologies, but *Pan* has by far the largest effect. This suggests that perhaps the hominin topology is still affected by bias from *Pan*, which may be ameliorated by the inclusion of more Miocene ape OTUs in future character matrices (e.g., Post et al., 2023). However, as with all test trees in this study, a higher topological distance between the baseline and test trees is not indicative of a lower degree of accuracy in the test tree in relation to the baseline tree. Rather, it indicates that the information coded in a given OTU's character states is more critical to the inference of a particular topology (in this study, the baseline topology).

Among the hominins, *K. platyops* has one of the highest TRRF values despite having the largest number of missing traits among all OTUs. This taxon and its potentially taphonomically-affected morphology are somewhat contentious (e.g., White, 2003, Alemseged, 2023). When *K. platyops* is removed from the character matrix, *Homo* becomes monophyletic, *Au. garhi* is inferred as the immediate sister group to *Paranthropus*, and *Au. africanus* as the sister taxon to the *Au. garhi-Paranthropus* clade. Despite the debate surrounding the validity of *K. platyops* as a taxon, its impact on the inferred topology is critical to understanding the origins and phylogenetic validity of our genus. More nuanced approaches may include treating *K. platyops* as evidence of

polymorphism in a variety of *Australopithecus* OTUs (e.g., amalgamating its characters with *Au. afarensis* following Alemseged, 2023). However, given its taphonomic damage, until more and better preserved fossil material is recovered for this OTU, any character matrix that includes *K. platyops* is inherently predicating their inferred topology on the validity of *K. platyops* as a hominin. We urge caution if including this taxon, and to a lesser extent *H. rudolfensis* (e.g., Bromage et al., 2008), in hominin phylogenetic inference until more specimens become available.

Conclusions

We assessed the phylogenetic information content of 107 craniodental characters compiled by Mongle et al. (2019) and found evidence that only 23 of them are phylogenetically informative and non-redundant. We document a disconnect in the dataset between a high cladistic information content — a predictor of potential phylogenetic information content (Cotton & Wilkinson, 2008) — and the direct assessment of what phylogenetic information a trait contributes to a given tree inference. However, we caution that phylogenetic information as defined here is not a valuation of utility but an assessment of a trait's contribution towards our specific baseline topology. On the other hand, we identified 22 traits that contribute no phylogenetic information whatsoever to the phylogenetic inference using this dataset, and recommend they be excluded.

We also assessed the amount of information contained in sets of traits composing anatomical units in the skull and found that only the maxilla contains significantly more information than would be expected. Nevertheless, even this relatively informative anatomical unit on its own is insufficient for accurate phylogenetic inference. Basing systematic arguments solely on isolated units can lead to substantial inconsistencies in the inferred topologies: isolated bones, such as mandibles, or teeth do not possess enough phylogenetic information to appropriately assign hominin specimens to a species, at least based on the taxa and traits used here. While it is plausible that other approaches such as maximum likelihood or Bayesian inference may allow more precise filtering of tree space, maximum parsimony alone is insufficient to distinguish between, in some cases, tens of thousands or even millions of most parsimonious phylogenetic trees.

We also showed that the hominin tree topology is heavily dependent on the inclusion of dubious and poorly preserved taxa, and argue that the inclusion of certain taxa in phylogenetic analyses may strongly bias the inferred topology. Our concerns are consistent with the challenges of taxonomic attribution of new fossil specimens in paleoanthropology and suggestive that most fossil

hominin taxa should be considered tentative (e.g., Martin, 2004). We suggest that phylogenetic placements of taxa that are represented primarily by isolated anatomical units or fragmentary finds (e.g., taxa such as *K. platyops* and *Au. garhi*) be regarded as putative until the following prerequisites for more robust phylogenetic inference are met. Firstly, phylogenetic inference accuracy can be improved via the inclusion of more specimens (or more complete specimens), and by not discretizing continuous traits, thus avoiding the loss of information and potential biases in how discrete codings are assigned. Secondly, we should incorporate new and more informative characters beyond those included in this dataset, including traits that are informative across the great apes rather than defined based on hominins alone, thus avoiding circularity. Lastly, we need to increase the application of likelihood-based methods to paleoanthropological systematics and have a better understanding of the information content of traits already in use, as we, and others (e.g., Gautney, 2023) have done here.

