Supporting Information

Wilde et al. 10.1073/pnas.1316513111

SI Methods

Samples and the Archaeological Background. Ancient DNA analyses were performed on 150 Eneolithic and Bronze Age individuals from 34 sites in the North Pontic steppe. The area under study stretches from southern Bulgaria across the present-day Republic of Moldova and Ukraine in the west of the Eurasian steppe belt up to the middle course of the Volga and the river Manych in the east. With the exception of several early Eneolithic graves all other burials were discovered in barrows, so-called *kurgans*. Fifty-eight percent of the samples were discarded in the course of the work because they did not yield any DNA (57 samples), were not preserved well enough to be reproduced sufficiently (26 samples), or were contaminated (4 samples). Samples from 63 individuals in 59 graves at 23 archaeological sites could be used in the population genetic analyses (Table S1). All samples were dated typologically and assigned to various cultures. In some cases radiocarbon dates were available to verify the relative dates by falling into the widely recognized time frames of the respective cultures.

The analyzed burials can be archaeologically attributed to the Eneolithic and the Early Bronze Age (in terms of the chronological terminology used in Eastern Europe). Whereas the latter can be dated across the entire fifth millennium B.P., the Eneolithic complexes scatter over a much broader time span. The oldest seem to be the graves from the flat necropolis near Smyadovo (1, 2). They are from the second half of the seventh millennium B.P. The Eneolithic burials from other sites could all be dated to the sixth millennium B.P. but were attributed to various archaeological cultures.

The youngest Eneolithic samples consisted of two individuals from the tumulus 8 burials near the village Mayaki. They already indicate the transition to the Bronze Age. Neither was radiocarbon-dated; instead, they were dated typologically. The ceramic inventory of both can be attributed to the Usatovo culture, the monuments of which are limited to the northwestern Pontic region. Archaeozoological data from two settlements suggest that at least in this particular area specialized sheep and goat stock-breeding was practiced, whereas in older horizons of Eneolithic settlements in adjacent regions no clear pattern of domesticated species was observed (3). A high percentage of wild animal bones is documented for the majority of settlements of the sixth millennium B.P. in the steppe.

The transition to Early Bronze Age took place around 5,100–5,000 B.P. and is characterized by burials of the Yamnaya (or Pit Grave) culture. In this time the practice of burying deceased in barrows was used predominantly, if not exclusively. The few archaeozoological data available from settlements of the Yamnaya culture suggest a change to specialized breeding of cattle, which continued as the most effective subsistence strategy in the northern Pontic region and other parts of the western Eurasian steppe belt for more than 2,000 y. Because only a few Yamnaya settlements have been discovered, a mobile form of pastoralism has been assumed (e.g., ref. 4). Twenty-eight of the samples analyzed here can be attributed to the Yamnaya culture and are disseminated from southern Bulgaria in the west over the entire steppe region north of the Black Sea; 11 of the 12 radiocarbon dates fall within the range of 5,000 and 4,500 calibrated years B.P. They confirm the typological classification of the samples as belonging to the Yamnaya culture. One sample from Pestchanka II was older than 5,000 y. It has been suggested to be the earliest Yamnaya grave in a burial mound erected on top of Eneolithic graves.

Between 4,700 and 4,500 B.P. early catacomb graves appeared in southern Russia and eastern Ukraine during the younger phase of the Yamnaya culture. By the middle of the fifth millennium the Yamnaya culture abandoned pit grave constructions in favor of building catacombs. This change stretched across the entire steppe region between the Volga in the East and the Carpathian mountains and the lower Danube in the West. Our sample includes 25 skeletons from the Catacomb culture, 10 of which fall into the early phase between 4,700 and 4,500 B.P. The others were assigned to the late/developed Catacomb culture dated between 4,500 and 4,000 BP. Ten radiocarbon dates confirmed the typological classification of the samples. As already mentioned, changes in grave construction and burial rites can be observed, but the subsistence strategy of the Yamnaya culture seems to be continued by the Catacomb culture (3, 5).

Anticontamination Measures, Sample Preparation, and Ancient DNA Extraction. Ancient DNA analyses were carried out using the facilities of the Institute of Anthropology in Mainz, Germany. Strict anticontamination measures and precautions were taken as previously described (6, 7). Additionally to the pre- and post-PCR laboratories a clean room was set up in a third building. It was used exclusively for the bar coding (tagging) of multiplex PCR products in preparation for 454 sequencing. Between visiting the general post-PCR area and entering the clean room technicians showered and donned freshly laundered clothes. Workspace and equipment in this laboratory were cleaned each workday with bleach or DNA ExitusPlus IF (AppliChem).

Sample preparation and DNA extraction followed the procedures described in Bramanti et al. (6), except that 50-kDA Amicon Ultra-15 Centrifugal Filter Units (Millipore) were used to desalt and concentrate the DNA in the final step, following phenol–chloroform extraction. From all but two samples two independent extracts were obtained. With PES7 one extraction failed, and there was not enough sample material of OLE7 for a second extraction.

mtDNA Amplification. Four hundred three base pairs of the hypervariable region I (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. The multiplex reaction was set up in a final volume of 50–70 μL, containing 1× Multiplex Master Mix (Qiagen), 0.05 μM each primer (Biospring) (12–13 pairs per setup), 0.4 μg/μL BSA (Roche), 6–8 μL DNA extract, and UV-irradiated HPLC water (Acros Organics/Fisher Scientific). Amplification was performed in a Mastercycler (Eppendorf) under the following conditions: initial denaturation at 95 °C for 15 min, 39–41 cycles of 30 s at 95 °C, 90 s at 56 °C, and 90 s at 72 °C, followed by a final elongation at 72° for 10 min.

The setup for the 50-μL single-locus amplification consisted of 1× PCR Gold Buffer, 2.5 mM $MgCl_2$, 2.5 U AmpliTaq Gold (Applied Biosystems), 0.2 mM dNTP mix (Qiagen), 0.1–0.2 μM each primer (Biospring), 0.4 μg/μL BSA (Roche), 3–8 μL DNA extract, and UV-irradiated HPLC water (Acros Organics). The PCR cycle program started with an initial denaturation of 3–6 min at 94 °C, followed by 39–41 cycles of 35 s at 94 °C, 35 s at 56 °C, 35 s at 72 °C, and a final elongation at 72 °C for 10 min. Besides the seven primer pairs from the multiplex, well-preserved samples were alternatively amplified with just four overlapping primer systems

(Table S4). PCR conditions were the same except that the annealing temperature was 58 °C.

Nuclear DNA Amplification. The primer systems to genotype rs12913832, rs16891982, and rs1042602 were designed with the software Primer Select (DNASTAR Lasergene 8) (Table S4). They were part of a multiplex PCR setup amplifying up to 23 nuclear loci. PCR reactions were carried out in a final volume of 80–100 μL, containing 1.5× PCR Gold Buffer, 6.25 mM $MgCl_2$, 8 U AmpliTaq Gold (Applied Biosystems), 0.4 mM dNTP mix (Qiagen), 0.025 μM each primer (Biospring), 0.6 μg/μL BSA (Roche), 10–18 μL DNA extract, and UV-irradiated HPLC water (Acros Organics). The amplification was performed in a Mastercycler gradient (Eppendorf) under the following conditions: initial denaturation at 94 °C for 3–6 min, 45–50 cycles of 40 s at each 94 °C, 58 °C, and 72 °C, and final elongation at 72 °C for 10 min.

