

# Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y

Sandra Wilde<sup>a</sup>, Adrian Timpson<sup>b,c</sup>, Karola Kirsanow<sup>a</sup>, Elke Kaiser<sup>d</sup>, Manfred Kayser<sup>e</sup>, Martina Unterländer<sup>a</sup>, Nina Hoffelder<sup>a,1</sup>, Inna D. Potekhina<sup>f</sup>, Wolfram Schier<sup>d</sup>, Mark G. Thomas<sup>b,2</sup>, and Joachim Burger<sup>a</sup>

<sup>a</sup>Institute of Anthropology, Johannes Gutenberg University Mainz, 55128 Mainz, Germany; <sup>b</sup>Research Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, United Kingdom; <sup>c</sup>Institute of Archaeology, University College London, London WC1H 0PY, United Kingdom; <sup>d</sup>Institute of Prehistoric Archaeology, Freie Universität Berlin, 14195 Berlin, Germany; <sup>e</sup>Department of Forensic Molecular Biology, Erasmus University Medical Center Rotterdam, 3000 CA, Rotterdam, The Netherlands; and <sup>f</sup>Institute of Archaeology, Academy of Science of the Ukraine, 04210 Kiev-210, Ukraine

Edited by Nina G. Jablonski, The Pennsylvania State University, University Park, Pennsylvania, and accepted by the Editorial Board February 1, 2014 (received for review September 4, 2013)

**Pigmentation is a polygenic trait encompassing some of the most visible phenotypic variation observed in humans. Here we present direct estimates of selection acting on functional alleles in three key genes known to be involved in human pigmentation pathways—*HERC2*, *SLC45A2*, and *TYR*—using allele frequency estimates from Eneolithic, Bronze Age, and modern Eastern European samples and forward simulations. Neutrality was overwhelmingly rejected for all alleles studied, with point estimates of selection ranging from around 2–10% per generation. Our results provide direct evidence that strong selection favoring lighter skin, hair, and eye pigmentation has been operating in European populations over the last 5,000 y.**

ancient DNA | computer simulations | natural selection | Neolithic/Bronze Age | Eastern Europe

Genomic signatures of natural selection in humans are usually obtained from modern population genetic data and take the form of patterns of variation outside those expected under neutrality (1), including strong correlations between allele frequencies and hypothesized ecological drivers of selection (2) and identifying alleles with unusually recent age estimates for their frequencies (1). All such indirect approaches have poor sensitivity and temporal resolution, most are confounded by past demographic processes, and many are insensitive to selection acting on standing variation (3). With advances in ancient DNA analysis techniques it is possible to obtain direct estimates of natural selection over specific time periods by estimating allele frequency change, permitting changes in selection intensity to be detected through time and a more detailed understanding of the forces shaping human evolution. However, to date no such estimates have been made.

Pigmentation is a particularly conspicuous human phenotypic variation and in the past has been misleadingly used as a proxy for deep biogeographical origins (4). Dark pigmentation is thought to be the ancestral state in humans and to have been maintained by purifying selection in low-latitude, high-UVR regions to protect against folate photolysis, UV radiation (UVR)-induced DNA damage, and possibly damage to immunoglobulins (5, 6). Continental-scale correlations between skin pigmentation and incident UVR levels strongly indicate positive ecological adaptation (5), although sexual selection—particularly in relation to eye and hair coloration (7)—and relaxation of selective constraints (6) may also have been important.

Melanin, a derivative of tyrosine, is the primary biopolymer responsible for constitutive animal pigmentation and is found in two forms in humans, eumelanin (black-brown) and pheomelanin (red-yellow). It is synthesized in melanosomes, organelles located in melanocytes in several tissues including the basal layer of the epidermis, hair follicles, and the iris. Variation in pigmentation depends mainly on differences in the amount and type of melanin synthesized and the shape and distribution of melanosomes in different tissues (8). The products of several genes

are involved in melanin synthesis and distribution, and known DNA polymorphism in those genes explains a substantial proportion of human pigmentation variation (9–12). Signatures of recent natural selection have been detected in some pigmentation genes, including both shared and regionally specific alleles associated with lighter skin pigmentation in eastern and western Eurasia (13, 14), using modern population genetic data (1, 12, 13, 15, 16).

To obtain direct estimates of the strength of natural selection driving depigmentation we analyzed three polymorphic sites in ancient and modern samples (Table 1) that had previously been identified through genome-wide association studies (GWAS) (1, 17) and fine-mapping SNP association (18) as influencing pigmentation in modern Europeans: *HERC2* (rs12913832 A > G) (18), *SLC45A2* (rs16891982 C > G) (13), and *TYR* (rs1042602 C > A) (11). The product of the *TYR* gene, tyrosinase, catalyzes the first two steps of the melanogenesis pathway (8), and its absence generates an epistatic mask on downstream pigment-coding genes, halting melanin production. The *TYR* SNP rs1042602 is highly polymorphic in Europeans, and the derived A allele has been associated with light skin (19) and eye color (20) and the

## Significance

Eye, hair, and skin pigmentation are highly variable in humans, particularly in western Eurasian populations. This diversity may be explained by population history, the relaxation of selection pressures, or positive selection. To investigate whether positive natural selection is responsible for depigmentation within Europe, we estimated the strength of selection acting on three genes known to have significant effects on human pigmentation. In a direct approach, these estimates were made using ancient DNA from prehistoric Europeans and computer simulations. This allowed us to determine selection coefficients for a precisely bounded period in the deep past. Our results indicate that strong selection has been operating on pigmentation-related genes within western Eurasia for the past 5,000 y.

