Received: 19 March 2021 | Revised: 23 July 2021 | Accepted: 15 September 2021

DOI: 10.1002/ajpa.24426

BRIEF COMMUNICATION

AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY

WILEY

Predicting skeletal stature using ancient DNA

Samantha L. Cox[1,2] | Hannah M. Moots[3] | Jay T. Stock[4,5] | Andrej Shbat[6] | Bárbara D. Bitarello[1] | Nicole Nicklisch[7] | Kurt W. Alt[7] | Wolfgang Haak[8] | Eva Rosenstock[9] | Christopher B. Ruff[10] | Iain Mathieson[1]

[1]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[2]Physical Anthropology Section, Penn Museum, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]Department of Anthropology, Stanford University, Stanford, California, USA

[4]Department of Anthropology, Western University, London, Ontario, Canada

[5]Department of Archaeology, Max Planck Institute for the Science of Human History, Jena, Germany

[6]Institute of Anatomy, First Faculty of Medicine, Charles University, Prague, Czech Republic

[7]Center of Natural and Cultural Human History, Danube Private University, Krems, Austria

[8]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

[9]Bonn Center for ArchaeoSciences, Institut für Archäologie und Kulturanthropologie, Universität Bonn, Bonn, Germany

[10]Center for Functional Anatomy and Evolution, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Correspondence
Samantha L. Cox and Iain Mathieson, Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
Email: coxsl@sas.upenn.edu; mathi@pennmedicine.upenn.edu

Funding information
Charles E. Kaufman Foundation, Grant/Award Number: KA2018-98559; German Research Foundation, Grant/Award Number: RO4148; Alfred P. Sloan Foundation, Grant/Award Number: FG-2018-10647

Abstract

Objectives: Ancient DNA provides an opportunity to separate the genetic and environmental bases of complex traits by allowing direct estimation of genetic values in ancient individuals. Here, we test whether genetic scores for height in ancient individuals are predictive of their actual height, as inferred from skeletal remains. We estimate the contributions of genetic and environmental variables to observed phenotypic variation as a first step towards quantifying individual sources of morphological variation.

Materials and methods: We collected stature estimates and femur lengths from West Eurasian skeletal remains with published genome-wide ancient DNA data ($n = 182$, dating from 33,000–850 BP). We also recorded genetic sex, genetic ancestry, date and paleoclimate data for each individual, and $\delta^{13}C$ and $\delta^{15}N$ stable isotope values where available ($n = 69$). We tested different methods of calculating polygenic scores, using summary statistics from four different genome wide association studies (GWAS) for height, and three methods for imputing missing genotypes.

Results: A polygenic score for height predicts 6.3% of the variance in femur length in our data ($n = 132$, SD = 0.0069%, $p = 0.001$), controlling for sex, ancestry, and date. This is consistent with the predictive power of height PRS in present-day populations and the low coverage of ancient samples. Comparatively, sex explains about 17% of the variance in femur length in our sample. Environmental effects also likely play a role in variation, independent of genetics, though with considerable uncertainty (longitude: $R^2 = 0.033$, SD = 0.008, $p = 0.011$). Genotype imputation did not improve polygenic prediction, and results varied based on the GWAS summary statistics we used.

Discussion: Polygenic scores explain a small but significant proportion of the variance in height in ancient individuals, though not enough to make useful predictions of individual phenotypes. However, environmental variables also contribute to phenotypic outcomes and understanding their interaction with direct genetic predictions will provide a framework with which to model how plasticity and genetic changes ultimately combine to drive adaptation and evolution.

KEYWORDS
genotype, human variation, phenotype, polygenic traits, population genetics, skeletal stature

1 | INTRODUCTION

One of the central goals of biological anthropology is to understand the environmental and genetic contributions to phenotypic change. Over short periods of time, the record of such change can illustrate human responses to stress, disease, and fluctuations in the physical environment. In the long term, if patterns persist, evolution occurs as genetic changes accumulate in the population, driven by both drift and selection. Research in humans generally aims to document and analyze signatures of historical trends in the physical bodies of individuals from the archeological and fossil records, as well as among living human populations, in order to infer causal relationships between environmental variables and biological patterns. However, while skeletal series covering long time periods and diverse environments are informative and recent analyses have become increasingly sophisticated, findings are often limited by an inability to separate the effects of genetics and environment and to further decompose genetic changes into the effects of genetic drift and natural selection. Some authors have indirectly estimated the genetic contribution to phenotypic variance and the relative contribution of natural selection and genetic drift by using patterns of neutral genetic variation, or proxies (e.g., Pomeroy et al., 2021; Roseman & Auerbach, 2015; Savell et al., 2016), but these are approximations which cannot fully separate these effects. This limitation can now be directly addressed. By combining data on the genetics of complex traits in living humans with ancient genomes, we are able to directly estimate the genetic contributions to stature in the archeological skeletal record as a first step towards quantifying sources of variation in human morphology.

The ability to generate genetic data from skeletal remains has had an enormous impact on studies of human history. By identifying genetic links among individuals and populations, ancient DNA allows us to reconstruct demographic histories on both large and small scales (Racimo et al., 2020; Skoglund & Mathieson, 2018), as well as the effects of natural selection (Marciniak & Perry, 2017). Genome-wide association studies (GWAS), involving hundreds of thousands of living individuals, have identified large numbers of single-nucleotide polymorphisms (SNPs) associated with hundreds of phenotypes (Visscher et al., 2017), and provide a foundation for studying the genetics of complex traits. Though the effect of any individual variant is typically small, the effects of many SNPs can be summed to produce a polygenic score (PRS) which can be thought of as an estimate of an individual's genetic potential or risk for a particular phenotype (Lewis & Vassos, 2020; Torkamani et al., 2018; Wray et al., 2007). Height is a polygenic trait with high heritability explained largely by common variants (Yang et al., 2010), and GWAS for height have directly identified thousands of significantly associated SNPs (Lango Allen et al., 2010; Wood et al., 2014; Yengo et al., 2018). Each one has only a tiny effect—±1–2 mm per SNP (Marouli et al., 2017)—but combined they explain around 25% of the phentoypic variance in present-day populations of European ancestry (Yengo et al., 2018).

To obtain phenotpyic data, stature estimation from long bones is relatively straightforward, assuming reasonable preservation of skeletal remains, and there is an excellent record of changes in stature in many parts of the world (Bogin & Keep, 1999; Cohen & Crane-Kramer, 2007; Rosenstock et al., 2019; Ruff, 2018). On a population level, changes in polygenic score in Europe computed from ancient DNA qualitatively track trends in stature in the European skeletal record (Cox et al., 2019). However, environmental effects on stature can still be large, as shown by 20th century secular trends (NCD Risk Factor Collaboration, 2016), and are not confined to recent history. For example, during the Bronze Age, genetic predictions suggest increasing stature, but estimated skeletal heights actually decrease (Cox et al., 2019). Many studies of stature variation focus on environmental effects such as nutritional level or socioeconomic status in either present-day (e.g., Silventoinen, 2003; Steckel, 1995) or ancient populations (e.g., Cohen & Crane-Kramer, 2007; Keep & Bogin, 1999; Koepke, 2016). Others focus on genetic effects, including those related to population movements and structure (e.g., Cox et al., 2019; Savell et al., 2016), but few have considered both types of effects simultaneously (Grasgruber et al., 2016). As these two variables are often correlated, we take a more comprehensive combined approach. By directly estimating the underlying genetic contributions to stature, we can begin to better understand the limits of phenotypic plasticity in response to environmental variables.

Polygenic scores have three main limitations. First, due to incomplete correction of population stratification in GWAS, they can capture environmental variation in present-day populations leading to ancestry-related biases (Berg et al., 2019; Sohail et al., 2019; Zaidi & Mathieson, 2020). Second, their accuracy decreases as genetic distance from the present-day European ancestry GWAS populations increases (Martin et al., 2019). Finally, their accuracy can be reduced by environmental differences, even within-population (Mostafavi et al., 2020). We therefore set out to test whether polygenic scores based on present-day GWAS predict height in ancient individuals by collecting data for individuals with both ancient DNA and skeletal measurements. This allows us to assess the validity of complex trait prediction in ancient individuals, and whether we can use this approach to understand the relationship between genetic and environmental components of stature and, perhaps, of other phenotypes.

2 | MATERIALS AND METHODS

2.1 | Data collection

We collected genetic data from published ancient DNA (aDNA) studies (31,000–850BP). Most individuals had pseudo-haploid genotypes at a set of 1.24 million SNPs (the "1240k array"; Haak et al., 2015; Fu et al., 2015) and for individuals with shotgun sequence data we randomly selected a single read at each of the covered 1240k sites. The majority of the available ancient DNA data comes from Western Eurasia and so we focused on this region, broadly defined (including individuals from up to 100°E longitude).

For each individual with published aDNA, we attempted to find data on their skeletal measurements. Some aDNA papers include stature or femur length measurements in their supplemental materials.

For other individuals, we searched archeological and anthropological literature for published data. However, the vast majority of published aDNA data come from skeletal individuals which are either unpublished or highly fragmented and therefore unmeasurable. We also report new measurements for 46 individuals (Table S1).