In case the authentication scheme (discussed below) could not be fulfilled by multiplex PCRs, individual loci were amplified via single-locus PCRs. Those were each set up in a total volume of 50 μL, with 1× PCR Gold Buffer, 2.5 mM $MgCl_2$, 2.5 U AmpliTaq Gold (Applied Biosystems), 0.2 mM dNTP mix (Qiagen), 0.2 μM each primer (Biospring), 0.4 μg/μL BSA (Roche), 3–5 μL DNA extract, and UV-irradiated HPLC water (Acros Organics). The PCR cycle program started with an initial denaturation of 3 min at 94 °C, followed by 48 cycles of 35 s at 94 °C, 35 s at 58 °C, 35 s at 72 °C, and a final elongation at 72 °C for 10 min.

The same nuclear loci were investigated in 60 anonymous modern Ukrainians (8). DNA extracts obtained from buccal mucosa were analyzed in the same multiplex PCR as the ancient samples. The PCRs were set up with 1.5× PCR Gold Buffer, 7.5 mM $MgCl_2$, 5 U AmpliTaq Gold (Applied Biosystems), 0.4 mM dNTP mix (Qiagen), 0.02 μM each primer (Biospring), 0.2 μg/μL BSA (Roche), 2 μL DNA extract, and HPLC water (Acros Organics) up to a final volume of 50 μL. PCR was performed by initial denaturation for 6 min at 94 °C, 33 cycles of 40 s at each 94 °C, 58 °C, and 72 °C, and a final elongation at 72 °C for 10 min. For all PCRs amplification success was determined by agarose gel electrophoresis.

The 454 Sequencing. MultiplexPCR products were tagged and pooled according to a protocol modified after Meyer et al. (9). The MinElute PCR Purification Kit (Qiagen) was used for the purification steps. The attempt to achieve equimolarity was abandoned because, particularly when working with ancient DNA, (*i*) despite optimization of the setup, it is highly unlikely that loci within a multiplex reaction are amplified equally well, and (*ii*) bacterial DNA and primer dimers can distort DNA quantification results.

Up to 60 mitochondrial and 60 nuclear PCR products were pooled using 60 8-nt barcodes with at least three pairwise nucleotide differences (9); 454 sequencing was carried out externally by GATC Biotech AG.

During the course of the experiments the two final steps of the original protocol, dephosphorylation and restriction digestion, were omitted. Their purpose is to exclude molecules from the 454 sequencing that carry no or only one barcode. For this measure to take effect the "end polishing" in the 454 library preparation has to be skipped and the workflow has to start directly with the adapter ligation. Because the workflow at GATC Biotech AG routinely includes "end polishing," completing the last two steps of original tagging protocol would have been obsolete and would have led to the loss of molecules that were tagged only one-sidedly.

Direct Sequencing. Single-locus PCR products were purified by enzymatic digestion with 2 U Exonuclease I (Fermentas/Thermo Scientific) and either 0.3 U Shrimp Alkaline Phosphatase (Fermentas) or 0.3 U FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific). Treatment consisted of incubation at 37 °C for 45 min (with Shrimp Alkaline Phosphatase) or 30 min (with FastAP), followed by inactivation at 85 °C for 15 min.

The BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) was used for sequencing. A 10-μL setup contained 0.75× BigDye Terminator 5× Sequencing Buffer (Applied Biosystems), 0.25× BigDye Terminator v1.1 Ready Reaction Mix (Applied Biosystems), 1 μM primer (Biospring), 1–4 μL PCR product (volume depending on band strength on agarose gel), and HPLC water (Acros Organics). Sequencing took 25 cycles of 30 s at 92 °C, 15 s at 56–58 °C, and 2.5 min at 60 °C in a Mastercycler gradient (Eppendorf).

DNA sequences were obtained by capillary electrophoresis on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems) using POP-6 polymer (Applied Biosystems).

Sequencing Analyses. The 454 raw data were first sorted by barcode and primer sequences of the multiplex PCRs using a perl script "sort3" written by B. Rieger. The program requires the 454 sequences file (*.fna), a barcode, and a primer file in *.txt format. Because 454 sequencing has high error rates in homopolymer regions the program optionally collapses those stretches to single nucleotides in all three files before searching for matches. The user determines how far into a 454 sequence the program will search for a tag or a primer. For each barcode and primer a separate output file is created that contains the sequences in *.fasta format, from which the barcode has been removed. Trimming of primers is optional.

Sorted and trimmed 454 data and sequences obtained by conventional Sanger sequencing were analyzed using SeqMan Pro (DNASTAR Lasergene 8, 9, and 10).

Hypervariable region 1 (HVR1) haplotypes were established using a two-step consensus building procedure. For each 454 run a consensus sequence of each amplicon was obtained by a two-thirds majority rule, requiring a minimum of three reads. Products with less than three reads were not considered. The same majority rule applied to base calling in the Sanger sequencing products. At least three such sequences were produced per primer pair from independent PCRs from two extracts. The final HVR1 consensus from np 16,011–16,413 [numbering according to the revised Cambridge Reference Sequence (rCRS) (10)] was built using these overlapping fragments and trimmed to np 16,011–16,400 for comparison with modern datasets from the literature. If there were discrepancies among the sequences, additional PCRs were performed. A nucleotide at a variable position was regarded as confirmed once it was present at a ratio of 3:1 or higher. If this could not be achieved, the sample was excluded from the analysis. As an additional confirmation, coding region SNPs from the multiplex PCRs were inspected for agreement with the inferred haplogroup assignment (Table S2).

Unlike the mitochondrial loci the autosomal SNPs are independent of each other because they are located on different chromosomes. That and diploidy make it harder to distinguish between authentic variation and that caused by postmortem damage, PCR or sequencing artifacts, carryover or other kinds of contamination, allelic dropout, and unspecific amplification, for example owing to sequence similarities in gene families. Therefore, the authentication scheme required the respective SNPs to be genotyped at least four times from independent PCR products from two extracts. For each 454 run genotypes were established by counting reads carrying the respective allele at the position of the SNP. A threshold was set that an allele had to be represented by at least 10 reads within a 454 run to be regarded as an authentic product of the respective PCR. An individual was deemed homozygous if all four independently obtained genotypes were consistent. If one of the four PCR products showed both alleles (above the threshold), whereas the inconsistent allele is presented

by fewer reads than the one present in the other PCRs, at least two additional PCRs were required to give homozygous results, so that a ratio of at least 11:1 in favor of one allele was achieved. Assuming that none of the alleles is amplified preferentially, the probability of allelic dropout occurring in five out of six PCRs is extremely low, ~0.3%. A sample was considered heterozygous if each allele is present at least twice and obtained at least once from each extract. If none of the above-mentioned criteria was fulfilled the sample was excluded from the analysis of that locus.

Validation of Ancient DNA Data. Blank controls were processed during milling (hydroxyl apatite), extraction, and PCR steps. After a multiplex PCR with amplicons as short as 72 bp, it is often not possible to distinguish between contamination or long secondary structures of unused primers. Therefore, 80% of the negative controls showing even a shadow of a band were sequenced. Contamination rate was established by dividing the number of loci giving sequences (after extrapolating for the 20% that were not sequenced) through the total number of sequenced loci. The overall mitochondrial and autosomal contamination rates including all primer systems used in the multiplex and single-locus PCR were 4.45% and 2.42% respectively. If only nuclear loci showing at least 10 reads are considered, then contamination rate drops to 0.38%. Narrowed down to the three pigmentation SNPs, contamination rate (≥10 reads) is 0.65%.

Within the mtDNA dataset contaminants in negative controls could be distinguished in almost all cases, carrying haplotypes that were not present in the ancient samples at all, or at least not in the samples that were processed in the same experimental setup. One extraction blank provided sequences of the A allele at the tyrosinase locus. All individuals that were processed in that extraction setup showed the C allele. Also, all samples amplified in the same PCR carried the C allele, except for RIL3, which showed the A allele. RIL3 was consistently heterozygous in three other PCRs, however. One PCR negative control contained tyrosinase C-allele sequences. Samples that gave the C allele in that PCR were consistently homozygous in a least three additional independent PCRs.