Author contributions: S.W., M.G.T., and J.B. designed research; S.W., A.T., M.U., N.H., and M.G.T. performed research; A.T. and M.G.T. contributed new reagents/analytic tools; E.K. and W.S. coordinated the acquisition of the archaeological sample material and provided background information; I.D.P. provided archaeological sample material and background information; M.G.T. and J.B. coordinated this study; S.W., A.T., N.H., and M.G.T. analyzed data; and S.W., A.T., K.K., E.K., M.K., M.U., N.H., I.D.P., W.S., M.G.T., and J.B. wrote the paper.

The authors declare no conflict of interest.

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

Freely available online through the PNAS open access option.

<sup>1</sup>Present address: Department of Ecology and Genetics, Evolutionary Biology, Uppsala University, 75236 Uppsala, Sweden.

<sup>2</sup>To whom correspondence should be addressed. E-mail: m.thomas@ucl.ac.uk.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1316513111/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1316513111/-DCSupplemental).

**Table 1. Allele frequencies of three functional SNPs associated with pigmentation in ancient and modern populations**

Gene	SNP	Polymorphism	Derived allele frequency				
			Europe	Asia	Africa	Modern Ukrainian sample	Ancient sample
<i>HERC2</i>	rs12913832 regulatory element ( <i>OCA2</i> )	A > G	0.710 [758]	0.002 [572]	0.000 [370]	0.651 (0.546–0.744) [86]	0.160 (0.099–0.247) [94]
<i>SLC45A2 (MATP)</i>	rs16891982 Leu374Phe	C > G	0.970 [758]	0.007 [572]	0.000 [370]	0.927 (0.849–0.965) [82]	0.432 (0.296–0.578) [44]
<i>TYR</i>	rs1042602 Ser192Tyr	C > A	0.368 [758]	0.002 [572]	0.000 [370]	0.367 (0.279–0.466) [98]	0.043 (0.018–0.106) [92]

Modern allele frequencies are from 1000 Genomes (<http://browser.1000genomes.org>) (65). Range in parentheses indicates the equal-tailed 95% confidence interval calculated as described using the *qbeta* function in R (66). Numbers in square brackets indicate  $2N$  individuals. The African American (ASW) data from 1000 Genomes were excluded because this population is admixed to an unknown extent.

absence of freckles (11). *SLC45A2 (MATP)* is involved in the distribution and intracellular processing of tyrosinase and other pigmentation enzymes (21). The derived rs16891982 G allele, which decreases in frequency along a north–south cline in Europe (13), is associated with lighter skin, hair, and eye pigmentation in modern populations (13, 22). The *HERC2* SNP rs12913832 A > G is the main determinant of iris pigmentation (brown/blue) (18, 23) and is also associated with skin and hair pigmentation and the propensity to tan (24). It is located within an intron of the Hect Domain and RCC1-like Domain2 (*HERC2*) gene 21 kb upstream from the *OCA2* promoter and serves as an enhancer for *OCA2* expression (23, 25). *OCA2* expression is increased in the melanocytes of carriers of the ancestral A allele, and attenuated in carriers of the derived G allele (25). *OCA2* encodes a melanosomal transmembrane protein, protein P, which is involved in the trafficking and processing of tyrosinase, the regulation of melanosomal pH, and glutathione metabolism (26). Selection favoring the *SLC45A2* rs16891982 G allele has been estimated to have begun between 11,000 and 19,000 y ago (14), well after the expansion of anatomically modern humans out-of-Africa. The age of the derived *TYR* rs1042602 allele has been estimated using two different methods: rho statistics place its emergence around 6,100 y ago, whereas the Bayesian coalescent approach using BEAST dates the allele to ~15,600 y ago (27).

However, these estimates were made by assuming that the selected allele arose at the time selection started and so do not accommodate the possibility of selection acting on standing variation (3). Furthermore, such estimates are themselves dependent on estimates of mutation and recombination rates. No estimates of the strength of selection or when it started to act on *HERC2/OCA2* and *TYR* are available at present.

