Testing the utility of dental morphological trait combinations for inferring human neutral genetic variation

Hannes Rathmann[a,1] and Hugo Reyes-Centeno[a]

[a]Deutsche Forschungsgemeinschaft Center for Advanced Studies "Words, Bones, Genes, Tools," University of Tübingen, 72070 Tübingen, Germany

Edited by Leslea J. Hlusko, University of California, Berkeley, CA, and accepted by Editorial Board Member C. O. Lovejoy March 24, 2020 (received for review August 17, 2019)

Researchers commonly rely on human dental morphological features in order to reconstruct genetic affinities among past individuals and populations, particularly since teeth are often the best preserved part of a human skeleton. Tooth form is considered to be highly heritable and selectively neutral and, therefore, to be an excellent proxy for DNA when none is available. However, until today, it remains poorly understood whether certain dental traits or trait combinations preserve neutral genomic signatures to a greater degree than others. Here, we address this long-standing research gap by systematically testing the utility of 27 common dental traits and >134 million possible trait combinations in reflecting neutral genomic variation in a worldwide sample of modern human populations. Our analyses reveal that not all traits are equally well-suited for reconstructing population affinities. Whereas some traits largely reflect neutral variation and therefore evolved primarily as a result of genetic drift, others can be linked to nonstochastic processes such as natural selection or hominin admixture. We also demonstrate that reconstructions of population affinity based on many traits are not necessarily more reliable than those based on only a few traits. Importantly, we find a set of highly diagnostic trait combinations that preserve neutral genetic signals best (up to $\tilde{x}_r = 0.580$; 95% r range = 0.293 to 0.758; $P = 0.001$). We propose that these trait combinations should be prioritized in future research, as they allow for more accurate inferences about past human population dynamics when using dental morphology as a proxy for DNA.

dental morphology | ASUDAS | genetic drift | bioarchaeology | biodistance

Human dental morphology is highly diverse and varies among individuals and populations. Teeth are the hardest tissue in the human body, and as such, their remains are generally well preserved after death and inhumation, even when associated skeletal and endogenous DNA preservation is poor. As a result, dental morphology is widely used for inferring the biogeographical origin of deceased individuals, particularly when no other biological markers are available. Typical applications in the study of dental morphology include ancestry identification of unknown individuals in forensic cases (1, 2), the assessment of past population structure and history in archaeological contexts (3–9), and the reconstruction of hominin phylogenies in paleontological studies (10–12).

Dental morphology is routinely characterized using nonmetric traits by reference to standardized scoring protocols such as the Arizona State University Dental Anthropology System (ASUDAS) (13, 14). The ASUDAS catalogs a large number of common crown and root shape variants for the permanent adult dentition, which have been found to be differentially expressed across modern human populations and thus useful for population comparisons. Examples of common dental variants include the number of cusps and roots, the relative size of cusps, or the pattern of fissures, ridges, and grooves on tooth crowns. It is widely assumed that ASUDAS tooth variants are highly heritable, selectively neutral, and evolutionarily conservative, and that human dental diversity worldwide was generated by random evolutionary processes consisting of founder effects and genetic drift (15). Indeed, recent research in population and quantitative genetics has shown that neutral genetic variation and dental morphological variation across modern human populations is significantly correlated, as expected under neutrality (16, 17). Additionally, within-population dental morphological variation decreases with increasing geographical distance from Africa (18), a signature also found in neutral genomic datasets as a result of the demographic expansion of modern humans originating in Africa (19).

However, it is debated whether certain dental traits preserve neutral genetic signatures to a greater degree than others (20–22), and until now, there is no definitive list of key dental traits that are most useful for adequately capturing neutral genomic variation (15). As a rule of thumb, researchers therefore assume that phenotypic analyses based on many dental traits are more reliable than those based on only a few traits (14, 15). This assumption, however, has never been formally tested empirically and might be problematic because reconstructions of human genetic affinities based on nonneutrally evolving dental traits

Significance

Scientists across disciplines rely on human tooth morphology to infer genetic affinities for a variety of research questions, ranging from ancestry identification in forensic cases, to the reconstruction of population history and hominin phylogeny in archaeological and paleontological studies. However, it remains unclear whether certain dental traits preserve neutral genomic signatures to a greater degree than others. By testing the association of millions of different dental traits and trait combinations with neutral genomic markers across modern humans worldwide, we identify a set of highly diagnostic combinations that preserve maximum amounts of neutral genetic signals. These trait combinations should be prioritized in future research as they allow for more accurate inferences about past human population dynamics when DNA is not available.

Author contributions: H.R. designed the research with input from H.R.-C; H.R. developed the programming code and performed the analyses; H.R.-C. processed the genomic data; and H.R. and H.R.-C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. L.J.H. is a guest editor invited by the Editorial Board.

Data deposition: The data and code used for analyses are publicly accessible from the Zenodo repository at https://zenodo.org/record/3713179.

[1]To whom correspondence may be addressed. Email: hannes.rathmann@uni-tuebingen.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1914330117/-/DCSupplemental.

First published May 6, 2020.

may erroneously reflect mechanisms unrelated to genetic drift, such as convergent adaptation in response to shared environments.

A promising approach to address these matters is to quantify the correlation of biological affinity measures across worldwide modern human populations, derived independently from neutral genomic markers, on the one hand, and different morphological regions, on the other hand (23). Such analyses have already been successfully applied in a range of anthropological studies that attempted to disentangle the differential neutral genetic signals preserved in various anatomical parts of the human cranium (24–28). However, to our knowledge, such approaches have not yet been applied to the various dental morphological traits of the ASUDAS. Moreover, whereas previous genotype–phenotype investigations on cranial elements used predefined functional and developmental modules, such study design might be suboptimal in light of the complex modularity, ontogeny, and inheritance of phenotypes in general, and dental traits in particular (21, 22, 29–32). We therefore propose that testing all possible combinations of dental traits in preserving neutral genetic signals is a more promising approach than restricting analyses to only individual traits or predefined trait combinations.

Here, we address these research gaps by systematically testing the utility of different dental morphological traits and trait combinations in reflecting neutral genomic patterns of variation using an exhaustive search algorithm. To assess the utility of a given trait or trait combination, we estimated dental phenotypic distances (D_P) between 20 worldwide modern human populations, and compared them to neutral genomic distances (D_G) among the same, or closely matched, populations (*SI Appendix*, Table S1 and Fig. S1). The congruence between D_P and D_G was quantified by linear regression of the off-diagonal values in the two distance matrices using Pearson's product-moment correlation coefficient (r). An r value close to 1 indicates that a trait or trait combination reliably reflects neutral genomic patterns of variation, whereas an r value close to 0 indicates that a trait or trait combination is less congruent with neutral expectations. To account for stochastic variation inherent to a neutral model of evolution, we calculated r for a given dental trait or trait combination 1,000 times, each time comparing the D_P matrix to different D_G matrices randomized by subsampling genomic loci. We then reported the median of the resulting distribution of r values as a point estimate and as the utility estimator for a given trait or trait combination ($\tilde{x}_r$). To measure the spread of r values around $\tilde{x}_r$, we constructed an interpercentile range accounting for 95% of the distribution of r values. We also calculated P values by permutation under the null hypothesis of no association between D_P and D_G, which permitted us to assess how frequently the utility estimate $\tilde{x}_r$ was produced by chance alone. Our analysis is based on a large microsatellite loci database (33) and the hitherto largest available ASUDAS dental trait database (15), enabling us to quantify the utility of 27 dental traits and all 134,217,700 possible combinations of these traits.

Results

Fig. 1 displays the utility of 27 dental morphological traits considered in the ASUDAS for reconstructing neutral genetic variation across worldwide modern human populations (*SI Appendix*, Table S2). We found that the various traits exhibit disparate levels of utility, with median utility estimates ($\tilde{x}_r$) ranging from −0.039 (95% r range = −0.167 to 0.192; $P = 0.576$) to 0.108 (95% r range = −0.107 to 0.471; $P = 0.129$). None of the $\tilde{x}_r$ utility estimates is statistically significant at $\alpha = 0.05$. The $\tilde{x}_r$ utility estimates are neither correlated with the average frequency of traits across populations (*SI Appendix*, Fig. S2) nor with the range of trait frequencies across populations (*SI Appendix*, Fig. S3).