Consensus HVR1 sequences were successfully assembled from 60 of the 63 individuals (Table S1). With the exception of five nucleotide positions these were always based on at least three independent PCR products. The first base pair of the first HVR1 fragment, np 16,011, was only reproduced twice in eight and np 16,012 in five samples. In 25 individuals two or three base pairs from np 16,272–16,274 were also only covered twice. No variation was observed at those nucleotide positions.

From three samples a consensus haplotype could not be obtained. Therefore, they were excluded from analyses and simulations that were based on mitochondrial DNA: OVI3 was successfully genotyped at the *HERC2* and the tyrosinase locus. For methodological reasons most samples were genotyped before the mtDNA was investigated. Preservation of OVI3 was not good to begin with and deteriorated, most likely owing to repeated thawing and freezing. The sample showed a U5 type but because this is based on only one or two PCRs OVI3 was not included in the mitochondrial analysis. KAL2 showed a systematic contamination in the mitochondrial DNA. Three positions in the HVR1 (16224Y, 16258R, and 16311Y) were ambiguous, indicating an H and a K type. Whereas both types were present in the coding region (not presented here) as well, there the H type dominated strongly. The only genotype that could be recovered from this sample was from the *HERC2* locus. Because KAL2 was clearly homozygous and the contamination risk for nuclear DNA is lower than for mtDNA, the allele is in all likelihood authentic. TEM8 shows a U5 type. Whereas positions 16189C, 16256T, and 16270T were confirmed, the status of 16192, 16193.1, and 16234 could not be positively identified. 16192 is part of a C stretch and besides the transition there might also be an insertion. Whether this was caused by polymerase slippage or is an authentic polymorphism remains unclear.

Nuclear data were obtained from 48 of the 63 individuals (Table S1). For rs12913832 (*HERC2*) one sample, PES7, could not fulfill the above-mentioned authentication rules and a genotype could not be provided. Only one extract was available, and it was heterozygous in two PCRs and homozygous for the A allele in four other PCRs. At rs1042602 (*TYR*), two samples, KAL2 and VIN5, were excluded from further analysis. Only three inconsistent PCR products were available from KAL2, and none from VIN5. *SLC45A2* was prone to locus dropout in the multiplex PCR. Because of the rigorous authentication rule of a minimum of 10 reads per allele, more than two thirds of the sequencing results had to be rejected. By single-locus PCR in combination with Sanger sequencing, four consistent results from independent PCRs could be obtained for 22 individuals.

Modeling Methods

Population Continuity Test. To test for population continuity in the region encompassing Bulgaria, Romania, Ukraine, and the southwest of the Russian Federation from the Neolithic until the present we first calculated the molecular F_{ST} (11) between a combined sample of 246 modern mtDNA HVR1 sequences comprising 30 Bulgarians (12), 18 Ukrainians (13), 198 southwest Russians (14), and a combined sample of the 60 homologous sequences from prehistoric remains generated in this study. The prehistoric samples are from across the same geographic region and dated from *ca.* 7,500 y B.P. to circa 3,700 y B.P. (31 ^{14}C-dated samples with an average SD of 56 y and 29 typologically dated samples with an average date range of 479 y). The region of overlap for all ancient and modern HVR1 samples was 390 bp, from 16,011–16,400 on the rCRS (10).

The F_{ST} between the ancient and modern sequence samples ($F_{ST} = 0.00551$) was calculated using the amova function in the R (15) library ade4 (16) and a P value ($P = 0.0663$) was estimated using 100,000 permutations of the data. We then examined whether this observed F_{ST} was greater than expected under a model of population continuity (6) using coalescent simulation for a range of combinations of female effective population sizes at the Upper Paleolithic transition in Europe (N_{UP}) 45,000 y ago (17) and the start of the Neolithic in the region (N_N) 7,000 y ago and assumed a generation time of 25 y. We assumed a modern female effective population size of 5,444,812 (1/10 of the modern female population size estimated for the region from which the modern mtDNA sequences were sampled, by amalgamating data from censuses ranging from 2000 to 2004, sourced from the Google public data repository and Wikipedia), and a pre-Upper Paleolithic transition African female effective population size of 5,000 (6). Ten thousand coalescent simulations were performed for each of all 10,000 combinations of 100 equally spaced values of N_{UP}, ranging from 10 to 5,000, and 100 equally spaced values of N_N, ranging from 1,000–100,000, using Fastsimcoal (18). In each of the 100 million simulations the simulated ancient sample was generated by sampling 60 sequences, each of a different date. Those dates were themselves generated by randomly sampling one date from the date distribution of each of the 60 ancient samples. For the 31 ^{14}C-dated ancient samples this first required calibration using the R library Bchron (19) and the IntCal09 calibration curve (20) to generate 10,000 calendar date estimates for each [500,000 Markov chain Monte Carlo (MCMC) iterations, with every 45th retained; burn-in: 50,000]. Where multiple ^{14}C dates were available for a single sample only 10,000 were randomly selected from the pooled MCMC calibration estimates. Similarly, we generated 10,000 date estimates for each typologically dated sample by random sampling under a uniform distribution bounded by the date range. Dates were then divided by 25 (years per generation) and rounded. Simulated F_{ST} values were calculated using the Amova function in the R (15) library

ade4 (16) and the proportion of simulated F_{ST} values equal to or greater than the observed F_{ST} were plotted for each combination of N_{UP} and N_N. All combinations of N_{UP} and N_N yielded P values >0.05.

Rationale for Forward Simulations. To test whether changes in *HERC2* rs12913832 G, *TYR* rs1042602 A, or *SLC45A2* rs16891982 G allele frequencies between ancient and modern sampling periods (Fig. S1; for genotype frequencies, see Table S3) can be explained by genetic drift, or if natural selection needs to be invoked, and to estimate the strength of natural selection where appropriate, we used a forward simulation approach. To reflect allele frequency uncertainty in the ancient sample we use the Beta distribution (with parameters $n_p + 1$ and $n_q + 1$, where n_p and n_q were the number of ancestral and derived alleles observed in our ancient sample, respectively), which is the conjugate prior for the binomial distribution (e.g., a sample of alleles for a biallelic polymorphism). This allows us to generate random ancestral allele frequency draws from the possible frequency range, given the observed allele counts in the ancient sample, and an uninformative (flat) prior. Put another way, the Beta distribution estimates the probability distribution of true allele frequencies, given the observed allele counts in the ancient sample, and so reflects allele frequency uncertainty in the ancient sample. Following this we simulate the trajectory of allele frequencies through time considering genetic drift and natural selection. In each generation genetic drift is simulated by binomial sampling across generations, where the sample size is assumed to be the modeled population size in that generation, and the outcome probability is taken from the allele frequency in the previous generation. In addition, in each generation we apply a standard selection equation to model the effects of different selection coefficients on allele frequency change. To accommodate allele frequency estimate uncertainty in the modern sample, 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* $n = 98$). To measure the goodness of fit of simulated and observed allele frequencies in the modern sample we used the Eq. $1 - 2 \times |0.5 - P|$, 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 change for each combination of the demographic and natural selection parameters considered (21) (Fig. 2 and Fig. S2).

Exponential population growth was modeled from a range of values of N_e at the time 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). The number of generations forward-simulated was drawn at random from a pool of 600,000 date estimates for the ancient samples, as reported above and in the main text, generated by pooling each set of 10,000 date estimates for all 60 ancient samples.

