Variation in dental morphology and inference of continental ancestry in admixed Latin Americans

Miguel Delgado[1,2], Luis Miguel Ramírez[3], Kaustubh Adhikari[4], Macarena Fuentes-Guajardo[4,5], Clément Zanolli[6], Rolando Gonzalez-José[7], Samuel Canizales[8], Maria-Catira Bortolini[9], Giovanni Poletti[10], Carla Gallo[10], Francisco Rothhammer[11,12], Gabriel Bedoya[13] and Andres Ruiz-Linares[14,15]

Total pages: 27
Pages of manuscript text: 22
Pages of literature cited: 5
Footnotes: 0
Tables: 7
Figures: 4

[1] División Antropología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, La Plata, República Argentina

[2] Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Buenos Aires, República Argentina.

[3] Facultad de Odontología, Universidad de Antioquia, Medellín, Colombia.

[4] Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, London, United Kingdom

[5] Departamento de Tecnología Médica, Facultad de Ciencias de la Salud, Universidad de Tarapacá, Arica, 1000009, Chile

[6] UMR 5288 CNRS, Université Toulouse III - Paul Sabatier, France

[7] Instituto Patagónico de Ciencias Sociales y Humanas, Centro Nacional Patagónico, CONICET, Puerto Madryn, U9129, Argentina

[8] Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, UNAM-Instituto Nacional de Medicina Genómica, México City, México

[9] Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, 91501-970, Brasil

[10] Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Perú

[11] Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile

[12] Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad de Chile, Santiago, Chile.

[13] GENMOL (Genética Molecular), Universidad de Antioquia, Medellín, Colombia

[14] Ministry of Education Key Laboratory of Contemporary Anthropology and Collaborative Innovation Center of Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai, China.

[15] Laboratory of Biocultural Anthropology, Law, Ethics, and Health (Centre National de la Recherche Scientifique and Etablissement Français du Sang, UMR-7268), Aix-Marseille University, Marseille, France

Abbreviated title: Dental morphology and genetic ancestry in Latin Americans

Key words: dental morphology, admixture, ancestry inference, Latin America

Corresponding authors:

Miguel Delgado. División Antropología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, La Plata (B1900FWA), República Argentina. E-mail: medelgado@fcnym.unlp.edu.ar.

Andrés Ruiz-Linares. Ministry of Education Key Laboratory of Contemporary Anthropology and Collaborative Innovation Center of Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai, China. Email: andresruiz@fudan.edu.cn

ABSTRACT **Objectives**: To investigate the variation in dental nonmetric traits and to evaluate the utility of this variation for inferring genetic ancestry proportions in a sample of admixed Latin Americans. **Materials and Methods**: We characterized a sample from Colombia (N=477) for 34 dental traits and obtained estimates of individual Native American, European and African ancestry using genome-wide SNP data. We tested for correlation between dental traits, genetic ancestry, age and sex. We carried out a biodistance analysis between the Colombian sample and reference continental population samples using the mean measure of divergence statistic calculated from dental trait frequency. We evaluated the inference of genetic ancestry from dental traits using a regression approach (with 10-fold cross-validation) as well as by testing the correlation between estimates of ancestry obtained from genetic and dental data. **Results**: Latin Americans show intermediate dental trait frequencies when compared to Native Americans, Europeans and Africans. Significant correlations were observed for several dental traits, genetic ancestry, age and sex. The biodistance analysis displayed a closer relationship of Colombians to Europeans than to Native Americans and Africans. Mean ancestry estimates obtained from the dental data are similar to the genetic estimates (Native American: 32% v 28%, European: 59% v 63% and African: 9% v 9%, respectively). However, dental features provided low predictive power for genetic ancestry of individuals in both approaches tested ($R^2 < 5\%$ for all genetic ancestries across methods). **Discussion**: The frequency of dental traits in Latin Americans reflects their admixed Native American, European and African ancestry and can provide reasonable average estimates of genetic ancestry. However, the accuracy of individual genetic ancestry estimates is relatively low, probably influenced by the continental differentiation of dental traits, their genetic architecture, and the distribution of genetic ancestry in the individuals examined.

1. INTRODUCTION

Due to the great preservation of teeth and their considerable morphological variation dental traits have been used extensively in human evolutionary studies, including the analysis of archaeological remains and the diversity of contemporary populations (Scott and Turner, 1997; Irish and Scott, 2016; Guatelli-Steinberg, 2016). The differentiation of dental traits among continental populations has also been exploited, usually in a forensic setting, for the purpose of assignment of ancestry to human remains of unknown origin (Scott and Turner, 1997; Alsoleihat, 2013; Edgar, 2005, 2009, 2013, 2015; Irish, 2015; Scott et al., 2018). Such studies have mostly focused on establishing individual ancestry with reference to discrete population categories, such as those defined in the US census (e.g. European-Americans and African-Americans). However, to our knowledge, no attempt has so far been reported to infer genetic ancestry proportions from dental data in individuals of mixed continental ancestry.

The population of Latin America has a history of extensive admixture between Native Americans, Europeans and Africans and therefore represents an ideal setting in which to evaluate the informativity of dental traits to estimate genetic ancestry and individual admixture proportions. Here we report an analysis of dental nonmetric trait variation in contemporary Latin Americans and we evaluate the informativity of these dental traits for inferring admixture proportions. Consistent with their historical admixture, Latin Americans present dental nonmetric traits that are common in Native Americans, Europeans and Africans, and some of these traits correlate with genetic ancestry. We observe that although dental traits provide mean ancestry estimates similar to those obtained by genetic data, the

informativity of these traits to estimate individual ancestry in the sample examined is relatively low.

2. MATERIALS AND METHODS

2.1 Subjects of study

We studied a sample of 477 individuals of both sexes (women/men = 259/218) aged 18-40 (mean = 23.4) recruited in Medellín, Colombia (the sample is denoted MED in Tables and Figures). This sample is part of the CANDELA cohort (Consortium for the Analysis of the Diversity and Evolution of Latin America, http://www.ucl.ac.uk/silva/candela) (Ruiz-Linares et al., 2014). This research was approved by the ethics committees of Universidad de Antioquia (Colombia), Universidad de Tarapacá (Chile), and University College London (UK). All participants provided written informed consent.

2.2 Characterization of dental morphology

Intraoral digital photographs were obtained (by LMR) using an IE3 Canon camera (at 12 megapixels resolution) under standardized conditions, including: captures in frontal, lateral and occlusal norms, constant light, proportion and distance. Using these photographs, we examined a total of 34 dental features (corresponding to 86 traits across teeth; Supplementary Table 1) in each individual (by M.D.) following ASUDAS trait definitions (Turner et al., 1991) with the exception of elongated mandibular premolars (Edgar and Sciulli, 2004) and lower premolar accessory ridges (Delgado, 2015) (Supplementary Table 1). This scoring was performed blindly with regards to age, sex or genetic ancestry of the individuals examined. The intra-observer concordance rate for these traits has been shown to be high (Delgado, 2015). We only scored teeth with no caries, no apparent wear and no

dental restorations. We retained the score of the antimere with strongest expression of the trait when asymmetric expression was evident (Scott, 1980). Subsequent to scoring, traits with extreme frequencies (<0.3% or >98%), or >5% missing values were excluded. This resulted in 28 traits being retained for subsequent analyses.