The utility results for all 134,217,700 possible combinations of dental traits are listed in a comprehensive table publicly available on Zenodo (34) at https://zenodo.org/record/3713179. The different trait combinations yielded vastly disparate $\tilde{x}_r$ utility estimates ranging from −0.036 (95% r range = −0.183 to 0.305; $P = 0.475$) to 0.580 (95% r range = 0.293 to 0.758; $P = 0.001$). Most of the $\tilde{x}_r$ utility estimates (99.4%) are statistically significant at $\alpha = 0.05$.

To survey which dental trait combinations were more useful and which ones were less informative, we plotted the proportional composition of traits involved in trait combinations yielding different $\tilde{x}_r$ utility estimates. For this, we first apportioned the generated range of $\tilde{x}_r$ values (−0.036 to 0.580) into 20 equally sized utility windows (resulting in a width of 0.031 each). We then quantified the number of times that a trait was represented in each window (Dataset S1 and *SI Appendix*, Fig. S4) and visualized the proportional composition of traits in each window using a stacked bar chart (Fig. 2). We found that dental trait combinations falling into the highest $\tilde{x}_r$ utility window (0.549 to 0.580) predominantly comprise the following six traits, at a frequency of >90% each: mesial ridge (UC), distal accessory ridge (UC), protostylid (LM1), lingual cusp number (LP2), cusp 6 (LM1), and cusp 7 (LM1). Dental trait combinations in the lowest $\tilde{x}_r$ utility window (−0.036 to −0.005) comprise the following eight traits: cusp number (LM2), tuberculum dentale (UI2), Carabelli trait (UM1), root number (LC), root number (LM2), Tomes' root (LP1), root number (UM2), and hypocone (UM2). Overall, the utility estimate $\tilde{x}_r$ of individual traits is a general indication of its effect on trait combination, where high-utility traits appear more frequently in high-utility combinations and vice versa (*SI Appendix*, Table S4).

To find dental trait combinations that performed best, we first searched for the top-performing trait combination that achieved the highest $\tilde{x}_r$ utility estimate ($n_{traits} = 19$; $\tilde{x}_r = 0.580$; 95% r range = 0.293 to 0.758; $P = 0.001$). We then compared the distribution of r values of the top-performing trait combination to the distribution of r values of all other 134,217,699 trait combinations. In total, we found a set of 267 combinations that all performed equally well in capturing maximum amounts of neutral genomic variation (Dataset S2). These 267 combinations consist of trait batteries ranging from 14 to 20 traits and always comprise the following five traits: mesial ridge (UC), distal accessory ridge (UC), protostylid (LM1), lingual cusp number (LP2), and cusp 6 (LM1).

Fig. 3 highlights the superior utility of the top-performing trait combination (Dataset S2; $n_{traits} = 19$; $\tilde{x}_r = 0.580$; 95% r range = 0.293 to 0.758; $P = 0.001$) in comparison to the full trait battery ($n_{traits} = 27$; $\tilde{x}_r = 0.428$; 95% r range = 0.146 to 0.688; $P = 0.001$). All plots convey how the 19-trait combination captures neutral genomic affinities across populations better than the full 27-trait battery, both in the superimposition of D_G on D_P in Procrustes ordination space (Fig. 3*A* versus Fig. 3*C*) and in yielding lower residual values in the D_G–D_P regression (Fig. 3*B* versus Fig. 3*D*). Notably, the Procrustes plot based on the 19-traits combination clearly shows major continental clusters of populations (Fig. 3*A*), in comparison to the full 27-trait battery (Fig. 3*C*).

Finally, in order to explore whether phenotypic inferences about neutral genetic variation based on many dental traits are more useful than those based on only a few traits, we plotted the distribution of $\tilde{x}_r$ and associated P values resulting from trait batteries of different sizes (from single traits to the full 27-trait battery) using violin plots (Fig. 4). On average, increasing the number of traits leads to a logarithmic increase in median $\tilde{x}_r$ values within a trait battery size class until $\tilde{x}_r$ approximates a plateau value of 0.428 (Fig. 4*A*). At the same time, increasing the number of traits reduces the variance of $\tilde{x}_r$ values within a trait battery size class. Nevertheless, the highest $\tilde{x}_r$ utility estimates were achieved by using a rather limited number of traits, ranging from 14 to 20 traits (Dataset S2), indicated by a red box in Fig. 4*A*. On average, increasing the number of traits also leads to a logarithmic decrease in median P values within a trait battery size class, and trait combinations comprising >15 traits (no

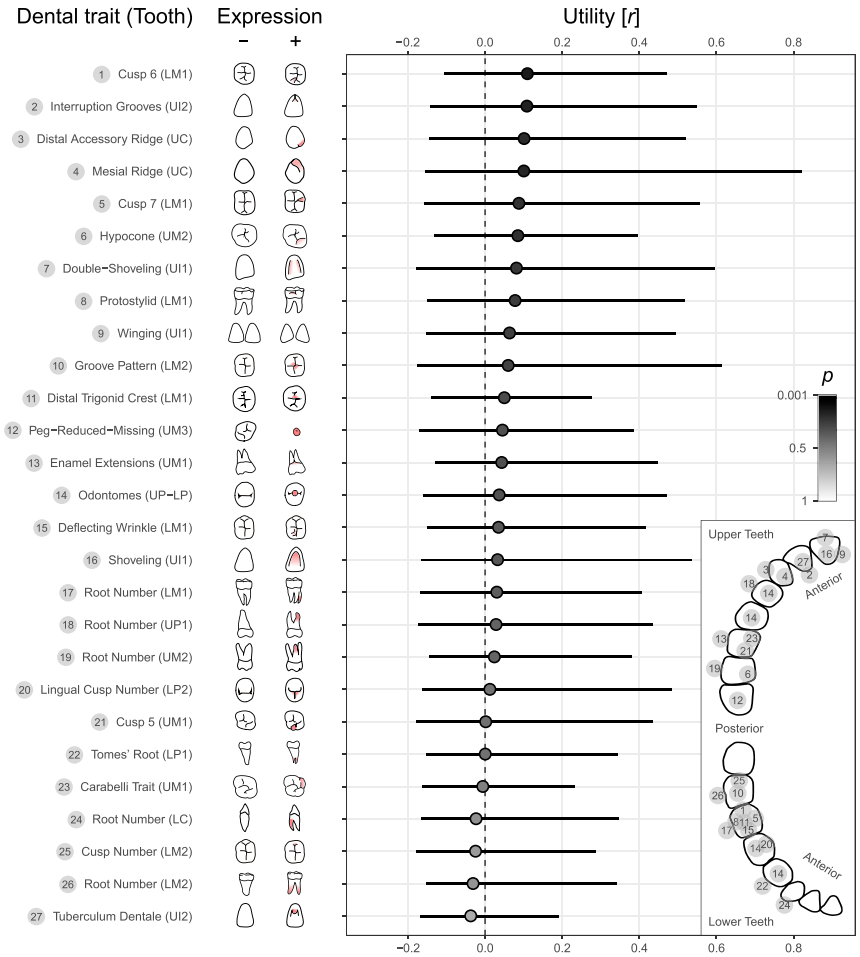

Fig. 1. Dental nonmetric traits in the ASUDAS and their estimated utility in reflecting human neutral genetic variation. Trait utility is calculated as correlation (r) between neutral genetic and dental phenotypic distances across modern human populations, with 1,000 loci-resampled iterations providing a distribution of r values to account for stochastic variation inherent to a neutral model of evolution (*Materials and Methods*). Dots display the median utility estimate ($\tilde{x}_r$). Bars indicate 95% of the distribution of r values around $\tilde{x}_r$. Grayscale of dots indicate P values for $\tilde{x}_r$ under the null hypothesis of no association between genetic and phenotypic variation, calculated as the proportion of r values from permuted data that are equally high or higher than $\tilde{x}_r$ obtained from the observed data (*Materials and Methods*). None of the $\tilde{x}_r$ utility estimates is statistically significant at $\alpha = 0.05$ (see grayscale legend on the *Right*). Traits are shown in descending order of $\tilde{x}_r$. Anatomical trait descriptions are provided in ref. 14. Abbreviations in brackets denote key tooth scored: C, canine; I, incisor; L, lower mandibular dentition; M, molar; P, premolar; U, upper maxillary dentition; Number, tooth positioning. Tooth figures schematically display trait expressions dichotomized into trait absence (–) and trait presence (+). The box in the *Lower Right* corner displays the location of traits within the dentition.

matter which traits of the 27 traits are chosen) are always significant at $\alpha = 0.05$ (Fig. 4*B*).