1. Chokhadzhiev S, Venelinova S (2007) Arkheologicheskо prouchvane na praistoricheski nekropol v m. Gorlomova Koriya krai Smyadovo. *Arkheologicheski otkritiya i razkopki prez* 2007:100–101.
2. Chokhadzhiev S, Venelinova S (2006) Arkheologicheski prouchvaniya na khalkolitniya nekropol v m. Gorlomova Koriya krai grad Smyadovo. *Arkheologicheski otkritiya i razkopki prez* 2006:65–70.
3. Kaiser E (2010) Der Übergang zur Rinderzucht im nördlichen Schwarzmeerraum. *Godišnjak Centar za balkanološka ispitivanja* 39:23–34.
4. Merpert NY (1974) *Drevneishie Skotovody Volzhsko-Uralskogo Mezhdurechya* (Nauka, Moscow).
5. Rassamakin Y (1999) The Eneolithic of the Black Sea Steppe: Dynamics of cultural and economic development 4500–2300 BC. *Late Prehistoric Exploitation of the Eurasian Steppe* (McDonald Institute for Archaeological Research, Cambridge, UK), pp 59–182.
6. Bramanti B, et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326(5949):137–140.
7. Haak W, et al. (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310(5750):1016–1018.
8. 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.
9. Meyer M, Stenzel U, Hofreiter M (2008) Parallel tagged sequencing on the 454 platform. *Nat Protoc* 3(2):267–278.
10. Andrews RM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.
11. 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.
12. 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.
13. 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.
14. Malyarchuk BA, et al. (2002) Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet* 66(Pt 4):261–283.
15. R Core Team (2012) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
16. Dray S, Dufour AB (2007) The ade4 package: Implementing the duality diagram for ecologists. *J Stat Softw* 22(4):1–20.
17. Powell A, Shennan S, Thomas MG (2009) Late Pleistocene demography and the appearance of modern human behavior. *Science* 324(5932):1298–1301.
18. Excoffier L, Foll M (2011) fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27(9):1332–1334.
19. Haslett J, Parnell A (2008) A simple monotone process with application to radiocarbon-dated depth chronologies. *J. R. Stat. Soc. Series C. Appl Stat* 57(4): 399–418.
20. Heaton TJ, Blackwell PG, Buck CE (2009) A Bayesian approach to the estimation of radiocarbon calibration curves: The IntCal09 methodology. *Radiocarbon* 51(4): 1151–1164.
21. Voight BF, et al. (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci USA* 102(51):18508–18513.

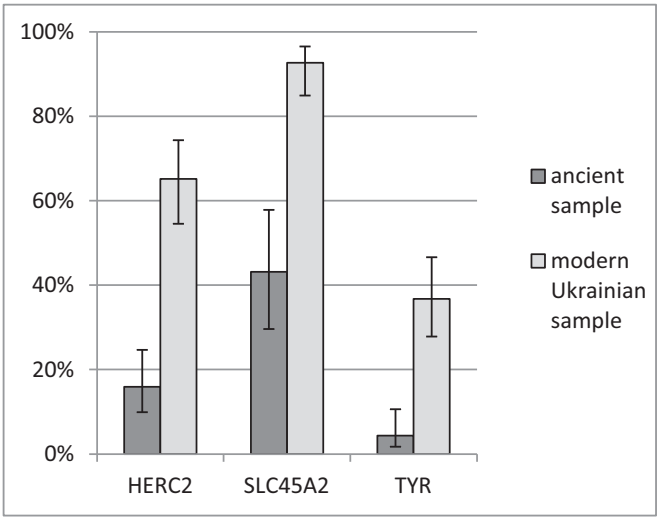

Fig. S1. Frequencies of the derived alleles in the ancient and modern dataset. Dark gray bars (*Left*) represent derived allele frequencies in the combined ancient dataset from the Eneolithic to Middle Bronze Age. Light gray bars (*Right*) show frequencies in modern Ukrainians. Error bars span the equal-tailed 95% confidence interval calculated as described using the qbeta function in R (15).

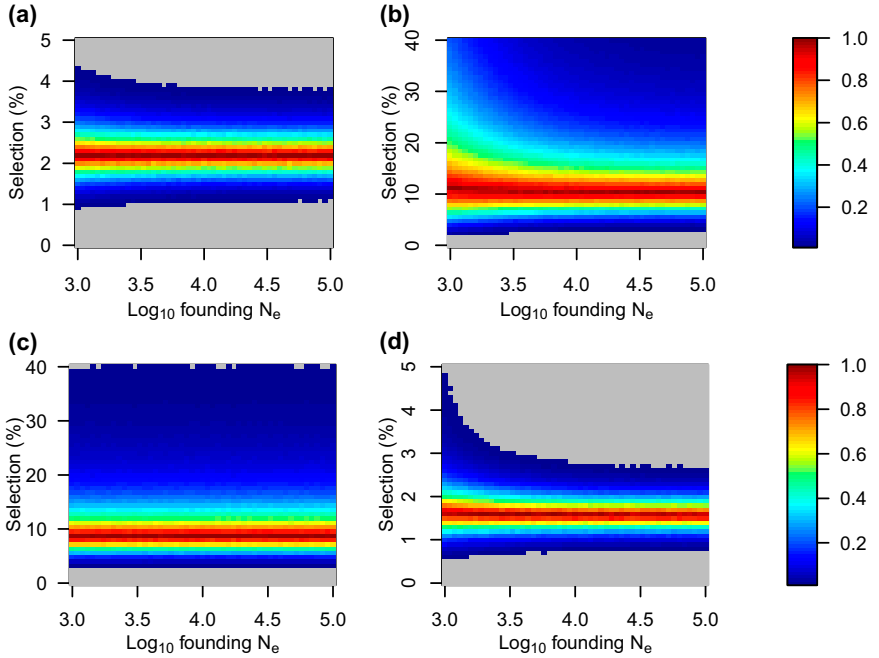

Fig. S2. Two-tailed empirical *P* values for obtaining the observed allele frequency increase. (*A*) *SLC45A2* rs16891982 G alleles and (*B*) *TYR* rs1042602 A alleles, both assuming recessivity. (*C*) *SLC45A2* rs16891982 G allele and (*D*) *TYR* rs1042602 A alleles, both assuming dominance (values less than 0.01 are shaded gray).