2.3. Biodistance analysis

We used C. A. B Smith´s mean measure of divergence (MMD) statistic (Sjøvold, 1977; Irish, 2010) to estimate biological distances from the frequency of the dental traits in the Colombian and reference population samples, using the dichotomies of dental traits proposed for ASUDAS (Supplementary Table 1) (Sjøvold 1977, Harris and Sjøvold, 2004; Irish, 2010). The reference population dataset used in these analyses (Appendix 2; Scott and Turner, 1997) consisted of frequencies reported for: Native Americans (American Arctic [AA, N=1,022], North-South America [NSA, N=3,276] and Northwest North America [NNA, N=741); Africans (West Africa [WA, N=92], Khoisan [K, N=155] and South Africa [SA, 531]) and Europeans (Western Europeans [WE, N=371], Northern Europeans [NE, N=319] and North Africans [NA, N=545]). Following Harris and Sjøvold (2004) we used Ascombe´s transformation of the MMD since it is slightly better than the Freeman-Tukey formula at asymptotically stabilizing sample variance and is more appropriate for the relatively large sample size examined here (Green and Suchey, 1976). We excluded traits significantly correlated with sex (since the MMD assumes lack of sexual dimorphism) or with other traits (as the MMD also assumes independence between traits) (Nikita, 2015), and those traits that showed no significant variation between samples. Three traits (SSUI1, DSUI1 and ODOUP1) that have been extensively studied and are highly differentiated between reference population were retained, despite SSUI1 and DSUI1 showing a

significant correlation and ODOUP1 a low frequency (<0.3%) in the Colombian sample investigated here. This resulted in 16 traits being retained for this analysis (Supplementary Tables 2 and 3). The MMD calculation was performed in R 3.4.3 (R Development Core Team, 2017) using a script written by M.D based on equations presented in Sjøvold (1977) and Harris and Sjøvold (2004).

2.4. Estimation of continental genetic ancestry

The individuals examined here have been previously genotyped on Illumina's HumanOmniExpress chip (Adhikari et al., 2015, 2016a,b), which includes over 700,000 SNPs. After pruning for Linkage Disequilibrium 93,328 autosomal SNPs were retained. Genotype data from the admixed samples was combined with genotype data of reference samples from three continental populations to estimate European, African and Native American ancestry proportions using ADMIXTURE (Alexander and Lange, 2011). Reference parental populations included in the ADMIXTURE analyses consisted of Africans and Europeans from 1000 Genomes (The 1000 Genomes Project Consortium, 2015) and selected Native American samples (Reich et al., 2012).

2.5 Correlation analyses

For all correlation analyses, the dental trait scores were considered ordinal variables. The justification for doing so is that we assume an underlying continuous variable (Scott and Turner, 1997), and the convention that for an ordinal variable with several categories there is little difference between fitting a linear regression or an ordered probit model (Harvati and Weaver, 2006). We confirmed that both approaches produced similar results. Simple correlation analysis was performed among dental traits. To evaluate the effect of covariates

(genetic ancestry, age, and sex) on the dental traits, we used partial correlation analysis. In these tests, the Bonferroni-adjusted P-value threshold for significance was $P < 0.001$.

2.6 Inference of individual genetic ancestry from dental morphology

We explored two approaches to infer individual continental ancestry proportions from dental traits, both implemented in Matlab (R2017b) by K.A.

2.6.1 Using reference population data

Data on full trait distributions was obtained from the literature (Scott and Irish (2017) for fourteen reference population samples from areas that contributed extensively to admixture in Latin America: one from Sub-Saharan Africa (West Africa); three from Western Europe (Spain, Netherlands and England) and ten from Central and South America (Native Americans from Mexico, Ecuador, Brazil and Chile). Of the 29 traits described in Scott and Irish (2017), 20 were scored in the Colombian sample. From these 20 we excluded traits that were missing in any reference population or had a low frequency in the Colombian sample (<1%). This resulted in the following 16 traits being retained: WINGUI1, SSUI1, DSUI1, IGUI2, TDUI2, MRUC, ODOUP1, HYPUM2, C5UM2, CTUM1, LCVLP2, GPLM2, CNLM1, CNLM2, PTSLM1 and C7LM1 (see Supplementary Table 1 for trait descriptions and Supplementary Tables 2 and 3 for the trait scores in the Colombian sample). To obtain estimates of ancestry in the Colombian sample based on these traits we followed the approach described below.

For a trait (*j*) with possible states 1, 2, …, *c,* the trait frequencies (*f*) in reference population *r* (with values *a*, *e* or *n* for African, European and Native, respectively) can be represented by: $\left(f_1^{jr}, f_2^{jr}, \cdots, f_c^{jr}\right)$, where the frequencies in each reference population sum up to one. These continental frequencies can be combined with ancestry proportions to construct a trait frequency distribution in individuals (*i*) of mixed ancestry (*m*) as:

$$f_{ki}^{jm} = p_i^a \cdot f_k^{ja} + p_i^e \cdot f_k^{je} + p_i^n \cdot f_k^{jn} \text{ for any trait value } k \text{ from 1 to } c.$$

Where (p_i^a, p_i^e, p_i^n) refer to the proportions of African, European and Native American ancestries in the admixed individuals (*i*). These 'mixed' trait frequencies also sum to one for any trait.

For a given ancestry proportion, a score can be constructed for each trait indicating how probable or improbable the observed trait value is, given the frequency distribution of this trait. For example, the score S_{ij} of an individual *i* of mixed ancestry with value t_i for a trait *j* when compared to the frequency distribution (f_i^{jm}) for this trait can be calculated as:

$$S_{ij} = \sum_{k=1}^{c} \left| t_{ij} - k \right| \cdot f_k^{jm}$$

This score represents the mean absolute deviation (Rao, 1973) of the trait value in an individual relative to a frequency distribution. For any individual *i*, and any trait *j*, these scores can be calculated for each trait given ancestry proportions (p_i^a, p_i^e, p_i^n). A composite score can then be constructed by summing across traits. For example, the composite score for person *i* over all *T* traits is:

$$S_i = \sum_{j=1}^{T} \sqrt{w_j} \cdot S_{ij}$$

To adjust for correlation between traits included in the analysis, we use weights (w_j) inversely proportional to their total correlation with other traits. As correlation values from the reference data was not available, correlation values in the Colombian dataset were used for this step. A commonly used weight (Zou et al., 2010) to scale the contribution of trait *j* is

$w_j = \frac{1}{\sum_{k=1}^{T} r_{jk}^2}$ where r_{jk} is the correlation between traits *j* and *k*.

To estimate ancestry for individual *i* we find the ancestry proportions (p_i^a, p_i^e, p_i^n) which minimize the composite score S_i using a grid search over all possible ancestry proportions and evaluating the score at each combination. The individual ancestry estimates obtained with genetic and dental data were compared using intraclass correlations (Shrout and Fleiss, 1979). The squared correlations, estimating the proportion of variance explained, were taken as estimates of the accuracy of estimation of genetic ancestry from dental morphology.

2.6.2 Estimation within the Colombian sample data

In the second approach we regressed genetic ancestry on each dental trait separately, or on multiple traits simultaneously, solely within the Colombian dataset. In the regression models we examined prediction accuracy based on 10-fold cross-validation (CV) (Hastie et al., 2009). Thus, for each of 10 random subsets of these data, we trained models based on 90% of the subset and predicted genetic ancestry values in the remaining 10%. Prediction accuracy was evaluated by the fraction of trait variance explained by a model (R^2_{CV}), averaged over the 10 CV sets.