Discussion

Here, we assessed the utility of different ASUDAS dental traits and trait combinations in reflecting global patterns of modern human neutral genetic variation. We did so by developing an exhaustive search algorithm that systematically tested all possible combinations of dental traits while accounting for stochastic variation inherent to a neutral model of evolution, drawing on the largest dental morphological and microsatellite genomic datasets currently available. Our results clearly show that not all dental traits and trait combinations are equally well-suited for inferring neutral genetic affinities. They highlight that phenotypic inferences about neutral genetic variation are better when based on trait combinations rather than individual traits. Importantly, we were able to isolate a set of 267 highly diagnostic trait combinations that preserve neutral genetic signals best (Dataset S2). These trait combinations always comprise the following five traits: mesial ridge (UC), distal accessory ridge (UC), protostylid (LM1), lingual cusp number (LP2), and cusp 6 (LM1). The combinatorial power of these traits can be explained by the fact that together they reflect major components of phenotypic structure at a global level. Whereas mesial ridge (UC) and distal accessory ridge (UC) partition global dental diversity into African and non-African components, lingual cusp number (LP2), protostylid (LM1), and cusp 6 (LM1) differentiate East Asians and Native Americans from other populations (15). These geographical patterns are consistent with observations from genomic structure at microsatellite loci, where population clusters are anchored by populations from Africa and America as a result of high diversity in the former and low diversity in the latter (19, 33, 35, 36). The addition of other dental traits serves to capture more subtle variation at lower geographic scales. We propose that any of the 267 trait combinations in Dataset S2 should be prioritized in future research, as they allow for more accurate inferences about global human population history when using dental morphology as a proxy for neutral DNA.

In contrast, we found several dental traits that were comparatively less informative about neutral genetic variation, most notably traits with near-zero $\tilde{x}_r$ utility estimates (Fig. 1) and traits that were never represented in trait combinations falling into the

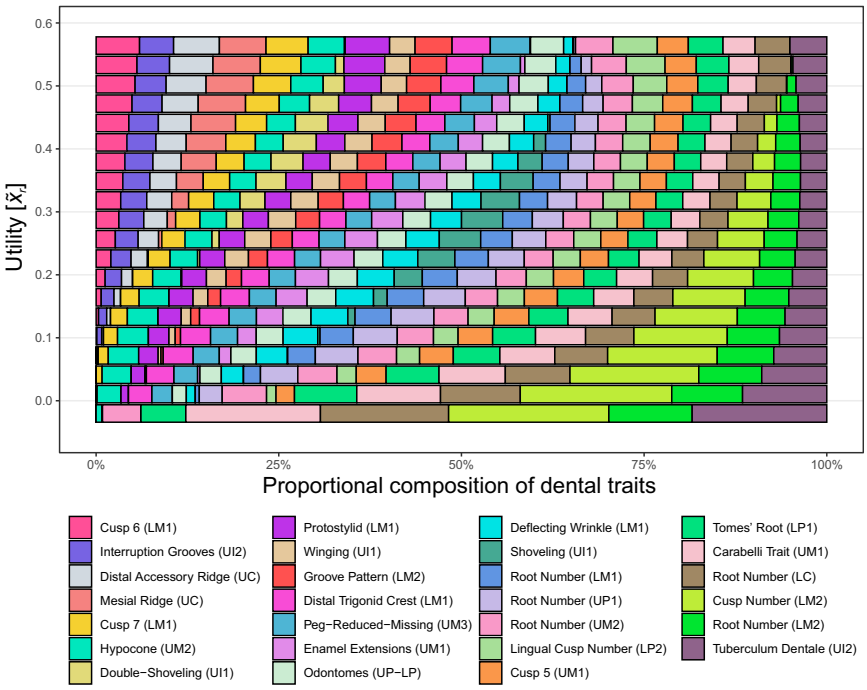

Fig. 2. Stacked bar chart showing the proportional composition of 27 dental nonmetric traits involved in 134,217,700 possible trait combinations yielding different $\tilde{x}_r$ utility estimates (from −0.036 to 0.580) apportioned into 20 equally sized $\tilde{x}_r$ utility windows (with a width of 0.031 each). The $\tilde{x}_r$ utility is calculated as the correlation between neutral genetic and dental phenotypic distances across modern human populations, while accounting for stochastic variation inherent to a neutral model of evolution (*Materials and Methods*). Color-coding denotes dental traits of the ASUDAS. Anatomical trait descriptions are provided in ref. 14. Abbreviations in brackets denote key tooth scored: C, canine; I, incisor; L, lower mandibular dentition; M, molar; P, premolar; U, upper maxillary dentition; Number, tooth positioning.

highest $\tilde{x}_r$ utility window (Fig. 2 and *SI Appendix*, Fig. S4). We reason that a large portion of the morphological variation in these traits is most likely linked to nonneutral evolutionary factors, such as natural selection. This interpretation is consistent with previous functional adaptation hypotheses relating shoveling (UI1) to enhanced biting performance (37), Carabelli trait (UM1) and cusp 5 (UM1) to improved chewing (38–40), and root number (UM1) to better molar retention in populations with high masticatory loading (41). Shoveling (UI1) and Carabelli trait (UM1) have also been found to be associated with environmental factors, suggesting that they reflect adaptations to selective pressures rather than being a result of genetic drift (20). Shoveling (UI1), double-shoveling (UI1), and cusp number (LM2) have been found to be associated with the ectodysplasin A receptor gene (*EDAR*) (42–45), which is a functional genomic region under positive selection (46). The high-utility mesial ridge (UC) was found to be linked to *EDAR* as well (44), although the association was only significant when tested in combination with other traits and not when tested individually after Bonferroni correction for multiple testing. Interestingly, *EDAR* has a range of pleiotropic effects on ectodermally derived structures, such as hair, mammary glands, and teeth (47). It is, therefore, likely that some of the dental traits linked to this gene are not direct targets of selection but rather "hitchhiking" when selection acts on other phenotypes (43). We propose that other dental traits that were comparatively less informative about neutral genetic variation in our study could likewise be linked to functional genomic regions under selection. Alternatively, dental traits that do not follow neutral expectations could also be linked to hominin admixture. For example, it has been suggested that the high prevalence of root number (LM1) in modern Asian populations is the result of Denisovan introgression into modern *Homo sapiens* (48, 49). In sum, the inclusion of dental traits that do not follow neutral expectations should be carefully reviewed or, better still, omitted in future phenotypic analyses aimed at reconstructing neutral genetic affinities in modern humans.

When we explored whether phenotypic inferences about neutral genomic variation based on many dental traits are more useful than those based on only a few traits (Fig. 4), we found that a larger battery of traits leads to phenotypic inferences that are, on average, increasingly congruent with neutral genomic expectations, which confirms standard assumptions in dental anthropological research (14, 15). However, we found that the increase in $\tilde{x}_r$ is logarithmic and not linear, with a gradual tipping point at which adding more traits only adds little new neutral genetic information. We observe trait combinations providing the highest $\tilde{x}_r$ utility estimates among trait battery sizes of 14 to 20 traits (see red box in Fig. 4*A*). Using trait batteries with >20 traits results in increasingly homogenous $\tilde{x}_r$ utility estimates with increasing minima, but also in decreasing maxima. Thus, more traits do not necessarily provide higher concordance with neutral expectations, since combinations with more traits will likely contain traits that are less informative about neutral genomic variation. As a result, we expect that many previous studies following the standard recommendation of using the maximum number of traits available are biased by traits that have not differentiated in a neutral fashion. Nevertheless, regardless of the $\tilde{x}_r$ effect size, our significance test results indicate that using at least 16 of any of the ASUDAS traits will reliably capture neutral genomic variation at $\alpha = 0.05$ (see red line in Fig. 4*B*).