The relationship between traits and phylogenies explored here ultimately carries implications across paleoanthropology, from taxonomic identification to the systematics of the entire hominin clade. We suggest that phylogenetic inference that aims to infer the most accurate species tree should only occur after having identified and removed problematic or uninformative traits and that systematic hypotheses be tempered not solely based on the confidence in the resolved tree (e.g., node support values), but also by confidence in the character matrix itself. The novel methodology we present here is flexible and generalizable to non-hominin clades, and may allow similar analyses in other extinct or mostly-extinct clades.

Data Availability

All statistical analyses and figures were performed in R. Scripts developed specifically to perform the data analyses reported in this work are available at: https://github.com/bitarellolab/hominin-morpho-phylogen.

Electronic Supplementary Material

Raskin_et_al_2024_SOM_table_S5.xlsx

Supplementary Figures

Figures S1 to S23

Supplementary Tables

Tables S1 to S8

Acknowledgments

We are grateful to Theodore Schurr for his comments on an early version of this manuscript, Nicholas Post and Ward Wheeler for their helpful comments about heuristic *vs.* non-heuristic algorithms. We are especially appreciative of Carrie Mongle and collaborators for their commitment to sharing their 2019 dataset openly, and we thank Carrie Mongle for her helpful clarifications about the minutiae of the 2019 dataset used in our study.

Bibliography

Alatorre Warren, J. L., Ponce de León, M. S., Hopkins, W. D., & Zollikofer, C. P. E. (2019). Evidence for independent brain and neurocranial reorganization during hominin evolution. *Proceedings of the National Academy of Sciences*, *116*(44), 22115–22121. https://doi.org/10.1073/pnas.1905071116

Alemseged, Z. (2023). Reappraising the palaeobiology of *Australopithecus*. *Nature*, *617*(7959), Article 7959. https://doi.org/10.1038/s41586-023-05957-1

Almécija, S., Tallman, M., Alba, D. M., Pina, M., Moyà-Solà, S., & Jungers, W. L. (2013). The femur of *Orrorin tugenensis* exhibits morphometric affinities with both Miocene apes and later hominins. *Nature Communications*, *4*(1), 2888. https://doi.org/10.1038/ncomms3888

Alroy, J. (1994). Four Permutation Tests for the Presence of Phylogenetic Structure. *Systematic Biology*, *43*(3), 430–437. https://doi.org/10.2307/2413678

Athreya, S. (2009). A comparative study of frontal bone morphology among Pleistocene hominin fossil groups. *Journal of Human Evolution*, *57*(6), 786–804. https://doi.org/10.1016/j.jhevol.2009.09.003

Berger, L. R., de Ruiter, D. J., Churchill, S. E., Schmid, P., Carlson, K. J., Dirks, P. H. G. M., & Kibii, J. M. (2010). *Australopithecus sediba*: A New Species of *Homo* -Like Australopith from South Africa. *Science*, *328*(5975), 195–204. https://doi.org/10.1126/science.1184944

Berger, L. R., Hawks, J., de Ruiter, D. J., Churchill, S. E., Schmid, P., Delezene, L. K., Kivell, T. L., Garvin, H. M., Williams, S. A., DeSilva, J. M., Skinner, M. M., Musiba, C. M., Cameron, N., Holliday, T. W., Harcourt-Smith, W., Ackermann, R. R., Bastir, M., Bogin, B., Bolter, D., … Zipfel, B. (2015). *Homo naledi*, a new species of the genus Homo from the Dinaledi Chamber, South Africa. *eLife*, *4*, e09560. https://doi.org/10.7554/eLife.09560

Bermúdez de Castro, J. M., Martinón-Torres, M., Rosell, J., Blasco, R., Arsuaga, J. L., & Carbonell, E. (2016). Continuity versus discontinuity of the human settlement of Europe between the late Early Pleistocene and the early Middle Pleistocene. The mandibular evidence. *Quaternary Science Reviews*, *153*, 51–62. https://doi.org/10.1016/j.quascirev.2016.10.010