Table S1. Results of the genetic analysis of the ancient samples

Archaeological site	Latitude	Longitude	Kurgan	Grave	Archaeological date	Laboratory code	HVR1 haplotype (np 16,011–16,413)	rs1291 3832	rs1689 1982	rs1042 602
Benkovski	42.41	25.90	2	2	EBA-Yam	BEN3	188T			
Durankulak	43.69	28.53	Flat grave	15	Late ENL	DUR1	114A 192T 256T 270T 294T			
Kalinovka I	53.83	51.29	1	4	EBA-Yam	KAL1	147G 172C 223T 248T 344T 355T	A/G	C/C	A/A
			1	6	EBA-Yam	KAL2	n/a	G/G		
Krasnorechensk	49.21	38.21	13	1	EBA-Cat1	KNO4	356C	A/G		C/C
Lisichansk	48.91	38.44	1	2	EBA-Cat1	LIS1	256T 270T 399G	A/A		C/C
			2	1	EBA-Cat1	LIS2	356C	A/A		C/C
			3	2	EBA-Cat1	LIS3	354T	A/G	C/C	C/C
Mayaki	46.41	30.28	1	7	EBA-Yam	MAJ3	256T 270T 399G	A/A		C/C
			1	13	EBA-Yam	MAJ4	192T 311C	A/A		C/C
			1	15	EBA-Yam	MAJ5	136C 189C 223T 278T	G/G	C/C	C/C
			8	2	Late ENL	MAJ8	126C 294T 296T 304C			
			8	4	Late ENL	MAJ9	223T 292T	A/A	C/C	C/C
Molyukhov Bugor	49.13	32.46	Flat grave	3	Late ENL	MOB1	192T 256T 270T 399G	A/A		C/C
			Flat grave	4.2	Late ENL	MOB3	192T 256T 270T 399G			
Nevskoe	49.17	37.98	2	10.1	EBA-Cat2	NEV1	192T 256T 270T 399G	A/A	C/C	C/C
			5	9	EBA-Cat2	NEV3	362C	A/A		C/C
Nikolaevka III	53.09	50.31	1	2	EBA-Yam	NIK1	126C 163G 186T 189C 294T	A/A		C/C
			5	1	EBA-Yam	NIK7	rCRS	A/A	C/G	C/C
Novozvanovka II	48.58	38.36	1	2	EBA-Cat2	NOZ1	356C	G/G	C/G	C/C
			1	3	EBA-Cat2	NOZ2	356C	G/G		C/C
Olennii	45.60	38.53	1	9	EBA-Yam	OLE1	126C 292T 294T	A/A		C/C
			3	14a	EBA-Yam	OLE7	048A 069T 126C 193T			
Ovchartsi	42.20	26.07	The Big Tumulus	10	EBA-Yam	OVI2	224C 290T 311C			
			2	1	EBA-Yam	OVI3	n/a	A/G		C/C
Peschanyi	46.56	43.68	1	3	EBA-Yam	PEJ1	192T 256T 270T 399G	A/A		C/C
			1	4	EBA-Cat2	PEJ2	234T			
			2	3	EBA-Cat2	PEJ3	362C	A/A		C/C
Peschanyi V			2	3	EBA-Cat2	PEJ4	362C	A/A		C/C
			2	3	EBA-Cat2	PEJ5	356C	A/A	C/C	C/C
Pestchanka II	48.60	35.31	1	15	EBA-Yam	PES7	042A 209C		G/G	C/C
Podlesnyj	53.14	50.67	3	3	EBA-Yam	POD1	192T 223T 292T 325C	A/G	G/G	C/C
			3	5	EBA-Yam	POD2	126C 292T 294T	A/A	G/G	C/C
Popovo	42.20	26.73	Golyamata Mogila	29	EBA-Yam	POP1	126C 294T 296T 324C	A/G	C/G	C/C
			Golyamata Mogila	36	EBA-Yam	POP3	051G 129C 362C	A/A	G/G	C/C
			Golyamata Mogila	37	EBA-Yam	POP4	256T 270T 399G	A/A		C/C
Riltsi	43.59	27.78	264	5	EBA-Yam	RIL3	192T 224C 311C	A/A	G/G	C/A
Shakhta Stepnaya	48.38	36.37	3	3	EBA-Cat1	SAC2	069T 126C 193T	A/A		C/C
Smyadovo	43.05	27.00	Flat grave	10	Early ENL	SMY3	rCRS			
			Flat grave	11b	Early ENL	SMY4	304C			
			Flat grave	22	Early ENL	SMY9	261T	A/A		C/C
			Flat grave	24	Early ENL	SMY11	126C 153A 269G 294T 296T			
Kirovograd Sugokleya	48.52	32.25	Sugokleya	8	EBA-Yam	SUG2	129A 172C 223T 311C 391A	A/A	C/C	C/C
			Sugokleya	13	EBA-Cat2	SUG5	362C	A/A		C/C
			Sugokleya	16	EBA-Yam	SUG6	362C	A/A	C/C	C/C
			Sugokleya	20	EBA-Yam	SUG7	rCRS			
			Sugokleya	24	EBA-Yam	SUG8	rCRS			
Temrta III	46.55	43.66	1	2	EBA-Cat2	TEM1	356C	A/A		C/C
			1	4	EBA-Cat1	TEM2	rCRS	A/A	G/G	C/C
			1	6	EBA-Cat2	TEM3	069T 126C 145A 172C 222T 261T			
Temrta V	46.55	43.66	1	2.1	EBA-Cat1	TEM4	256T 270T 399G	A/A	C/G	C/C
			1	2.2	EBA-Cat1	TEM5	311C	A/A		C/C
			1	3.1	EBA-Cat1	TEM6	311C	A/A		C/C
			1	3.2	EBA-Cat1	TEM7	356C	A/A		C/C
			1	4.1	EBA-Cat2	TEM8	n/a	A/G		C/C

Table S1. Cont.

Archaeological site	Latitude	Longitude	Kurgan	Grave	Archaeological date	Laboratory code	HVR1 haplotype (np 16,011–16,413)	rs1291 3832	rs1689 1982	rs1042 602
Tetcani	48.18	26.98	3	22	EBA-Cat2	TET1	223T 391A			
			5	5	EBA-Yam	TET2	134T 356C			
Vinogradnoe	47.20	35.56	3	15	Early ENL	VIN1	rCRS	A/A	C/G	C/C
			3	25	EBA-Yam	VIN2	126C 163G 186T 189C 294T	A/A	C/G	C/C
			3	36.2	EBA-Cat2	VIN3	256T 270T 399G	A/A	C/G	C/C
			24	8.2	EBA-Yam	VIN5	126C 163G 186T 189C 294T	A/A		
			24	17	EBA-Cat2	VIN8	069T 126C 145A 172C 222T 261T	A/A	C/C	C/A
			24	31	EBA-Yam	VIN12	126C 163G 294T	A/A		C/C

Cat1, early Catacomb culture (*ca*. 4,700–4,500 y ago); Cat2, developed Catacomb culture (*ca*. 4,500–4,000 y ago); Early ENL, Early Eneolithic (*ca*. 6,500–5,000 y ago); EBA, Early Bronze Age; Late ENL, Late Eneolithic (*ca*. 5,500–4,000 y ago); Yam, Yamnaya/Pit Grave culture (*ca*. 5,000–4,500 y ago). Haplotypes are presented according to the rCRS positions (1) (minus 16,000). SNPs of the coding region are presented in Table S2.

