Article

A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes

Qiaomei Fu,[1,3,19] Alissa Mittnik,[2,19] Philip L.F. Johnson,[4] Kirsten Bos,[2,5] Martina Lari,[6] Ruth Bollongino,[7] Chengkai Sun,[8] Liane Giemsch,[9,10] Ralf Schmitz,[9] Joachim Burger,[7] Anna Maria Ronchitelli,[11] Fabio Martini,[12] Renata G. Cremonesi,[13] Jiří Svoboda,[14,15] Peter Bauer,[16] David Caramelli,[6] Sergi Castellano,[1] David Reich,[17,18] Svante Pääbo,[1] and Johannes Krause[2,*]

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany
[2]Institute for Archaeological Sciences, University of Tübingen, Rümelinstr. 23, 72070 Tübingen, Germany
[3]CAS-MPS Joint Laboratory for Human Evolution and Archaeometry, Institute of Vertebrate Paleontology and Paleoanthropology of Chinese Academy of Sciences, Beijing 100044, P.R. China
[4]Department of Biology, Emory University, Atlanta, GA 30322, USA
[5]Laboratoire de Paléoanthropologie, École Pratique des Hautes Études, UMR 5199 PACEA, CNRS-Université de Bordeaux, 33405 Bordeaux, France
[6]Dipartimento di Biologia Evoluzionistica, Università di Firenze, 4 - 50121 Firenze, Italy
[7]Palaeogenetics Group, Institute for Anthropology, Johannes Gutenberg-University, Saarstrasse 21, D-55099 Mainz, Germany
[8]Shandong Museum, 11899 Jing 10th Road, Jinan, Shandong 250014, P.R. China
[9]LVR-Landesmuseum Bonn, Bachstrasse 5-9, D-53115 Bonn, Germany
[10]Department of Prehistoric and Protohistoric Archaeology, Institute for Archaeology and Cultural Anthropology, University of Bonn, Regina-Pacis-Weg 7, 53113 Bonn, Germany
[11]Dipartimento di Scienze Fisiche, della Terra e dell'Ambiente, U.R. Ecologia Preistorica, Università degli Studi di Siena, Via Laterina, 8 - 53100 Siena, Italy
[12]Università di Firenze, Dipartimento di Scienze dell'Antichità, Medioevo e Rinascimento e Linguistica, Piazza Brunelleschi, 3-4 - 50121 Firenze, Italy
[13]Dipartimento di Scienze Archeologiche, Università di Pisa, via Galvani 1, 56126 Pisa, Italy
[14]Department of Anthropology, Faculty of Science, Masaryk University, Vinařská 5, 603 00 Brno, Czech Republic
[15]Institute of Archaeology, Academy of Science of the Czech Republic, Kralovopolska 147, 612 00 Brno, Czech Republic
[16]Human Genetics Department, Medical Faculty, University of Tübingen, 72070 Tübingen, Germany
[17]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[18]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

Summary

Background: Recent analyses of de novo DNA mutations in modern humans have suggested a nuclear substitution rate that is approximately half that of previous estimates based on fossil calibration. This result has led to suggestions that major events in human evolution occurred far earlier than previously thought.
Results: Here, we use mitochondrial genome sequences from ten securely dated ancient modern humans spanning 40,000 years as calibration points for the mitochondrial clock, thus yielding a direct estimate of the mitochondrial substitution rate. Our clock yields mitochondrial divergence times that are in agreement with earlier estimates based on calibration points derived from either fossils or archaeological material. In particular, our results imply a separation of non-Africans from the most closely related sub-Saharan African mitochondrial DNAs (haplogroup L3) that occurred less than 62–95 kya.
Conclusions: Though single loci like mitochondrial DNA (mtDNA) can only provide biased estimates of population divergence times, they can provide valid upper bounds. Our results exclude most of the older dates for African and non-African population divergences recently suggested by de novo mutation rate estimates in the nuclear genome.

Introduction

Differences in DNA sequences correspond to nucleotide substitutions that have accumulated since their split from a most recent common ancestor (MRCA). When the average number of substitutions occurring per unit of time can be determined, the "molecular clock" rate can be estimated. Under the assumption of constant rates of change among lineages, molecular clocks have been used to estimate divergence times between closely related species or between populations. Fossil evidence has been frequently used to estimate a date for the MRCA of two related groups, thus providing a calibration point for the molecular clock. The sparseness of the fossil record, however, poses limitations on the reliability of such estimates. For example, in human evolution, no fossil has yet been identified to represent the uncontested MRCA for humans and chimpanzees or other closely related primate species. As a consequence, the nuclear and mitochondrial mutation rates for the human lineage have been heavily debated [1].