Our results have implications for a wide range of previous bioarchaeological studies. For example, several studies have used dental nonmetric traits to test competing out-of-Africa dispersal models of modern humans during the Late Pleistocene, drawing on the observation that within-population morphological diversity decreases with increasing geographic distance from Africa (50, 51). These studies have supported models that are sometimes in conflict with other lines of evidence, including those applying

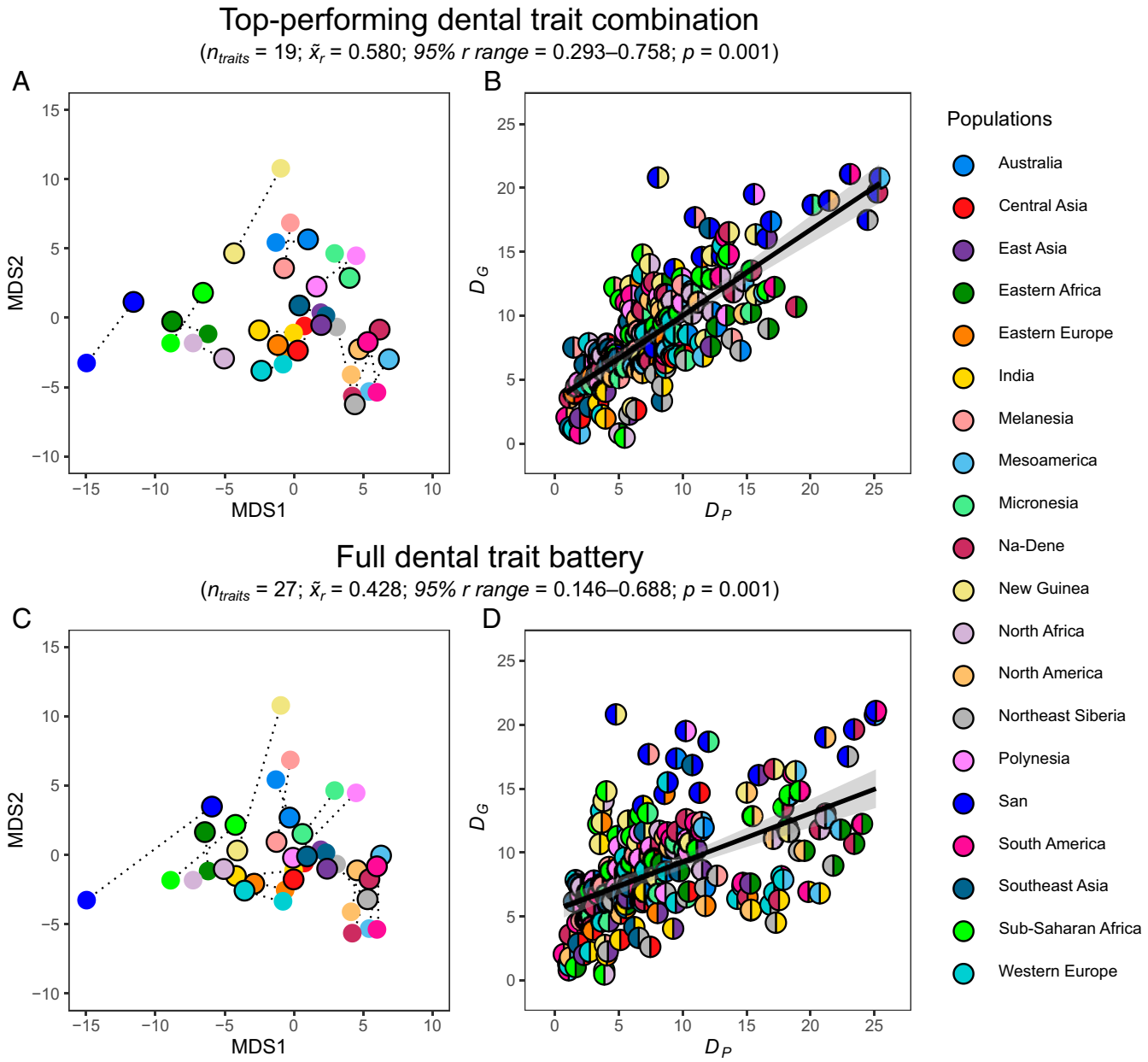

Fig. 3. Comparison of one of the most useful dental nonmetric trait combinations found in this study (n_{traits} = 19; $\bar{x}_r$ = 0.580; 95% r range = 0.293 to 0.758; P = 0.001) versus the utility of the full trait battery (n_{traits} = 27; $\bar{x}_r$ = 0.428; 95% r range = 0.146 to 0.688; P = 0.001). The $\bar{x}_r$ utility is calculated as the correlation between neutral genetic distances (D_G) and dental phenotypic distances (D_P) across modern human populations, while accounting for stochastic variation inherent to a neutral model of evolution (*Materials and Methods*). For visualization, we used the full genomic loci dataset (n_{loci} = 645). (*A*) Procrustes superimposition plot maximizing similarity between D_G and D_P in a 2D multidimensional scaling (MDS) map, calculated for the 19-trait combination. Color-coding of points denotes populations (see right-side legend). Points with no border are D_G estimates. Point with a black border are D_P estimates. Pairwise population residuals between D_G and D_P are displayed by black dotted lines. (*B*) Regression plot of pairwise population relationships between D_G and D_P displayed with a fitted linear regression line and estimated 95% confidence interval, calculated for the 19-trait combination. Color-coding of points denotes population pairs. (*C*) Procrustes superimposition plot for the full 27-trait battery. (*D*) Regression plot for the full 27-trait battery.

the same study design to cranial morphology (52). This discrepancy can be partly explained by the fact that the studies using dental traits may not have captured sufficient neutral genomic variation to allow for proper inference. Indeed, a reevaluation of the ASUDAS traits employed by refs. 50 and 51 indicates that the trait combination captures significant neutral genomic variation, but the strength of association is rather weak ($\bar{x}_r$ = 0.278; 95% r range = −0.025 to 0.585; P = 0.010). Thus, previous inferences about out-of-Africa dispersal models based on dental morphology should be treated with caution, and when possible, they should be reconsidered using one of the 267 highest-utility trait combinations reported here (Dataset S2).

Our results also have direct implications for the field of forensic death investigations. For example, the latest version of the rASUDAS program (2), a web-based application for estimating the ancestry of an unknown individual based on its suite of crown and root traits, utilizes a battery of 21 ASUDAS traits, following the standard assumption that using more traits results in better inferences on genetic affinities. Our results show that this