In the case of single-trait regression, ordinary multivariate linear regression was used. Values of one ancestry component were regressed on age, sex, and one dental trait. In the case of regression involving multiple traits, we explored reducing overfitting and collinearity (which could affect prediction performance, Chatterjee and Hadi, 2012), by applying the LASSO approach, which selectively includes only a few covariates in the regression model (Hastie, 2009). We also used two dimension-reduction techniques on the set of all dental traits: Principal Component Analysis (PCA) and Independent Component Analysis (ICA), which summarize the total dental trait variation into a reduced number of independent variables, each capturing a substantial fraction of the total variation (Hastie, 2009). In the case of PCA we retained for the regression analyses 20 PCs with relatively high eigenvalues and explaining 86% of the total variance. For ICA we retained 3 or 6 components, explaining 31.3% and 52.5% of the total variance, respectively.

We performed simulations to assess the performance of this prediction methodology under two scenarios. In the first scenario, we simulated a uniform ancestry distribution (between 0 to 100%) by sampling with replacement individuals from the Colombian data (Fuller, 2009), so as to obtain a simulated dataset with the same sample size and the same relationship between all variables and covariates as in the original dataset. We generated 100 simulated samples and obtained the average R^2 across samples to calculate prediction accuracy.

In the second scenario we divided the Colombian data into two subsets: one highly European (>95% European ancestry) and one highly Native (>95% Native American ancestry). We then sampled with replacement from these two subsets to create a simulated sample in which half of the individuals are highly European and the other half highly Native.

As before, this resampling maintains the same relationship between all variables and covariates. We generated 100 simulated samples. Since here the regression model is predicting dichotomous group labels, we used classification accuracy (% of correctly predicted group label) as an indicator of prediction accuracy (Hastie, 2009).

3. RESULTS

3.1 Distribution of individual genetic ancestry

Average estimates of genetic ancestry in the Colombian sample investigated were: 63% European, 28% Native American and 9% African. Individual estimates of Native ancestry show a relatively sharp peak at around 30% (Figure 1). As expected (since ancestry proportions are constrained by having to add up to 1) there is a strong negative correlation between European and non-European (Native or African) ancestry ($r < -0.65$). African ancestry presents a highly skewed distribution, with few individuals presenting >20% of African ancestry and no individual with >80% of such ancestry.

3.2 Correlation between dental traits and covariates

The frequencies of the dental traits examined in the Colombian sample are presented in Supplementary Tables 2-3. This sample shows low to moderate frequencies of traits common in Native Americans (e.g., WINGUI1, SSUI1, SSUI2, DSUI1, DSUI2, C5LM2; C6LM1), whereas traits with high frequencies in Africans (MRUC, DIASUI1 and C7LM1) presented low frequencies in the Colombians. Several features characterizing the Eurasian dental complex are also present in the Colombian sample (e.g., high frequencies of CTUM1, LCVLP2, 3CUM2 and CNLM1 and low frequencies of WINGUI1, DSUI1, GPLM2,

C6LM1, C7LM1 and DWLM1). Overall, the frequency of dental traits in the sample examined reflects its mixed ancestry.

Moderate to strong positive, and significant, correlations ($r>0.5$, p-value $<1.29E-31$) were observed between several of the traits examined. These usually represent the same trait scored in different teeth, including: SSUI1, SSUI2, SSLI1, SSLI2, DSUI1, DSUI2, DSLI1, DSLI2, DTUI1, DTUI2, ARUP1, ARUP2, AMTUP1, AMTUP2, AFLM1 and AFLM2. Across traits, several significant positive correlations were observed, with the strongest occurring between SS and DS, in both upper and lower incisors. Possible explanations for these correlations include admixture linkage disequilibrium (particularly for traits present with high frequencies in parental populations) and pleiotropic effects of certain genetic variants influencing dental development (Townsend et al., 2009; Hughes and Townsend, 2013; Hillson, 2014; Dhamo et al., 2018).

A number of weak, but significant, correlations were observed between dental traits and covariates (r values 0.12-0.23) (Supplementary Table 4). Nine traits showed a negative correlation with age (DARUC, ARUP1, ARUP2, AMTUP1, AMTUP2, METUM2, HYPUM1, ARLP2, AFLM2), and one trait (PTSLM1) was positively correlated with age. With the exception of PTSLM1 all dental traits correlated with age are related to cusps, ridges and foveae, which are structures that are very susceptible to wear. The relevance of dental wear as a proxy of biological age has been highly exploited in bioarchaeology (Lovejoy, 1985). Our findings suggest that in contemporary humans, despite the consumption of soft and processed foods, the effect of age on dental wear is considerable (Faillace et al., 2017). Finally, we found that five traits were correlated with sex: TDUI1, TDUC, MRUC, DARUC and DARLC. Four of these traits represent features of the canines,

the most sexually dimorphic teeth in humans, hominins and non-human primates (Plavcan, 2012). These observations underline the utility of canine morphology in the assignment of sex in undetermined samples from contemporary human populations (Tardivo et al., 2011). Positive correlations were observed between African ancestry and three traits: C5UM2, CNLM1 and C7LM1. In addition, fourteen additional traits showed positive correlations with Native American ancestry (and, correspondingly, negative correlations with European ancestry): SSUI1, SSUI2, DSUI1, DSUI2, METUM1, METUM2, SSLI1, SSLI2, DSLI1, DSLI2, AFLM2, CNLM2, C5LM2, C7LM2.

2.3 Inter-population differentiation based on dental trait frequency

Table 1 shows the MMD matrix between nine reference population samples and the sample from Colombia calculated on the basis of the frequency of 16 dichotomous dental traits. Figure 2 displays the frequency of the 16 dental traits used in calculating the MMD matrix in the samples investigated and Figure 3 shows a principal coordinate analysis (PCoA) of the MMD matrix. All MMD distances except the pairs WE-NE, NE-NA and AA-NWNA, NWNA-NSA were significant ($p<0.025$). In Figure 3 PCo1 differentiates Native Americans from Europeans, North Africans and Sub-Saharan Africans, while PCo2 discriminates Sub-Saharan Africans from Europeans, North Africans and Native Americans. Consistent with its mainly Native American-European but predominantly European ancestry, the Colombian sample lies on the European-Native American axis and is closest to Europeans.

2.4 Estimation of ancestry from dental traits

We explored two approaches to estimating ancestry from dental data. In the first approach, based on the frequency of dental traits in reference population samples, we estimated (as described in the Materials and Methods) average African, European and Native ancestry proportions as: 0.09, 0.59 and 0.32, respectively (with standard deviations of 0.26, 0.49, 0.47, respectively). These average ancestry estimates are similar to those obtained with genetic data (respectively, 0.09, 0.63, 0.28), although the values estimated from the dental data show more variation than the genetic estimates (which have standard deviations of 0.10, 0.13, 0.10, respectively). When contrasting individual genetic and dental ancestry estimates, the squared correlations (R^2) were low for the three ancestries: 0.7%, 0.7% and 0.5% (for African, European and Native American ancestries, respectively).

In the second approach, we predicted genetic ancestry values using regression models incorporating dental trait variation solely in the Colombian sample. Based on results from a 10-fold CV approach, we find that predictions of genetic ancestry from single dental traits have low R^2 values for all ancestries (Table 2: median R^2 of ~ 1% (range 0.7% - 1.4%) for African ancestry, ~1.1% for European (range 0.8% – 3.3%) and ~1.7 % for Native American (range 0.8% – 4.4%). Highest prediction scores for Native American ancestry are obtained with SSLI1 (4.4%), SSLI2 (4.4%), DSUI2 (3.3%), SSUI2 (2.9%), DSUI1 (2.5%) and SSUI1 (2.4%). The traits with the highest prediction scores for African ancestry are TDUC (1.6%) and C5UM1 (1.4%). The traits with the highest prediction scores for European ancestry are SSLI1 (3.3%), SSLI2 (2.9%), C6LM1 (2.4%) and CNLM2 (2.3%).