Recent analyses of de novo substitutions from genome sequencing of parent and offspring trios allow the direct calculation of nuclear substitution rates per generation. This alternative to the fossil calibration of the human molecular clock is arguably more accurate. Surprisingly, publications using this approach have recently pointed to de novo rates that are about half the value of those previously determined from fossil calibrations [1–5]. A slower substitution rate has important implications for inferring the timing of key events in human evolution such as our divergence from our common ancestor with chimpanzees, our divergence from Neanderthals and Denisovans, and the migration of modern humans from one region or habitat to another. Taking these new rates into consideration, most date estimates would be pushed back by a factor of two—for example, yielding a West African/non-African split date of 90–130 kya, which is up to 60 kya older than some previous estimates [1].

[19]These authors contributed equally to this work
*Correspondence: johannes.krause@uni-tuebingen.de

Table 1. Samples Analyzed in this Study

Sample	^{14}C Age (calBP)	Lab Number	nt Covered at 3-Fold Coverage (% of mtDNA)	Contamination (Based on [34])	Contamination (Bayesian Estimate)	C to T Misincorporation at 5′ End (%)	Hg
Tianyuan 1301 [29] [a]	39,475 ± 645	BA-03222	16,559 (99.9%)	0%–4.7%	0.9%–4.7%	31.5	B
Kostenki 14 [10] [a]	37,985 ± 665	OxA-X-2395-15	16,568 (100%)	0%–7%	1.7%–8.5%	38	U2
Dolni Vestonice 13 [a,b]	31,155 ± 85	GrN-14,831	16,570 (100%)	0%–2.3%	0.9%–2.4%	30.1	U8
Dolni Vestonice 14 [a,b]	31,155 ± 85	GrN-14,831	16,530 (99.8%)	0%–100%	1.9%–9.2%	24.4	U
Dolni Vestonice 15 [b]	31,155 ± 85	GrN-14,831	16,051 (96.9%)	0%–100%	0%–3.9%	20.1	U
Oberkassel 998 [a,b]	14,020 ± 150	OxA-4790	16,570 (100%)	0%–100%	0.5%–2.4%	36.0	U5b1
Oberkassel 999 [b]	13,430 ± 140	OxA-4792	16,560 (99.9%)	0 %–100%	3.8%–8.3%	31.5	U5b1
Continenza 7 [b]	N/A		16,478 (99.5%)	0.3%–2.8%	0.9%–7.2%	29.1	U5b2b1
Paglicci Accesso sala 2 Rimp [b]	N/A		16,548 (99.9%)	0%–5.3%	0.3%–2.8%	57.1	U2′3′4′7′8′9
Paglicci Str. 4b [b]	N/A		16,509 (99.6%)	0%–4.8%	1.5%–5.1%	8.8	H1
Boshan 11 [a,b]	8,180 ± 140	MAMS-13530	16,559 (99.9%)		0%–1.8%	47.9	B4c1a
Loschbour [a,b]	8,054 ± 127	OxA-7338	16,569 (100%)	0%–0.5%	1.3%–1.9%	27.6	U5b1a
Iceman [30] [a]	4,550	OxA-3371–6 OxA-3419-21	16,576 (100%)	N/A	N/A	N/A	K1
Saqqaq [31] [a]	3,600–4,170	OxA-20656	16,568 (100%)	N/A	N/A	N/A	D2a1
"Cro-Magnon 1" [a,b]	690 ± 39	OxA-V-2321-38	16,499 (99.6%)	0%–7.7%	0.3%–3.3%	N/A	T2b1

[a]Sequence used for estimating substitution rate via linear regression and Bayesian framework.
[b]Sequence presented in this study.

Attempts at calculating the human mitochondrial DNA (mtDNA) substitution rate have relied on either estimates derived from fossil calibration or archaeological evidence of founding migrations; however, reliance on a single calibration date can easily lead to a biased rate estimate [6]. Estimates of substitution rates for the coding region of the mtDNA have ranged from 1.26×10^{-8} substitutions per site per year [7], calibrated with a chimpanzee-human divergence time of 6.5 million years before present (BP), to 1.69×10^{-8}, inferred from an accepted date of 45 kya for the peopling of Oceania [8]. This 45 kya date is based on radiocarbon-dated archaeological material, thus providing a calibrated age for the haplogroup (hg) Q lineage that is unique to contemporary populations from this region. Soares et al. [9] considered both the coding and noncoding regions in their estimate to accommodate the effects of natural selection and arrived at a rate of 1.67×10^{-8} substitutions per site per year using a calibration point of 7 million years BP for the chimpanzee-human split.

An alternative approach for obtaining greater precision in measuring substitution rates is through the analysis of genetic data from ancient samples for which reliable radiocarbon dates are available. Ancient humans are well suited to provide calibration points for the human mitochondrial molecular clock: reliable radiocarbon dates are available for many specimens; hence, the number of substitutions that have accumulated among lineages can be directly translated into the number of substitutions per site per year. Branch shortening—the effect of fewer nucleotide substitutions on ancient branches in a phylogenetic tree as compared to modern—is commonly observed in phylogenetic studies of ancient humans [10]. The observed branch shortening reflects the comparatively shorter time since the common ancestor for the ancient human as compared to the present-day individual: a present-day lineage has had more time to accumulate nucleotide changes. If we know the age of the ancient sample (e.g., from a carbon date), we can thus infer the mutation rate necessary to produce the observed degree of branch shortening.