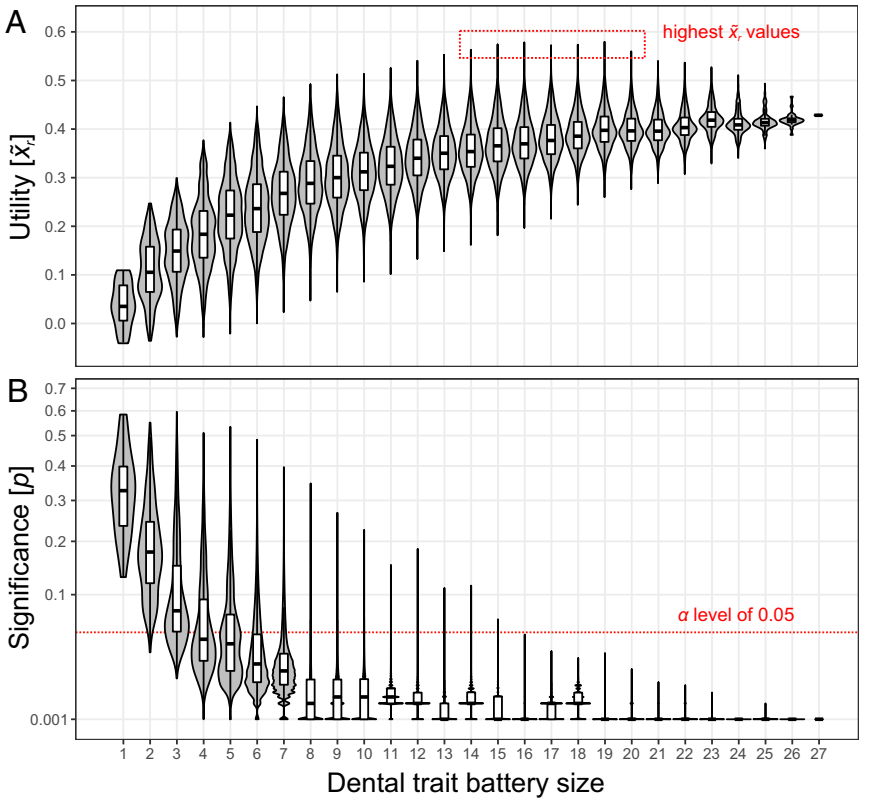

Fig. 4. Estimated utility for 27 dental nonmetric traits and 134,217,700 possible trait combinations apportioned into trait batteries of different size (from 1 to 27 traits). (*A*) Violin plots showing the distribution of $\tilde{x}_r$ utility estimates per trait battery size, where $\tilde{x}_r$ is calculated as the correlation between neutral genetic and dental phenotypic distances across modern human populations, while accounting for stochastic variation inherent to a neutral model of evolution (*Materials and Methods*). Box plots are superimposed to show median values and interquartile ranges. The red dotted box indicates the highest $\tilde{x}_r$ utility estimates found in our study, achieved by trait combinations of batteries ranging from 14 to 20 traits. (*B*) Violin plots showing the distribution of (square root transformed) *P* values associated with $\tilde{x}_r$ utility estimates per trait battery size under the null hypothesis of no association between genetic and phenotypic variation, where *p* is calculated as the proportion of correlations from permuted data that are equally high or higher than the utility estimator $\tilde{x}_r$ obtained from the observed data (*Materials and Methods*). The red dotted line represents the conventional α level of 0.05.

particular 21-trait combination is capturing neutral genomic variation significantly, but only moderately well ($\tilde{x}_r = 0.326$; 95% *r* range = 0.084 to 0.585; $P = 0.001$). This could partly explain the suboptimal rASUDAS ancestry classification accuracy ranging from 51.8 to 72.2% (2). We anticipate that using the highest-utility dental trait combinations (Dataset S2) would substantially increase classification accuracy, given the fact that the top-performing 19-trait combination ($\tilde{x}_r = 0.580$; 95% *r* range = 0.293 to 0.758; $P = 0.001$) successfully separates populations into broad geographical clusters (Fig. 3*A*). However, traits found to be of high utility in our study, such as mesial ridge (UC) and distal accessory ridge (UC), are not implemented in the current rASUDAS application, highlighting the necessity to further develop this important forensic tool. Nevertheless, we note that dental traits associated with functional genomic regions under selection (e.g., *EDAR* related traits such as shoveling) are still useful for forensic ancestry classification, given that they are found at extreme high or extreme low frequencies in different populations across the globe (15). Further research should clarify which dental traits perform best for discriminating between different ancestry groups, especially when investigations are performed at different geographic scales.

Following a long research tradition in biological anthropology that seeks to identify skeletal regions that preserve maximum amounts of neutral genomic signals (24–28, 53), our study adds an important and long overdue contribution to this domain—that of nonmetric traits of the dentition. The quantified degree of congruence between neutral genomic variation and dental trait combinations of highest utility reported here (e.g., n_{traits} = 19; $\tilde{x}_r = 0.580$; 95% *r* range = 0.293 to 0.758; $P = 0.001$) is comparable to the highest congruence for different anatomical regions of the cranium ($r = 0.563$ to 0.665; $P < 0.001$) found by a study using a methodological setup similar to ours (25). Thus, dental and cranial morphology appear to be equally well suited for inferring neutral genetic variation. However, we caution that previous studies on the association of cranial and genomic variation are not directly comparable to ours since none of the previous studies accounted for stochastic variation inherent to a neutral model of evolution. Moreover, different populations have been sampled, and different methodological approaches for quantifying morphology have been employed. Importantly, whereas previous investigations on cranial bones used predetermined anatomical regions (24–27), the study design used here tested all possible combinations of traits, something that has not yet been attempted for cranial data. While future work will serve to further clarify the combinatorial utility of cranial regions, the better state of preservation of teeth and their higher recovery in forensic, archaeological, and paleontological contexts ultimately provides the advantage of larger sample sizes and more robust statistical analyses.

We note that the $\tilde{x}_r$ utility estimates reported here are biased toward not finding significant associations between neutral genetic and dental morphological variation. First, we compared matched but unpaired datasets, with dental samples coming from

different individuals than those sampled for genomic loci. Although it is a well-established procedure to compare unpaired data at a global scale (16, 24–28), any comparison of genetic and morphological affinities in unpaired samples tends to reduce the magnitude of their association given that between-population variation is low compared to within-population variation (54). Second, it is possible that the dichotomized dental trait data employed in this study are not capturing adequate morphological variation. Trait dichotomization is a well-established approach with the advantage of minimizing observer error (14, 15); however, it also reduces information about variation in trait expressivity and may skew phenotypic distance estimates (55). Third, we used a phenotypic distance statistic for measuring between-population variation that is assuming complete independence among traits. Although it has been repeatedly shown that correlations among dental traits recorded on key teeth are low (8, 9, 14, 15, 44), even modest trait correlations may lead to overrepresented variation from traits that co-occur. Given the limitations of our study, the reported $\tilde{x}_r$ utility estimates must be considered as minimum and not as exact estimates of the strength of correlation between neutral genetic and dental morphological variation. Nevertheless, because it is likely that the above-mentioned limitations apply to all generated $\tilde{x}_r$ utility estimates in a similar manner, they may not bias our conclusions since we are interested in the relative utility of the different dental trait combinations to each other.

We anticipate that the results of our study will serve as an important reference for a wide range of future dental morphological investigations, allowing researchers to select dental trait combinations that most reliably reflect neutral genetic signatures in modern humans. For this, we advise relying on trait combinations with highest $\tilde{x}_r$ utility estimates (Dataset S2), or, alternatively, to remove traits with near-zero $\tilde{x}_r$ utility estimates (Fig. 1). The generated table of $\tilde{x}_r$ utility estimates for all possible 134,217,700 trait combinations (34) can also be used to validate the performance of a particular trait combination employed in previous studies. Finally, yet importantly, we emphasize that researchers should continue to collect and report on the full battery of dental morphological traits. Continuing to collect as many traits as possible is important because the ASUDAS is continuously growing as new traits are proposed for inclusion (15), including those that capture dental variation across hominin taxa (10, 12). We caution that because our results reflect genomic and phenotypic variation in recent modern humans, further work is necessary to apply them to the fossil record. Nevertheless, future research pairing ancient genomic and phenotypic data has great potential to further fine-tune our results at deeper time depths using the conceptual template we have laid out here, paving the way for testing new dental trait combinations useful in reconstructing human evolutionary history. More broadly, we propose that dental traits that were comparatively less informative about neutral genetic variation in our study could be linked to functional genomic regions, and we recommend that future genome-wide association studies should further investigate these potential dental trait candidates under selection, leading to exciting new research directions.

Materials and Methods

Matching Population Samples. Materials for this study comprise two different types of data: 1) dental nonmetric traits and 2) single-tandem repeat (STR) alleles of microsatellite loci across the autosomal genome. All data were taken from existing databases (15, 33). We matched datasets for 20 globally distributed modern human populations for which both morphological and genetic data were available (*SI Appendix*, Table S1 and Fig. S1). Populations were chosen for inclusion in this study based on three criteria: 1) availability of dental nonmetric trait data; 2) availability of STR allele data; and 3) sample antiquity such that none of the samples consists of exclusively archaeological material dated older than 2,000 y, so as to control for temporal bias. In instances where exact population matches could not be achieved, a geographically proximate population with ethnolinguistic affinities was selected.