Bermúdez de Castro, J.-M., Martinón-Torres, M., Gómez-Robles, A., Prado, L., & Sarmiento, S. (2007). Comparative analysis of the Gran Dolina-TD6 (Spain) and Tighennif (Algeria) hominin mandibles. *Bulletins et mémoires de la Société d'Anthropologie de Paris. BMSAP*, *19*(3–4), Article 3–4. https://doi.org/10.4000/bmsap.4623

Blomberg, S. P., Garland JR., T., & Ives, A. R. (2003). Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution*, *57*(4), 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

Braga, J., Wood, B. A., Zimmer, V. A., Moreno, B., Miller, C., Thackeray, J. F., Zipfel, B., & Grine, F. E. (2023). Hominin fossils from Kromdraai and Drimolen inform *Paranthropus robustus* craniofacial ontogeny. *Science Advances*, *9*(18), eade7165. https://doi.org/10.1126/sciadv.ade7165

Bromage, T. G., Schrenk, F., & Zonneveld, F. W. (1995). Paleoanthropology of the Malawi Rift: An early hominid mandible from the Chiwondo Beds, northern Malawi. *Journal of Human Evolution*, *28*(1), 71–108. https://doi.org/10.1006/jhev.1995.1007

Bromage, T., McMahon, J., Thackeray, J. F., Kullmer, O., Hogg, R., Rosenberger, A., Schrenk, F., & Enlow, D. (2008). Craniofacial architectural constraints and their importance for reconstructing the early Homo skull KNM-ER 1470. *Journal of Clinical Pediatric Dentistry*, *33*(1), 43–54. https://doi.org/10.17796/jcpd.33.1.8168115j12103nut

Bruner, E., Preuss, T. M., Chen, X., & Rilling, J. K. (2017). Evidence for expansion of the precuneus in human evolution. *Brain Structure and Function*, *222*(2), 1053–1060. https://doi.org/10.1007/s00429-015-1172-y

Brunet, M., Guy, F., Pilbeam, D., Mackaye, H. T., Likius, A., Ahounta, D., Beauvilain, A., Blondel, C., Bocherens, H., Boisserie, J.-R., De Bonis, L., Coppens, Y., Dejax, J., Denys, C., Duringer, P., Eisenmann, V., Fanone, G., Fronty, P., Geraads, D., … Zollikofer, C. (2002). A new hominid from the Upper Miocene of Chad, Central Africa. *Nature*, *418*(6894), Article 6894. https://doi.org/10.1038/nature00879

Carter, K., Worthington, S., & Smith, T. (2014). News and views: Non-metric dental traits and hominin phylogeny. *Journal of Human Evolution*, *69*. https://doi.org/10.1016/j.jhevol.2014.01.003

Collard, M., & Wood, B. (2001). Homoplasy and the early hominid masticatory system: Inferences from analyses of extant hominoids and papionins. *Journal of Human Evolution*, *41*(3), 167–194. https://doi.org/10.1006/jhev.2001.0487

Cotton, J. A., & Wilkinson, M. (2008). Quantifying the potential utility of phylogenetic characters. *TAXON*, *57*(1), 131–136. https://doi.org/10.2307/25065954

de Ruiter, D. J., DeWitt, T. J., Carlson, K. B., Brophy, J. K., Schroeder, L., Ackermann, R. R., Churchill, S. E., & Berger, L. R. (2013). Mandibular Remains Support Taxonomic Validity of *Australopithecus sediba*. *Science*, *340*(6129), 1232997. https://doi.org/10.1126/science.1232997

Dembo, M., Matzke, N. J., Mooers, A. Ø., & Collard, M. (2015). Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20150943. https://doi.org/10.1098/rspb.2015.0943

Dembo, M., Radovčić, D., Garvin, H. M., Laird, M. F., Schroeder, L., Scott, J. E., Brophy, J., Ackermann, R. R., Musiba, C. M., De Ruiter, D. J., Mooers, A. Ø., & Collard, M. (2016). The evolutionary relationships and age of *Homo naledi:* An assessment using dated Bayesian phylogenetic methods. *Journal of Human Evolution*, *97*, 17–26. https://doi.org/10.1016/j.jhevol.2016.04.008