For each individual, we recorded maximum femur length, when available; otherwise, we recorded the estimated stature and estimation method. For individuals with published femur lengths, we estimated stature using the method of Ruff et al. (2012). We restricted analyses to adult individuals free from reported major pathology, specifically those that could have affected growth (including cribra orbitalia, porotic hyperostosis and indications of leprosy) or prevented accurate measurement (including femoral fractures, osteomalacia and severe osteoarthritis). Ultimately, from approximately 4000 published aDNA samples (Antonio et al., 2019; Brace et al., 2019; de Barros Damgaard et al., 2018; Fernandes et al., 2018; Fu et al., 2016; Furtwängler et al., 2020; González-Fortes et al., 2017; Krzewińska et al., 2018; Lipson et al., 2017; Margaryan et al., 2020; Martiniano et al., 2016; Mathieson et al., 2015, 2018; Mittnik et al., 2018, 2019; Narasimhan et al., 2019; Olalde et al., 2018, 2019; Schiffels et al., 2016; Schroeder et al., 2019; Sikora et al., 2017), we compiled metric data for 182 individuals (Alciati, 1967; Alpaslan-Roodenberg, 2001; Andrews & Thompson, 2016; Auerbach, 2004; Auerbach & Ruff, 2004; Berthold et al., 2008; Boroneant, 2010; Caffell & Holst, 2012; Cairns, 2015; Dunwell, 2007; Fokkens et al., 2017; Frei et al., 2019; Kitti, 2008; Kjellström, 2005; Köhler et al., 2017; Malmström et al., 2019; Massy, 2018; Pardini, 1977; Price et al., 2016; Rosenstock et al., 2019; Saag et al., 2020; Schiffels et al., 2016; Szczepanek, 2013; Tebelskis & Jankauskas, 2002). We removed 28 samples with more than 95% missing genetic data and one individual with an unusually short femur (AITI_95, see below), bringing the sample size to 153. Finally, 21 stature estimates used unknown methods and we removed them from the majority of tests, bringing the final sample size for most analyses to 132 (Table S1).

There are many methods for estimating living statures based on skeletal measurements. Population-specific methods are the most accurate, but are not available for every population (Ruff et al., 2012). Researchers ideally choose a method which has been developed on a population similar to that under investigation in terms of ancestry or body proportions; however, this is often not possible and there are a few methods which are most frequently used, even if not population-specific (e.g., Ruff et al., 2012; Sjøvold, 1990; Trotter & Gleser, 1952). Ideally, we would compare statures estimated using the same equations but that was not feasible since underlying data were not available in many cases.

We dealt with this issue in two ways. First, we included stature estimation method as a discrete factor in the linear regression to predict stature, but we were concerned about the statistically significant proportion of variance attributable to the estimation method in our models ($R^2 = 0.065$, $p = 0.020$) and the lack of population-specific equations. Due to this, we took a second approach: since the single bone which offers the most accurate stature estimate is the femur (Ruff, 2018; Ruff et al., 2012; Trotter & Gleser, 1952) we tried predicting maximum femur length rather than stature. Femur lengths were only published for 78 out of our 182 individuals. For the other individuals, since statures are estimated as linear functions of long bone lengths, we inverted the estimation equations to retrieve femur measurements corresponding to each individual stature (Table 1; for further discussion on this approach see Kopke and Baten (2005); Rosenstock et al. (2019)). For individuals for which stature had been estimated using a non-femur long bone, this procedure gives us the femur length which would have produced the originally estimated stature (Figure 1b). We confirmed that estimation method did not have a significant effect on femur lengths estimated using this approach ($p = 0.539$). We used femur length rather than estimated stature for most analyses, omitting 21 individuals for which the stature estimation method was not cited. Finally, the slope of the regression line from which the Bach (1965) formula is derived deviates from those of the other estimation methods and produces outlier femur measurements (Figure 1b). The applicability of the Bach (1965) method has also been questioned by other researchers (Sládek et al., 2015). This method was only used for a few individuals ($n = 3$), and as inclusion of these individuals did not affect prediction results, all were included in the analysis except for one. This individual (DNA sample ID: AITI_95, a genetic female from the Bronze Age German site of Kleinaitingen-Gewerbegebiet Nord) was estimated to have had an unusually short femur (estimated femur = 36 cm, estimated stature = 154 cm, Figure 1b) and was excluded as we could not be confident in the accuracy of the estimate.

We also collected other variables for inclusion in our models: ancestry, date, sex, climate, and diet. We estimated genetic ancestry by multi-dimensional scaling (MDS, with $k = 4$ dimensions) of the genetic data—referred to collectively as the "ancestry" variable. We determined the date of each sample based on the calibrated radiocarbon dates reported in the original publication of the genetic data. For the few samples for which there was no direct date available, we used the mid-point of the archeological date range. Genetic sex was reported in the original publications. Climate variables of mean daily temperature and annual precipitation were obtained from the 5 min (medium resolution, data points every 10 km) paleoclimate dataset available at PaleoClim.org (Brown et al., 2018). We extracted the relevant data using the raster package (Hijmans & van Etten, 2012) in R. Using geographic coordinates, we calculated the distance from the sites where skeletons were excavated to the surrounding climate data points (using the gdist() function from the *lmap* package; Wallace (2012)), and chose the point closest to the site to represent its climate. For Western Europe, most sites are within a few kilometers of available climate data; however, there are a handful of sites in present-day Russia and the Middle East which are quite far from any available PaleoClim data (200–1500 km). Latitude and longitude were included as climate surrogates since they have been previously correlated with reconstructed statures in Holocene Europeans (Niskanen et al., 2018). Both latitude and longitude are associated with climatic variation across Europe, with a general southwest to northeast cline in winter temperatures, for example (Huijzer & Vandenberghe, 1998; Niskanen et al., 2018). Some aspects of diet can be reconstructed

TABLE 1 Inverted stature estimation equations

Reference		Stature equation	Error (cm)	Femur equation
Manouvrier (1892)[a]	Females	49.319 + 2.543 Femur		$\frac{\text{Stature}-49.319}{2.543}$
	Males	71.065 + 2.13 Femur		$\frac{\text{Stature}-71.065}{2.13}$
Pearson (1899)	Females	72.844 + 1.945 Femur		$\frac{\text{Stature}-81.306}{1.88}$
	Males	81.306 + 1.88 Femur		$\frac{\text{Stature}-81.306}{1.88}$
Breitinger (1937)	Males	94.31 + 1.64 Femur	±4.8	$\frac{\text{Stature}-94.31}{1.64}$
Bach (1965)	Females	106.69 + 1.313 Femur	±4.1	$\frac{\text{Stature}-106.69}{1.313}$
Trotter and Gleser (1952)	White females	54.10 + 2.47 Femur	±3.72	$\frac{\text{Stature}-54.10}{2.47}$
	White males	61.41 + 2.38 Femur	±3.27	$\frac{\text{Stature}-61.41}{2.38}$
Sjøvold (1990)	Independent	49.96 + 2.63 Femur	±4.52	$\frac{\text{Stature}-49.96}{2.63}$
Ruff et al. (2012)	Females	43.56 + 2.69 Femur	±2.92	$\frac{\text{Stature}-43.56}{2.69}$
	Males	49.85 + 2.72 Femur	±3.21	$\frac{\text{Stature}-42.85}{2.72}$

[a]These equations are interpolated from the data given in the tables of the original publication.

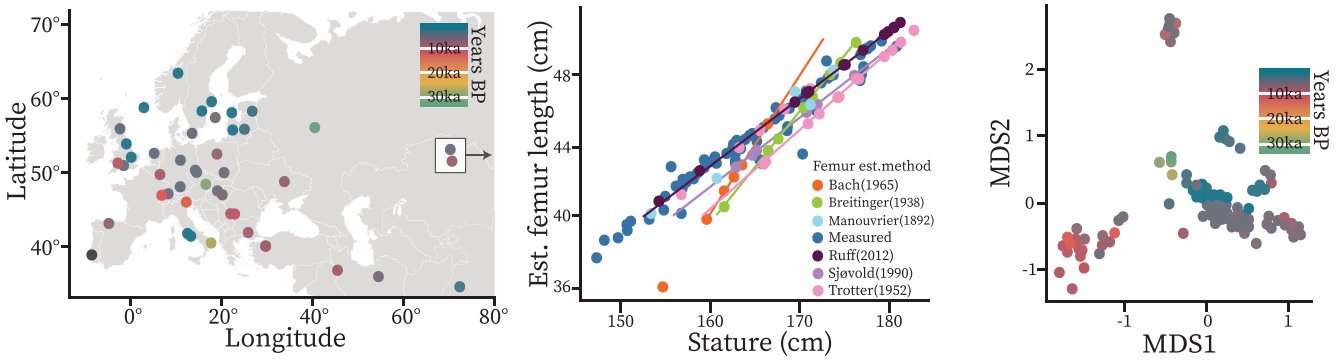

FIGURE 1 (a) Map of sites showing individuals with both ancient DNA and metric data; (b) the relationship between skeletal statures and femur length. Lines indicate the regression line for each stature estimation method. The stature for the outlying individual in the lower left was estimated using the Bach (1965) method, and was removed from further analysis. For all individuals with directly measured femora, stature was estimated using the method by Ruff et al. (2012); (c) the first two multi-dimensional scaling axes of the genetic data summarizing the genetic ancestry of the samples. Clusters correlate with time, and are largely associated with substantial shifts in genetic ancestry. The uppermost cluster represents the most eastern individuals in our sample, from present-day Siberia

from stable isotope data (O'Brien, 2015), so we searched for $\delta^{13}C$ and $\delta^{15}N$ values for the individuals in our study. We found data for about half the individuals ($n = 69$; Afshar et al., 2019; Antanaitis-Jacobs & Ogrinc, 2000; Antanaitis-Jacobs et al., 2009; Andrews & Thompson, 2016; Antonio et al., 2019; Berthold et al., 2008; Dunwell, 2007; Kjellström, 2005; Kjellström et al., 2009; Lehuray et al., 2006; Mathieson et al., 2018; Malmström et al., 2019; Müldner et al., 2011; Münster et al., 2018; Olalde et al., 2018; Price et al., 2016; Scheibner, 2016; Szczepanek, 2013; Stockhammer et al., 2015; Waterman et al., 2016; Weber et al., 2011, 2016).

