Current Biology *18*, 1687–1693, November 11, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2008.09.028

Report

Complete Mitochondrial Genome Sequence of the Tyrolean Iceman

Luca Ermini,[1,3,4] Cristina Olivieri,[1,4] Ermanno Rizzi,[2] Giorgio Corti,[2] Raoul Bonnal,[2] Pedro Soares,[3] Stefania Luciani,[1] Isolina Marota,[1] Gianluca De Bellis,[2] Martin B. Richards,[3] and Franco Rollo[1,*]

[1]Laboratorio di Archeo-Antropologia Molecolare/DNA Antico
Dipartimento di Biologia Molecolare, Cellulare e Animale
University of Camerino
62032 Camerino
Italy

[2]Istituto di Tecnologie Biomediche
Consiglio Nazionale delle Ricerche
20090 Milano
Italy

[3]Institute of Integrative and Comparative Biology
Faculty of Biological Sciences
University of Leeds
LS2 9JT Leeds
UK

Summary

The Tyrolean Iceman was a witness to the Neolithic–Copper Age transition in Central Europe 5350–5100 years ago, and his mummified corpse was recovered from an Alpine glacier on the Austro-Italian border in 1991 [1]. Using a mixed sequencing procedure based on PCR amplification and 454 sequencing of pooled amplification products, we have retrieved the first complete mitochondrial-genome sequence of a prehistoric European. We have then compared it with 115 related extant lineages from mitochondrial haplogroup K. We found that the Iceman belonged to a branch of mitochondrial haplogroup K1 that has not yet been identified in modern European populations. This is the oldest complete *Homo sapiens* mtDNA genome generated to date. The results point to the potential significance of complete-ancient-mtDNA studies in addressing questions concerning the genetic history of human populations that the phylogeography of modern lineages is unable to tackle.

Results

One of the world's most celebrated human mummies was discovered on September 19th, 1991, at 3270 m above sea level, in the Eastern Alps near the Austro-Italian border. The age of the remains, 5350–5100 years before present, corresponds to the Late Neolithic or the Copper Age. One of the most striking features of the discovery was the amazing state of preservation of the body and of the clothing and equipment found nearby. This is currently attributed to a spontaneous freeze-drying process. Scientific investigations carried out since the discovery have shown that the man, popularly known as the Tyrolean Iceman or "Ötzi," was approximately 46 years old, and they suggest that he had been severely wounded by an arrow and then finished off with a mace blow to the face [2].

The first genetic investigation of the Iceman's DNA was performed by collaborating teams in Munich and Oxford [3]. By independently analyzing soft-tissue and bone specimens taken from the body, they determined the DNA sequence of the first hypervariable segment (HVS-I) of the mitochondrial DNA (mtDNA) control region. Two nucleotide transitions, at positions 16224 and 16311, according to the reference sequence of Anderson et al. [4], were identified, indicating that the Iceman's mtDNA belonged to haplogroup K [5], a subclade of the major west Eurasian haplogroup U [6].

On September 25th, 2000, at the South Tyrol Archaeological Museum in Bolzano, Italy, the Iceman's body was completely defrosted for the first time, and samples of the intestinal contents were collected under sterile conditions [7]. They were used in animal- and plant-DNA investigations that ultimately led to the reconstruction of the man's last meals [8]. During these analyses, it was noted that, in addition to animal and higher-plant DNA, a relatively large fraction of the DNA from the intestines was of human origin and belonged to the man. This DNA was utilized to provisionally characterize the haplogroup of the mummy by PCR amplification and sequencing of limited portions of the control and coding region [9–11]. It was found that the Iceman's mtDNA lineage belonged to the K1 subhaplogroup but did not fit with any of the known branches into which the group is presently divided [12–15].

Recently developed high-throughput sequencing technologies provide innovative and powerful strategies for the sequencing of entire genomes, and the 454 pyrosequencing technology has recently been used to characterize up to one million base pairs (bp) of the Neanderthal nuclear genome [16, 17] and 13 million bp of nuclear DNA from a 28,000-year-old mammoth bone [18]. More recently, Gilbert et al. [19], using pyrosequencing technology, obtained ten complete mtDNA genomes from well-preserved mammoth hair shafts.

To shed further light on the genetics of the Iceman, we utilized 454 pyrosequencing technology to sequence his entire mitochondrial genome and used the ancient sequence, together with 115 published complete mtDNA sequences from modern individuals, to construct a phylogeny of the K haplogroup.