Dental Nonmetric Trait Data. The dental nonmetric trait data were obtained from the hitherto largest available global database comprising observations of 27 dental traits scored for more than 11,000 individuals from several populations of modern humans (15). Most of the individuals come from archaeological and historical skeletal series dated to a few hundred years old. The majority of the samples were collected by C. G. Turner II and were later enlarged by work of G. R. Scott, J. D. Irish, and D. E. Hawkey. All workers used the ASUDAS (13) to collect dental trait observations. The ASUDAS comprises a reference set of dental casts illustrating expression levels for various traits alongside specific instructions that ensure a standardized scoring procedure, which minimizes intraobserver and interobserver error. Scoring followed the individual count method (56), where a trait was counted only once per dentition, regardless of whether or not the trait appeared bilaterally. In cases where a trait was expressed asymmetrically, the side with the highest expression level was scored. Dental trait expression scores were collapsed into simplified binary dichotomies of absence or presence in order to calculate trait frequencies per population. Dichotomization is based on established breakpoints that best represent easily recognizable and replicable points along the trait expression scale (*SI Appendix*, Table S3). While dichotomization reduces information about variation in trait expressivity, it has the advantage of further minimizing observer error (14, 15). In addition, trait frequencies are expected to be correlated with the level of trait expressivity within a population under a threshold model of quasicontinuous variation (57). Dental traits listed in the ASUDAS have little or no sexual dimorphism (14, 15); therefore, it is a standard procedure to pool sexes (4, 6, 8, 12). Population comparisons based on ASUDAS dental traits typically focus on key teeth (usually the most mesial member of a tooth district) because these are considered the most stable members in terms of development and evolution (15) and are largely independent from each other (8, 9, 14, 15, 44). Dental trait frequencies per population are calculated as the average of several trait frequencies estimated for various groups in each population. The average trait frequencies of the 20 populations used for analysis are based on a total of 185 groups, with population representation varying from 3 to 25 groups. Ranges of group trait frequencies for each population are provided in ref. 15. Average trait frequencies for the 20 populations used for analysis are provided in Dataset S3.

STR Allele Data. The STR allele data were obtained from a global dataset compiling modern human microsatellite genotypes at 645 common loci (33). The dataset comprises several published studies, including the global samples of the Human Genome Diversity Project deposited at Centre d'Etude du Polymorphisme Humain (HGDP-CEPH Human Genome Diversity Cell Line Panel; ref. 58), as well as several regional population studies. Data filtering of the different datasets consisted of removing microsatellites with >10% missing data, individuals with >27.5% missing data, data duplicates, and first- and second-degree relative pairs (33). Data from regional populations were merged with the global HGDP-CEPH dataset, aligning allele sizes to the latter. Matching this dataset with the 20 dental populations resulted in more than 4,000 individuals from 213 groups (*SI Appendix*, Table S1 and Fig. S1). For each population, we extracted mean allele sizes (Dataset S4).

Testing the Utility of Different Dental Trait Combinations. We performed an exhaustive search to systematically test the utility of different ASUDAS dental traits and trait combinations for phenotypic analysis. With the ASUDAS dataset used in our analysis, we tested 27 single dental traits and all 134,217,700 dental trait combinations possible. The utility of a given trait or trait combination was assessed by estimating dental phenotypic distance values (D_P) between 20 worldwide modern human populations, and by comparing them to neutral genomic distance values (D_G) among the same, or closely matched, populations.

For each dental trait or trait combination, pairwise D_P values among all sampled populations were calculated using the Euclidean squared (D^2) distance formula as follows:

$$D_{ij}^2 = \sum_{k=1}^{n} (z_{ik} - z_{jk})^2,$$

where D_{ij}^2 is the Euclidean squared distance between the two populations i and j; z_{ik} and z_{jk} are threshold values of the dental trait k for populations i

and j, respectively; and n is the number of analyzed dental traits. The threshold values z_{ik} and z_{jk} were estimated using a probit function as $z_{ik} = \text{probit}(p_{ik})$ and $z_{jk} = \text{probit}(p_{jk})$, where p is the percentage of dental trait k present in populations i and j, respectively. Under a threshold model of quasicontinuous variation, z_{ik} and z_{jk} are analogous to means because binary dichotomies employed for dental nonmetric traits code an underlying normally distributed continuous variable with unit SD (59).

Pairwise D_G values among all sampled populations were calculated using the delta-mu squared ($\delta\mu^2$) distance equation (60) as follows:

$$\delta\mu_{ij}^2 = \left(\sum_{k=1}^{n}\left(\mu_{ik} - \mu_{jk}\right)^2\right) \Big/ n,$$

where $\delta\mu_{ij}^2$ is the delta-mu squared distance between two populations i and j; μ_{ik} and μ_{jk} are the means of allele sizes in locus k for populations i and j, respectively; and n is the number of analyzed loci. Both distance statistics, D_G and D_P, are comparable to each other because both measure squared pairwise differences in mean values among populations and their distance values are expected to increase with time in diverging populations (61).

The congruence between D_P and D_G was assessed by linear regression of the off-diagonal values in the two distance matrices using the Pearson product-moment correlation coefficient (r). An r value close to 1 indicates that a trait or trait combination reliably reflects neutral genomic patterns of variation, whereas an r value close to 0 indicates that a trait or trait combination is less congruent with neutral expectations.

To account for stochastic variation inherent to a neutral model of evolution, we calculated r for a given dental trait or trait combination 1,000 times, each time comparing the D_P matrix to different D_G matrices arrived at by resampling the microsatellite loci data (23, 62). In each resampling iteration, we randomly subsampled the same number of loci as there are dental traits in a given trait combination. This sampling strategy is consistent with population and quantitative genetics theory, where a completely heritable, additive, and selectively neutral phenotypic trait is approximately as informative about population differentiation as a single neutral genomic locus, regardless of how many loci influence the phenotypic trait (63, 64). We then reported the median r value from the resulting distribution of r values as a point estimate and as the utility estimator for a given trait or trait combination ($\tilde{x}_r$). To measure the spread of r values around $\tilde{x}_r$, we constructed an interpercentile range from the 2.5th to the 97.5th percentile accounting for 95% of the distribution of r values.

To assess statistical significance of the $\tilde{x}_r$ utility estimate, we first estimated a null distribution of r values by comparing 1,000 permuted D_P matrices (where rows and columns were randomly rearranged) to loci-resampled D_G matrices. We then calculated the P value as the proportion of r values from the null distribution that are equally high or higher than the utility estimator $\tilde{x}_r$ obtained from the observed data. This permutation test permitted us to assess how frequently the $\tilde{x}_r$ utility estimate from the observed data were produced by chance alone. To account for multiple testing, we ran a Benjamini–Hochberg P value adjustment that controls for the false-discovery rate at 5%, which is the expected proportion of false discoveries among the rejected null hypotheses of no association (65).

Visualizing the Utility of Dental Traits and Trait Combinations. To visualize the differential utility of the 27 individual dental traits for inferring neutral genetic variation, we plotted $\tilde{x}_r$ utility estimates for each trait using a scatterplot with error bars displaying interpercentile ranges accounting for 95% of the distribution of r values (Fig. 1). To survey the differential utility of dental trait combinations, we plotted the proportional contribution of traits involved in trait combinations yielding different $\tilde{x}_r$ utility estimates. For this, we first apportioned the generated range of $\tilde{x}_r$ values (−0.036 to 0.580) into 20 equally sized utility windows (resulting in a width of 0.031 each). We then quantified the number of times that a trait was represented in each window (Dataset S1 and *SI Appendix*, Fig. S4) and visualized the proportional contribution of traits in each window using a stacked bar chart (Fig. 2).

Finding the Most Useful Dental Trait Combinations. To find dental trait combinations that are the most useful for inferring neutral genetic variation, we first selected the top-performing trait combination achieving the highest $\tilde{x}_r$ utility estimate in our study ($n_{traits} = 19$; $\tilde{x}_r = 0.580$; 95% r range = 0.293 to 0.758; $P = 0.001$). We then compared the distribution of r values of this top-performing trait combination to the distribution of r values of all other 134,217,699 trait combinations using multiple Mann–Whitney U tests with a Benjamini–Hochberg P value adjustment. This allowed us to extract a set of 267 trait combinations whose r value distributions were not significantly different from the r value distributions of the top-performing trait combination (Dataset S2). We considered these 267 dental trait combinations as all equally useful for inferring maximum amounts of neutral genetic variation.