Here we use the complete or nearly complete mitochondrial genomes from ten ancient modern humans for which reliable radiocarbon dates are available to calculate the human mtDNA substitution rate directly. This strategy circumvents the limitations imposed by the use of indirect measures of substitution rates such as those obtained via fossil calibration. The samples used in this analysis span 40,000 years of human history and originate from Europe and Eastern Asia. We use our substitution rate to estimate the dates of major human evolutionary events in the last 200,000 years. Of particular note, our rate suggests an upper bound on the split between non-Africans and sub-Saharan Africans of less than 95 kya. Even though this estimate is a conservative upper bound because the genetic divergence at any particular locus in the genome is by definition older than the population divergence time, this date, is on the extreme lower end of population split times estimated from de novo substitution rates for nuclear DNA [1].

Results

DNA Preservation and Contamination

High-throughput sequencing of the enriched libraries yielded between 4,898 and 27,022,382 sequences for each sample. These sequences were used as inputs for the iterative mapping assembler (MIA) [11]. For each sample, between 0 and 74,435 unique human mtDNA fragments were mapped against the revised Cambridge Reference Sequence (rCRS). Complete or nearly complete mitochondrial genomes with at least 3-fold coverage could be reconstructed for 11 individuals (Table 1), including the sample from China. The average length of the obtained mtDNA fragments ranged from 54 to 77 bp (Table S6, available online). Using a previously published contamination estimation method [11], four samples showed a low percentage of inconsistent fragments, suggesting that the DNA originated from a single biological source, whereas the other sequences did not contain enough unique diagnostic substitutions to assess contamination using this method. Our more powerful Bayesian contamination estimate (Supplemental Information, section 5) suggests that all samples with the exception of Oberkassel 999 are uncontaminated to within the limits of our resolution (Table 1). On this basis, we excluded Oberkassel 999 from subsequent analyses. To further evaluate the authenticity of the ancient DNA we calculated the proportion of nucleotide misincorporations arising from DNA

Table 2. Inferred TMRCA for All Modern Humans and Mean Substitution Obtained for Various Subsets of the mtDNA Genome Assuming a Constant Population Size and Relaxed Molecular Clock

mtDNA Partition	TMRCA			μ/ Site / Year (Units of 10^{-8})		
	Best Estimate	Lower 2.5%	Upper 2.5%	Best Estimate	Lower 2.5%	Upper 2.5%
Whole mtDNA	157,000	120,000	197,000	2.67	2.16	3.16
Coding region	178,000	126,000	236,000	1.57	1.17	1.98
1st-2nd Codon	207,000	78,800	382,000	0.82	0.30	1.37
3rd Codon	233,000	134,000	356,000	3.27	1.94	4.62

Table 3. TMRCA of Different Haplogroups

mtDNA Hg Group	Whole Molecule TMRCA (KBP)	95% HPD		Coding Region TMRCA (KBP)	95% HPD	
		Lower	Upper		Lower	Upper
A	33.7	22.4	45.1	28.6	17.0	41.2
C	24.2	16.6	32.5	24.2	14.7	34.9
D	37.8	27.6	49.1	45.3	28.7	62.7
H	23.9	16.6	32.3	20.7	12.5	29.6
J	26.0	16.4	36.4	27.1	13.9	42.2
L3	78.3	62.4	94.9	89.7	66.8	116.4
M+N	77.0	61.4	93.2	88.2	64.5	114.5
P	53.6	43.1	65.5	60.9	41.6	83.2
Q	42.0	30.0	54.9	39.8	25.6	55.7
T	21.1	13.0	29.8	17.4	9.0	26.8
U5	29.6	22.7	37.2	34.4	22.8	47.7
U6	35.7	24.7	47.6	31.8	20.4	44.6
U	52.1	44.3	60.9	56.1	44.0	70.5
V	13.5	7.1	20.3	15.5	7.5	24.5

damage, a quantity that is known to increase over time after the death of an individual [12] and has been used as an indication of authenticity in previous work [10]. It was suggested that bone samples 100 years and older have a minimum of 20% C to T misincorporations concentrated at the 5′ end of the molecule [13]. Using this criterion, we excluded Paglicci Str. 4b from further analysis as the rate of C to T misincorporation at the 5′ end was only 8.8%, thus making an ancient origin for the DNA in this sample uncertain [14].