Foley, R. A. (2016). Mosaic evolution and the pattern of transitions in the hominin lineage. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1698), 20150244. https://doi.org/10.1098/rstb.2015.0244

Gautney, J. R. (2023). A new approach to exploratory data analysis in hominin phylogenetic reconstruction. *Journal of Human Evolution*, *182*, 103412. https://doi.org/10.1016/j.jhevol.2023.103412

Haile-Selassie, Y., Gibert, L., Melillo, S. M., Ryan, T. M., Alene, M., Deino, A., Levin, N. E., Scott, G., & Saylor, B. Z. (2015). New species from Ethiopia further expands Middle Pliocene hominin diversity. *Nature*, *521*(7553), Article 7553. https://doi.org/10.1038/nature14448

Hawks, J., de Ruiter, D. J., & Berger, L. R. (2015). Comment on "Early *Homo* at 2.8 Ma from Ledi-Geraru, Afar, Ethiopia." *Science*, *348*(6241), 1326–1326. https://doi.org/10.1126/science.aab0591

Hendy, M. D., & Penny, D. (1982). Branch and bound algorithms to determine minimal evolutionary trees. *Mathematical Biosciences*, *59*(2), 277–290. https://doi.org/10.1016/0025-5564(82)90027-X

Huelsenbeck, J. P., & Bull, J. J. (1996). A Likelihood Ratio Test to Detect Conflicting Phylogenetic Signal. *Systematic Biology*, *45*(1), 92–98. https://doi.org/10.1093/sysbio/45.1.92

Kimbel, W. H. (2013). Hesitation on hominin history. *Nature*, *497*, 573–574.

Laurin, M. (2004). The Evolution of Body Size, Cope's Rule and the Origin of Amniotes. *Systematic Biology*, *53*(4), 594–622. https://doi.org/10.1080/10635150490445706

Ledogar, J. A., Senck, S., Villmoare, B. A., Smith, A. L., Weber, G. W., Richmond, B. G., Dechow, P. C., Ross, C. F., Grosse, I. R., Wright, B. W., Wang, Q., Byron, C., Benazzi, S., Carlson, K. J., Carlson, K. B., Pryor McIntosh, L. C., van Casteren, A., & Strait, D. S. (2022). Mechanical compensation in the evolution of the early hominin feeding apparatus. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1977), 20220711. https://doi.org/10.1098/rspb.2022.0711

Liu, W., Athreya, S., Xing, S., & Wu, X. (2022). Hominin evolution and diversity: A comparison of earlier-Middle and later-Middle Pleistocene hominin fossil variation in China. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1847), 20210040. https://doi.org/10.1098/rstb.2021.0040

Lockwood, C. A., & Tobias, P. V. (2002). Morphology and affinities of new hominin cranial remains from Member 4 of the Sterkfontein Formation, Gauteng Province, South Africa. *Journal of Human Evolution*, *42*(4), 389–450. https://doi.org/10.1006/jhev.2001.0532

Macho, G. (1994). Variation in enamel thickness and cusp area within human maxillary molars and its bearing on scaling techniques used for studies of enamel thickness between species. *Archives of Oral Biology*, *39*, 783–792. https://doi.org/10.1016/0003-9969(94)90008-6

Martin, J. M., Leece, A. B., Baker, S. E., Herries, A. I. R., & Strait, D. S. (2024). A lineage perspective on hominin taxonomy and evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, e22018. https://doi.org/10.1002/evan.22018

Mongle, C. S., Strait, D. S., & Grine, F. E. (2019). Expanded character sampling underscores phylogenetic stability of *Ardipithecus ramidus* as a basal hominin. *Journal of Human Evolution*, *131*, 28–39. https://doi.org/10.1016/j.jhevol.2019.03.006

Mongle, C. S., Strait, D. S., & Grine, F. E. (2023). An updated analysis of hominin phylogeny with an emphasis on re-evaluating the phylogenetic relationships of *Australopithecus* sediba. *Journal of Human Evolution*, *175*, 103311. https://doi.org/10.1016/j.jhevol.2022.103311