2.2 | Polygenic scores

We calculated PRS using standing height summary statistics from the Neale Lab (2018) based on 337,000 people of British ancestry in UK Biobank. After intersecting these sites with those on the 1240k array, we tested a variety of PRS constructions. For the main analysis, we further restricted to HapMap3 SNPs (leaving 405,502 remaining) and estimated SNP weights using the infinitesimal model of *LDpred* (Vilhjalmsson et al., 2015) with an LD reference panel made up of 8000 randomly chosen individuals of British ancestry from the UK Biobank. We then computed polygenic scores using the --score command in *plink* (Chang et al., 2015). As an alternative approach, we also calculated PRS using a simpler clumping/thresholding method. Clumping parameters included all combinations of: 10^{-2}, 10^{-6}, and 10^{-8} *p*-value cut-offs; 100, 250, and 500 kb windows; and r^2 cutoffs of 0.1, 0.3, and 0.5. We used plink 1.9 (Chang et al., 2015) to clump (--clump) SNPs using these parameters with an LD reference panel made up of 500 individuals from five European populations (1000 Genomes Project Consortium, 2015), and to compute polygenic scores (--score). Missing genotypes are ignored in the score

calculation, equivalent to replacing them with the sample mean frequency—a conservative approach that shrinks scores towards the sample mean. To test how predictions are affected by the constraints of different GWAS, we computed *LDpred* PRS using summary statistics from three additional GWAS analyses. First, a different analysis of UK Biobank using *fastGWA* (Jiang et al., 2019) (456,000 people of European ancestry from UK Biobank); second, the GIANT consortium meta-analysis of 253,000 European individuals (Wood et al., 2014); finally, a GWAS of 192,000 Japanese individuals from Biobank Japan (Akiyama et al., 2019).

2.3 | Genotype imputation

The genetic data are relatively low coverage (haploid median = 0.607, range = 0.001–1). Therefore, it is not possible to infer diploid genotypes, and we use pseudo-haploid data that represents a single allele at each site. This limits performance of the PRS because effectively we only see at most half of each individual's genotype. In practice, samples often perform worse because many sites are missing data entirely. We attempted to improve individual prediction by testing three different methods of genotype imputation to infer diploid genotypes and missing sites. In each case, we restrict analysis to 1240k sites (i.e., we do not attempt to impute genotypes at sites not included on the 1240k array).

- First, we used a one-step approach using Beagle4 (Browning & Browning, 2016). We extracted reference/alternative read counts at each of the 1240k sites, and computed genotype likelihoods based on a binomial distribution of reads with a 2% rate of error. We ran Beagle4 (version 4.1) with these genotype likelihoods as input and a reference panel made up of the European ancestry populations from the 1000 Genomes Project Consortium (2015).
- Second, we used the pipeline described in Hui et al. (2020). Briefly, this uses a two-stage approach. The first stage uses Beagle4 with the European 1000 Genomes reference panel as in our first imputation method, but then implements a hard cutoff of genotype probability >0.99 to fix genotypes before running Beagle5 (Browning et al., 2018) with the entire 1000 Genomes reference panel to impute other sites. We use the default pipeline described in Hui et al. (2020), using ATLAS (Link et al., 2017) to compute genotype likelihoods.
- Finally, we ran GLIMPSE (Rubinacci et al., 2021) using its default pipeline. Unlike Beagle, this method is explicitly designed for use on low coverage sequence data, though not specifically for ancient DNA. We used the European 1000 Genomes sample as a reference panel.

In order to test the accuracy of the three methods, we down-sampled the high coverage (22×) individual NE1 (Gamba et al., 2014) 40-fold (to approximately 0.55× coverage), and included it in our imputation sample, and compared the imputed genotypes to diploid calls made on the complete data by Günther et al. (2018).

2.4 | Statistical analysis

We fit linear models of femur length (and stature) as a function of sex, PRS, genome-wide ancestry, and date, and also included stable isotope and climate variables when appropriate. The ancestry component includes 4 MDS axes. Date includes both date (years before present) and date squared, to allow for nonlinear effects (Figure 2b). We evaluated the contribution of each term based on the difference in R^2 between the full model and a reduced model without the term being tested; we refer to this difference simply as R^2. To test the significance of a particular variable, we permuted its value within the dataset, keeping constant those variables which were not being tested. We permuted each test variable 10,000 times. We computed p-values as the proportion of times that the R^2 value in the original data was greater than the R^2 of the permuted distribution. When permuting terms with multiple components, for example ancestry, we ensured that the relationship between each of the permuted components was maintained.

3 | RESULTS

The LDpred polygenic score predicts 6.3% of the variation in femur length in our data (R^2=0.063, SD = 0.007, p = 0.000) (Table 2), showing that PRS explains a small but statistically significant proportion of the variance in femur length , once the other variables are taken into account. For comparison, this is less than half of the variance explained by sex (R^2 = 0.17). The effect of PRS on femur length is similar in both sexes (Figure 2a), consistent with theoretical predictions (Rogers & Mukherjee, 1992) and empirical observations (Randall et al., 2013) of very limited sex differences in genetic effects on height. Ancestry had R^2 = 0.051 (SD = 0.0114, p = 0.016); a contribution that may also partially reflect genetic effects. In a model without ancestry, the R^2 of the PRS term increases to 0.082, indicating systematic variation in PRS with ancestry. Date has R^2 = 0.072 although this largely reflects differences between the Early Upper Paleolithic and later populations (Figure 2b). When we predicted stature, rather than femur length, including a constant term for estimation method, we found consistent results, though with a slightly lower R^2, possibly due to the error introduced by the variation in estimation methods (PRS R^2 = 0.052, SD = 0.0051, p = 0.000) (Table 2).

We also tested the predictive power of PRS computed using different GWAS summary statistics (Table S2). The Jiang et al. (2019) summary statistics performed better than the Neale lab statistics (R^2 = 0.084 vs. 0.063). The GIANT summary statistics (Wood et al., 2014) had very low R^2 (0.009), though when ancestry was removed from the model, GIANT predictions improved (R^2 = 0.044, p = 0.007). Biobank Japan summary statistics (Akiyama et al., 2019) also had very low predictive power (R^2 = 0.008), consistent with the low R^2 of East Asian ancestry PRS in present-day European populations (Martin et al., 2019).

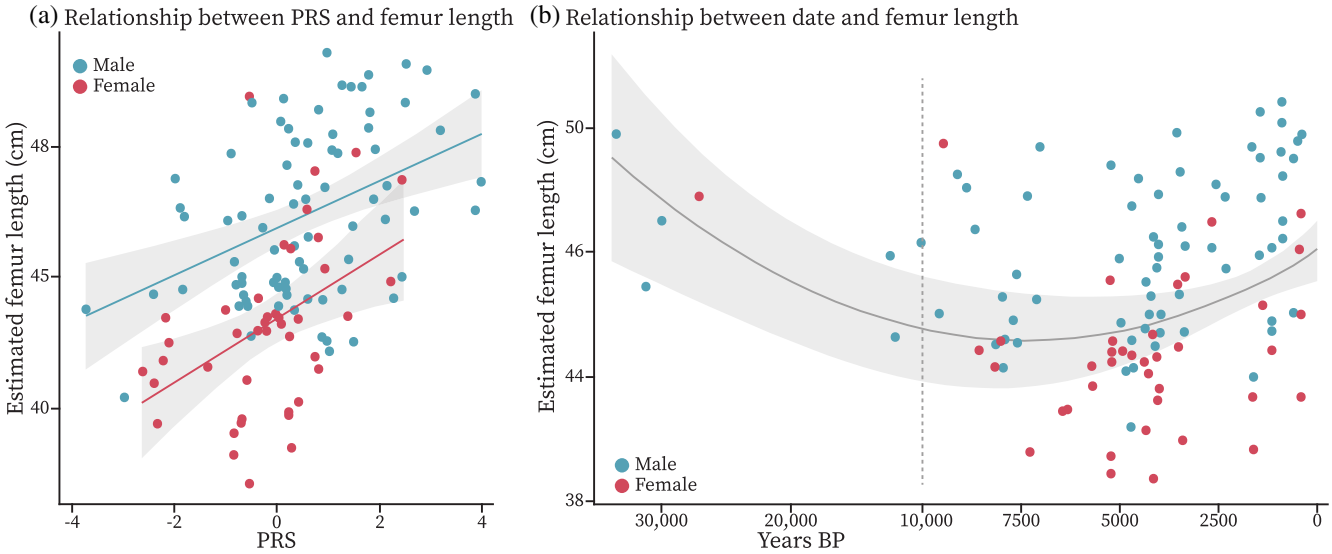

FIGURE 2 (a) Plot of the linear relationship between polygenic score (PRS) and femur length. Higher PRS values are associated with longer femur lengths in the data. Colors indicate sex, the lines are the regression lines for males and females separately, and the gray shadows are the 95% confidence intervals. For our main results, we assume the slope of this regression is identical between sexes. The R^2 of PRS is 0.063 and of sex is 0.17. (b) Plot of the fitted quadratic relationship between date and femur length ($R^2 = 0.072$). Colors indicate sex, the solid gray line is the quadratic fit line for the pooled-sex group, the gray shadow is the 95% confidence interval, and the vertical dashed line indicates the change in *x*-axis plotting scale