Evolutionary Analysis

All but one of the ancient modern human sequences from Europe belonged to mtDNA hg U, thus confirming previous findings that hg U was the dominant type of mtDNA before the spread of agriculture into Europe [15]. The exception was the Cro-Magnon 1 sample, which belonged to the derived hg T2b1, an unexpected hg given its putative age of 30,000 years [16]. Since the radiocarbon date for this specimen was obtained from an associated shell [16], we dated the sample itself using accelerator mass spectrometry (AMS). Surprisingly, the sample had a much younger age of about 700 years, suggesting a medieval origin. Consequently, this bone fragment has now been removed from the Cro-Magnon collection at the Musée de l'Homme in Paris. Attempts to directly date other remains from the Cro-Magnon type collection unfortunately failed. The good molecular preservation of our sample for both DNA and AMS dating, in contrast, suggests that this particular bone has a different origin from the other remains in the collection.

For the remaining eight ancient Europeans, we built a phylogenetic tree for hg U, which included 63 contemporary mtDNAs from this haplogroup. The tree clearly shows that all four Paleolithic pre-Last Glacial Maximum (LGM) samples display a short branch compared to the four ancient post-LGM samples (Figure S2). Predictably, the older samples Dolni Vestonice 14 and 15 fall in a basal position relative to the contemporary mtDNA hg U5. The Tianyuan sequence from Eastern China falls basal to the contemporary hg B, common in most parts of Eastern Asia, Oceania, and the Americas. The mtDNA genetic diversity that we measure in early modern Europeans is about 2-fold less than the mtDNA diversity in today's Europeans, but about 1.5 times higher than that measured in Neanderthals contemporary with these early modern humans [17] (excluding the older Mezmaiskaya individual) (Table S3). Although these measurements provisionally suggest that a higher population size might have contributed to early modern humans outcompeting Neanderthals after their arrival in Europe, there are caveats to this analysis. First, we have a limited sample size of ancient specimens. Second, we have sampled from several different time periods, a practice that overestimates actual genetic diversity [17]. Third, our sampling is nonrandom; for example, we included several

individuals from within the same burial site (Dolni Vestonice), where maternal relatedness is possible and would give an underestimate of true diversity. More data are necessary to provide a definitive assessment of the genetic diversity of these prehistoric populations.

Substitution Rate Estimates

For the linear regression approach we estimate a substitution rate of 1.92×10^{-8} per site and year ($1.16–2.68 \times 10^{-8}$, 95% CI) for the whole mtDNA and $1.25 \pm 0.68 \times 10^{-8}$ per site and year ($0.57–1.93 \times 10^{-8}$, 95% CI) for the coding region.

For the Bayesian approach, the final model was chosen based on a Log_{10} of the Bayes factor (BF) being more than 1.3. The best-fit comparison of the results of the Bayesian Markov Chain Monte Carlo (MCMC) analysis, calibrated with the fossil ages, favors the constant population size model over the exponential growth model (Log_{10} BF = 1087.2 > 1.3). Although values obtained with the relaxed clock model fit the data better (Log_{10} BF = 3.6 > 1.3), the maximum likelihood (ML) test does not reject the null hypothesis of a constant substitution rate across the tree topology. Using the constant size model and a relaxed clock we thus estimate a substitution rate of 2.67×10^{-8} substitutions per site per year ($2.16–3.16 \times 10^{-8}$, 95% HPD) for the whole mtDNA genome and 1.57×10^{-8} substitutions per site per year for the coding region ($1.17–1.98 \times 10^{-8}$, 95% HPD) (Table 2). The substitution rates for the mtDNA coding region and whole mtDNA largely overlap with the above results when the four radiocarbon-dated Neanderthals are included alongside the ten ancient modern humans (Supplemental Information, section 6 and Table S5). In theory, mitochondrial substitution rates could have changed between Neanderthals and modern humans, though we do not detect evidence of this because inclusion of Neanderthal data does not lead to a rejection of the molecular clock. We use only the substitution rates calculated with radiocarbon-dated ancient modern humans to calculate modern human mtDNA divergence times, given that substitution rates among modern humans are most relevant to estimating divergence times among modern humans and that the Neanderthal data add little extra statistical precision.

Haplogroup Divergence Time Estimates

Using the substitution rate for the whole mtDNA genome obtained by Bayesian estimation, we estimated the time of the

MRCA for all modern humans at 157 kya (120–197 kya, 95% HPD) (Table 2). Our rate also implies a split of all non-African hgs from the closest widespread sub-Saharan African hg (L3) of 78.3 kya (62.4–94.9 kya) (Table 3). The MRCA of hg Q, often referred to as a maximum age for the settlement of Australia, was calculated at 42 kya (30–54.9 kya) (Table 3). The time to most recent common ancestor (TMRCA) of hg U5, often argued to have evolved within the first early modern humans in Europe [18], was calculated at 29.6 kya (22.7–37.2 kya) (Table 3).

To test whether the inferred mutation rates are dependent on a single directly dated mtDNA sequence (which in principle could have an inaccurate carbon date), we carried out a Bayesian MCMC analysis for the coding regions with a constant size model and relaxed clock using each mtDNA sequence older than 4,000 years independently as separate tip calibrations. The results range from 1.14×10^{-8} substitutions per site per year for the Kostenki specimen to 4.5×10^{-8} substitutions per site per year for the Boshan specimen (Table S1). The confidence intervals for all samples overlap the value obtained for the whole data set, suggesting that no single sample is driving our overall mutation rate estimate.