We also evaluated prediction of genetic ancestry from regression models incorporating all dental traits examined or components from two dimension-reduction methods (ICA and PCA) (Table 3). When including all traits, R^2 value were 0.8% for

African, 2.1% for European and 3.4% for Native American ancestry. Although still low, prediction accuracy improved somewhat when using ICA or PCA. For ICA, using 3 components we obtain R^2 of 1.8%, 4.5% and 3.4% for African, European and Native American ancestry, respectively. Using 6 ICA components we obtain R^2 of 2.5%, 4.2% and 3.4% for African, European and Native American ancestry, respectively. When using 20 PCs we obtained R^2 of 0.8%, 4.1% and 4.6% for African, European and Native American ancestry, respectively. Finally, for the combined analysis using both ICs and PCs we obtained the R^2 of 1.2%, 4.1% and 4.9% for African, European and Native American ancestry, respectively.

4. DISCUSSION

The diversity in dental morphology observed in the Colombian sample studied here is consistent with the history of admixture between Native Americans, Europeans and Africans that characterizes Latin Americans. Traits common in those three continental populations are prevalent in the Colombian sample. Furthermore, the biodistance analysis and the average estimates of ancestry in the Colombian sample obtained from dental morphology data are consistent with the genetic estimates of ancestry: a predominant European ancestry, with substantial Native American ancestry and a relatively small African contribution. The correlation of certain dental features with specific genetic ancestries suggests that aspects of tooth morphology are likely to be influenced by alleles differentiated in frequency between continental populations (see Hubbard et al., 2015; Rathmann et al., 2017), probably involving loci impacting on tooth development (Edgar and Ousley, 2016; Dhamo et al., 2018). The three traits correlated with African ancestry detected here (C5UM2, CNLM1 and C7LM1) have been reported to show markedly higher frequencies in

Sub-Saharan Africans relative to other continental populations (Irish, 1997, 2013; Scott and Turner, 1997; Scott and Irish, 2017). However, certain traits that have been described as characterizing a "Sub-Saharan African dental complex" (such as DIASUI1, MRUC, LCUI2 and GPLM2; Irish, 2013) showed no significant correlation with genetic estimates of African ancestry. Similarly, most of the 14 traits showing positive correlation with Native American ancestry have been extensively documented as common in East Asians and Native Americans (Scott and Turner, 1997 and references therein). Noticeably, we did not detect any dental trait positively correlated with European ancestry. However, the distribution of dental traits in the Colombian sample shares some common features with the so-called "Western Eurasian dental complex" (*sensu* Scott et al., 2013), that is, low frequencies of WINGUI1, GPLM2, C6LM1, C7LM1 and DWLM1 and moderate to high frequencies of CTUM1, LCVLP2, 3CUM2 and four-cusped LM2 (CNLM2). So-called "Classic" European traits, such as the Carabelli tubercle, showed positive although not significant correlations ($r<0.05$) with European genetic ancestry. The lack of a significant correlation with ancestry of certain dental traits could be related to a lack of power to detect such effects, for instance due to the relatively low Sub-Saharan African ancestry and the rather narrow spread of individual ancestry estimates in the Colombian sample studied here (Figure 1). Furthermore, the difference in frequency of some traits between the parental populations involved in admixture in Latin America might have been lower than what has been documented in available reference population data.

Despite dental data providing relatively good estimates of average genetic ancestry in the Colombian sample, our prediction analyses indicate that the dental traits examined are relatively poor predictors of individual genetic ancestry in this sample. The prediction

accuracy estimated is likely influenced by a range of factors, including: (i) the differentiation in trait frequency between the populations contributing to the admixture, (ii) the genetic architecture of the traits used (i.e., number of loci and allele frequencies at these loci, additive/dominant/recessive genetic effects), (iii) categorization of the traits in the ASUDAS system, including how it relates to the underlying dental morphology, (iv) the magnitude of the ancestry components being estimated, and (v) the distribution of individual ancestry values in the study sample.

As an illustration of the importance of the distribution of individual genetic ancestry on prediction accuracy, we examined data for the SSUI1 trait characterized in the CANDELA sample from Chile (N=1,792; MFG unpublished). This sample is on average 49.4% European, 48.3% of Native American and 2.2% African (Ruiz-Linares et al., 2014) and shows a wider individual ancestry distribution than the sample from Colombia (s.d. of 0.10 and 0.16, in Colombia and Chile, respectively, Supplementary Figure 1). Consistent with the relatively larger spread of individual ancestry along the European-Native American axis, the Chilean sample shows both a stronger correlation of Native American ancestry with SSUI1 than seen in the Colombian sample (r of 0.33 v. 0.16, respectively) and higher ancestry prediction scores: 2.3%, 12.8% and 12.4% for African, European and Native American ancestry, respectively. For comparison, we evaluated the impact of the distribution of ancestry values on the accuracy of individual genetic ancestry prediction using SSUI1 by simulations based on resampling the Colombian data. A simulation changing the Colombian ancestry distribution to uniform (resulting in ancestry s.d. increasing from 0.10 to 0.29) resulted in an increase in R^2 for prediction based on SSUI1 from 2.4% to 61.9%. In a second simulation we replicated the setting usually examined in the literature, which evaluates

prediction of ancestry as pertains to discrete population categories to which individuals are assigned (Edgar, 2013; Irish, 2015). Using the Colombian data-set we generated simulated sets of individuals either with high (>95%) European or Native American ancestry and then tested how accurately SSUI1 can assign individuals to these two sets. Classification accuracy was very high, at 93.5%.

In conclusion, our study shows that the dental characteristics of Latin Americans reflect their history of admixture involving Native Americans, Europeans and Africans. However, despite dental morphology reflecting past admixture, the use of dental traits for inferring genetic ancestry components in admixed Latin Americans is a considerably more difficult task than the ancestry assignments usually performed in studies using discrete categories, such as those often used in US study samples. The correlation of certain dental traits with genetic ancestry suggests that aspects of tooth morphology could be influenced by specific alleles differentiated in frequency between continental populations. Further study of Latin American populations could provide a fruitful approach to the identification of such dental morphology loci.

ACKNOWLEDGMENTS

We thank the volunteers for their enthusiastic support for this research. We are very grateful to Universidad de Antioquia, Medellín, Colombia and Universidad de Tarapacá, Arica, Chile for kindly providing facilities for the assessment of volunteers. We are indebted to Lina Maria Parra (Universidad Libre de Colombia, Cali), Bernardo Salazar and Camilo Preciado (Facultad de Odontología, Universidad de Antioquia, Medellín) for their support during sample collection. We thank G.R Scott and J.D. Irish for their clarifications on the reference population data used in the present research. We thank Peter Ellison (AJPA Associate Editor) and two anonymous reviewers for their helpful comments provided during the reviewing process. Work leading to this publication was funded by grants from the: Leverhulme Trust (F/07 134/DF), BBSRC (BB/I021213/1) (to ARL), Excellence Initiative of Aix-Marseille University - A*MIDEX (a French “Investissements d’Avenir” programme) (to ARL), National Natural Science Foundation of China (31771393) (to ARL), Wenner-Gren Foundation for Anthropological Research (to MD), Universidad de Antioquia (CODI 2014-1124 to MD and LMR), and UCL Global Engagement Fund (to KA). KA was also supported by a Wellcome Investigator Award WT107055AIA (to Prof. C.D. Stern).