TABLE 2 Linear model results

Model	Variables	R^2	β	SE	Permuted values	
					R^2 SD	*p*-value
Femur length (cm) with ancestry ($n = 132$)	PRS	0.063	0.89 cm/SD	0.23	0.0065	0.000
	Ancestry	0.051			0.0114	0.016
	Sex (M)	0.170	2.59 cm	0.41	0.0074	0.000
	Date	0.072			0.0085	0.000
Femur length, no ancestry ($n = 132$)	PRS	0.082	0.97 cm/SD	0.23	0.0073	0.000
	Sex (M)	0.184	2.67 cm	0.42	0.0079	0.000
	Date	0.043			0.0085	0.005
Stature ($n = 153$)	PRS	0.052	2.34 cm/SD	0.57	0.0051	0.000
	Ancestry	0.041			0.0094	0.015
	Sex (M)	0.199	8.06 cm	1.06	0.0066	0.000
	Date	0.036			0.0068	0.005
	Estimation method	0.065			0.0141	0.020

Most of the predictive power of the PRS comes from samples above median coverage (haploid median coverage = 0.63, Figure 3b, c). To test how PRS construction might affect prediction results, we also constructed PRS using a more traditional clumping and thresholding method. For this, we computed PRS from the Neale lab summary statistics using p-value cut-offs of 10^{-2}, 10^{-6}, and 10^{-8}, r^2 of 0.1, 0.3, and 0.5, with 100, 250, and 500 kb windows. LDpred provides the highest predicted R^2 values, though there are sets of clumping parameters which perform similarly. In Figure 3a, we report the R^2 values for all tested clumping parameter values.

Attempting to improve prediction, we imputed diploid genotypes, including missing SNPs, in our dataset (imputed diploid median coverage = 0.987) using three different imputation methods (Table 3, Table S3). All three imputation methods performed similarly in terms of genotype accuracy. Prediction using GLIMPSE and Hui et al. (2020) was comparable in R^2 to the unimputed data, while the one-step

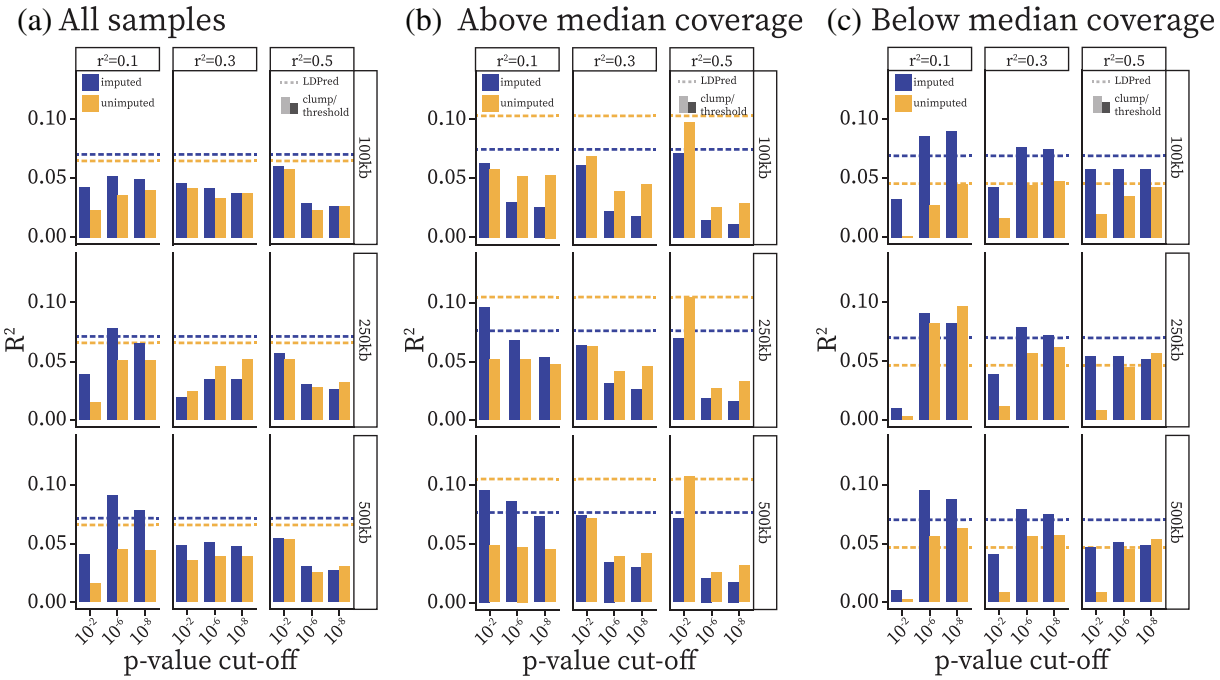

FIGURE 3 Bar plots showing the R^2 value for each set of clumping and thresholding parameter combinations (window size, r^2, p-value cut-off) used for PRS construction. Horizontal dashed lines show the R^2 values from the LDpred polygenic score. Regression models include sex, date, and ancestry as covariates

TABLE 3 Accuracy of imputation methods

Method	Precision (all)	Recall (all)	Precision (het)	Recall (het)	R^2
Beagle4 1-step	89.8%	89.4%	88.6%	86.3%	0.050
Hui et al.	88.9%	89.8%	86.7%	87.8%	0.068
GLIMPSE	89.4%	88.8%	87.7%	86.0%	0.067

Precision and recall for 1240k sites in the NE1 high coverage genome using the Günther et al. (2018) calls as a truth dataset (including all sites, and heterozygous (het) sites only). R^2 is for the first model in Table 2, using PRS computed with each imputed dataset.

TABLE 4 Linear model results for femur length, climate and stable isotope variables

					Permuted values	
Model	Variables	R^2	β	SE	R^2 SD	p-value
Femur length with climate and geographic variables ($n = 132$)	PRS	0.057	0.86 cm/SD	0.23	0.0069	0.001
	Ancestry	0.064			0.0125	0.003
	Sex (M)	0.131	2.34 cm	0.41	0.0066	0.000
	Date	0.087			0.0093	0.000
	Latitude	0.001	0.02 cm/°	0.04	0.0076	0.628
	Longitude	0.033	0.08 cm/°	0.03	0.0083	0.011
	Avg temp	0.013	0.13 cm/°	0.07	0.0057	0.071
	Annual precip	0.009	0.00 cm/cm	0.00	0.0059	0.132
Femur length with diet variables ($n = 69$)	PRS	0.004	0.29 cm/SD	0.40	0.0110	0.488
	Ancestry	0.044			0.0219	0.277
	Sex (M)	0.288	3.51 cm	0.58	0.0173	0.000
	Date	0.041			0.0164	0.081
	$\delta^{13}C$	0.009	−0.38 cm/‰	0.36	0.0108	0.287
	$\delta^{15}N$	0.017	0.28 cm/‰	0.19	0.0110	0.147

Beagle4 approach performed slightly worse. In general, the GLIMPSE and Hui et al. (2020) methods produce an improvement in prediction for low coverage samples, while the Beagle4 and unimputed predictions are similar for all methods of PRS calculation. In the high coverage samples, imputation generally improved prediction for the clumping/thresholding PRS, but actually decreased the R^2 for the LDpred PRS (Figure 3). Overall, genotype imputation provides little improvement over the unimputed data, but might improve predictions when the majority of samples are of low coverage. If we perform MDS on the imputed genotypes, the ancestry term is no longer significant ($R^2 = 0.006$, $p = 0.58$ using the one-step Beagle4 imputed genotypes), suggesting some loss of information.

When we include geographic and climate variables in the model, the contributions of annual precipitation, mean daily temperature and latitude, are non-significant (Table 4, all R^2 < 1%). Longitude has an R^2 of 3% ($p = 0.011$; an increase of 0.08 cm/°), but its contribution to the model changes depending on which other variables are included. For instance, when longitude is removed from the model, the effect of ancestry decreases (Table S4), suggesting a complex relationship which may obscure its effects. It is likely that this variable is a proxy for others that are not included in our models, and so this significant value is difficult to interpret. Finally, the effects of stable isotope values are not significant (Table 4), though this might be due to the relatively small sample size ($n = 69$). Indeed, neither PRS nor ancestry is significant in this subset. Nonetheless, our results suggest the possibility that increasing δ^{15}N and decreasing δ^{13}C could be associated with increased femur length (increase of 0.3 cm/‰δ^{15}N; decrease of 0.4 cm/‰δ^{13}C).