DNA was extracted from a 70 mg portion of one of the mummy's intestinal content (rectum) specimens collected in 2000, via a phenol-chloroform method previously adapted for work with this kind of material [8]. Purified DNA was PCR amplified with a set of 235 primer-pair systems (Table S1, available online), covering portions of the mtDNA sequence ranging in size from 82 bp to 188 bp. Amplification products were checked by gel electrophoresis, diluted to equal concentrations, then pooled and used as a substrate for the pyrosequencing reaction.

The Complete Sequencing of the Iceman's mtDNA

The GS-FLX genome sequencer (FLX Roche 454 LifeSciences) yielded a total of 45,829 reads. These sequences were further processed with the GS Amplicon Variant Analyzer application (AVA by Roche). This software initially operates by identifying

*Correspondence: francougo.rollo@unicam.it
[4]These two authors contributed equally to this work

Table 1. Transitions Displayed by the Iceman's Sequence Relative to the Revised Cambridge Reference Sequence

Position	Transition
73	A→G
263	A→G
750	A→G
1189	T→C
1438	A→G
1811	A→G
2706	A→G
3480	A→G
3513	C→T
4769	A→G
7028	C→T
8137	C→T
8860	A→G
9055	G→A
9698	T→C
10398	A→G
10550	A→G
11299	T→C
11467	A→G
11719	G→A
12308	A→G
12372	G→A
14167	C→T
14766	C→T
14798	T→C
15326	A→G
16224	T→C
16311	T→C
16362	T→C
16519	T→C

those sequences (reads) showing the PCR primers at their termini and eliminating those lacking them. This initial screening yielded 42,695 sequences (93.2% of the total). Subsequently, the software trims the sequences to remove the portions corresponding to the primers. The resulting sequences were further aligned (Lasergene 7 Software by DNASTAR) into distinct clonal groups, which were assembled together on the basis of the overlapping tracts and their position in relation to the revised Cambridge Reference Sequence (rCRS) [20]. At this point, we produced consensus sequences by applying a filtering threshold of 95%. Ninety-six percent of the nucleotide positions passed the selection. The remaining 4% were manually processed (BioEdit 7.0.9). In this case, the filtering threshold never fell below 71%.

During the assembly step, it was noted that the L2334/H2436, L8323/H8414, and L6244/H6360 groups of reads were missing (Table S2), possibly due to operators' errors in the pooling of the GS-FLX substrates. Therefore, the corresponding PCR products were cloned into vectors and sequenced with conventional Sanger technology. The resulting sequences were checked for homogeneity (95% filtering threshold) and used to fill the gaps.

The alignment of the final assembly generated a consensus sequence. When compared with the rCRS, the consensus sequence showed 30 mtDNA transitions (Table 1). None of them is reported to be in association with any form of disease (www.mitomap.org).

Phylogenetic Analysis of the Iceman's Complete mtDNA Sequence

To test the consistency of the mtDNA assembly and determine the Iceman's phylogenetic position, we performed a reduced-median-network analysis [21] using all 115 haplogroup K complete sequences currently available (Figure 1). This confirmed that the Iceman's sequence falls within haplogroup K, with all of the mutational-motif positions that identify the haplogroup, including coding-region transitions at positions 3480, 9055, 9698, 10550, 11299, 14167, and 14798. This phylogenetic position is confirmed if we refer to the control-region HVS-I motif 16519-16311-16224 [12, 22]. The network further indicates that the Iceman's mtDNA, with transitions at positions 1189 and 10398, belongs to the K1 subhaplogroup but not to any of the three subclades into which K1 is currently further subdivided (K1a, K1b, and K1c).

We checked carefully for the 497, 498, and 5913 mutations, diagnostic of the K1a, K1b, and K1c subhaplogroups, respectively, in the corresponding alignments of reads (L459/H580, L5822/H5936, L5897/H5960, L5882/H5936; Table S2) and found a complete absence of the K1a marker (zero out of 33 reads) and only four out of 613 reads (0.65%) carrying the K1c marker. On the other hand, we found three out of 33 reads (9%) carrying the K1b marker. This relatively high proportion of K1b reads, however, is most likely due to a pyrosequencing artifact [23]. This conclusion is supported by Sanger sequencing across the 498 position with the use of three different PCR systems, all of which consistently failed to evidence the deletion [11].

The Iceman's mtDNA, therefore, seems to belong to a novel branch of K1, defined by transitions at nucleotide positions 3513 and 8137. To assess the reliability of these two transitions, we checked their distribution throughout the total available complete-mtDNA genome database (as of April 2008). Both mutations, synonymous transitions, occur at low frequency throughout the mtDNA phylogeny and are part of the motif that defines some subhaplogroups, reducing the likelihood of their being phantom mutations [24].