1. Andrews RM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.

Table S2. Mitochondrial coding region SNPs

Primer system	M01	M01	M01	M02	M02	M03	M04	M04	M05	M06	M07	M08	M08	M09	M10	M11	M11	M11	M12	M13	M14	M14	M14	M14	M14	M15	M15	M16	M16	M17	M17	M18	M19	M19	M20	M20	M21	M21	M22	M22	M23	M24	M24	M25	M26	M26	M26	M27	M27	M27	M28	M28	M29	M29	M30	M30	
SNP position	456	462	481	663	669	3010	4194	4216	4529	4580	4833	4917	4937	5178	5843	6366	6371	6392	6776	7028	8251	8260	8269	8272–80 Del	8388	8392	8935	8994	9090	9117	10034	10115	10118	10398	10400	10873	10876	11719	11732	11947	12308	12346	12705	13263	13269	13626	13637	13680	13708	13734	14766	14769	14793	14798		HG*
rCRS	C	C	C	A	T	G	C	T	A	G	A	A	T	C	A	G	C	T	T	C	G	T	G	-	T	G	C	G	T	T	T	T	A	C	T	A	G	T	A	A	C	C	A	A	C	A	C	G	T	C	A	A	T		
BEN3	-	-	-	-	-	-	-	-	-	—	-	-	C	-	-	—	—	—	-	-	-	-	-	-	-	-	T	-	-	-		-	-	—	—	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	H
DUR1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	-	T	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	G	-	U
KAL1	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-		-	-	G	-	-	-	A	-	-	-	-	T	-	-	-	-	-	-	-	T	-	—	—	N1a
KAL2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
KNO4	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	T	-	-	-	-	-	—	—	—	-	-		-	-	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	—	—	U
LIS1	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	T	-	-	-	-	-	—	—	-	-	-		—	—	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	G	-	U
LIS2	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	—	T	-	-	-	-	-	—	—	—	-	-		—	—	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	A	-	T	-	—	—	U
LIS3	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	-	-	-	-	-	-	—	—	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
MAJ3				-	-		—	—	-		-	-	-	—	-	-	-	-		T																		A	-	-			-	-	-	—	—	—	-	-	T	—				?
MAJ4	—	—	—	-	-	—	Y	-	-	—	-	—	—	—	-	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—		—	—	-	-	-	-	A	-	-	G	—	-	-	-	—	G	—	-	-	T	—	—	—	U
MAJ5	—	—	—	-	-	-	-	-	-	—	-	—	—	—	-	-	T	—	—	T	—	—	—	—	—	—	—	—	—	—		—	—	-	-	-	-	A	-	-	-	-	T	-	-	-	-	-	-	-	T	-	—	—	X
MAJ8	-	-	-	-	-	-	-	C	-	-	-	G	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	A	-	-	-	-	-	—	-	-	-	-	-	-	T	-	-	-	T
MAJ9	—	—	—	-	-	-	—	—	-	—	-	-	-	-	-	—	—	—	—	T	A	-	-	-	-	—	—	—	—	-		—	—	-	-	-	-	A	-	G	-	-	Y	-	-	-	-	-	-	-	T	-	—	—	W
MOB1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	G	-	U
MOB3	—	—	—	-	-	-	-	-	-		-	-	-	-	-	—	—	—	-	T	R	-	-	-	-	-	-	—	—	-		-	-	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	—			U/K
NEV1	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	-	T	-	-	-	-	-	—	—	—	-	—		—	—	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	G	—	U
NEV3	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	—	-	-	-	-	-	-	—	—	—	—	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
NIK1	-	-	-	-	-	-	-	C	-	-	-	G	-	-	-	—	—	—	-	T	-	-	-	-	-	-	-	-	-	—		-	-	-	-	-	-	A	-	-	-	T	-	-	-	-	-	-	-	-	T	-	-	-	T
NIK7	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	-	-	-	-	-	-	—	—	—	—	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	H
NOZ1	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	—	—	—	—	-		-	-	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	—	—	U
NOZ2	—	—	—	-	-	—	-	-	-	—	-	-	-	-	-	-	-	-	—	T	-	-	-	-	-	—	—	—	—	-		—	—	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	—	—	U
OLE1	—	—	—	-	-	-	-	C	-	-	-	G	-	-	-	-	-	-	—	T	-	-	-	-	-	-	-	—	—	-		—	—	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	T	-	—	—	T
OLE7	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	—	—	—	-	T	-	-	-	-	-	-	-	-	-	-		-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	A	-	T	G	—	—	J
OVI2	—	—	—	-	-	—	-	-	-	—	-	-	-	-	-	-	-	-	—	T	-	-	-	-				—	—	-		-	-	G	-	-	-	A	-	-	G	—	-	-	-	-	-	-	-	-	T	-	—	C	K
OVI3				—	—	—	-	-	-	—	-			—	—	—	—	—	—		-	-	-	-				—	—	—	—				-	-	-	-	A	-	—	G	—	-	-	-				-	-	T	—				U/K
PEJ1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	-	T	-	-	-	-	-	-	-	-	-	-		—	—	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	G	-	U
PEJ2	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	-	-	-	-	-	-	-	-	—	—	—		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
PEJ3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	—	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
PEJ4	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	—	—	-	-	-	-	-	—	—	—	—	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
PEJ5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	—	—	-	-	-		-	-	—	—	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	-	-	U
PES7	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	-	-	-	-	-	-	-	—	—	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
POD1	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	A	-	-	-			—	A	-	-		-	-	—	—	-	-	A	-	G	-	-	T	-	-	-	-	-	-	-	T	-	—	—	W
POD2	-	-	-	-	-	-	-	C	-	-	-	G	-	-	-	-	-	-	-	T	-	-	-	-	-	—	—	-	-	-		-	-	—	—	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	T
POP1	—	—	—	-	-	-	-	C	-	-	-	G	-	-	-	—	—	—	—	T	-	-	-	-	-	—	—	-	-	C		-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	T
POP3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	—	T	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	G	A	C	-	G	-	-	-	-	-	-	-	-	C	T	-	—	—	U
POP4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	—	-	-		-	-	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	G	-	U
RIL3	—	—	—	-	-	-	-	-	-	—	-	-	-	-	-	-	-	-	—	T	-	-	-	-	-	—	—	-	-	-		-	-	G	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	T	-	—	C	K
SAC2	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-		-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	A	-	T	-	-	-	J
SMY3	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	—	-	-	-	-	-	-	-	-	—	—	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
SMY4	Y	-	-	-	-	-	-	-	-	—	-	—	—	-	-	—	—	—	-	-	—	—	—	—	—	—	—	—	—	-		—	—	—	—	-	-	-	-	-	-	-	-	-	—	-	-	-	-	-	—	—	-	-	H

Table S2. Cont.

Primer system	M01	M01	M01	M02	M02	M03	M04	M04	M05	M06	M07	M08	M08	M09	M10	M11	M11	M11	M12	M13	M14	M14	M14	M14	M15	M15	M16	M16	M17	M17	M18	M19	M19	M20	M20	M21	M21	M22	M22	M23	M24	M24	M25	M26	M26	M27	M27	M27	M28	M28	M29	M29	M30	M30	
SNP position	456	462	481	663	669	3010	4194	4216	4529	4580	4833	4917	4937	5178	5843	6366	6371	6392	6776	7028	8251	8260	8269	8272–80 Del	8388	8392	8935	8994	9090	9117	10034	10115	10118	10398	10400	10873	10876	11719	11732	11947	12308	12346	12705	13263	13269	13626	13637	13680	13708	13734	14766	14769	14793	14798		HG*
SMY9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	-	-	-	-	-	-	—	—	-	-	-	-		-	-	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		H
SMY11	—	—	—	-	-	-	-	**C**	-	-	-	G	-	-	-	—	—	—	-	T	-	-	-	-	-	-	-	-	-	-		—	—	—	—	-	-	A	—	-	-	-	-	-	-	-	-	-	-	**T**	-	-	-		T
SUG2	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	—	—	—	-	**T**	**A**	**C**	-	-	-	-	-	-	-	-		-	-	G	-	-	-	**A**	-	-	-	-	**T**	-	-	—	—	—	-	-	**T**	-	-	-		I
SUG5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		H
SUG6	-	-	-	-	-	-	-	-	-	—	-	-	-	-	-	—	—	—	-	Y	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		H
SUG7	—	—	—	-	-	-	—	—	—	—	-	—	—	-	-	-	-	-	-	—	—	—	—	—	—	—	—	—	-	-		—	—	-	-	-	-	—	-	-	-	-	—	-	-	-	-	-	-	—	—	-	-		H
SUG8	—	—	—	-	-	**R**	—	—	-	—	-	-	-	—	—	-	-	-	-	-	-	-	-	-			-	-	-	-		—	—	-	-	—	—	—	-	-	-	-	—	—	-	-	-	—	—	—	—	-	-	-		H
TEM1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**T**	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	**A**	-	-	**R**	-	-	-	-	-	-	-	-	**T**	-	-	-		U
TEM2	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		H
TEM3	—	T	—	-	-	**A**	-	**C**	-	-	-	-	-	-	-	-	-	—	T	-	-	**A**	-	-	-	-	—	—	—	—		—	—	**G**	-	-	-	**A**	-	-	-	-	-	-	-	-	-	-	-	**A**	-	**T**	-	—	—		J1
TEM4	—	—	—	-	-	-	-	-	-	—	-	-	-	-	-	—	—	—	—	T	-	-	-	-	—	—	-	-	—	—		—	—	-	-	-	-	**A**	-	-	G	-	-	—	-	-	-	-	—	-	**T**	**G**				U/K
TEM5	—	—	—	-	-	-	-	-	-	-	-	**G**	-	-	—	-	-	-	—	T	-	-	-	-	C	—	—	—	—	—		—	—	-	-	-	-	**A**	-	-	-	-	-	-	-	-	-	-	-	T	—	—	—		R1
TEM6	-	-	-	-	-	-	-	-	-	-	-	G	—	-	-	—	—	—	-	T	-	-	-	-	C	—	-	-	—	—		-	-	—	—	-	-	**A**	-	-	-	-	-	—	—	-	-	-	-	-			-	-		R1
TEM7	—	—	—	-	-	-	-	-	-	—	-	-	-	-	-	-	-	-	—	T	-	-	-	-	—	—	—	—	—	—		—	—	-	-	-	-	**A**	-	-	**G**	-	-	—	—	-	-	-	-	-	**T**	-	—	—		U
TEM8	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	T	-	-	-	-	-	-	—	—	-	-		-	-	-	-	-	-	**A**	-	-	**G**	-	-	-	-	-	-	-	-	-	**T**	-	G	—		U
TET1	-	-	-	-	-	-	-	-	**T**	-	-	-	-	-	-	—	—	—	-	**T**	**A**	-	-	-	—	—	-	—	-	-		-	-	**G**	-	-	-	**A**	-	-	-	-	**T**	-	-	—	—	—	-	-	T	—	-	-		I
TET2	—	—	—	-	-	-	-	-	-	—	-	-	-	-	-	-	-	-	—	T	-	-	-	-	—	—	—	—	—	—		—	—	-	-	-	-	**A**	-	-	**G**	-	-	-	-	-	-	-	-	-	**T**	-	—	—		U
VIN1	—	—	—	-	-	—	-	-	-	—	-	-	-	-	-	-	-	-	—	—	-	-	-	-	—	—	—	—	—	—		—	—	-	-	-	-	-	-	-	—	—	—	-	-	-	-	-	-	-	-	—	—		H
VIN2				-	-	-	-	**C**	-		-	**G**	-	-	-	-	-	-	—	T	-	-	-	-			—	—	—	—				-	-	-	-	**A**	-	-	-	-	-	-	-	—	—	—	-	-	**T**	-				T
VIN3	—	—	—	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	—	T	-	-	-	-	-	—	-	-	—	—		—	—	-	-	-	-	**A**	-	-	**G**	-	-	-	-	-	-	-	-	-	**T**	-	G	—		U
VIN5	—	—	—	-	-	—	-	**C**	-	—	-	G	—	-	-	—	—	—	—	T	-	-	-	-	—	—	—	—	—	—		—	—	—	—	—	—	**A**	-	-	—	—	-	—	—	—	—	—	-	-	T	—	—	—		T
VIN8	—	T	—	-	-	**A**	-	**C**	-	—	-	-	-	-	-	-	-	—	—	T	-	-	**A**	-	—	—	—	—	—	—		—	—	**G**	-	-	-	**A**	-	-	-	-	-	-	-	-	-	-	**A**	-	**T**	-	—	—		J1
VIN12	—	—	—	-	-	—	-	**C**	-	—	-	**G**	-	-	-	-	-	-	—	T	-	-	-	-	—	—	—	—	-	-		—	—	—	—	-	-	**A**	-	-	—	—	-	-	-	—	—	—	-	-	T	—	—	—		T