Discussion

We were able to reconstruct three complete and six nearly complete mitochondrial genomes from ancient human remains that were found in Europe and Eastern Asia and span 40,000 years of human history. All Paleolithic and Mesolithic European samples belong to mtDNA hg U, as was previously suggested for pre-Neolithic Europeans [15]. Two of the three individuals from the Dolni Vestonice triple burial associated with the pre-ice age Gravettian culture, namely, 14 and 15, show identical mtDNAs, suggesting a maternal relationship. Furthermore, both individuals display a mitochondrial sequence that falls basal in a phylogenetic tree compared to the post-ice age hunter-gatherer samples from Italy and central Europe, as well as the contemporary mtDNA hg U5 (Figure 1). It has been argued that hg U5 is the most ancient subhaplogroup of the U lineage, originating among the first early modern humans in Europe [18]. Our results support this hypothesis because we find that the two Dolni Vestonice individuals radiocarbon dated to 31.5 kya carry a type of mtDNA that is as yet uncharacterized, sits close to the root of hg U, and carries two mutations that are specific to hg U5. With our recalibrated molecular clock, we date the age of the U5 branch to approximately 30 kya, thus predating the LGM. Because the majority of late Paleolithic and Mesolithic mtDNAs analyzed to date fall on one of the branches of U5 (see also [15]), our data provide some support for maternal genetic continuity between the pre- and post-ice age European hunter-gatherers from the time of first settlement to the onset of the Neolithic. U4, another hg commonly found in Mesolithic hunter-gatherers [15], has so far not been sequenced in a Paleolithic individual, and we find hgs U8 and U2 in pre-LGM individuals but not in later hunter-gatherers. At present, the genetic data on Upper Paleolithic, and especially pre-ice age, populations are too sparse to comment on whether or not this is representative of a change in the genetic structure of the population, perhaps caused by a bottleneck during the LGM and a subsequent repopulation from glacial refugia.

Using ancient mtDNA sequences from securely dated archaeological samples as calibration points has allowed us to obtain an estimate of the mtDNA substitution rate that is complementary to the existing estimates based on calibration from the fossil and archaeological records. We arrive at a rate of 1.57×10^{-8} substitutions per site per year for the coding region and 2.67×10^{-8} substitutions per site per year for the whole molecule, which is approximately 1.6-fold higher than the fossil-calibrated rate [7]. Our inferred substitution rate from the whole mtDNA implies a coalescence date for all modern human mtDNAs of 120–197 kya and of 62–95 kya for hg L3, the lineage from which all non-African mtDNA hgs descend. This places a conservative upper bound of 95 kya for the time of the last major gene exchange between non-African and sub-Saharan African populations. It is important to recognize that this divergence time may merely represent the most recent gene exchanges between the ancestors of non-Africans and the most closely related sub-Saharan Africans and thus may reflect only the most recent population split in a long, drawn-out process of population separation [1]. Nevertheless, the fact that hg L3 is currently so widespread within Africa suggests that the population split dated by L3 is likely to be one of the most important ones in that history of separation, giving rise to lineages that contributed substantial fractions to the ancestry of both present-day sub-Saharan African and present-day non-African populations.