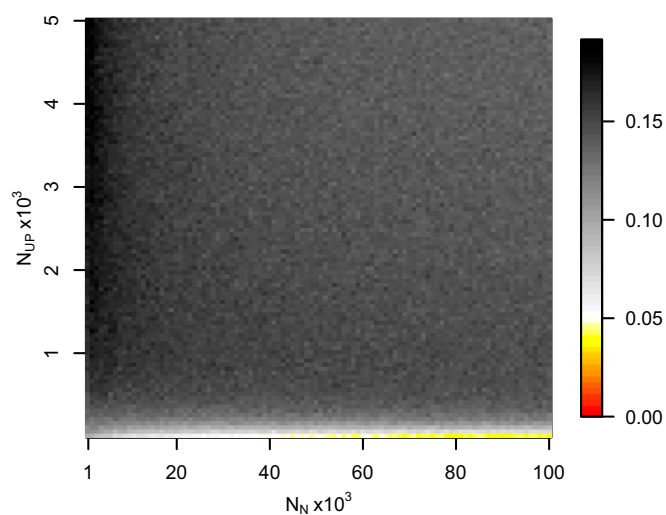
## Results

Ancient DNA was retrieved from 63 out of 150 Eneolithic (ca. 6,500–5,000 y ago) and Bronze Age (ca. 5,000–4,000 y ago) samples from the Pontic–Caspian steppe, mainly from modern-day Ukraine. We used multiplex-PCR enrichment and next-generation sequencing to genotype the three pigmentation-associated SNPs (rs12913832, rs16891982, and rs1042602) and mtDNA hypervariable region 1 (HVR1) sequences plus 32 mtDNA coding region SNPs and a 9-bp-indel from these individuals (Tables S1 and S2). Consensus HVR1 sequences were successfully assembled from 60 individuals. Pigmentation gene data were obtained from 48 samples. We also genotyped the three pigmentation-associated SNPs in a sample of 60 modern Ukrainians (28) and observed an increase in frequency of all derived alleles between the ancient and modern samples from the same geographic region (Table 1 and Fig. S1). This implies that the pigmentation of the prehistoric population is likely to have differed from that of modern humans living in the same area. Modern frequencies of the derived alleles within all of Europe and outside of Europe are provided for comparison (Table 1).

Inferring natural selection based on temporal differences in allele frequency requires the assumption of population continuity.

To this end we compared the 60 mtDNA HVR1 sequences obtained from our ancient sample to 246 homologous modern sequences (29–31) from the same geographic region and found low genetic differentiation ( $F_{ST} = 0.00551$ ;  $P = 0.0663$ ) (32). Coalescent simulations based on the mtDNA data, accommodating uncertainty in the ancient sample age, failed to reject population continuity under a wide range of assumed ancestral population size combinations (Fig. 1).

Conversely, continuity between early central European farmers and modern Europeans has been rejected in a previous study (33). However, the Eneolithic and Bronze Age sequences presented here are ~500–2,000 y younger than the early Neolithic and belong to lineages identified both in early farmers and late hunter–gatherers from central Europe (33). A plausible explanation for this is that the prehistoric populations sampled in this study are a product of admixture between in situ hunter–gatherers and immigrant early farmers during the centuries after the arrival of farming, and that this admixture was a major process shaping modern patterns of mtDNA variation (34) and possibly



**Fig. 1.** Probabilities of obtaining  $F_{ST}$  equal to or greater than that observed (0.00551) between 60 Eneolithic (ca. 6,500–5,000 y ago) and Bronze Age (ca. 5,000–4,000 y ago) samples from the Pontic–Caspian steppe, and a combined sample of 246 homologous modern sequences from the same geographic region, across a range of assumed ancestral population size combinations. Two phases of exponential growth were modeled, the first after the initial colonization of Europe 45,000 y ago, of assumed effective female population size  $N_{UP}$  (y axis), and ending when farming began in the region considered 7,000 y ago, when the assumed effective female population size was  $N_N$  (x axis), and the second leading up to the present, when the assumed effective female population size is 5,444,812. The initial colonizers of Europe were sampled from a constant ancestral African population of 5,000 effective females. Gray shaded areas indicate  $P$  values  $>0.05$ .

also the variability observed in European hair, eye, and skin color.

To test whether the observed increases in the three light pigmentation-associated alleles can be explained by genetic drift alone or whether natural selection needs to be invoked, we performed forward computer simulations of drift plus selection, accommodating uncertainty in ancient and modern allele frequency, population size, and ancient sample age. We assumed codominance for both *SLC45A2* rs16891982 G and *TYR* rs1042602 A alleles (22, 35) and that the derived *HERC2* rs12913832 G allele is recessive (36). Using these simulations, neutrality ( $S = 0$ ) was rejected under all assumed ancestral effective population sizes—ranging from  $10^3$  to  $10^5$  at the time of the ancient sample (*SLC45A2*  $P < 1 \times 10^{-5}$ , *TYR*  $P < 2 \times 10^{-5}$ , and *HERC2*  $P < 1 \times 10^{-5}$ ). The values of selection acting on the *SLC45A2* rs16891982 G allele, the *TYR* rs1042602 A allele, and the *HERC2* rs12913832 G allele that best explained the observed derived allele frequency changes were 0.030, 0.026, and 0.036, respectively (Fig. 2). Whereas there is strong evidence that the derived *HERC2* rs12913832 G allele is recessive (36), it is less clear whether the *SLC45A2* and *TYR* derived alleles are codominant, recessive, or dominant (22, 35). Under the assumption that both *SLC45A2* and *TYR* derived alleles are recessive, the selection values that best explain the observed changes in frequency are 0.022 and 0.104, respectively, and under the assumption that they are dominant the selection values are 0.088 and 0.016, respectively; again, neutrality was rejected for all three alleles ( $P < 4 \times 10^{-5}$ ) under all ancestral population sizes modeled and all assumptions of dominance/codominance/recessivity (Fig. S2).