Mounier, A., Marchal, F., & Condemi, S. (2009). Is *Homo heidelbergensis* a distinct species? New insight on the Mauer mandible. *Journal of Human Evolution*, *56*(3), 219–246. https://doi.org/10.1016/j.jhevol.2008.12.006

Neaux, D., Bienvenu, T., Guy, F., Daver, G., Sansalone, G., Ledogar, J. A., Rae, T. C., Wroe, S., & Brunet, M. (2017). Relationship between foramen magnum position and locomotion in extant and extinct hominoids. *Journal of Human Evolution*, *113*, 1–9. https://doi.org/10.1016/j.jhevol.2017.07.009

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Parins-Fukuchi, C. (2018). Use of Continuous Traits Can Improve Morphological Phylogenetics. *Systematic Biology*, *67*(2), 328–339. https://doi.org/10.1093/sysbio/syx072

Parins-Fukuchi, C., Greiner, E., MacLatchy, L. M., & Fisher, D. C. (2019). Phylogeny, ancestors, and anagenesis in the hominin fossil record. *Paleobiology*, *45*(2), 378–393. https://doi.org/10.1017/pab.2019.12

Parravicini, A., & Pievani, T. (2019). Mosaic evolution in hominin phylogeny: Meanings, implications, and explanations. *Journal of Anthropological Sciences*, *97*, 45–68. https://doi.org/10.4436/JASS.97001

Penny, D., & Hendy, M. D. (1985). The Use of Tree Comparison Metrics. *Systematic Zoology*, *34*(1), 75–82. https://doi.org/10.2307/2413347

Post, N. W., Gilbert, C. C., Pugh, K. D., & Mongle, C. S. (2023). Implications of outgroup selection in the phylogenetic inference of hominoids and fossil hominins. *Journal of Human Evolution*, *184*, 103437. https://doi.org/10.1016/j.jhevol.2023.103437

Prendini, L., & Wheeler, W. C. (2005). Scorpion higher phylogeny and classification, taxonomic anarchy, and standards for peer review in online publishing. *Cladistics*, *21*(5), 446–494. https://doi.org/10.1111/j.1096-0031.2005.00073.x

Püschel, H. P., Bertrand, O. C., O'Reilly, J. E., Bobe, R., & Püschel, T. A. (2021). Divergence-time estimates for hominins provide insight into encephalization and body mass trends in human evolution. *Nature Ecology & Evolution*, *5*(6), Article 6. https://doi.org/10.1038/s41559-021-01431-1

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Ramirez Rozzi, F. V., Bromage, T., & Schrenk, F. (1997). UR 501, the Plio-Pleistocene hominid from Malawi. Analysis of the microanatomy of the enamel. *Comptes Rendus de l'Académie Des Sciences - Series IIA - Earth and Planetary Science*, *325*(3), 231–234. https://doi.org/10.1016/S1251-8050(97)88294-8

Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, *53*(1), 131–147. https://doi.org/10.1016/0025-5564(81)90043-2

Russo, G. A., & Kirk, E. C. (2013). Foramen magnum position in bipedal mammals. *Journal of Human Evolution*, *65*(5), 656–670. https://doi.org/10.1016/j.jhevol.2013.07.007

Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592–593. https://doi.org/10.1093/bioinformatics/btq706

Schmittbuhl, M., Rieger, J., Le Minor, J.-M., Schaaf, A., & Guy, F. (2007). Variations of the mandibular shape in extant hominoids: Generic, specific, and subspecific quantification using elliptical fourier analysis in lateral view. *American Journal of Physical Anthropology*, *132*(1), 119–131. https://doi.org/10.1002/ajpa.20476

Schrenk, F., Bromage, T. G., Betzler, C. G., Ring, U., & Juwayeyi, Y. M. (1993). Oldest *Homo* and Pliocene biogeography of the Malawi Rift. *Nature*, *365*(6449), 833–836. https://doi.org/10.1038/365833a0

Shao, Y., Zhou, L., Li, F., Zhao, L., Zhang, B.-L., Shao, F., Chen, J.-W., Chen, C.-Y., Bi, X., Zhuang, X.-L., Zhu, H.-L., Hu, J., Sun, Z., Li, X., Wang, D., Rivas-González, I., Wang, S., Wang, Y.-M., Chen, W., … Wu, D.-D. (2023). Phylogenomic analyses provide insights into primate evolution. *Science*, *380*(6648), 913–924. https://doi.org/10.1126/science.abn6919

Smith, M. R. (2019). TreeTools: Create, modify and analyse phylogenetic trees. *Comprehensive R Archive Network*. https://doi.org/10.5281/zenodo.3522725

Steel, M., & Penny, D. (2005). Maximum parsimony and the phylogenetic information in multistate characters. In V. A. Albert (Ed.), *Parsimony, phylogeny, and genomics*. Oxford University Press.