Although our estimate for the population divergence of non-African and sub-Saharan Africans has a small overlap with those calculated from the de novo genomic rates, which range from 90 to 130 kya [1], the ~30,000 year difference in the mean divergence time obtained via the different methods is notable. We believe this discrepancy is unlikely to be explained by differences in the inheritance patterns between the two parts of the genome (exclusively maternal versus biparental) or by differences in generation times of males and females [19]. In particular, de novo rates appear to scale with parental age (per year) rather than with generation time [4], and our calculation from branch shortening directly uses units of years without making any assumptions about generation times. We note that our calculated dates are more consistent with some interpretations of the fossil record: for example, the low nuclear mutation rates from the de novo investigations imply a date of human-orangutan speciation that is at least ten million years older than what is supported by the fossil record [20], whereas our dates are in accord with a more recent speciation [21, 22]. One possible reason for the discrepancies between our inferences and those that have been made based on de novo mutation rates in the nuclear genome is a substantial rate of false-negative mutations in the de novo data sets (due to the intense filtering that these studies need to apply to discriminate false-positive mutations from true positives). It is also possible that the filtering applied in the de novo mutation rate estimation studies has excluded subsets of the genome that are more mutable and that have been included in sequence divergence calculations; it is important to estimate mutation rate and sequence divergence in the same subsets of the genome to properly calibrate time estimates and this has not been done as far as we are aware. An important direction for future research is to compare the new de novo rates with those estimated based on patterns of substitutions observed over time in the nuclear genome as determined from ancient sequences [23], using methods similar to those employed here.
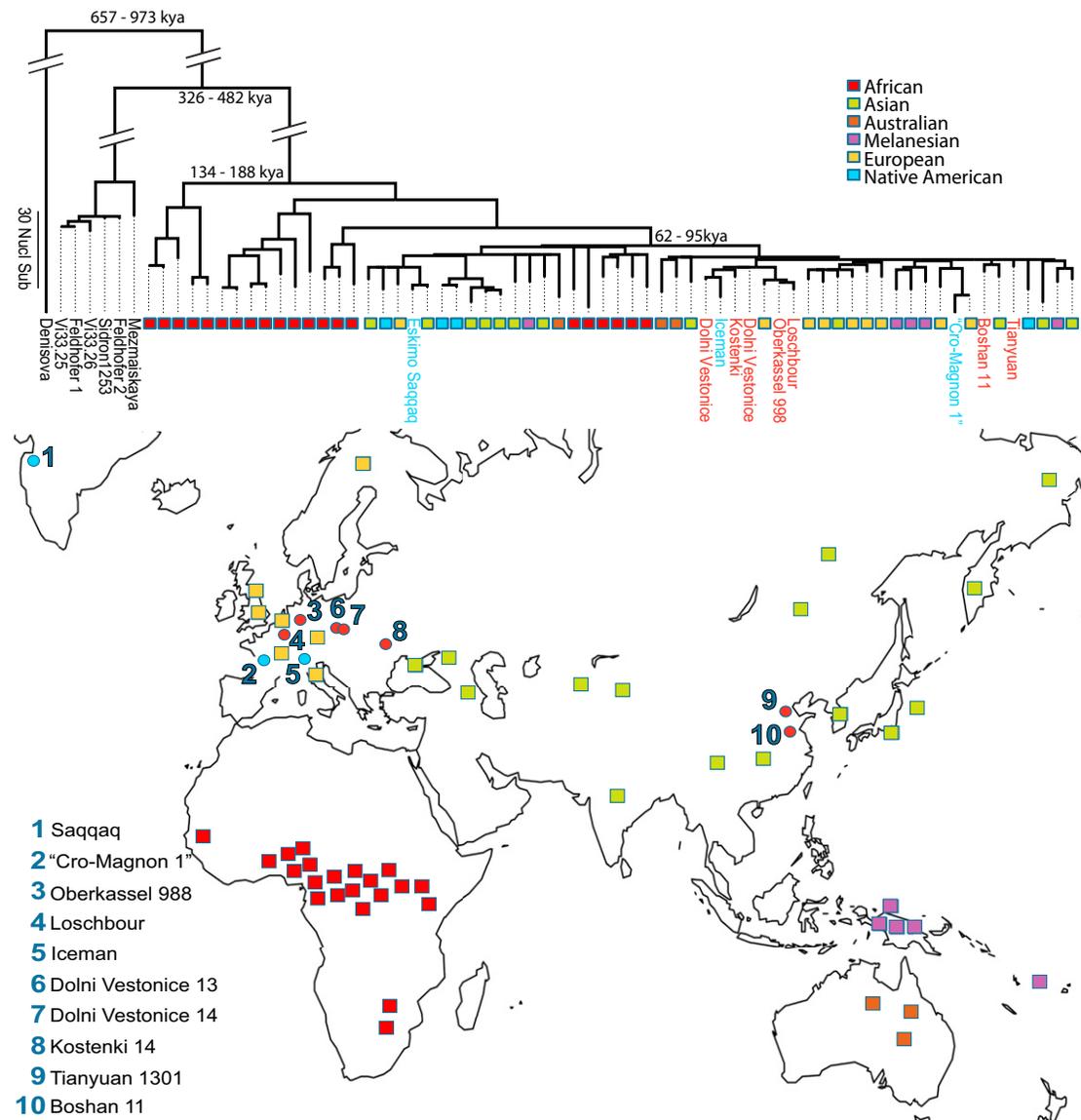
Figure 1. Tree for 54 Present-Day Humans, Ten Ancient Modern Humans, and Seven Archaic Humans

The phylogeny in the top panel was constructed using Maximum Parsimony and rooted using midpoint rooting. The branches for present-day humans do not all end at the same point, giving a sense of the inherent uncertainty in time measurements based on mtDNA due to its limited sequence span. However, the consistent shortening of the branches of ancient humans relative to their closest present-day human relatives is apparent in the figure. This is the basis for our clock calibration. Pre- and post-Neolithic ancient samples are indicated as red and blue circles, respectively, and colored squares indicate the geographical origin of 54 present-day humans that we coanalyzed with them. Date estimates for major divergence events are shown at the nodes. In the bottom panel we show a map giving geographical origin of the samples.

Experimental Procedures

Samples, DNA Extraction, and Molecular Processing

DNA extraction was performed on skeletal remains from 53 humans from Europe. Descriptions for these can be found in the Supplemental Information, section 1, as well as in Tables 1 and S6. Details of DNA extraction, enrichment, and Illumina sequencing are available in the Supplemental Information, sections 2 and 3. The GenBank accession numbers of the mtDNA consensus sequences determined in this study are KC521454 (Boshan 11), KC521455 (Loschbour), KC521456 (Cro-Magnon 1), KC521457 (Oberkassel 998), KC521458 (DolniVestonice 14), and KC521459 (DolniVestonice 13).