## Discussion

Our analysis indicates that positive selection on pigmentation variants associated with depigmented hair, skin, and eyes was still ongoing after the time period represented by our archaeological population, 6,500–4,000 y ago. This finding suggests that either the selection pressures that initiated the selective sweep during the Late Pleistocene or early Holocene were still operative or that a new selective environment had arisen in which depigmentation was favored for a different reason.

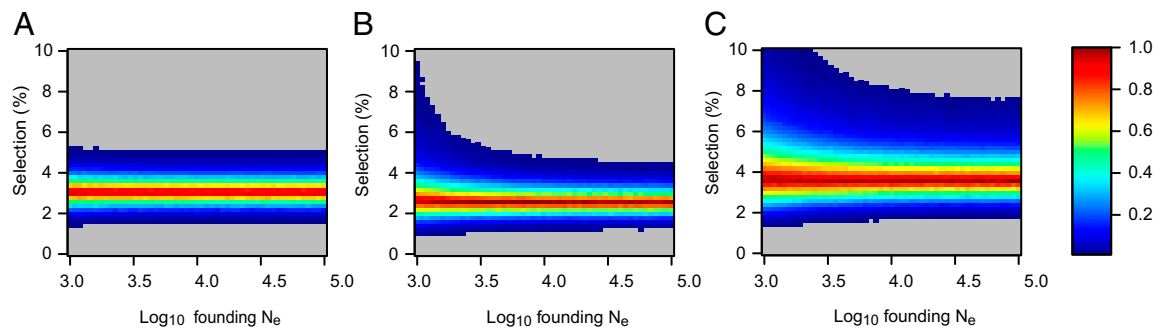
The high selection coefficients estimated for pigmentation genes *HERC2*, *SLC45A2*, and *TYR* are best understood in the context of estimates obtained for other recently selected loci. Using spatially explicit simulation and approximate Bayesian computation, selection on the *LCT* -13,910\*T allele—which is strongly associated with lactase persistence in Europeans and southern Asians—was inferred to fall in the range 0.0259–0.0795 and to have begun around 7,500 y ago in the region between the Balkans and central Europe (37). However, another simulation-

based study incorporating latitudinal effects on selection resulted in a lower estimate of  $S$  (0.008–0.018) (38). The selective advantage of the *G6PD* A- and Med deficiency alleles conferring resistance to malaria have been estimated at 0.019–0.048 and 0.014–0.049, respectively, in regions where malaria is endemic (39). These alleles are estimated to have arisen ~6,357 y ago (*G6PD* A-) and 3,330 y ago (*G6PD* Med) (39). Thus, the estimates of  $S$  for the three pigmentation genes examined in this study are comparable to those for the most strongly selected loci in the human genome.

Although these estimated selection coefficients are high, they are comparable to previous estimates for genes in the pigmentation complex. The selective sweeps favoring the *SLC45A2* derived allele, as well as the derived alleles of SNPs in *SLC24A5* and *TYRP1*, which are also implicated in the lightening of skin pigmentation, are estimated to have begun between 11,000 and 19,000 y ago, after the separation of the ancestors of modern Europeans and East Asians (the ages of the selective sweeps affecting *HERC2* and *TYR* have not yet been estimated) (14, 40). Beleza et al. (14) recently estimated the coefficient of selection at the *SLC45A2* locus to be 0.05 under a dominant model of inheritance and 0.04 under an additive model. Selection favoring the derived alleles of SNPs in *SLC24A5* and *TYRP1* was found to be similarly strong.

Estimating selection coefficients using the ancient DNA-based simulation approach presented here offers considerable advantages over traditional methods based on allele age and frequency estimates (1): Selection coefficients are estimated over a defined period; selection acting on standing variation can be accommodated; and our approach is insensitive to the frequently unaccounted for uncertainties associated with allele age estimation using molecular or recombination clocks. This latter advantage is likely to result in considerable improvements in precision. However, our approach does require the assumption of population continuity and will not provide direct estimates of when a selective sweep began.

Although the strength of the selection coefficients in a certain time window can be estimated with improved precision using our ancient DNA-based simulation approach, the actual nature of the selection pressure remains unknown. However, temporal and geographical information from the prehistoric skeletal population under study can help in formulating reasonable hypotheses. Geographic variation in many functional skin pigmentation gene polymorphisms (13), and lighter skin pigmentation more generally, correlate strongly with distance from the equator in long-established populations, suggesting that selective pressure also occurred along a latitudinal gradient. The samples in our study were from between 42°N and 54°N, a latitudinal belt in which



**Fig. 2.** Two-tailed empirical  $P$  values for obtaining the observed (A) *SLC45A2* G allele, (B) *TYR* A allele, and (C) *HERC2* G allele frequency increase.  $P$  values were obtained by forward simulation of drift and natural selection across a range of assumed ancestral population sizes and selection coefficients, assuming exponential growth to a modern  $N_e$  of 4,845,710. The *SLC45A2* rs16891982 G allele and the *TYR* rs1042602 A allele were assumed to be codominant. The *HERC2* rs12913832 G allele was assumed to be recessive (values less than 0.01 are shaded gray).

yearly average UVR is insufficient for vitamin D3 photosynthesis in highly melanized skin (4, 41). Constraints on the ability to photosynthesize vitamin D3 imposed by low incident UVR intensity may have provided significant selective pressure favoring lighter pigmentation populations in high-latitude regions such as the northern Pontic steppe belt. The need to admit UVB radiation to catalyze the synthesis of vitamin D3, together with the decreased danger of folate photolysis at higher latitudes, may account for the observed skin depigmentation from prehistoric to modern times in this region (5).