REFERENCES

Adhikari, K., Reales, G., Smith, A., Konka, E., Palmen, J., Quinto-Sanchez, M., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Fuentes-Guajardo, M., Pizarro, M., Barquera Lozano, R., Macín Pérez, G., Gómez-Valdés, J., Villamil-Ramírez, H., Hunemeier, T., Ramallo, V., Silva de Cerqueira, C., Hurtado, M., Villegas, V., Granja, V., Gallo, C., Poletti, G., Schuler-Faccini, L., Salzano, F., Bortolini, C., Canizales-Quinteros, S,, Rothhammer, F., Bedoya, G., Calderón, R., Rosique, J., Cheesman, M., Bhutta, M., Humpries, S., Gonzalez-José, R., Headon, D., Balding, D., & Ruiz-Linares, A. (2015). A genome-wide association study identifies multiple loci variation in ear morphology. *Nature Communications*, 6, 7500 DOI: 10.1038/ncomms8500

Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacón-Duque, J.C., Al-Saadi, F., Johansson, J., Quinto-Sanchez, M., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Barquera Lozano, R., Macín Pérez, G., Gómez-Valdés, J., Villamil-Ramírez, H., Hunemeier, T., Ramallo, V., Silva de Cerqueira, C., Hurtado, M., Villegas, V., Granja, V., Gallo, C., Poletti, G., Schuler-Faccini, L., Salzano, F., Bortolini, M., Canizales-Quinteros, S., Rothhammer, F., Bedoya, G., Gonzalez-José, R., Headon, D., López-Otín, C., Tobin, D., Balding, D., & Ruiz-Linares, A (2016a). A genome-wide scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nature Communications*, 7, 10815 DOI: 10.1038/ncomms10815.

Adhikari, K., Fuentes-Guajardo, M., Quinto-Sanchez, M., Mendoza-Revilla, J., Chacón-Duque, J.C., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Barquera Lozano, R., Macín Pérez, G., Gómez-Valdés, J., Villamil-Ramírez, H., Hunemeier, T., Ramallo, V., Silva de Cerqueira, C., Hurtado, M., Villegas, V., Granja, V., Gallo, C., Poletti, G., Schuler-Faccini, L., Salzano, F., Bortolini, C., Canizales-Quinteros, S., Cheesman, M., Rosique, J., Bedoya, G., Rothhammer, F., Headon, D., Gonzalez-José, R., Balding, D., & Ruiz-Linares, A. (2016b). A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nature Communications*, 7, 11616. DOI: 10.1038/ncomms11616.

Alexander, D., Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, 12, 246.

Alsoleihat, A. (2013). A new quantitative method to predict forensic racial identity based on dental morphological trait analysis. *International Journal of Morphology*, 31, 418-424.

Chatterjee, S., & Hadi, S.A. (2012). *Regression analysis by example*. John Willey & Sons, Inc.

Dahmo, B., Kragt, L., Grgic, O., Vucic, S., Medina-Gomez, C., Rivadeneira, F., Jaddoe, V., Wolvius, E., & Ongkosuwito, E. (2018). Ancestry and dental development: a geographic and genetic perspective. *American Journal of Physical Anthropology* 165: 299-308.

Delgado, M.E. (2015). *Variación dental y craneofacial en el Norte de los Andes durante el Pleistoceno y el Holoceno. Su relevancia para la discusión de la colonización temprana de Sudamérica* (Unpublished doctoral dissertation). Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, La Plata, Buenos Aires.

Edgar, H.J.H. (2005). Prediction of race using characteristics of dental morphology. *Journal of Forensic Sciences*, 50, 269-273.

Edgar H.J.H. (2009). Testing the utility of dental morphological traits commonly used in the forensic identification of ancestry. In Koppe, T. Meyer G & K Alt (Eds.) Comparative dental morphology. *Frontiers of Oral Biology*, 13, 49–54.

Edgar, H.J.H. (2013). Estimation of ancestry using dental morphological characteristics. *Journal of Forensic* Sciences, 58, S3–S8.

Edgar, H.J.H. (2015). Dental morphological estimation of ancestry in forensic contexts. In Berg, GE & SC Ta'ala (Eds.) *Biological Affinity in Forensic Identification of Human Skeletal Remains.* (pp 191–208).Boca Raton, FL: CRC Press.

Edgar, HJ.H., & Sciulli, P. (2004). Elongated premolar: a new morphological variant. *Dental Anthropology*, 17, 24-27.

Edgar, H.J.H., & Ousley, S. (2016). Dominance in dental morphological traits: implications for biological distance studies. In M.A. Pilloud & J.T. Hefner (Eds.), *Biological distance analysis*. *Forensic and bioarchaeological perspectives* (pp. 317-329). Academic Press.

Faillace, K.E., Bethard, J.D., & Marks, MK. (2017). The applicability of dental wear in age estimation for a modern American population. *American Journal of Physical* Anthropology, 164, 776-787.

Fuller, W.A (2009). *Sampling statistics*. John Wiley & Sons, Inc.

Green, R,F., & Suchey, J.M. (1976). The use of inverse sine transformations in the analysis of non-metric cranial data. *American Journal of Physical Anthropology*, 45, 61–68.

Guatelli-Steinberg, D. (2016). *What teeth reveal about human evolution*. Cambridge: Cambridge University Press.

Harris, E.F., & Sjøvold, T. (2004). Calculation of Smith's mean measure of divergence for intergroup comparisons using nonmetric data. *Dental Anthropology*, 17, 83–93.

Harvati, K., & Weaver, TD. (2006). Human cranial anatomy and the differential preservation of population history and climate signatures. *Anatomical Record*, 288A,1225-1233.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning. Data mining, inference and prediction*. Springer.

Hillson, S. (2014). *Tooth development in human evolution and bioarchaeology*. Cambridge: Cambridge University Press

Hughes, T., & Townsend, G. (2013). Twin and family studies of human dental crown morphology: genetic, epigenetic and environmental determinants of the modern human dentition. In G.R. Scott & J.D. Irish (Eds.), *Anthropological Perspectives on Tooth Morphology, Genetics, Evolution, Variation* (pp. 31-68). Cambridge: Cambridge University Press.

Hubbard, A., Guatelli-Steinberg, D., & Irish, J.D. (2015). Do nuclear and dental nonmetric data produce similar reconstructions of regional population history? An example from modern coastal Kenya. *American Journal of Physical Anthropology*, 157, 295-304.

International HapMap Consortium (2005). A haplotype map of the human genome. *Nature*. 437: 1299–1320.

Irish, J.D. (1997). Characteristic high- and low-frequency dental traits in Sub-Saharan African populations. *American Journal of Physical Anthropology,* 102, 455-467.

Irish, J.D. (2010). The mean measure of divergence: its utility in model-free and model-bound analyses relative to the Mahalanobis D^2 distance for nonmetric traits. *American Journal of Human Biology*, 22, 378–395.

Irish, J,D. (2013). Afridonty: "the Sub-Saharan African dental complex" revisited. In G.R. Scott & J.D. Irish (Eds.), *Anthropological Perspectives on Tooth Morphology, Genetics, Evolution, Variation* (pp. 278-295). Cambridge: Cambridge University Press.