4 | DISCUSSION AND CONCLUSION

We find that that we can predict a small but statistically significant proportion of individual height variation using polygenic scores in ancient individuals. The analysis also gives us the first indication of the proportions of phenotypic variation attributable to some of the major determinants of morphology. While the present analysis only focuses on one trait, this gives us a baseline for evaluating the contributions of variation in other traits and to begin to quantify the relationship between genetics and environment in phenotypic outcomes. Given that our data are pseudo-haploid, meaning we observed only one chromosome of each chromosome pair, we would automatically expect our PRS predictions to perform approximately half as well as they would on diploid data. From there, on average, approximately half of the SNPS are missing in each individual, decreasing our expected predictions by another half. Therefore, we would expect to be able to predict about one quarter of the variation that can be predicted in present-day Europeans, which is roughly consistent with our findings. Our results show that polygenic scores cannot accurately predict individual traits, but do support their application to the quantitative study of evolutionary trends and environmental relationships on a population level (Cox et al., 2019).

Though we used Neale Lab summary statistics for the main analysis, the PRS calculated from the Jiang et al. (2019) GWAS give better predictions. While these both used UK Biobank data, the Jiang et al. (2019) analysis includes all 456,000 European ancestry individuals while the Neale Lab analysis restricts to the subset of 337,000 British ancestry individuals. The increased predictive power is likely due to the larger sample size, despite the less homogeneous sample. Despite its sample size of 223,000 individuals, predictions based on the GIANT summary statistics were quite poor. It is possible that the relatively high residual population stratification in GIANT (Berg et al., 2019; Sohail et al., 2019) obscures any individual-level predictive power. Surprisingly, we find that genotype imputation does not increase predictive accuracy overall. One possibility is that imputation biases genetic variation to be similar to samples in the reference panel, leading to no increase in predictive power. This could cause the decrease in predictive power seen in the high coverage samples as they become more homogenized and similar to the reference panel.

There is a correlation between PRS and ancestry: genome-wide ancestry explains a similar amount of variation in height as the PRS, but when ancestry is removed, the variation explained by the PRS increases. There are two non-exclusive explanations for this observation. One is that genetic height varies systematically with ancestry—consistent with the observation that, on a population level, stature tracks genetically predicted height through time (Cox et al., 2019). A large portion of the predicted genetic change in stature is attributable to major admixture events, which may therefore make a substantial contribution to changes in stature over time. Differences in genetic height among populations do not necessarily indicate directional selection—substantial differences can also arise under neutrality or even stabilizing selection (Harpak & Przeworski, 2020). A second possible explanation is that ancestry is spuriously correlated with environmental variables from the GWAS population. Known as population stratification, this is a common and potentially strong source of bias in GWAS analysis, and while measures are taken to reduce its impact, there can still be evidence of residual population stratification in the GWAS results (Berg et al., 2019; Sohail et al., 2019). However, for this to affect our study it would also require a somewhat coincidental correlation between ancient and present-day stratification. With current methods and data, the signatures of residual structure and ancestry-linked variation would appear identical. However, even if the contribution of genome-wide ancestry is entirely driven by stratification, the polygenic score still explains a significant proportion of phenotypic variation beyond its interaction with ancestry. Similarly, the date term in our model could represent both genetic contributions not captured by the PRS, and correlated changes in nongenetic factors.

Beyond the genetic component, dietary variables can have a substantial impact on height outcomes. Nitrogen values are mainly associated with dietary protein intake from both plant and animal sources, but are also correlated to factors such as climate (O'Brien, 2015; Scheibner, 2016), and there is an established link between protein malnutrition/undernourishment and stunting of linear growth in children (Ghosh, 2016). Given this, we would expect to see a positive

trend between nitrogen values and femur length which is present, though not significant, in our data. Carbon values are more indicative of dietary plant resources, and of the terrestrial versus marine versus limnic provenance of food (O'Brien, 2015; Scheibner, 2016). C^4 plants, such as millet, lead to lower $\delta^{13}C$ values and became widespread in Western Eurasia after ca. 3000 BCE. While Paleolithic diets were mainly terrestrial, increased variance of $\delta^{13}C$ values around 10,000 BCE reflect the increased exploitation of aquatic food resources (Scheibner, 2016). Hence, our expectations for the effect of $\delta^{13}C$ on height are unclear. Moreover, stable isotope values in general may co-vary to some extent with both date and climate. Another issue is that we did not control for the bone element from which collagen samples originated. Samples might not necessarily reflect the diet of the individual during the developmental period that is relevant for the establishment of stature. Thus, we consider the interpretation of isotope values in our study as generally representative of subsistence patterns, rather than quantitative assays of relevant diet.

Previous work found a relationship between latitude and height in Europe which we do not observe in our sample. Cox et al. (2019) suggested that the observed latitudinal trend might be genetically driven by post-Neolithic Steppe migrations; however, even if we remove the ancestry term from our model, latitude is still not significant. However, our sample is biased towards Northern European collections for which we found more published metrics on DNA sampled individuals. Lack of a substantial Southern European sample might explain why we do not see a relationship.

Longitude has also previously been shown to correlate with stature in the European pre-Bronze Age periods (Cox et al., 2019; Ruff, 2018), as have climate variables (Ruff, 2018). We do replicate this observation. However, this is partly driven by the relatively tall individuals from the Danube Gorges region of Southeastern Europe (12 individuals in our sample). It has been well documented that the populations of this region do not follow the same height decreases that affect the rest of the continent through history. Since the nutritional status and general environment of the Danube Gorges has been considered less than ideal in recent times, some authors have suggested that the consistently tall stature in this region has a genetic basis (see citations in Ruff and Holt (2018) for further discussion). In our analysis, however, neither genetic ancestry nor polygenic score predict this variation although there may be genetic factors that we do not capture. The trend might also be driven by environmental factors, though we do not have the data here to speculate about what factors might be involved. This motivates further study of the basis of the distinct trends in the Danube Gorges.

It is not currently practical to use genetic (or environmental) data to predict individual phenotypes for height or other complex traits. However, our study shows how polygenic scores can begin to separate the effects of genetics and environment on a population level. We have shown that genetic variation can independently predict stature, validating the use of polygenic scores to track evolutionary changes (Cox et al., 2019). Future work should therefore focus on compiling anthropometric, genetic, and environmental data, as our results show promise for the application of this approach on more comprehensive data. With larger samples and more detailed information about environmental covariates, more accurate quantification of the role of environment and therefore of the relative importance of genetics and environment should be possible.

DATA AVAILABILITY STATEMENT

All ancient DNA data used in this paper are publicly available, see respective citations for the sources. Summary statistics used for the PRS calculations are publicly available from the Neale Lab (2018) or as described in the original sources. Skeletal metrics, polygenic scores, dietary isotopes climate variables, and relevant citations for each individual analyzed in this paper are provided in Supplementary Table S1.

ACKNOWLEDGMENTS

This research was supported by a Research Fellowship from the Alfred P. Sloan Foundation (FG-2018-10647) and a New Investigator Research Grant from the Charles E. Kaufman Foundation (KA2018-98559). Some isotope and skeletal data were collected in collaboration with Alisa Scheibner and Julia Ebert, within the Emmy-Noether-Projekt "LiVES" funded by German Research Foundation Grant RO4148/1 (PI Eva Rosenstock). This research made use of the UK Biobank Resource under Application 33923.

AUTHOR CONTRIBUTIONS

Samantha L. Cox: Conceptualization (equal); data curation (lead); formal analysis (lead); investigation (lead); methodology (equal); project administration (equal); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Hannah M. Moots:** Resources (supporting); writing – review and editing (equal). **Jay T. Stock:** Resources (supporting); writing – review and editing (equal). **Andrej Shbat:** Resources (supporting); writing – review and editing (equal). **Bárbara D. Bitarello:** Formal analysis (equal); methodology (supporting); writing – review and editing (equal). **Nicole Nicklisch:** Resources (equal); writing – review and editing (supporting). **Kurt W. Alt:** Resources (supporting); writing – review and editing (equal). **Wolfgang Haak:** Resources (supporting); writing – review and editing (equal). **Eva Rosenstock:** Resources (supporting); writing – review and editing (equal). **Christopher B. Ruff:** Conceptualization (supporting); writing – review and editing (equal). **Iain Mathieson:** Conceptualization (equal); data curation (supporting); formal analysis (equal); funding acquisition (lead); methodology (equal); supervision (equal); writing – review and editing (equal).

CONFLICT OF INTEREST

The authors have no conflicts of interest.

ORCID

Samantha L. Cox https://orcid.org/0000-0003-2557-6711
Andrej Shbat https://orcid.org/0000-0001-8866-5207
Nicole Nicklisch https://orcid.org/0000-0001-7423-8343
Kurt W. Alt https://orcid.org/0000-0001-6938-643X

Eva Rosenstock https://orcid.org/0000-0002-6693-4303
Christopher B. Ruff https://orcid.org/0000-0002-2932-3634

REFERENCES

1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74.

Afshar, Z., Millard, A., Roberts, C., & Gröcke, D. (2019). The evolution of diet during the 5th to 2nd millennium BCE for the population buried at Tepe Hissar, north-eastern central Iranian plateau: The stable isotope evidence. *Journal of Archaeological Science: Reports*, *27*, 101983.