Dietary change during the Neolithization process may have reinforced selection pressure favoring depigmented skin. The individuals analyzed in this study lived ~500–2,000 y after the arrival of farming in the region north of the Black Sea (42, 43). In many parts of Europe, the Mesolithic–Neolithic transition is associated with a switch from a vitamin D-rich aquatic or game-based hunter–gatherer diet (44) to a vitamin D-poor agriculturalist diet. In low-UV regimes such as the one prevailing in our study region, it is difficult to meet vitamin D requirements without the consumption of significant quantities of oily fish or animal liver (45, 46). The vitamin D recommended dietary allowance of 800–1,000 IU for adults requires the daily consumption of the equivalent of 100 g of wild salmon (the dietary input with the greatest measured vitamin D concentration). Isotopic evidence suggests that the populations sampled in our study continued to access aquatic resources, primarily river fish, in the Neolithic, Eneolithic, and Bronze Age, although there was considerable heterogeneity in fish consumption within the study region (47–50). However, any diminution in fish consumption may have been sufficient to generate additional selective pressure favoring depigmentation at this low-incident-UVR latitude.

Although ecological and environmental factors may be sufficient to explain the observed change in European skin pigmentation, these explanations are unlikely to hold for eye and hair color. The geographic distribution of iris and hair pigmentation variation does not conform as well to a latitudinal cline model, with much of the observed phenotypic variation restricted to Europe and closely related neighboring populations (51, 52). The blue iris phenotype characteristic of the *HERC2* rs12913832 G allele, for example, is almost completely restricted to western Eurasia and some adjacent regions, its descendant populations, and populations containing European admixture (51, 52). It is possible that depigmented irises or the various human hair color morphs in Europeans are by-products of selection on skin pigmentation. There is evidence for gene–gene interaction within the polygenic system governing complex pigmentation traits; interactions between *HERC2*, *OCA2*, and *MC1R*, in particular, have been found to have a statistically significant effect on hair, iris, and skin color (36). There is also evidence for epistatic interactions between components of the melanin synthesis pathway in other mammalian model systems, including interactions between the products of *ASIP*, *MC1R*, and *TYR* (53). Additionally, many pigmentation genes, including *TYR*, *HERC2*, and *SLC45A2* have pleiotropic effects on skin, hair, and eye color (11, 36).

Given that intraspecific pigmentation variability in other taxa, particularly avians, has been attributed to signaling and other factors associated with mate choice (54) it is possible that depigmented irises and the various hair colors observed in Europeans arose through sexual selection (7). Frequency-dependent sexual selection in favor of rare variants has been observed in vertebrates (55, 56), and such selection favoring rare pigmentation morphs could have driven alleles associated with lighter hair and eye colors to higher frequency. Once lighter hair and eye pigmentation phenotypes reached appreciable frequencies in European populations, these novel traits may have continued to be preferred as indicators of group membership, facilitating assortative mating. Assortative mating based on coloration is common in vertebrates

(57), and skin pigmentation has been observed as a criterion for endogamy in modern human populations (58, 59). In addition, there is some evidence that lighter iris colors, because of their recessive mode of inheritance, may be preferred by males in assortative mating regimes to improve paternity confidence (60). Consistent with positive assortative mating, an exact test of Hardy–Weinberg equilibrium reveals an excess of *HERC2* rs12913832 homozygotes in both the modern ( $P = 0.0543$ ) and ancient ( $P = 0.0084$ ) East European samples genotyped here (Table S3), despite the relatively small sample sizes.

The observed excess of *HERC2* rs12913832 homozygotes in the ancient sample might be explained by population stratification in a temporally heterogeneous population sample. Although we do not observe any chronological or spatial patterning of the pigmentation markers in our prehistoric sample, we cannot exclude population stratification in the absence of additional neutral SNPs. However, we note that neither the *TYR* nor the *SLC45A2* SNPs investigated here, nor three additional SNPs investigated in the same ancient and modern samples, showed any significant observable excess of homozygotes (Table S3), suggesting that the excess of *HERC2* rs12913832 homozygotes is less likely to be due to population stratification.

In sum, a combination of selective pressures associated with living in northern latitudes, the adoption of an agriculturalist diet, and assortative mating may sufficiently explain the observed change from a darker phenotype during the Eneolithic/Early Bronze age to a generally lighter one in modern Eastern Europeans, although other selective factors cannot be discounted. The selection coefficients inferred directly from serially sampled data at these pigmentation loci range from 2 to 10% and are among the strongest signals of recent selection in humans.