SNP positions are numbered according to the rCRS (1). The SNPs in the coding region are part of an interdependent hierarchical system and were not used in the actual population genetic analysis, but were merely a support and confirmation of the HVR1 results. Therefore, the strict triple reproduction criterion that was used for the HVR1 was not enforced on the coding region. Gray shading indicates additional polymorphic sites in the amplicons besides the targeted SNP for haplogroup assignments. Bold SNPs and short dashes (if the positions match the rCRS) are based on at least two PCRs (at least three reads each). Nonbold SNPs and long dashes are based either on one PCR, or at least two PCRs with one giving an ambiguous result. Empty fields indicate that amplification was unsuccessful. HG*, haplogroup assignment based on the coding region SNPs.

1. Andrews RM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.

Table S3. Relative genotype frequencies and exact Hardy–Weinberg equilibrium (HWE) testing for the three pigmentation SNPs and three SNPs in the *ABCB1* gene in the ancient and modern Ukrainian dataset

Population	*N*	Genotype frequencies			HWE test, *P* value
HERC2 rs12913832		*AA*	*AG*	*GG*	
Ancient	47	0.766 (36)	0.149 (7)	0.085 (4)	0.0084
Modern	43	0.186 (8)	0.326 (14)	0.488 (21)	0.0543
SLC45A2 rs16891982		*CC*	*CG*	*GG*	
Ancient	22	0.409 (9)	0.318 (7)	0.273 (6)	0.0899
Modern	41	0.000 (0)	0.146 (6)	0.854 (35)	1.0000
TYR rs1042602		*CC*	*CA*	*AA*	
Ancient	46	0.935 (43)	0.043 (2)	0.022 (1)	0.0656
Modern	49	0.408 (20)	0.449 (22)	0.143 (7)	0.4957
ABCB1 rs1128503		*CC*	*CT*	*TT*	
Ancient	24	0.333 (8)	0.417 (10)	0.250 (6)	0.3137
Modern	44	0.205 (9)	0.568 (25)	0.227 (10)	0.8730
ABCB1 rs2032582		*GG*	*GT*	*TT*	
Ancient	28	0.414 (12)	0.310 (9)	0.241 (7)	0.0668
Modern	41	0.159 (7)	0.545 (24)	0.227 (10)	0.9162
ABCB1 rs1045642		*CC*	*CT*	*TT*	
Ancient	31	0.161 (5)	0.355 (11)	0.484 (15)	0.1946
Modern	48	0.125 (6)	0.500 (24)	0.375 (18)	0.7570

The genotypes of the three SNPs in the ABCB1 gene were established by a minimum of three PCR products from two independent extractions. *P* value is a one-tailed test for excess of homozygotes, using the HWExact function, as implemented in the R library HardyWeinberg (1). Numbers in parentheses are absolute number of observed genotypes. *N*, total number of genotyped individuals.

1. R Core Team (2012) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).

Table S4. Primer for the amplification of the mitochondrial HVR1 and genotyping of three pigmentation SNPs

Multiplex	Amplicon	Primer	Primer sequence 5′ > 3′	Source
MP-A	HVR1-M31	16011U	AGC ACC CAA AGC TAA GAT TCT AAT TT	M.U.
		16088L	GTG GCT GGC AGT AAT GTA CGA AAT AC	M.U.
MP-B	HVR1-M32	16071U	GGG TAC CAC CCA AGT ATT GAC TCA	M.U.
		16153L	TGA TGT GGA TTG GGT TTT TAT GTA CTA	M.U.
MP-C	HVR1-M33	16119U	GTA CAT TAC TGC CAG CCA CCA TG	M.U.
		16207L	TGA TAG TTG AGG GTT GAT TGC TGT AC	M.U.
MP-A	HVR1-M34	16185U	TAC ATA AAA ACC CAA TCC ACA TCA AAA C	M.U.
		16271L	GGT GGG TAG GTT TGT TGG TAT CCT	M.U.
MP-B	HVR1-M35	16233U	AGT ACA GCA ATC AAC CCT CAA CTA TC	M.U.
		16305L	TGT ACG GTA AAT GGC TTT ATG TAC TAT G	M.U.
MP-C	HVR1-M36	16274U	AAA GCC ACC CCT CAC CCA CTA G	M.U.
		16345L	TGG GGA CGA GAA GGG ATT TGA C	M.U.
MP-A	HVR1-M37	16340U	ACA TAA AGC CAT TTA CCG TAC ATA GCA C	M.U.
		16413L	CAC TCT TGT GCG GGA TAT TGA TTT C	M.U.
—	HVR1-I	L15996	CTC CAC CAT TAG CAC CCA AAG C	1
		H16142	ATG TAC TAC AGG TGG TCA AG	2
—	HVR1-II	L16117	TAC ATT ACT GCC AGC CAC CAT	3
		H16233	GCT TTG GAG TTG CAG TTG ATG TGT	3
—	HVR1-III	L16209	CCC CAT GCT TAC AAG CAA GT	4
		H16348	ATG GGG ACG AGA AGG GAT TTG	3
—	HVR1-IV	L16287	CAC TAG GAT ACC AAC AAA CC	4
		H16410	GCG GGA TAT TGA TTT CAC GG	4
NP	HERC2 rs12913832	HER_U	CCA AGA GGC GAG GCC AGT TTC A	S.W.
		HER_L	AAG CCT CGG CCC CTG ATG ATG	S.W.
NP	SLC45A2 rs16891982	S45_U2	AGA ATA AAG TGA GGA AAA CA	S.W.
		S45_L2	GAA AGA GGA GTC GAG GTT GGA	S.W.
NP	TYR rs1042602	TYR_U	TTT GTC TGG ATG CAT TAT TAT GTG TCA	S.W.
		TYR_L	CTT CAT GGG CAA AAT CAA TGT CTC	S.W.