Phylogenetic Analysis

The consensus sequences were assigned to haplogroups (Table S2) according to Phylotree.org [24] with a custom PERL script. They were

aligned using the software MUSCLE [25]. MEGA 5 [26] was used to calculate mean pairwise differences and to generate a Maximum Parsimony tree, which included the sequences obtained here along with previously published early modern human mtDNAs with radiocarbon dates, 54 contemporary modern human mtDNAs from a worldwide distribution [27], six Neanderthal mtDNAs, and one Denisovan mtDNA [17, 28] (Figure 1).

Estimation of Substitution Rates

For the direct calculation of the human mtDNA substitution rate we used ten samples. These included six samples (out of a total of 54 ancient modern human remains mentioned above) that were reliably dated and for which complete or nearly complete mtDNA sequences had been generated at a minimum of 3-fold coverage. The remaining four samples came from previously published early modern human mtDNA sequences. The specific sequences we analyzed were Dolni Vestonice 13 and 14, Tianyuan [29], Boshan, Cro-Magnon 1, Oberkassel 998, Kostenki [10], Iceman [30], Saqqaq

[31], and Loschbour (Table 1). Radiocarbon dates were calibrated with Ox-Cal 4.1 with the data set INTCAL09.

The complete mtDNAs were aligned using the software MUSCLE with a worldwide data set of 311 contemporary mtDNAs [25]. Both ancient Italian samples had no reliable 14C date and were therefore not used to estimate the rate of mtDNA substitutions in humans.

To estimate the substitution rate, we used a linear regression model as well as a Bayesian analysis. For the linear regression, nucleotide distances (Figures S1A and S1B) to the common ancestor of all mtDNAs that fall into haplogroup R were calculated for all applicable ancient and modern-day sequences using the software MEGA5. For this analysis, we included all ancient samples that were descendents of the R lineages and for which 99.5% of mitochondrial positions were covered by at least 3 reads; this excluded the Saqqaq individual. The number of substitutions per year was then obtained as the slope of the regression of radiocarbon age and nucleotide distance (Figures S1C and S1D). The obtained rate was divided by the number of positions to calculate the rate per site per year. Although the regression approach is straightforward, it does not make the most efficient use of the available information as it weights all samples equally despite their shared evolutionary history. Our second approach avoids this problem by calculating substitution rates in a Bayesian framework. Using the software package BEAST [32], we explicitly accounted for the shared phylogeny of the 311 mtDNAs and ten ancient human mtDNAs. The general time-reversible sequence substitution model with a fixed fraction of invariable sites and gamma distributed rates (GTR+I+G) was used because this model is the best fit to our data according to Modeltest and PAUP* [33]. In addition, two different models of rate variation among branches were investigated: a strict clock and an uncorrelated lognormal-distributed relaxed clock. For these two models, both a constant population size coalescent and an exponential growth coalescent were used as tree priors. Thus, we analyzed 4 = 2 × 2 models in total. For each model, two MCMC runs were carried out with 30,000,000 iterations each, sampling every 1,000 steps. The first 6,000,000 iterations were discarded as burn-in. For each model both independent runs were combined resulting in 48,000,000 iterations.

To estimate substitution rates for various partitions of the mtDNA, we carried out MCMC runs for all four models on four subsets of the mtDNA alignments. The first corresponded to the whole mtDNA sequence, the second to the coding region (positions 577–16,023), the third to the first and second codon positions of the protein-coding genes, and the fourth to the third codon position.

Archaic Humans

Because we do not know the radiocarbon age for some of the complete Neanderthal mtDNAs and the Denisovan individual, the TMRCA of modern and archaic human mtDNAs was estimated from our inferred substitution rate. For this approach we aligned 54 modern human mtDNAs from a worldwide data set with six Neanderthal mtDNAs, one Denisovan, one chimpanzee, and one bonobo using MUSCLE. A molecular clock test was performed using MEGA5 by comparing the maximum likelihood value for the given topology with and without the molecular clock constraints under the general time reversible model (GTR+I+G). Differences in evolutionary rates among sites were modeled using a discrete Gamma (G) distribution that allowed for invariant (I) sites. The null hypothesis of equal evolutionary rate throughout the tree was not rejected (p = 0.181, Supplemental Information, section 4 and Table S4). Hence, we used the estimated substitution rate to extrapolate the TMRCA of modern and archaic human mtDNAs. All positions containing gaps and missing data were eliminated. There were a total of 16,518 positions in the final data set. To calculate mtDNA divergence times, we used the determined substitution rate computed for the whole mtDNA genome (from the previous section, based on ten ancient mtDNAs) as a prior in the Bayesian software package BEAST.