## Methods

**Ancient DNA Extraction, Amplification, and Sequencing.** Sample information, detailed descriptions of DNA extraction, amplification, and sequencing methods as well as validation of the ancient DNA data are provided in [Supporting Information](#) owing to the extensive nature of the experimental setup.

Skeletal material from 150 Eneolithic and Bronze Age individuals from the west and north Pontic region were available for ancient DNA analyses. From all but one sample DNA was extracted twice independently using 0.5–1.0 g bone powder; 403 bp of the hypervariable region 1 [nucleotide positions (np) 16,011–16,413] were amplified using seven overlapping primer pairs (Table S4). They were integrated in a triple multiplex setup that included 32 clade-determining coding region SNPs and a 9-bp-indel (Table S5), as well as used in single-locus PCRs. rs12913832, rs16891982, and rs1042602 were amplified in a multiplex PCR together with 18 other nuclear loci (Table S4).

Sequencing of the PCR products was primarily done by 454 sequencing by GATC Biotech. Before that, samples were pooled according to a protocol modified after Meyer et al. (61). Raw data (454) were sorted by barcode and primer sequences of the multiplex PCRs and then analyzed with SeqMan ProTM (DNASTAR Lasergene 8, 9, and 10). For authentication purposes mitochondrial haplotypes are based on at least three, and SNP genotypes on at least four, independent amplification products from two extracts. In cases where the authentication scheme could not be fulfilled using the available 454 data, those loci were additionally amplified in single-locus PCRs, followed by direct sequencing after Sanger.

**Population Genetic Analyses.** Continuity between Neolithic and present-day populations in the geographic region encompassing Bulgaria, Romania, Ukraine, and the southwest of the Russian Federation was tested by calculating the  $F_{ST}$  between ancient and modern observed mtDNA HVR1 samples (29–32) and comparing this with  $F_{ST}$ s between ancient and modern DNA samples generated by coalescent simulation (33). These simulated DNA samples were generated using the program Fastsimcoal (62) under the null model of a single continuous population, using plausible population parameter ranges, and serial ancient and modern samples that replicated the observed sample numbers and dates. Details of the continuity test are provided in [Supporting Information](#).

To test whether changes in *HERC2* rs12913832 G, *TYR* rs1042602 A, or *SLC45A2* rs16891982 G allele frequencies (Fig. S1) can be explained by genetic drift, or whether natural selection needs to be invoked, and to

estimate the strength of natural selection where appropriate, we used a forward simulation approach. In each forward simulation we first drew the ancestral allele frequency estimate from a random Beta ( $n_p+1$ ,  $n_q+1$ ) distribution, where  $n_p$  and  $n_q$  were the number of ancestral and derived alleles observed in our ancient sample, respectively, to reflect uncertainty in ancestral allele frequencies. Then drift and natural selection were forward-simulated by binomial sampling across generations and using a standard selection equation (63), respectively. We assumed codominance for both *SLC45A2* rs16891982 G and *TYR* rs1042602 A alleles (22, 35) and that the derived *HERC2* rs12913832 G allele is recessive (36) (Fig. 2). However, although there is strong evidence that the derived *HERC2* rs12913832 G allele is recessive (36), it is less clear whether the *SLC45A2* and *TYR* derived alleles are codominant, recessive, or dominant (22, 35). For this reason we also performed the same simulations assuming both dominance and recessivity for the *SLC45A2* and *TYR* derived alleles (Fig. S2). Exponential population growth was modeled from a range of values of  $N_e$  at the time of the ancient sample (50 equally spaced  $\log_{10}$  values between 1,000 and 100,000) to a modern  $N_e$  of 4,845,710 (1/10 of the census population size of Ukraine in 2001, the year that the modern Ukrainian sample was collected; [http://en.wikipedia.org/wiki/Demographics\\_of\\_Ukraine](http://en.wikipedia.org/wiki/Demographics_of_Ukraine)). The number of generations forward-simulated was drawn from a pool of 600,000 date estimates for the ancient samples reported here, generated by pooling each set of 10,000 date estimates for all 60 ancient samples. In the final generation of each forward simulation, simulated modern sample allele frequencies were

picked from a random binomial with  $N$  equal to the modern sample size (*HERC2*  $n = 86$ , *SLC45A2*  $n = 82$ , and *TYR*  $n = 98$ ). Forward simulations were repeated 100,000 times for each combination of the 50 assumed  $N_e$  values at the time the ancient sample and 50 selection coefficient ( $S$ ) values, starting at zero. Finally, the simulated distribution of modern derived allele frequencies was compared with those observed using the equation  $1 - 2 \times |0.5 - P|$ , where  $P$  is the proportion of simulated modern allele frequencies that are greater than that observed. This yielded a two-tailed empirical  $P$  value for the observed allele frequency increase for each combination of the demographic and natural selection model parameters (64) (Fig. 2 and Fig. S2).