Irish J.D. (2015). Dental nonmetric variation around the world: using key traits in populations to estimate ancestry in individuals. In Berg G.E., & S.C Ta'ala (Eds). *Biological Affinity in Forensic Identification of Human Skeletal Remains.* (pp 165–190). Boca Raton, FL: CRC Press.

Irish J.D., & Scott G.R. (2016). *A companion to dental anthropology*. Chichester: Wiley Blackwell.

Lovejoy, C.O. (1985). Dental wear in the Libben population: its functional pattern and role in the determination of adult skeletal age. *American Journal of Physical Anthropology*, 68, 47-56.

MATLAB and Statistics Toolbox Release 2017b, The MathWorks, Inc., Natick, Massachusetts, United States.

Nikita, E. (2015). 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. *American Journal of Physical Anthropology* 157: 284-94.

Plavcan, M. (2012). Sexual size dimorphism, canine dimorphism, and male-male competition in primates: where do humans fit in? *Human Nature*, 23, 45-67.

R Core Team. (2017). *R Foundation for statistical computing*. Vienna, Austria. Version 3.4.3. Available from: http://www.R-project.org.

Rao, C.R. (1973). *Linear statistical inference and its applications*: Wiley New York

Rathmann, H., Reyes-Centeno, H., Ghirotto, S., Creanza, N., Hanihara, T., & Harvati, K. (2017). Reconstructing human population history from dental phenotypes. *Scientific Reports*, 7, 12495 DOI:10.1038/s41598-017-12621-y

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N,, Parra, M.V., Rojas, W., Duque, C., Mesa, M., García, L., Triana, O., Blair, S., Maestre, A., Dib, J., Bravi, C., Bailliet, G., Corach ,D., Hünemeier, T., Bortolini, M., Salzano, M., Petzl-Erler, ML., Acuña-Alonzo, V., Aguilar-Salinas, C., Canizales-Quinteros, S., Tusié-Luna, T., Riba, L., Rodríguez-Cruz, M., Lopez-Alarcón, M., Coral-Vazquez, R., Canto-Cetina, T., Silva-Zolezzi, I., Fernandez-Lopez, J.C., Contreras, A.V., Jimenez-Sanchez, G., Gómez-Vázquez, M.J., Molina, J., Carracedo, A., Salas, A., Gallo, C., Poletti, G., Witonsky, D.B., Alkorta-Aranburu, G., Sukernik, R.I., Osipova, L., Fedorova, S., Vasquez, R., Villena, M., Moreau, C., Barrantes, R., Pauls, D., Excoffier, L., Bedoya, G., Rothhammer, F., Dugoujon, J.M., Larrouy, G., Klitz, W., Labuda, D., Kidd, J., Kidd, K., Di Rienzo, A., Freimer, N.B., Price, A.L., & Ruiz-Linares A. (212). Reconstructing Native American population history. *Nature*, 488, 370-374.

Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Snchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de Avila, F., Gómez-Valdés, J., León-Mimila, P., Hunemeier, T., Ramallo, V., Silva de Cerqueira, C.C., Burley, M.W., Konka, E., Zagonel de Oliveira, M., Veronez, M.R., Rubio-Codina, M., Attanasio, O., Gibbon, S., Gallo C., Poletti, G., Rosique, J., Schuler-Faccini, L., Salzano, F.M., Bortolini, M.C., Canizales-Quinteros, S., Rothhammer, F., Bedoya, G., Balding, D., & Gonzalez-José R. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genetics* 10:e1004572.

Scott, R.G. (1980). Populations variation of Carabelli´s trait. *Human Biology*, 52, 63-78.

Scott, G.R., & Turner, C.G. (1997). *The Anthropology of Modern Human Teeth: Dental Morphology and its Variation in Recent Human Populations, 2nd Edition*. Cambridge: Cambridge University Press.

Scott, G.R., Irish, J.D. (2017). *Human tooth crown and root morphology. The Arizona State University Dental Anthropology System*. Cambridge: Cambridge University Press.

Scott, G.R., Anta, A., Schomberg, R., & De la Rúa C. (2013). Basque dental morphology and the "Eurodont" dental pattern. In G.R. Scott & J.D. Irish (Eds.), *Anthropological*

Perspectives on Tooth Morphology, Genetics, Evolution, Variation (pp. 296-318). Cambridge: Cambridge University Press.

Scott, G.R., Pilloud, M., Navega, D., d´Oliveira, J., Cunha, E., & Irish J. (2018). rASUDAS. A new web-based application for estimate ancestry from tooth morphology. *Forensic Anthropology* 1:18-31.

Sjøvold, T. (1977). Nonmetrical divergence between skeletal populations: the theoretical foundation and biological importance of C. A. B. Smith's mean measure of divergence. *Ossa* 4 (Suppl. 1):1–133.

Shrout, P.E., & Fleiss J.L (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.

Tardivo, D., Sastre, S., Ruquet, M., Thollon, L., Adalian, P., Leoneti, G., & Foti B. (2011). Three-dimensional modeling of the various volumes of canines to determine age and sex: a preliminary study. *Journal of Forensic Sciences*, 56, 766-770.

The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation, *Nature*, 526, 68-74.

Townsend, G., Hughes, T., Luciano, M., Bockmann, M., & Brook A. (2009). Genetic and environmental influences on human dental variation: a critical evaluation of studies involving twins, *Archives of Oral Biology*, 54, 45-51.

Turner, II C.G., Nichol, C.R & Scott, G.R. (1991). Scoring procedures for key morphological traits of the permanent dentition: the Arizona State University Dental Anthropology System. In M.A. Kelley & C.S. Larsen (Eds.), *Advances in Dental Anthropology* (pp. 13-31). New York, Wiley-Liss.

Zou, F., Lee, S., Knowles, M.R & Wright, F.A. (2010). Quantification of population structure using correlated SNPs by shrinkage principal components. *Human Heredity*, 70, 9-22.

Tables

TABLE 1. Mean Measure of Divergence (MDD) matrix showing distances between the Colombian and reference population samples based on dichotomized trait frequencies.

	WE	NE	NA	WA	SA	KHO	AA	NWNA	NSA	MED
WE	0.000									
NE	**0.036**	0.000								
NA	**0.020**	**0.035**	0.000							
WA	0.403	0.392	0.294	0.000						
SA	0.187	0.192	0.101	0.117	0.000					
KHO	0.343	0.430	0.255	0.118	0.117	0.000				
AA	0.550	0.573	0.497	0.523	0.404	0.563	0.000			
NWNA	0.711	0.718	0.618	0.604	0.515	0.699	**0.043**	0.000		
NSA	0.843	0.833	0.724	0.688	0.598	0.818	0.117	**0.023**	0.000	
MED	0.228	0.204	0.188	0.469	0.368	0.430	0.388	0.427	0.468	0.000

Note: WE, Western Europe; NE, Northern Europe; NA, North Africa; WA, Western Africa; SA, South Africa; Kho, Khoisan; AA, American Arctic; NWNA, Northwest North America; NSA; North and South Native Americans; MED, Colombia. Values in bold are not significant at $P<0.025$.

TABLE 2. R^2 and weight values from regression analysis of genetic ancestry on each dental trait in the Colombian sample.