Strait, D. S. (2001). Integration, phylogeny, and the hominid cranial base. *American Journal of Physical Anthropology*, *114*(4), 273–297. https://doi.org/10.1002/ajpa.1041

Strait, D. S., Constantino, P., Lucas, P. W., Richmond, B. G., Spencer, M. A., Dechow, P. C., Ross, C. F., Grosse, I. R., Wright, B. W., Wood, B. A., Weber, G. W., Wang, Q., Byron, C., Slice, D. E., Chalk, J., Smith, A. L., Smith, L. C., Wood, S., Berthaume, M., … Ledogar, J. A. (2013). Viewpoints: Diet and dietary adaptations in early hominins: The hard food perspective: Diet and Dietary Adaptations in Early Hominins. *American Journal of Physical Anthropology*, *151*(3), 339–355. https://doi.org/10.1002/ajpa.22285

Strait, D. S., & Grine, F. E. (2001). The systematics of *Australopithecus garhi*. *Ludus Vitalis*, *9*(15), 109–135.

Strait, D. S., & Grine, F. E. (2004). Inferring hominoid and early hominid phylogeny using craniodental characters: The role of fossil taxa. *Journal of Human Evolution*, *47*(6), 399–452. https://doi.org/10.1016/j.jhevol.2004.08.008

Strait, D. S., Grine, F. E., & Moniz, M. A. (1997). A reappraisal of early hominid phylogeny. *Journal of Human Evolution*, *32*(1), 17–82. https://doi.org/10.1006/jhev.1996.0097

Thorley, J. L., Wilkinson, M., & Charleston, M. (1998). The Information Content of Consensus Trees. In A. Rizzi, M. Vichi, & H.-H. Bock (Eds.), *Advances in Data Science and Classification* (pp. 91–98). Springer. https://doi.org/10.1007/978-3-642-72253-0_12

Villmoare, B. A., & Kimbel, W. H. (2011). CT-based study of internal structure of the anterior pillar in extinct hominins and its implications for the phylogeny of robust *Australopithecus*. *Proceedings of the National Academy of Sciences*, *108*(39), 16200–16205. https://doi.org/10.1073/pnas.1105844108

Villmoare, B., Kimbel, W. H., Seyoum, C., Campisano, C. J., DiMaggio, E. N., Rowan, J., Braun, D. R., Arrowsmith, J. R., & Reed, K. E. (2015). Early *Homo* at 2.8 Ma from Ledi-Geraru, Afar, Ethiopia. *Science*, *347*(6228), 1352–1355. https://doi.org/10.1126/science.aaa1343

von Cramon-Taubadel, N., & Lycett, S. J. (2014). A comparison of catarrhine genetic distances with pelvic and cranial morphology: Implications for determining hominin phylogeny. *Journal of Human Evolution*, *77*, 179–186. https://doi.org/10.1016/j.jhevol.2014.06.009

White, T. (2003). Early Hominids—Diversity or Distortion? *Science*, *299*(5615), 1994–1997. https://doi.org/10.1126/science.1078294

Wood, B., & Constantino, P. (2007). *Paranthropus boisei*: Fifty years of evidence and analysis. *American Journal of Physical Anthropology*, *134*(S45), 106–132. https://doi.org/10.1002/ajpa.20732

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zollikofer, C. P. E., Lieberman, D. E., Guy, F., Pilbeam, D., Likius, A., Mackaye, H. T., Vignaud, P., & Brunet, M. (2005). Virtual cranial reconstruction of *Sahelanthropus tchadensis*. *Nature*, *434*.