Akiyama, M., Ishigaki, K., Sakaue, S., Momozawa, Y., Horikoshi, M., Hirata, M., Matsuda, K., Ikegawa, S., Takahashi, A., Kanai, M., Suzuki, S., Matsui, D., Naito, M., Yamaji, T., Iwasaki, M., Sawada, N., Tanno, K., Sasaki, M., Hozawa, A., … Kamatani, Y. (2019). Characterizing rare and low-frequency height-associated variants in the japanese population. *Nature Communications*, *10*(1), 4393.

Alciati, G. (1967). *I resti ossei umani delle necropoli dello Swat, W. Pakistan, Volume 1. Reports and Memoirs* (Vol. VIII). ISIAO.

Alpaslan-Roodenberg, S. (2001). Newly found human remains from Manteşe in the Yenişehir plain: The season of 2000. *Anatolica*, *27*, 1–14.

Andrews, P., & Thompson, S. (2016). A early beaker funerary monument at Porton down, Wiltshire. *Wiltshire Archaeological & Natural History Magazine*, *109*, 38–82.

Antanaitis-Jacobs, I., & Ogrinc, N. (2000). Chemical analysis of bone: Stable isotope evidence of the diet of Neolithic and Bronze Age people in Lithuania. *Istorija*, *45*, 3–12.

Antanaitis-Jacobs, I., Richards, M., Daugnora, L., Jankauskas, R., & Orginc, N. (2009). Diet in early Lithuanian prehistory and the new stable isotope evidence. *Archaeologia Baltica*, *12*, 12–30.

Antonio, M. L., Gao, Z., Moots, H. M., Lucci, M., Candilio, F., Sawyer, S., Oberreiter, V., Calderon, D., Devitofranceschi, K., Aikens, R. C., Aneli, S., Bartoli, F., Bedini, A., Cheronet, O., Cotter, D. J., Fernandes, D. M., Gasperetti, G., Grifoni, R., Guidi, A., … Pritchard, J. K. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science*, *366*(6466), 708–714.

Auerbach, B. M. (2004). The Goldman Dataset. https://web.utk.edu/auerbach/GOLD.htm.

Auerbach, B. M., & Ruff, C. B. (2004). Human body mass estimation: A comparison of "morphometric" and "mechanical" methods. *American Journal of Physical Anthropology*, *125*(4), 331–342.

Bach, H. (1965). Zur Berechnung der Köperhöhe aus den langen Gliedmaßenknocken weiblicher Skelette. *Anthropologischer Anzeiger*, *20*, 12–21.

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., Boyle, E. A., Zhang, X., Racimo, F., Pritchard, J. K., & Coop, G. (2019). Reduced signal for polygenic adaptation of height in UKbiobank. *eLife*, *8*, e39725.

Berthold, B., Alt, K., Bramanti, B., Drings, S., Kranzbühler, J., Meyer, C. (2008). *Die Totenhütte von Benzingerode. Archäologie und Anthropologie*. Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt, Landesmuseum für Vorgeschichte.

Bogin, B., & Keep, R. (1999). Eight thousand years of economic and political history in Latin America revealed by anthropometry. *Annals of Human Biology*, *26*, 333–351.

Boroneant, A. (2010). Ostrovul Corbului - o discuţe asupra mormintelor mezolitice şi neolitice timpurii. *Buletinui Muzeului Judeţean Teleorman, Serie Arheologie*, *2*, 5–27.

Brace, S., Diekmann, Y., Booth, T. J., van Dorp, L., Faltyskova, Z., Rohland, N., Mallick, S., Olalde, I., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Stewardson, K., Martiniano, R., Walsh, S., Kayser, M., Charlton, S., Hellenthal, G., Armit, I., … Barnes, I. (2019). Ancient genomes indicate population replacement in early Neolithic Britain. *Nature Ecology & Evolution*, *3*(5), 765–771.

Breitinger, E. (1937). Zur Berechnung der Köperhöhe aus den langen Gliedmaßenknocken. *Anthropologischer Anzeiger*, *14*, 249–274.

Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C., & Haywood, A. M. (2018). PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data*, *5*(1), 180254.

Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. *American Journal of Human Genetics*, *98*(1), 116–126.

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *American Journal of Human Genetics*, *103*(3), 338–348.

Caffell, A., & Holst, M. (2012). Osteological Analysis 3 and 6 Driffield Terrace, York, North Yorkshire. Tech. Rep. 0212, York Osteoarchaeology Ltd.

Cairns, A. M. (2015). *Health in Medieval and Early Modern Norway: A comparative analysis of the impact of social, economic, and environmental change on skeletal remains* [Unpublished Ph.D. thesis], Australian National University.

Chang, C., Chow, C., Tellier, L., Vattikuti, S., Purcell, S., & Lee, J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7.

Cohen, N., & Crane-Kramer, G. M. (2007). Editors' summation. In N. Cohen, G. M. Crane-Kramer (Eds.) *Ancient health: Skeletal indicators of agricultural and economic intensification*, (pp. 320–343), University Press of Florida.

Cox, S. L., Ruff, C. B., Maier, R. M., & Mathieson, I. (2019). Genetic contributions to variation in human stature in prehistoric Europe. *Proceedings of the National Academy of Sciences*, *116*(43), 21484–21492.

de Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J. V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., Baimukhanov, N., Zaibert, V., Merz, V., Biddanda, A., Merz, I., Loman, V., Evdokimov, V., Usmanova, E., Hemphill, B., Seguin-Orlando, A., … Willerslev, E. (2018). The first horse herders and the impact of early bronze age steppe expansions into Asia. *Science*, *360*(6396), eaar7711.

Dunwell, A. (2007). Cist burials and an iron age settlement at Dryburn bridge, Innerwick, East Lothian. *Scottish Archaeological Internet Report*, *24*. http://soas.is.ed.ac.uk/index.php/sair/article/view/1139

Fernandes, D. M., Strapagiel, D., Borówka, P., Marciniak, B., Żądzińska, E., Sirak, K., Siska, V., Grygiel, R., Carlsson, J., Manica, A., & Lorkiewicz, W. (2018). A genomic Neolithic time transect of hunter-farmer admixture in Central Poland. *Scientific Reports*, *8*(1), 14879.

Fokkens, H., Veselka, B., Bourgeois, Q., Olalde, I., & Reich, D. (2017). Excavations of late Neolithic arable, burial mounds and a number of well-preserved skeletons at Oostwoud-Tuithoorn: A reanalysis of old data. In H. Kamermans & C. Bakels (Eds.), *Excerpta Archaeologica Leidensia II* (pp. 97–150). Leiden University.

Frei, K. M., Bergerbrant, S., Sjögren, K.-G., Jørkov, M. L., Lynnerup, N., Harvig, L., Allentoft, M. E., Sikora, M., Price, T. D., Frei, R., & Kristiansen, K. (2019). Mapping human mobility during the third and second millennia BA in present-day Denmark. *PLoS One*, *14*(8), e0219850.

Fu, Q., Hajdinjak, M., Moldovan, O., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., & Viola, B. (2015). An early modern human from Romania with a recent neanderthal ancestor. *Nature*, *524*, 216–219.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., Nickel, B., Peltzer, A., Rohland, N., Slon, V., Talamo, S., Lazaridis, I., Lipson, M., Mathieson, I., Schiffels, S., … Reich, D. (2016). The genetic history of ice age Europe. *Nature*, *534*(7606), 200–205.

Furtwängler, A., Rohrlach, A. B., Lamnidis, T. C., Papac, L., Neumann, G. U., Siebke, I., Reiter, E., Steuri, N., Hald, J., Denaire, A., Schnitzler, B., Wahl, J., Ramstein, M., Schuenemann, V. J., Stockhammer, P. W.,

Hafner, A., Lösch, S., Haak, W., Schiffels, S., & Krause, J. (2020). Ancient genomes reveal social and genetic structure of late Neolithic Switzerland. *Nature Communications*, *11*(1), 1915.

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kovári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T. F. G., Hofreiter, M., Bradley, D. G., & Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of european prehistory. *Nature Communications*, *5*, 5257.

Ghosh, S. (2016). Protein quality in the first thousand days of life. *Food and Nutrition Bulletin*, *37*, S14–S21.

González-Fortes, G., Jones, E. R., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., Drak, L., Siska, V., Simalcsik, A., & Boroneanţ, A. (2017). Paleogenomic evidence for multi-generational mixing between Neolithic farmers and Mesolithic Hunter-gatherers in the lower Danube Basin. *Current Biology*, *27*(12), 1801–1810.

Grasgruber, P., Sebera, M., Hrazdfra, E., Cacek, J., & Kalina, T. (2016). Major correlates of male height: A study of 105 countries. *Economics & Human Biology*, *21*, 172–195.

Günther, T., Malmstrom, H., Svensson, E. M., Omrak, A., Sanchez-Quinto, F., Kilinc, G. M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., & Munters, A. R. (2018). Population genomics of mesolithic scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biology*, *16*(1), e2003703.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., & Fu, Q. (2015). Massive migration from the steppe was a source for indo-european languages in europe. *Nature*, *522*(7555), 207–211. https://www.ncbi.nlm.nih.gov/pubmed/25731166

Harpak, A., & Przeworski, M. (2020). The evolution of group differences in changing environments. *PLoS Biology 19*(1), e3001072.