Trait code	Africa	Weight	Europe	Weight	America	Weight
SSUI1	0.009	0.002	0.013	**0.896**	0.024	**1**
SSUI2	0.009	0.004	0.019	**0.992**	0.029	**1**
DSUI1	0.008	**0.538**	0.009	**0.134**	0.025	**1**
DSUI2	0.008	**0.122**	0.010	**0.746**	0.033	**1**
CAUI2	0.008	**0.156**	0.010	**0.876**	0.009	**0.112**
PSUI2	0.009	0.008	0.008	**0.5**	0.018	**0.994**
WINUI1	0.009	0.004	0.009	**0.148**	0.009	**0.256**
LCUI1	0.009	0.002	0.010	0.042	0.009	0.032
LCUI2	0.008	**0.120**	0.011	**0.874**	0.009	**0.608**
DIASUI1	0.009	0.012	0.009	**0.14**	0.009	**0.628**
IGUI1	0.009	0	0.010	0.058	0.009	0.068
IGUI2	0.009	0	0.009	0.002	0.009	0
TDUI1	0.009	0.02	0.009	0.016	0.009	0.002
TDUI2	0.008	**0.208**	0.009	**0.296**	0.010	0.018
TDUC	0.016	**0.972**	0.008	**0.652**	0.009	0
MRUC	0.009	0.044	0.008	**0.416**	0.008	**0.278**
DARUC	0.009	0.006	0.009	**0.128**	0.009	**0.112**
ARUP1	0.008	**0.34**	0.010	**0.756**	0.009	**0.126**
ARUP2	0.008	**0.13**	0.009	0.014	0.009	0.002
ODOUP1	0.008	0.006	0.010	0.022	0.009	0.006
AMTUP1	0.009	0	0.009	0.01	0.009	0.006
AMTUP2	0.007	**0.228**	0.010	0.008	0.009	0.002
CAUP2	0.008	**0.528**	0.011	**0.824**	0.009	**0.258**
X3CUP1	0.008	0.008	0.010	0.02	0.009	0.01
MetUM1	0.008	**0.556**	0.015	**0.96**	0.008	**0.454**

MetUM2	0.009	**0.776**	0.019	**0.994**	0.010	**0.704**
X3CUM2	0.008	**0.684**	0.010	0.002	0.012	**0.85**
HipUM1	0.008	**0.732**	0.009	**0.158**	0.010	0
HipUM2	0.008	**0.716**	0.009	0.014	0.009	**0.1**
C5UM1	0.014	**0.94**	0.010	**0.676**	0.009	0
C5UM2	0.012	**0.692**	0.009	0.07	0.008	**0.224**
CarUM1	0.009	0.004	0.008	**0.152**	0.009	**0.17**
CarUM2	0.008	0.014	0.009	0.004	0.010	0.004
ParUM1	0.009	0.002	0.009	0.02	0.009	0.014
ParUM2	0.008	**0.454**	0.009	0.012	0.009	0.012
SSLI1	0.009	0.01	0.033	**1**	0.044	**1**
SSLI2	0.009	0.002	0.029	**1**	0.044	**1**
DSLI1	0.009	0.022	0.016	**0.958**	0.018	**0.942**
DSLI2	0.008	0.004	0.014	**0.92**	0.019	**0.968**
CALI1	0.008	0.042	0.009	0.054	0.009	0.026
DARLC	0.009	0.002	0.009	0.004	0.010	0
LCVLP1	0.008	**0.278**	0.009	0	0.008	**0.2**
LCVLP2	0.008	0.094	0.010	0.006	0.009	0.008
EPLP1	0.008	0	0.010	0.012	0.009	0.086
EPLP2	0.008	0.01	0.009	0.004	0.009	0
ARPrLP1	0.008	**0.398**	0.010	0.012	0.009	0.042
ARPrLP2	0.008	**0.602**	0.009	**0.526**	0.010	0
ODOUL1	0.008	**0.278**	0.009	0.08	0.009	0.028
AFLM1	0.007	**0.162**	0.012	**0.936**	0.009	**0.57**
AFLM2	0.008	0.002	0.010	**0.812**	0.023	**1**
GPLM1	0.009	0.028	0.009	0.006	0.009	0.004
GPLM2	0.008	**0.108**	0.008	**0.414**	0.009	0.068

CNLM1	0.012	**0.888**	0.020	**0.998**	0.008	**0.312**
CNLM2	0.008	**0.32**	0.023	**0.992**	0.016	**0.892**
C5LM1	0.009	**0.622**	0.008	**0.344**	0.009	0.004
C5LM2	0.008	**0.218**	0.016	**0.962**	0.012	**0.804**
C6LM1	0.008	**0.384**	0.024	**0.996**	0.010	**0.628**
C6LM2	0.009	0.01	0.010	0.006	0.009	0.01
C7LM1	0.009	**0.724**	0.007	**0.72**	0.009	0.002
C7LM2	0.009	0.046	0.011	**0.854**	0.021	**0.916**
DWLM1	0.008	0.018	0.009	**0.672**	0.009	**0.218**
DWLM2	0.009	0.014	0.010	0	0.009	0.008
DTCLM1	0.008	0.028	0.009	0.008	0.010	0.018
DTCLM2	0.008	**0.244**	0.010	0.006	0.009	0.008
PrtostLM1	0.009	0.002	0.010	0.01	0.010	0.014
PrtostLM2	0.009	0.096	0.008	**0.204**	0.009	0.022

Note: Boldfaced values denote higher weights (> 0.1). Trait codes as in Supplementary Table 1

TABLE 3. R^2 values from a regression analysis of continental genetic ancestry on components obtained from Independent Components Analysis (ICA) or Principal Components Analysis (PCA) of the dental traits examined.

Variables used	African	European	Native American
All traits	0.008	0.021	0.034
ICA, 3 Components (31.3% of total variance)	0.018	0.045	0.034
ICA, 6 Components (52.5% of total variance)	0.025	0.042	0.037
PCA, 20 Components (86% of total variance)	0.008	0.041	0.046
ICA+PCA combined	0.012	0.041	0.046

Figures

Figure 1. Distribution of individual African (A), European (B) and Native American (C) ancestry obtained using genome-wide SNP data in the Colombian sample examined here.

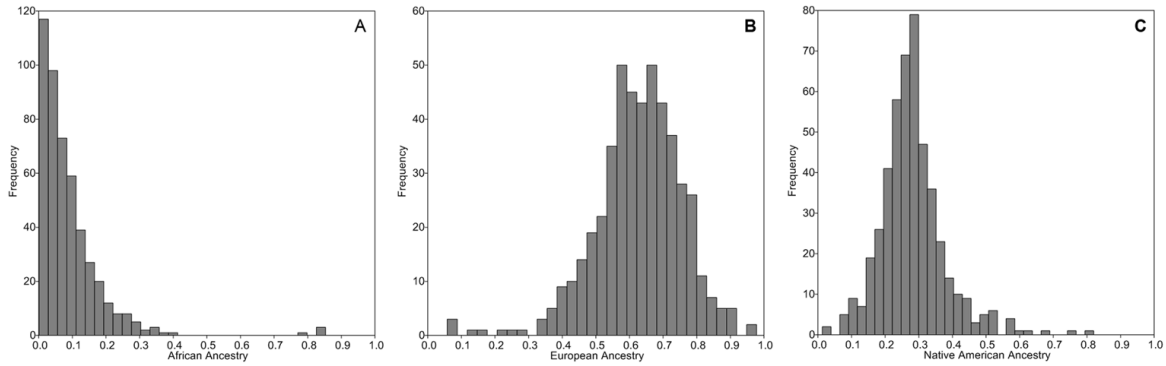


Figure 2 Frequency of the 16 dental traits used in the biodistance analysis in the Colombian and reference population samples. Sample codes: MED: Colombia; WA: Western Africans; KHO; Khoisan; SA: Southern Africans; WE: Western Europeans; NE: Northern Europeans; NA: Northern Africans; AA: American Artic; NWNA: Northwest North America; NSA: North and South Native Americans. Trait codes are as in Supplementary Table 1.