**ACKNOWLEDGMENTS.** We thank all colleagues who contributed their archaeological knowledge and samples to this project, notably D. Agre, St. Alexandrov, V. Bubulici, A. N. Gei, A. A. Khokhlov, I. N. Klyuchneva, A. Kozak, N. Neradenko, A. V. Nikolova, V. P. Petrenko, Yu. Ya. Rassamakin, V. A. Romashko, S. N. Sanzharov, E. N. Sava, N. N. Shishlina, and D. Ya. Teslenko. We also thank Tom Gilbert and colleagues for giving us a hands-on introduction to bar coding and 454 sequencing workflow, and providing us with protocols. Thanks to Benjamin Rieger for his "sort3" perl script. The radiocarbon dates were provided by the Research Laboratory for Archaeology and the History of Art (University of Oxford), and financially supported by the Excellence Cluster 264 Topoi, Berlin. The authors acknowledge the use of the UCL Legion High Performance Computing Facility and associated support services in the completion of this work. The project was funded by German Federal Ministry of Education and Research Grant 01UA0809A.

- Sabeti PC, et al.; International HapMap Consortium (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449(7164): 913–918.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185(4):1411–1423.
- Peter BM, Huerta-Sanchez E, Nielsen R (2012) Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genet* 8(10): e1003011.
- Jablonski NG, Chaplin G (2000) The evolution of human skin coloration. *J Hum Evol* 39(1):57–106.
- Jablonski NG, Chaplin G (2010) Colloquium paper: Human skin pigmentation as an adaptation to UV radiation. *Proc Natl Acad Sci USA* 107(Suppl 2):8962–8968.
- Harding RM, et al. (2000) Evidence for variable selective pressures at MC1R. *Am J Hum Genet* 66(4):1351–1361.
- Darwin C (1871) *The Descent of Man, and Selection in Relation to Sex* (John Murray, London).
- Sturm RA, Teasdale RD, Box NF (2001) Human pigmentation genes: Identification, structure and consequences of polymorphic variation. *Gene* 277(1-2):49–62.
- Rees JL, Harding RM (2012) Understanding the evolution of human pigmentation: Recent contributions from population genetics. *J Invest Dermatol* 132(3 Pt 2):846–853.
- Myles S, Somel M, Tang K, Kelso J, Stoneking M (2007) Identifying genes underlying skin pigmentation differences among human populations. *Hum Genet* 120(5):613–621.
- Sulem P, et al. (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* 39(12):1443–1452.
- Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M (2007) Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann Hum Genet* 71(Pt 3):354–369.
- Norton HL, et al. (2007) Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol Biol Evol* 24(3):710–722.
- Beleza S, et al. (2013) The timing of pigmentation lightening in Europeans. *Mol Biol Evol* 30(1):24–35.
- Alonso S, et al. (2008) Complex signatures of selection for the melanogenic loci *TYR*, *TYRP1* and *DCT* in humans. *BMC Evol Biol* 8:74.
- Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective sweeps. *Genome Res* 20(3):393–402.
- Sulem P, et al. (2008) Two newly identified genetic determinants of pigmentation in Europeans. *Nat Genet* 40(7):835–837.
- Eiberg H, et al. (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Hum Genet* 123(2):177–187.
- Shriver MD, et al. (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet* 112(4):387–399.
- Frudakis T, et al. (2003) Sequences associated with human iris pigmentation. *Genetics* 165(4):2071–2083.
- Costin GE, Valencia JC, Vieira WD, Lamoreux ML, Hearing VJ (2003) Tyrosinase processing and intracellular trafficking is disrupted in mouse primary melanocytes carrying the underwhite (*uw*) mutation. A model for oculocutaneous albinism (*OCA*) type 4. *J Cell Sci* 116(Pt 15):3203–3212.
- Cook AL, et al. (2009) Analysis of cultured human melanocytes based on polymorphisms within the *SLC45A2/MATP*, *SLC24A5/NCKX5*, and *OCA2/P* loci. *J Invest Dermatol* 129(2):392–405.
- Sturm RA, et al. (2008) A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *Am J Hum Genet* 82(2):424–431.
- Han J, et al. (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet* 4(5):e1000074.
- Visser M, Kayser M, Palstra R-J (2012) *HERC2* rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the *OCA2* promoter. *Genome Res* 22(3):446–455.
- Donnelly MP, et al. (2012) A global view of the *OCA2-HERC2* region and pigmentation. *Hum Genet* 131(5):683–696.
- Hudjashov G, Vilems R, Kivisild T (2013) Global patterns of diversity and selection in human tyrosinase gene. *PLoS ONE* 8(9):e74307.
- Nadkarni NA, Weale ME, von Schantz M, Thomas MG (2005) Evolution of a length polymorphism in the human *PER3* gene, a component of the circadian system. *J Biol Rhythms* 20(6):490–499.
- Malyarchuk BA, Derenko MV (2001) Mitochondrial DNA variability in Russians and Ukrainians: Implication to the origin of the Eastern Slavs. *Ann Hum Genet* 65(Pt 1): 63–78.
- Malyarchuk BA, et al. (2002) Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet* 66(Pt 4):261–283.
- Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: Mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* 60(Pt 1):35–49.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131(2):479–491.
- Bramanti B, et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326(5949):137–140.
- Bollongino R, et al. (2013) 2000 years of parallel societies in Stone Age Central Europe. *Science* 342(6157):479–481.
- Stokowski RP, et al. (2007) A genome-wide association study of skin pigmentation in a South Asian population. *Am J Hum Genet* 81(6):1119–1132.
- Branicki W, Brudnik U, Wojas-Pelc A (2009) Interactions between *HERC2*, *OCA2* and *MC1R* may influence human pigmentation phenotype. *Ann Hum Genet* 73(2): 160–170.
- Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG (2009) The origins of lactase persistence in Europe. *PLoS Comput Biol* 5(8):e1000491.
- Gerbault P, Moret C, Currat M, Sanchez-Mazas A (2009) Impact of selection and demography on the diffusion of lactase persistence. *PLoS ONE* 4(7):e6369.
- Tishkoff SA, et al. (2001) Haplotype diversity and linkage disequilibrium at human *G6PD*: Recent origin of alleles that confer malarial resistance. *Science* 293(5529): 455–462.
- Soejima M, Tachida H, Ishida T, Sano A, Koda Y (2006) Evidence for recent positive selection at the human *AIM1* locus in a European population. *Mol Biol Evol* 23(1):179–188.
- Jablonski NG, Chaplin G (2012) Human skin pigmentation, migration and disease susceptibility. *Philos Trans R Soc Lond B Biol Sci* 367(1590):785–792.
- Kotova N (2009) The Neolithization of Northern Black Sea area in the context of climate changes. *Documenta Praehistorica* 36:159–174.
- Kotova N, Makhortkykh S (2010) Human adaptation to past climate changes in the northern Pontic steppe. *Quat Int* 220(1-2):88–94.
- Richards MP, Schulting RJ, Hedges RE (2003) Archaeology: Sharp shift in diet at onset of Neolithic. *Nature* 425(6956):366.
- Lu Z, et al. (2007) An evaluation of the vitamin D3 content in fish: Is the vitamin D content adequate to satisfy the dietary requirement for vitamin D? *J Steroid Biochem Mol Biol* 103(3-5):642–644.
- Hollick MF (2004) Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. *Am J Clin Nutr* 80(6, Suppl): 1678S–1688S.
- Lillie M, Budd C, Potekhina I (2011) Stable isotope analysis of prehistoric populations from the cemeteries of the Middle and Lower Dnieper Basin, Ukraine. *J Archaeol Sci* 38(1):57–68.