Hijmans, R. J., & van Etten, J. (2012). *raster: Geographic analysis and modeling with raster data*. R package version 2.0-12. http://CRAN.R-project.org/package=raster

Hui, R., D'Atanasio, E., Cassidy, L. M., Scheib, C. L., & Kivisild, T. (2020). Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Scientific Reports*, *10*(1), 18542.

Huijzer, B., & Vandenberghe, J. (1998). Climatic reconstruction of the Weichselian Pleniglacial in northwestern and Central Europe. *Journal of Quaternary Science: Published for the Quaternary Research Association*, *13*(5), 391–417.

Jiang, L., Zheng, Z., Qi, T., Kemper, K. E., Wray, N. R., Visscher, P. M., & Yang, J. (2019). A resource-efficient tool for mixed model association analysis of large-scale data. *Nature Genetics*, *51*(12), 1749–1755.

Kitti, K. (2008). The physical anthropological characterization of the population connected to the Baden culture in Hungary. In M. Furholt, M. Szmyt, & A. Zastawny (Eds.), *The Baden complex and the outside world* (pp. 95–110). SAO/SPES 4.

Kjellström, A. (2005). *The Urban Farmer: Osteoarchaeological analysis of skeletons form Medieval Sigtuna interpreted in a socioeconomic perspective* [Unpublised Ph.D. thesis], Theses and Papers in Osteoarchaeology No. 2, Stockholm University

Kjellström, A., Storå, J., Possnert, G., & Linderholm, A. (2009). Dietary patterns and social structures in medieval Sigtuna, Sweden, as reflected in stable isotope values in human skeletal remains. *Journal of Archaeological Science*, *36*(12), 2689–2699.

Koepke, N. (2016). The biological standard of living in Europe from the late iron age to the little ice age. In *The Oxford handbook of economics and human biology* (pp. 70–108). Oxford University Press.

Köhler, K., Marcsik, A., Kondor, K., & Hajdu, T. (2017). The results of the physical anthropological study of the skeletal remains from the late copper age mass graves at the site of Abony, Turjanyos-dulo, Hungary. *Tagungen des Landesmuseums für Vorgeschichte Halle*, *16*, 377–386.

Kopke, N., & Baten, J. (2005). Biological standard of living in Europe during the last two millennia. *European Review of Economic History*, *9*(1), 61–95.

Krzewińska, M., Kjellström, A., Günther, T., Hedenstierna-Jonson, C., Zachrisson, T., Omrak, A., Yaka, R., Kılınç, G. M., Somel, M., Sobrado, V., & Evans, J. (2018). Genomic and strontium isotope variation reveal immigration patterns in a Viking age town. *Current Biology*, *28*(17), 2730–2738.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S., & Ferreira, T. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*(7317), 832–838. https://www.ncbi.nlm.nih.gov/pubmed/20881960

Lehuray, J. D., Schutkowski, H., & Richards, M. P. (2006). La Téne dietary variation in Central Europe: A stable isotope study of human skeletal remains from Bohemia. In C. Knüsel & B. Gowland (Eds.), *The social archaeology of funeral remains* (pp. 99–121). Oxbow Books.

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Medicine*, *12*(1), 44.

Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., & Wegmann, D. (2017). Atlas: Analysis tools for low-depth and ancient samples. *bioRxiv*. https://doi.org/10.1101/105346

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Harney, E., Nordenfelt, S., Llamas, B., Gusztáv Mende, B., Köhler, K., Oross, K., Bondár, M., … Reich, D. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature*, *551*(7680), 368–372.

Malmström, H., Günther, T., Svensson, E. M., Juras, A., Fraser, M., Munters, A. R., Pospieszny, L., Tõrv, M., Lindström, J., Götherström, A., Storå, J., & Jakobsson, M. (2019). The genomic ancestry of the Scandinavian Battle axe culture people and their relation to the broader corded ware horizon. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1912), 20191528.

Manouvrier, L. (1892). Determination de la taille d'apres les grands os des membres. *Rev Men de l'Ecole d'Anthrop*, *2*, 227–233.

Marciniak, S., & Perry, G. H. (2017). Harnessing ancient genomes to study the history of human adaptation. *Nature Reviews: Genetics*, *18*(11), 659–674.

Margaryan, A., Lawson, D. J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I., Cassidy, L. M., Jørsboe, E., Ingason, A., Pedersen, M. W., Korneliussen, T., Wilhelmson, H., Bus, M. M., de Barros Damgaard, P., Martiniano, R., Renaud, G., Bhérer, C., Moreno-Mayar, J. V., Fotakis, A. K., … Willerslev, E. (2020). Population genomics of the Viking world. *Nature*, *585*(7825), 390–396.

Marouli, E., Graff, M., Medina-Gomez, C., Lo, K. S., Wood, A. R., Kjaer, T. R., Fine, R. S., Lu, Y., Schurmann, C., Highland, H. M., Rüeger, S., Thorleifsson, G., Justice, A. E., Lamparter, D., Stirrups, K. E., Turcot, V., Young, K. L., Winkler, T. W., Esko, T., … Lettre, G. (2017). Rare and low-frequency coding variants alter human adult height. *Nature*, *542*(7640), 186–190.

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591.

Martiniano, R., Caffell, A., Holst, M., Hunter-Mann, K., Montgomery, J., Müldner, G., McLaughlin, R. L., Teasdale, M. D., van Rheenen, W., Veldink, J. H., van den Berg, L. H., Hardiman, O., Carroll, M., Roskams, S., Oxley, J., Morgan, C., Thomas, M. G., Barnes, I., McDonnell, C., … Bradley, D. G. (2016). Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nature Communications*, *7*(1), 10326.

Massy, K. (2018). *Die Gräber der Frühbronzezeit im südlichen Bayern - Untersuchungen zu den Bestattungs- und Beigabensitten sowie gräberfeldimmamenten Strukturen*. Kallmünz/Opf.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., & Fernandes, D. (2018). The genomic history of southeastern Europe. *Nature*, *555*(7695), 197–203.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M. B., Carbonell, E., … Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503.

Mittnik, A., Massy, K., Knipper, C., Wittenborn, F., Friedrich, R., Pfrengle, S., Burri, M., Carlichi-Witjes, N., Deeg, H., Furtwängler, A., Harbeck, M., von Heyking, K., Kociumaka, C., Kucukkalipci, I., Lindauer, S., Metz, S., Staskiewicz, A., Thiel, A., Wahl, J., … Krause, J. (2019). Kinship-based social inequality in bronze age Europe. *Science*, *366*(6466), 731–734.

Mittnik, A., Wang, C.-C., Pfrengle, S., Zarina, G., Hallgren, F., Allmäe, R., et al. (2018). The genetic prehistory of the Baltic Sea region. *Nature Communications*, *442*, 1–11.

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *Elife*, *9*. e48376. https://www.ncbi.nlm.nih.gov/pubmed/31999256

Müldner, G., Chenery, C., & Eckardt, H. (2011). The 'headless romans': Multi-isotope investigations of an unusual burial ground from Roman Britain. *Journal of Archaeological Science*, *38*(2), 280–290.

Münster, A., Knipper, C., Oelze, V. M., Nicklisch, N., Stecher, M., Schlenker, B., Ganslmeier, R., Fragata, M., Friederich, S., Dresely, V., & Hubensack, V. (2018). 4000years of human dietary evolution in Central Germany, from the first farmers to the first elites. *PLoS One*, *13*(3), e0194862. https://www.ncbi.nlm.nih.gov/pubmed/29584767

Narasimhan, V. M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., Kim, A. M., Olivieri, L. M., Coppa, A., Vidale, M., Mallory, J., Moiseyev, V., Kitov, E., Monge, J., Adamski, N., … Reich, D. (2019). The formation of human populations in south and Central Asia. *Science*, *365*(6457), eaat7487.

NCD Risk Factor Collaboration. (2016). A century of trends in adult human height. *eLife*, *5*, e13410.

Neale Lab, T. (2018). UK Biobank GWAS Results. http://www.nealelab.is/uk-biobank.

Niskanen, M., Ruff, C. B., Holt, B., Sládek, V., & Berner, M. (2018). Temporal and geographic variation in body size and shape of europeans from the late pleistocene to recent times. In C. B. Ruff (Ed.), *Skeletal variation and adaptation in Europeans: Upper Paleolithic to the twentieth century* (pp. 49–89). Wiley Online Library.

O'Brien, D. M. (2015). Stable isotope ratios as biomarkers of diet for Health Research. *Annual Review of Nutrition*, *35*, 565–594.

Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., Altena, E., Lipson, M., Lazaridis, I., Harper, T. K., Patterson, N., Broomandkhoshbacht, N., Diekmann, Y., Faltyskova, Z., Fernandes, D., … Reich, D. (2018). The beaker phenomenon and the genomic transformation of Northwest Europe. *Nature*, *555*(7695), 190–196.

Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C. J., Gandini, F., Pala, M., Soares, P., Ferrando-Bernal, M., Adamski, N., Broomandkhoshbacht, N., Cheronet, O., Culleton, B. J., Fernandes, D., Lawson, A. M., Mah, M., … Reich, D. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. *Science*, *363*(6432), 1230–1234.