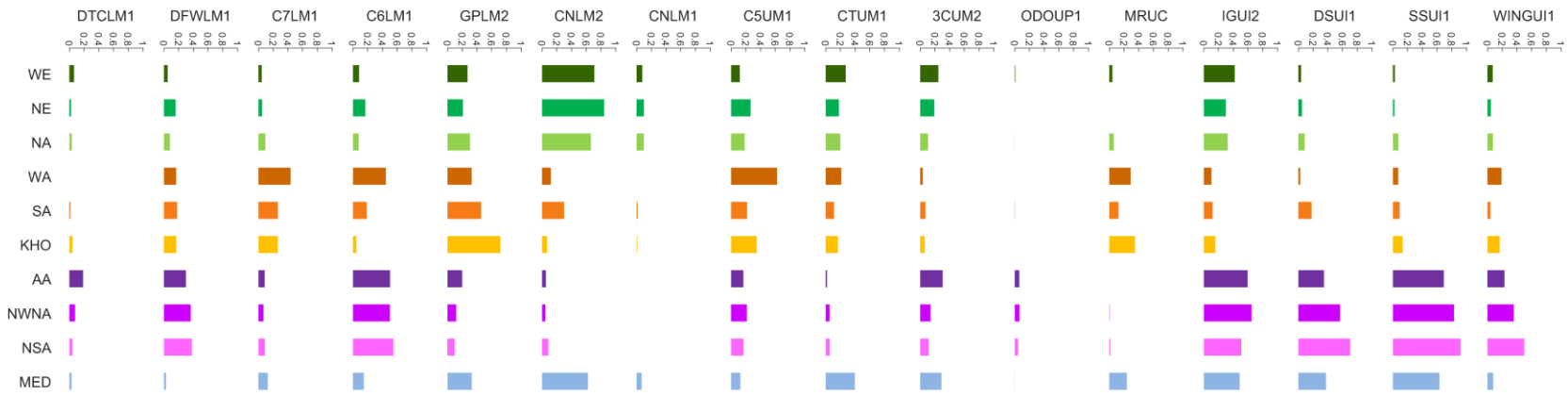
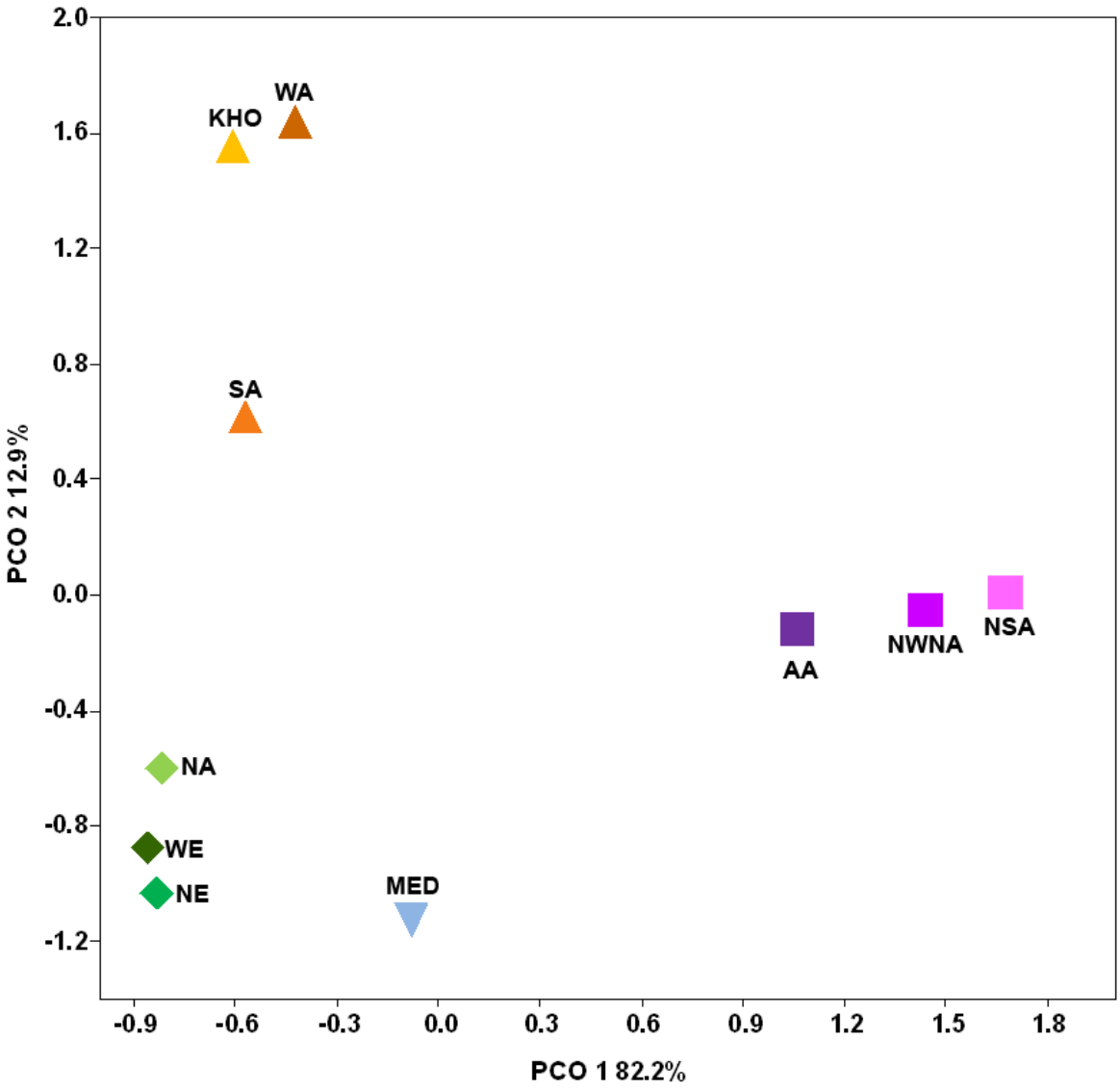

DTCLM1
DFWLM1
C7LM1
C6LM1
GPLM2
CNLM2
CNLM1
C5UM1
CTUM1
3CUM2
ODOUP1
MRUC
IGUI2
DSUI1
SSUI1
WINGUI1
0
0.2
0.4
0.6
0.8
1
WE
NE
NA
WA
SA
KHO
AA
NWNA
NSA
MED

Figure 3. Principal Coordinate Analysis (95.1% of the variance explained) of the MMD matrix (Table 1) displaying the relatedness of the population samples examined. MED: Colombia; WA: Western Africans; KHO; Khoisan; SA: Southern Africans; WE: Western Europeans; NE: Northern Europeans; NA: Northern Africans; AA: American Artic; NWNA: Northwest North America; NSA: North and South Native Americans.



Supplementary Figure 1. Distribution of individual African (A), European (B) and Native American (C) ancestry obtained using genome-wide SNP data in the Chilean sample examined.