Accession Numbers

The GenBank accession numbers of the mtDNA consensus sequences determined in this study are KC521454 (Boshan 11), KC521455 (Loschbour), KC521456 (Cro-Magnon 1), KC521457 (Oberkassel 998), KC521458 (Dolni Vestonice 14), and KC521459 (Dolni Vestonice 13).

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures, two figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.02.044.

References

1. Scally, A., and Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. Nat. Rev. Genet. 13, 745–753.

2. Awadalla, P., Gauthier, J., Myers, R.A., Casals, F., Hamdan, F.F., Griffing, A.R., Côté, M., Henrion, E., Spiegelman, D., Tarabeux, J., et al. (2010). Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. Am. J. Hum. Genet. 87, 316–324.

3. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A.; 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073.

4. Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature 488, 471–475.

5. Roach, J.C., Glusman, G., Smit, A.F., Huff, C.D., Hubley, R., Shannon, P.T., Rowen, L., Pant, K.P., Goodman, N., Bamshad, M., et al. (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328, 636–639.

6. Endicott, P., Ho, S.Y., Metspalu, M., and Stringer, C. (2009). Evaluating the mitochondrial timescale of human evolution. Trends Ecol. Evol. 24, 515–521.

7. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., et al. (2003). Natural selection shaped regional mtDNA variation in humans. Proc. Natl. Acad. Sci. USA 100, 171–176.

8. Friedlaender, J., Schurr, T., Gentz, F., Koki, G., Friedlaender, F., Horvat, G., Babb, P., Cerchio, S., Kaestle, F., Schanfield, M., et al. (2005). Expanding Southwest Pacific mitochondrial haplogroups P and Q. Mol. Biol. Evol. 22, 1506–1517.

9. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., and Richards, M.B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. Am. J. Hum. Genet. 84, 740–759.

10. Krause, J., Briggs, A.W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A., and Pääbo, S. (2010). A complete mtDNA genome of an early modern human from Kostenki, Russia. Curr. Biol. 20, 231–236.

11. Green, R.E., Malaspinas, A.S., Krause, J., Briggs, A.W., Johnson, P.L., Uhler, C., Meyer, M., Good, J.M., Maricic, T., Stenzel, U., et al. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell 134, 416–426.

12. Briggs, A.W., Stenzel, U., Johnson, P.L., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., and Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. USA 104, 14616–14621.

13. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS ONE 7, e34131.

14. Stoneking, M., and Krause, J. (2011). Learning about human population history from ancient and modern genomes. Nat. Rev. Genet. 12, 603–614.

15. Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.J., et al. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science *326*, 137–140.

16. Henry-Gambier, D. (2002). Les fossiles de Cro-Magnon (Les Eyzies-de-Tayac, Dordogne): Nouvelles données sur leur Position chronologique et leur attribution culturelle. Bull. Mem. Soc. Anthropol. Paris *14*, 89–112.

17. Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., et al. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science *325*, 318–321.

18. Richards, M.B., Macaulay, V.A., Bandelt, H.J., and Sykes, B.C. (1998). Phylogeography of mitochondrial DNA in western Europe. Ann. Hum. Genet. *62*, 241–260.

19. Fenner, J.N. (2005). Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am. J. Phys. Anthropol. *128*, 415–423.

20. Sun, J.X., Helgason, A., Masson, G., Ebenesersdóttir, S.S., Li, H., Mallick, S., Gnerre, S., Patterson, N., Kong, A., Reich, D., and Stefansson, K. (2012). A direct characterization of human mutation based on microsatellites. Nat. Genet. *44*, 1161–1165.

21. Hobolth, A., Dutheil, J.Y., Hawks, J., Schierup, M.H., and Mailund, T. (2011). Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res. *21*, 349–356.

22. Wood, B., and Harrison, T. (2011). The evolutionary context of the first hominins. Nature *470*, 347–352.

23. Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226.

24. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. *30*, E386–E394.

25. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

26. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. *28*, 2731–2739.

27. Ingman, M., Kaessmann, H., Pääbo, S., and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. Nature *408*, 708–713.

28. Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., and Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature *464*, 894–897.

29. Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc Natl Acad Sci U S A. *110*, 2223–2337.

30. Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M.B., and Rollo, F. (2008). Complete mitochondrial genome sequence of the Tyrolean Iceman. Curr. Biol. *18*, 1687–1693.

31. Gilbert, M.T., Kivisild, T., Grønnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Götherström, A., Campos, P.F., et al. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science *320*, 1787–1789.

32. Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. *7*, 214.

33. Posada, D. (2003). Using MODELTEST and PAUP* to select a model of nucleotide substitution. Curr. Protoc. Bioinformatics, Chapter 6, Unit 6.5.

34. Green, R.E., Briggs, A.W., Krause, J., Prüfer, K., Burbano, H.A., Siebauer, M., Lachmann, M., and Pääbo, S. (2009). The Neandertal genome and ancient DNA authenticity. EMBO J. *28*, 2494–2502.