Pardini, E. (1977). The human remains from Aligrama settlement (swat, Pakistan). *East and West*, *27*(1/4), 207–226.

Pearson, K. (1899). On the reconstruction of stature in prehistoric races. Mathematical contribution to the theory of evolution. *Transactions of the Royal Society A*, *192*, 169–244.

Pomeroy, E., Stock, J. T., & Wells, J. C. K. (2021). Population history and ecology, in addition to climate, influence human stature and body proportions. *Scientific Reports*, *11*(1), 274.

Price, T. D., Peets, J., Allmäe, R., Maldre, L., & Oras, E. (2016). Isotopic provenancing of the Salme ship burials in pre-Viking age Estonia. *Antiquity*, *90*(352), 1022–1037.

Racimo, F., Sikora, M., Vander Linden, M., Schroeder, H., & Lalueza-Fox, C. (2020). Beyond broad strokes: Sociocultural insights from the study of ancient genomes. *Nature Reviews Genetics*, *21*(6), 355–366.

Randall, J. C., Winkler, T. W., Kutalik, Z., Berndt, S. I., Jackson, A. U., Monda, K. L., Kilpeläinen TO, Esko, T., Mägi, R., Li, S., & Workalemahu, T. (2013). Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet*, *9*(6), e1003500. https://www.ncbi.nlm.nih.gov/pubmed/23754948

Rogers, A. R., & Mukherjee, A. (1992). Quantitative genetics of sexual dimorphism in human body size. *Evolution*, *46*(1), 226–234. https://www.ncbi.nlm.nih.gov/pubmed/28564974

Roseman, C. C., & Auerbach, B. M. (2015). Ecogeography, genetics, and the evolution of human body form. *Journal of Human Evolution*, *78*, 80–90.

Rosenstock, E., Ebert, J., Martin, R. D., Hicketier, A., Walter, P., & Groß, M. (2019). Human stature in the near east and Europe ca. 10,000-1,000 BC: Its spatiotemporal development in the Bayesian errors-in-variables mode. *Archaeological and Anthropological Sciences*, *11*, 5657–5690.

Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J., & Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Publishing Group*, *53*(1), 120–126.

Ruff, C., & Holt, B. (2018). The Balkans. In C. B. Ruff (Ed.), *Skeletal variation and adaptation in Europeans: Upper Paleolithic to the twentieth century*. John Wiley & Sons.

Ruff, C. B., Holt, B. M., Niskanen, M., Sladék, V., Berner, M., Garofalo, E., Garvin, H. M., Hora, M., Maijanen, H., Niinimäki, S., Salo, K., Schuplerová, E., & Tompkins, D. (2012). Stature and body mass estimation from skeletal remains in the European Holocene. *American Journal of Physical Anthropology*, *148*(4), 601–617.

Saag, L., Vasilyev, S. V., Varul, L., Kosorukova, N. V., Gerasimov, D. V., Oshibkina, S. V., Griffith SJ, Solnik A, Saag L, D'Atanasio E, Metspalu E, (2020). Genetic ancestry changes in stone to bronze age transition in the east European plain. bioRxiv.

Savell, K. R., Auerbach, B. M., & Roseman, C. C. (2016). Constraint, natural selection, and the evolution of human body form. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(34), 9492–9497.

Scheibner, A. (2016). *Prähistorische Ernährung in Vorderasien und Europa. Eine kulturgeschichtliche Synthese auf der Basis ausgewählter Quellen. Schriften zum Lebensstandard in der Vorgeschichte 1*. Leidorf: Rahden: Berliner Archäologische Forschungen.

Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., Tyler-Smith, C., Cooper, A., & Durbin, R. (2016). Iron age and Anglo-Saxon genomes from East England reveal British migration history. *Nature Communications*, *7*(1), 2041–1723.

Schroeder, H., Margaryan, A., Szmyt, M., Theulot, B., Wlodarczak, P., Rasmussen, S., Gopalakrishnan, S., Szczepanek, A., Konopka, T., Jensen, T. Z. T., Witkowska, B., Wilk, S., Przybyla, M. M., Pospieszny, L., Sjögren, K. G., Belka, Z., Olsen, J., Kristiansen, K., Willerslev, E., … Allentoft, M. E. (2019). Unraveling ancestry, kinship, and violence in a late Neolithic mass grave. *Proceedings of the National Academy of Sciences*, *116*(22), 10705–10710.

Sikora, M., Seguin-Orlando, A., Sousa, V. C., Albrechtsen, A., Korneliussen, T., Ko, A., Rasmussen, S., Dupanloup, I., Nigst, P. R., Bosch, M. D., Renaud, G., Allentoft, M. E., Margaryan, A., Vasilyev, S. V., Veselovskaya, E. V., Borutskaya, S. B., Deviese, T., Comeskey, D., Higham, T., … Willerslev, E. (2017). Ancient genomes show social and reproductive behavior of early upper Paleolithic foragers. *Science*, *358*(6363), 659–662.

Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of Biosocial Science*, *35*(2), 263–285. https://www.ncbi.nlm.nih.gov/pubmed/12664962

Sjøvold, T. (1990). Estimation of stature from long bones utilizing the line of organic correlation. *Human Evolution*, *5*(5), 431–447.

Skoglund, P., & Mathieson, I. (2018). Ancient genomics of modern humans: The first decade. *Annual Review of Genomics and Human Genetics*, *19*, 381–404.

Sládek, V., Macháček, J., Ruff, C. B., Schuplerova, E., Prichystalova, R., & Hora, M. (2015). Population-specific stature estimation from long bones in the early medieval Pohansko (Czech Republic). *American Journal of Physical Anthroplogy*, *158*(2), 312–324.

Sohail, M., Maier, R. M., Ganna, A., Bloemendal, A., Martin, A. R., Turchin, M. C., Chiang, C. W. K., Hirschhorn, J., Daly, M. J., Patterson, N., Neale, B., Mathieson, I., Reich, D., & Sunyaev, S. R. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*, *8*, e39702.

Steckel, R. H. (1995). Stature and the standard of living. *Journal of Economic Literature*, *33*(4), 1903–1940.

Stockhammer, P. W., Massy, K., Knipper, C., Friedrich, R., Kromer, B., Lindauer, S., Radosavljevic, J., Wittenborn, F., & Krause, J. (2015). Rewriting the central European early bronze age chronology: Evidence from large-scale radiocarbon dating. *PLoS One*, *10*(10), e0139705.

Szczepanek, A. (2013). Analiza antropologiczna szkieletów. In M. M. Przybyla, A. Szczepanek, & P. Wlodarczak (Eds.), *Koszyce, stanowisko 3 Przemoc i rytual u schylku neolitu* (pp. 65–92). Profil-Archaeo Publishing House and Archaeological Studio.

Tebelskis, P., & Jankauskas, R. (2002). The late Neolithic grave at Gyvakarai in Lithuania in the context of archaeological and anthropological knowledge. *Archaeologia Baltica*, *6*, 8–20.

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews. Genetics*, *19*(9), 581–590.

Trotter, M., & Gleser, G. C. (1952). Estimation of stature from long bones of American whites and negroes. *American Journal of Physical Anthropology*, *10*(4), 463–514.

Vilhjalmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindstrom, S., Ripke, S., Genovese, G., Loh, P. R., Bhatia, G., Do, R., & Hayeck, T. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *American Journal of Human Genetics*, *97*(4), 576–592.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *American Journal of Human Genetics*, *101*(1), 5–22.

Wallace, J. R. (2012). *Imap: Interactive Mapping*. R package version 1.32. https://cran.r-project.org/package=Imap

Waterman, A. J., Tykot, R. H., & Silva, A. M. (2016). Stable isotope analysis of diet-based social differentiation at late prehistoric collective burials in South-Western Portugal. *Archaeometry*, *58*(1), 131–151.

Weber, A. W., Schulting, R. J., Ramsey, C. B., & Bazaliiskii, V. I. (2016). Biogeochemical data from the Shamankaa II early Neolithic cemetery on southwest Baikal: Chronological and dietary patterns. *Quaternary International*, *405*, 233–254.

Weber, A. W., White, D., Bazaliiskii, V. I., Goriunova, O. I., Savel'ev, N. A., & Anne Katzenberg, M. (2011). Hunter–gatherer foraging ranges, migrations, and travel in the middle Holocene Baikal region of Siberia: Insights from carbon and nitrogen stable isotope signatures. *Journal of Anthropological Archaeology*, *30*(4), 523–548.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., Chu, A. Y., Estrada, K., Kutalik, Z., Amin, N., & Buchkovich, M. L. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173–1186.

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research*, *17*(10), 1520–1528.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common snps explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., Visscher, P. M., & the GIANT Consortium. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Human Molecular Genetics*, *27*(20), 3641–3649.

Zaidi, A. A., & Mathieson, I. (2020). Demographic history mediates the effect of stratification on polygenic scores. *eLife*, *9*, e61548.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Cox, S. L., Moots, H. M., Stock, J. T., Shbat, A., Bitarello, B. D., Nicklisch, N., Alt, K. W., Haak, W., Rosenstock, E., Ruff, C. B., & Mathieson, I. (2022). Predicting skeletal stature using ancient DNA. *American Journal of Biological Anthropology*, *177*(1), 162–174. https://doi.org/10.1002/ajpa.24426