The first column states in which multiplex-PCR setups the primers have been used. Primer names of the seven fragments amplifying the HVR1 (M31–M37) present the first base after the 3′ end of the primer according to the rCRS (5). Primer names of the four overlapping systems (I–IV) indicate the last base of the 3′ end of the primer. For primers for SNPs in the coding region see Table S5.

1. Endicott P, et al. (2003) The genetic origins of the Andaman Islanders. *Am J Hum Genet* 72(1):178–184.
2. Stone AC, Stoneking M (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *Am J Hum Genet* 62(5):1153–1170.
3. Haak W, et al. (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310(5750):1016–1018.
4. Handt O, Krings M, Ward RH, Pääbo S (1996) The retrieval of ancient human DNA sequences. *Am J Hum Genet* 59(2):368–376.
5. Andrews RM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.

Table S5. Primer for the amplification of mitochondrial coding region SNPs

Multiplex	Amplicon	Targeted SNP	Primer	Primer sequence 5′ > 3′
MP-A	M01	456	423U	AAT TTT ATC TTT TGG CGG TAT GCA CTT
			485L	GAT GGG CGG GGG TTG TAT TG
MP-B	M02	663	654U	CTC ACA TCA CCC CAT AAA CAA ATA GG
			699L	AAC TCA CTG GAA CGG GGA TGC T
MP-C	M03	3010	2992U	CAA CAA TAG GGT TTA CGA CCT CGA T
			3057L	CTC CGG TCT GAA CTC AGA TCA CGT A
MP-A	M04	4216	4155U	TAC CCC CGA TTC CGC TAC GA
			4221L	ATG CTG GAG ATT GTA ATG GGT ATG GA
MP-B	M05	4529	4499U	CTG GCC CAA CCC GTC ATC TA
			4554L	GCA TGT TTA TTT CTA GGC CTA CTC AGG
MP-C	M06	4580	4549U	ACA GCG CTA AGC TCG CAC TGA T
			4617L	ATG GCA GCT TCT GTG GAA CGA G
MP-A	M07	4833	4815U	GAA TAG CCC CCT TTC ACT TCT GAG TC
			4864L	TGA GAT GGG GGC TAG TTT TTG TCA T
MP-B	M08	4917	4871U	GGC CTG CTT CTT CTC ACA TGA CA
			4940L	ACT GCC TGC TAT GAT GGA TAA GAT TGA
MP-C	M09	5178	5163U	CCA GCA CCA CGA CCC TAC TAC TAT CT
			5179L	GGA TGG AAT TAA GGG TGT TAG TCA TGT T
MP-A	M10	5843	5836U	AAA TCA CCT CGG AGC TGG TAA AAA G
			5875L	GGG GTG AGG TAA AAT GGC TGA GT
MP-B	M11	6371	6336U	CAC CCT GGA GCC TCC GTA GAC
		6392	6403L	ATG GCA GGG GGT TTT ATA TTG ATA ATT
MP-C	M12	6776	6764U	CAA TTG GCT TCC TAG GGT TTA TCG T
			6814L	GAT GAT TAT GGT AGC GGA GGT GAA A
MP-A	M13	7028	6975U	GGT GGC CTG ACT GGC ATT GTA
			7046L	TAT GAT GGC AAA TAC AGC TCC TAT TGA
MP-B	M14	8272–8280 (9bp-del)	8226U	CAT GCC CAT CGT CCT AGA ATT AA
			8287L	GCT AAG TTA GCT TTA CAG TGG GCT CTA
MP-C	M15	8392	8385U	TAC AGT GAA ATG CCC CAA CTA AAT ACT A
			8417L	TTT AGT TGG GTG ATG AGG AAT AGT GTA A
MP-A	M16	8994	8932U	ACT TCT TAC CAC AAG GCA CAC CTA CA
			8996L	AGT AAT GTT AGC GGT TAG GCG TAC G
MP-B	M17	9090	9072U	GAA GCG CCA CCC TAG CAA TAT C
			9124L	TAA GGC GAC AGC GAT TTC TAG GAT AG
MP-C	M18	10034	10000U	CAT CTA TTG ATG AGG GTC TTA CTC TTT TA
			10048L	AAA TTA AGG CGA AGT TTA TTA CTC TTT TT
MP-A	M19	10115, 10118	10105U	TTA ATA ATC AAC ACC CTC CTA GCC TTA C
			10166L	GGT CGA AGC CGC ACT CGT A
MP-B	M20	10398, 10400	10387U	TCT GGC CTA TGA GTG ACT ACA AAA AG
			10451L	AGG GGC ATT TGG TAA ATA TGA TTA TC
MP-C	M21	10873	10865U	CAA CCA CCC ACA GCC TAA TTA TTA GC
			10895L	TGG GGA ACA GCT AAA TAG GTT GTT GT
MP-A	M22	11719	11700U	AGC TTC ACC GGC GCA GTC A
			11743L	GTG CGT TCG TAG TTT GAG TTT GCT AG
MP-B	M23	11947	11935U	ACC ACG TTC TCC TGA TCA AAT ATC AC
			11983L	CCC CAT TGT GTT GTG GTA AAT ATG TA
MP-C	M24	12308	12303U	GAT AAC AGC TAT CCA TTG GTC TTA GGC
			12352L	GGA AGT CAG GGT TAG GGT GGT TAT AG
MP-A	M25	12705	12692U	CAG ACC CAA ACA TTA ATC AGT TCT TCA
			12754L	GCC CTC TCA GCC GAT GAA CA
MP-B	M26	13263	13231U	GCG CCC TTA CAC AAA ATG ACA TC
			13275L	GGT TGG TTG ATG CCG ATT GTA ACT AT
MP-C	M27	13626	13620U	AAG CGC CTA TAG CAC TCG AAT AAT TCT
			13683L	CCA GGC GTT TAA TGG GGT TTA GTA G
MP-A	M28	13708	13701U	ACC CCA CCC TAC TAA ACC CCA TTA A
			13740L	GAT GCG GGG GAA ATG TTG TTA GT
MP-B	M29	14766	14717U	CAA CCA CGA CCA ATG ATA TGA AAA AC
			14784L	GGA GGT CGA TGA ATG AGT GGT TAA TT
MP-C	M30	14798	14783U	ATA CGC AAA ACT AAC CCC CTA ATA AAA
			14839L	GCC AAG GAG TGA GCC GAA GTT

Coding region primers designed by M.U. were used in a triple multiplex-PCR setup with HVR1 primers M31–M37. Multiplex indicates in which of the three mitochondrial multiplex-PCR setups (A, B, or C) the respective locus was amplified. Targeted SNP indicates the position of the SNP under investigation according to rCRS numbering (1). Primer names indicate the last base of the 3′ end of the primer according to rCRS numbering (1).

1. Andrews RM, et al. (1999) Ranalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.