48. Lillie MC, Richards M (2000) Stable isotope analysis and dental evidence of diet at the Mesolithic–Neolithic transition in Ukraine. *J Archaeol Sci* 27(10):965–972.
49. Shishlina NI, et al. (2009) Paleoecology, subsistence, and C-14 chronology of the Eurasian Caspian Steppe Bronze Age. *Radiocarbon* 51(2):481–499.
50. Hollund HI, Higham T, Belinskij A, Korenevskij S (2010) Investigation of palaeodiet in the North Caucasus (South Russia) Bronze Age using stable isotope analysis and AMS dating of human and animal bones. *J Archaeol Sci* 37(12):2971–2983.
51. Sturm RA, Frudakis TN (2004) Eye colour: Portals into pigmentation genes and ancestry. *Trends Genet* 20(8):327–332.
52. Walsh S, et al. (2013) The HirisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet* 7(1):98–115.
53. Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9(11):855–867.
54. Dale J (2006) Intraspecific variation in coloration. *Bird Coloration* 2:36–86.
55. Hughes KA, Houde AE, Price AC, Rodd FH (2013) Mating advantage for rare males in wild guppy populations. *Nature* 503(7474):108–110.
56. Farr JA (1977) Male rarity or novelty, female choice behavior, and sexual selection in the guppy, *Poecilia reticulata* Peters (Pisces: Poeciliidae). *Evolution* 31(1):162–168.
57. Hofreiter M, Schöneberg T (2010) The genetic and evolutionary basis of colour variation in vertebrates. *Cell Mol Life Sci* 67(15):2591–2603.
58. Banerjee S (1985) Assortative mating for colour in Indian populations. *J Biosoc Sci* 17(2):205–209.
59. Roberts DF, Kahlon DP (1972) Skin pigmentation and assortative mating in Sikhs. *J Biosoc Sci* 4(1):91–100.
60. Laeng B, Mathisen R, Johnsen J-A (2007) Why do blue-eyed men prefer women with the same eye color? *Behav Ecol Sociobiol* 61(3):371–384.
61. Meyer M, Stenzel U, Hofreiter M (2008) Parallel tagged sequencing on the 454 platform. *Nat Protoc* 3(2):267–278.
62. Excoffier L, Foll M (2011) fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27(9):1332–1334.
63. Maynard Smith J (1998) *Evolutionary Genetics* (Oxford Univ Press, New York), 2nd Ed.
64. Voight BF, et al. (2005) Interrogating multiple aspects of variation in a full re-sequencing data set to infer human population size changes. *Proc Natl Acad Sci USA* 102(51):18508–18513.
65. The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.
66. R Core Team (2012) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).