Visualizing the Utility of the Top-Performing Dental Trait Combination versus the Utility of the Full Trait Battery. We highlight the superior utility of one of the dental trait combinations listed in Dataset S2 (the top-performing combination with the highest $\tilde{x}_r$ utility estimate: $n_{traits} = 19$; $\tilde{x}_r = 0.580$; 95% r range = 0.293 to 0.758; $P = 0.001$) in comparison to the full trait battery ($n_{traits} = 27$; $\tilde{x}_r = 0.428$; 95% r range = 0.146 to 0.688; $P = 0.001$). For simplicity, we used the full 645 genetic loci dataset for comparison. For each of the two different trait combinations, we visualized the congruence between dental phenotypic (D_P) and neutral genetic (D_G) distances among sampled populations using two complementary techniques: regression plots and Procrustes superimposition plots (Fig. 3). For the regression plots, we visualized the pairwise relationship between D_P and D_G in a scatterplot with a fitted linear regression line and an estimated 95% confidence interval. For the Procrustes superimposition plots, we first subjected the D_P and D_G distance matrices to nonmetric multidimensional scaling (MDS) in order to generate a two-dimensional (2D) representation of the relative affinities among populations. The stress level for the D_G matrix was 0.066. The stress levels for the two different D_P matrices were 0.088 and 0.063, respectively. These low stress levels indicate that two dimensions capture the overall among-population variation of the different datasets well and are below the acceptable threshold of 0.15 (66). Thereafter, the lower-dimensional MDS ordination datasets were subjected to Procrustes superimposition to scale and rotate the two different D_P matrices to maximum similarity with the target D_G matrix by minimizing the overall sum of squared differences among populations. For each of the two different dental trait combinations, the two Procrustes superimposed D_P and D_G distance matrices were then visualized in a single MDS plot.

Visualizing the Utility of Different Numbers of Dental Traits. To explore whether phenotypic inferences about neutral genetic variation based on many dental traits are more useful than those based on only a few traits, we plotted the distribution of $\tilde{x}_r$ and associated P values resulting from trait batteries of different size. For this, we first portioned the generated range of $\tilde{x}_r$ and P values based on the number of traits employed in each combination (from a single trait to the combined total of 27 traits). We then plotted the 27 resulting distributions of $\tilde{x}_r$ and P values using violin plots (Fig. 4).

We additionally assessed whether $\tilde{x}_r$ utility estimates for the 27 dental traits (*SI Appendix*, Table S2) were correlated with either average frequency of a trait across populations (Dataset S3) or with the range of trait frequencies across populations (Dataset S3). For both tests, we used linear regressions utilizing the Pearson product-moment correlation coefficient r, and we estimated P values under the null hypothesis of no association (*SI Appendix*, Figs. S2 and S3). We also explored the effect of individual traits on trait combinations by determining the correlation between $\tilde{x}_r$ utility estimates for individual traits (*SI Appendix*, Table S2) and their frequency within different $\tilde{x}_r$ utility windows (Dataset S1) using Pearson's r and respective significance test as described above (*SI Appendix*, Table S4)

All analyses were performed in R, version 3.6.1 (67). The raw data and R script for the exhaustive search algorithm are publicly available on Zenodo (34) at https://zenodo.org/record/3713179. The Benjamini–Hochberg correction was calculated using the *p.adjust* function (method = "BH") in the R package *stats*, version 3.6.1 (67). The R package *vegan*, version 2.5.2 (68), was used to conduct the Procrustes analyses and MDS calculations, using the *procrustes* and *metaMDS* functions, respectively. All graphics were created using the R package *ggplot2*, version 3.0.0 (69).

ACKNOWLEDGMENTS. This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft [DFG] FOR 2237: Project "Words, Bones, Genes, Tools: Tracking Linguistic, Cultural, and Biological Trajectories of the Human Past"). We are grateful to Katerina Harvati for support and advice in providing the necessary computational resources for our analyses, performed at the Laboratory of Virtual Anthropology and Morphometrics and the Paleoanthropology High-Resolution Computing Tomography Laboratory, University of Tübingen, funded in part by the Senckenberg Nature Research Society and the German Research Foundation (DFG Major Instrumentation Grant INST 37/706-1). We thank Patrícia Santos for assistance with data preparation, as well as Andrea Benazzo and Silvia Ghirotto for discussing the programming code.

1. H. J. H. Edgar, Estimation of ancestry using dental morphological characteristics. *J. Forensic Sci.* **58** (suppl. 1), S3–S8 (2013).
2. G. R. Scott *et al*., rASUDAS: A new web-based application for estimating ancestry from tooth morphology. *Forencic Anthropol.* **1**, 18–31 (2018).
3. C. S. Ragsdale, H. J. H. Edgar, Cultural interaction and biological distance in postclassic period Mexico. *Am. J. Phys. Anthropol.* **157**, 121–133 (2015).
4. H. Matsumura, M. F. Oxenham, Demographic transitions and migration in prehistoric East/Southeast Asia through the lens of nonmetric dental traits. *Am. J. Phys. Anthropol.* **155**, 45–65 (2014).
5. M. A. Pilloud, C. S. Larsen, "Official" and "practical" kin: Inferring social and community structure from dental phenotype at Neolithic Çatalhöyük, Turkey. *Am. J. Phys. Anthropol.* **145**, 519–530 (2011).
6. J. D. Irish, L. Konigsberg, The ancient inhabitants of Jebel Moya redux: Measures of population affinity based on dental morphology. *Int. J. Osteoarchaeol.* **17**, 138–156 (2007).
7. K. S. Paul, C. M. Stojanowski, M. M. Butler, Biological and spatial structure of an early classic period cemetery at Charco Redondo, Oaxaca. *Am. J. Phys. Anthropol.* **152**, 217–229 (2013).
8. H. Rathmann, B. Kyle, E. Nikita, K. Harvati, G. Saltini Semerari, Population history of southern Italy during Greek colonization inferred from dental remains. *Am. J. Phys. Anthropol.* **170**, 519–534 (2019).
9. J. D. Irish, Population continuity vs. discontinuity revisited: Dental affinities among Late Paleolithic through Christian-era Nubians. *Am. J. Phys. Anthropol.* **128**, 520–535 (2005).
10. M. Martinón-Torres *et al*., Dental evidence on the hominin dispersals during the Pleistocene. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13279–13282 (2007).
11. J. D. Irish, D. Guatelli-Steinberg, S. S. Legge, D. J. de Ruiter, L. R. Berger, Dental morphology and the phylogenetic "place" of *Australopithecus sediba*. *Science* **340**, 1233062 (2013).
12. J. D. Irish, S. E. Bailey, D. Guatelli-Steinberg, L. K. Delezene, L. R. Berger, Ancient teeth, phenetic affinities, and African hominins: Another look at where Homo naledi fits in. *J. Hum. Evol.* **122**, 108–123 (2018).
13. C. Turner, C. Nichol, G. R. Scott, "Scoring procedures for key morphological traits of the permanent dentition: The Arizona State University dental anthropology system" in *Advances in Dental Anthropology*, M. Kelley, C. Larsen, Eds. (Wiley-Liss, 1991), pp. 13–32.
14. G. R. Scott, J. D. Irish, *Human Tooth Crown and Root Morphology*, (Cambridge University Press, 2017).
15. G. R. Scott, C. G. Turner, G. C. Townsend, M. Martinón-Torres, *The Anthropology of Modern Human Teeth*, (Cambridge University Press, 2018).
16. H. Rathmann *et al*., Reconstructing human population history from dental phenotypes. *Sci. Rep.* **7**, 12495 (2017).
17. A. R. Hubbard, D. Guatelli-Steinberg, J. D. Irish, Do nuclear DNA and dental nonmetric data produce similar reconstructions of regional population history? An example from modern coastal Kenya. *Am. J. Phys. Anthropol.* **157**, 295–304 (2015).
18. T. Hanihara, Morphological variation of major human populations based on nonmetric dental traits. *Am. J. Phys. Anthropol.* **136**, 169–182 (2008).
19. S. Ramachandran *et al*., Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15942–15947 (2005).
20. Y. Mizoguchi, "Significant among-population associations found between dental characters and environmental factors" in *Anthropological Perspectives on Tooth Morphology: Genetics, Evolution, Variation*, G. R. Scott, J. D. Irish, Eds. (Cambridge University Press, 2013), pp. 108–125.
21. C. M. Stojanowski, K. S. Paul, A. C. Seidel, W. N. Duncan, D. Guatelli-Steinberg, Heritability and genetic integration of anterior tooth crown variants in the South Carolina Gullah. *Am. J. Phys. Anthropol.* **167**, 124–143 (2018).
22. C. M. Stojanowski, K. S. Paul, A. C. Seidel, W. N. Duncan, D. Guatelli-Steinberg, Quantitative genetic analyses of postcanine morphological crown variation. *Am. J. Phys. Anthropol.* **168**, 606–631 (2019).
23. T. Leinonen, R. J. S. McCairns, R. B. O'Hara, J. Merilä, Q(ST)-F(ST) comparisons: Evolutionary and ecological insights from genomic heterogeneity. *Nat. Rev. Genet.* **14**, 179–190 (2013).
24. H. Reyes-Centeno, S. Ghirotto, K. Harvati, Genomic validation of the differential preservation of population history in modern human cranial anatomy. *Am. J. Phys. Anthropol.* **162**, 170–179 (2017).
25. K. Harvati, T. D. Weaver, Human cranial anatomy and the differential preservation of population history and climate signatures. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* **288**, 1225–1233 (2006).
26. N. von Cramon-Taubadel, Congruence of individual cranial bone morphology and neutral molecular affinity patterns in modern humans. *Am. J. Phys. Anthropol.* **140**, 205–215 (2009).
27. H. F. Smith, Which cranial regions reflect molecular distances reliably in humans? Evidence from three-dimensional morphology. *Am. J. Hum. Biol.* **21**, 36–47 (2009).
28. C. C. Roseman, Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12824–12829 (2004).
29. L. J. Hlusko, Elucidating the evolution of hominid dentition in the age of phenomics, modularity, and quantitative genetics. *Ann. Anat.* **203**, 3–11 (2016).
30. L. J. Hlusko, C. A. Schmitt, T. A. Monson, M. F. Brasil, M. C. Mahaney, The integration of quantitative genetics, paleontology, and neontology reveals genetic underpinnings of primate dental evolution. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9262–9267 (2016).
31. T. Hughes, G. C. Townsend, "Twin and family studies of human dental crown morphology: Genetic, epigenetic, and environmental determinants of the modern human dentition" in *Anthropological Perspectives on Tooth Morphology: Genetics, Evolution, Variation*, G. R. Scott, J. D. Irish, Eds. (Cambridge University Press, 2013), pp. 31–68.
32. C. R. Nichol, Complex segregation analysis of dental morphological variants. *Am. J. Phys. Anthropol.* **78**, 37–59 (1989).
33. T. J. Pemberton, M. DeGiorgio, N. A. Rosenberg, Population structure in a comprehensive genomic data set on human microsatellite variation. *G3 (Bethesda)* **3**, 891–907 (2013).
34. H. Rathmann, Data and code for publication "Testing the utility of dental morphological trait combinations for inferring human neutral genetic variation." Zenodo. https://zenodo.org/record/3713179. Deposited 17 March 2020.
35. N. A. Rosenberg *et al*., Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).
36. S. A. Tishkoff *et al*., The genetic structure and history of Africans and African Americans. *Science* **324**, 1035–1044 (2009).
37. Y. Mizoguchi, *Shovelling: A Statistical Analysis of its Morphology*, (University of Tokyo Press, 1985).
38. A. A. Dahlberg, Dental evolution and culture. *Hum. Biol.* **35**, 237–249 (1963).
39. J. D. Cadien, "Dental variation in man" in *Perspectives on Human Evolution 2*, S. L. Washburn, P. Dolhinow, Eds. (Holt, Rinehart and Winston, 1972), pp. 199–222.
40. G. Townsend, H. Yamada, P. Smith, The metaconule in Australian aboriginals: An accessory tubercle on maxillary molar teeth. *Hum. Biol.* **58**, 851–862 (1986).
41. C. G. Turner 2nd, Late Pleistocene and Holocene population history of East Asia based on dental variation. *Am. J. Phys. Anthropol.* **73**, 305–321 (1987).
42. R. Kimura *et al*., A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *Am. J. Hum. Genet.* **85**, 528–535 (2009).
43. J.-H. Park *et al*., Effects of an Asian-specific nonsynonymous EDAR variant on multiple dental traits. *J. Hum. Genet.* **57**, 508–514 (2012).
44. J. Tan *et al*., Characteristics of dental morphology in the Xinjiang Uyghurs and correlation with the EDARV370A variant. *Sci. China Life Sci.* **57**, 510–518 (2014).
45. Q. Peng *et al*., EDARV370A associated facial characteristics in Uyghur population revealing further pleiotropic effects. *Hum. Genet.* **135**, 99–108 (2016).
46. J. Bryk *et al*., Positive selection in East Asians for an EDAR allele that enhances NF-kappaB activation. *PLoS One* **3**, e2209 (2008).
47. L. J. Hlusko *et al*., Environmental selection during the last ice age on the mother-to-infant transmission of vitamin D and fatty acids through breast milk. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4426–E4432 (2018).
48. S. E. Bailey, J.-J. Hublin, S. C. Antón, Rare dental trait provides morphological evidence of archaic introgression in Asian fossil record. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14806–14807 (2019).
49. G. R. Scott, J. D. Irish, M. Martinón-Torres, A more comprehensive view of the Denisovan 3-rooted lower second molar from Xiahe. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 37–38 (2020).
50. T. Hanihara, "Geographic structure of dental variation in the major human populations of the world" in *Anthropological Perspectives on Tooth Morphology: Genetics, Evolution, Variation*, G. R. Scott, J. D. Irish, Eds. (Cambridge University Press, 2013), pp. 479–509.
51. H. Reyes-Centeno, H. Rathmann, T. Hanihara, K. Harvati, Testing modern human out-of-Africa dispersal models using dental non-metric data. *Curr. Anthropol.* **58**, 406–417 (2017).
52. H. Reyes-Centeno *et al*., Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7248–7253 (2014).
53. M. S. Ponce de León *et al*., Human bony labyrinth is an indicator of population history and dispersal from Africa. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4128–4133 (2018).
54. D. J. Witherspoon *et al*., Genetic similarities within and between human populations. *Genetics* **176**, 351–359 (2007).
55. E. Nikita, A critical review of the mean measure of divergence and Mahalanobis distances using artificial data and new approaches to the estimation of biodistances employing nonmetric traits. *Am. J. Phys. Anthropol.* **157**, 284–294 (2015).
56. C. Turner, G. R. Scott, "Dentition of easter islanders" in *Orofacial Growth and Development*, A. Dahlberg, T. Graber, Eds. (Mouton, 1977), pp. 229–249.
57. H. Grüneberg, Genetical studies on the skeleton of the mouse. IV. Quasi-continuous variations. *J. Genet.* **51**, 95–114 (1952).
58. L. L. Cavalli-Sforza, The human genome diversity project: Past, present and future. *Nat. Rev. Genet.* **6**, 333–340 (2005).
59. L. W. Konigsberg, Analysis of prehistoric biological variation under a model of isolation by geographic and temporal distance. *Hum. Biol.* **62**, 49–70 (1990).
60. D. B. Goldstein, A. Ruiz Linares, L. L. Cavalli-Sforza, M. W. Feldman, Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6723–6727 (1995).
61. T. Weaver, Neutral theory and the evolution of human physical form: An introduction to models and applications. *J. Anthropol. Sci.* **96**, 7–26 (2018).
62. M. C. Whitlock, Evolutionary inference from QST. *Mol. Ecol.* **17**, 1885–1896 (2008).
63. A. R. Rogers, H. C. Harpending, Population structure and quantitative characters. *Genetics* **105**, 985–1002 (1983).
64. M. D. Edge, N. A. Rosenberg, Implications of the apportionment of human genetic diversity for the apportionment of human phenotypic diversity. *Stud. Hist. Philos. Biol. Biomed. Sci.* **52**, 32–45 (2015).
65. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
66. P. Dugard, J. B. Todman, H. Staines, *Approaching Multivariate Analysis: A Practical Introduction*, (Routledge, 2010).
67. R Core Team, R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2019).
68. J. Oksanen *et al*, Vegan: Community ecology package. R package, Version 2.5-6 (2019). https://cran.r-project.org/web/packages/vegan/index.html. Accessed 1 May 2019.
69. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis*, (Springer, 2009).