The transition at position 3513 is rare but has been seen previously, having occurred twice in published sequences. It forms part of the subhaplogroup L1c1a2 motif [25] and also specifies, together with a transition at position 8607, the subhaplogroup M1a4 [26]. The Iceman's sequence shares only this mutation with both the L1c1a2 motif and M1a4, excluding any possible identification of mosaic haplotypes.

The transition at position 8137 has been found in samples from Pakistan, India, and the Iberian Peninsula. However, all of these samples belong to the major branch of U7 [22, 27, 28]. Therefore, again, the transition is known to have occurred during the evolution of the human mtDNA, and it is highly improbable that its presence in the Iceman's lineage is due to an artifact. As before, the Iceman's sequence does not show any of the other mutations that define U7, ruling out artificial recombination.

The Iceman's sequence also shares a transition at position 16362 with a single sequence (GenBank number: EU073969), belonging to the Family Tree database (www.familytreedna.com), which is the only other paraphyletic K1* complete mtDNA genome found to date: i.e., belonging to K1 but not falling within any of its defined subclades, K1a, K1b, and K1c. Curiously, this transition is not reported in the alignment presented by Handt and colleagues in their 1994 paper [3]. As a possible explanation for this discrepancy, we note that these authors utilized rather long (394 and 221 bp) PCR systems, potentially prone to amplify contaminants and generate chimeras, to cover the 16362 position. Furthermore, for the shorter (221 bp) and potentially more reliable of the two PCR systems employed, they sequenced only seven clones. In our data,

the 16362 transition is confirmed by a total of 335 reads, stemming from two PCR systems (L16303/H16407 and L16347/H16473), of, respectively, 143 and 167 bp in length (Table S4) and, thus, significantly shorter than those employed by Handt et al. Interestingly, among the 190 reads produced by the L16303/H16407 system, 14 (7.3%) correspond to the seven 221 bp clones of the 1994 paper in showing the 16311 mutation but not the 16362 mutation (Table S5).

The transitions at positions 3513 and 8137 establish a previously unidentified mtDNA branch within K1, which we have informally named "Ötzi's branch" or K1ö. Given the hypervariability of np 16362 [29], it would be unwise to classify the Iceman's lineage alongside the extant Family Tree sequence that also carries this variant. Nevertheless, it may share a common ancestor with a subset of lineages in the worldwide mtDNA HVS-I database that include the same three HVS-I variants, at positions 16224, 16311, and 16362. A comparison with 901 haplogroup K sequences from the worldwide database indicated that 11 individuals—from Austria, Germany, the Netherlands, Norway, Scotland, Greece, and Russia—display the same three transitions, with 12 more showing the motif plus further variants. About half of the matching HVS-I types are from Germany and Austria, suggesting a geographical focus in north-central Europe.

The Mummy's DNA Preservation and Contamination Issues

In generating the complete mtDNA sequence from an ancient European, addressing the issue of potential contamination is of paramount importance. Given that a certain amount of contamination seems unavoidable even in specimens that have been excavated with the use of special precautions [30], it has been appreciated since the end of the 1980s that the impact of this phenomenon varies as a function of the degree of preservation of the original DNA [31]. We may add that most of the debate concerning the reliability of claims of identification of ancient *H. sapiens* sequences arose as a consequence of the publication of papers describing the analysis of particularly old specimens, the age and preservation conditions of which render questionable the survival of the original DNA [32, 33].

This is not the case for the Iceman, a mummified human body of relatively modest (5350–5100 years before present) age that, additionally, underwent a spontaneous freeze-desiccation process and was preserved at low temperature both before and after it was recovered [34]. Indeed, several lines of evidence have indicated a relatively good preservation of the mummy's endogenous DNA:

1. Figures of D/L-Asp for the mummy and the organic remains from the equipment are in the range of 0.06–0.08 [8, 35] and, thus, potentially compatible with the survival of the native DNA, according to the studies of Poinar and associates [36].
2. The mummy's intestinal specimens utilized for the present mtDNA study have been shown to contain animal DNA from red deer (*Cervus elaphus*) and ibex (*Capra ibex*), as well as DNA from different kinds of land plants, such as cereals, pines, and ferns. In fact, phylogenetically reliable sequences of mtDNA for *C. elaphus* and *C. ibex* and for plant nuclear 18S ribosomal RNA (18S rRNA), as well as plastidial ribulose bisphosphate carboxylase large subunit (*rbc*L) gene sequences, were previously obtained [8].
3. The two human-mtDNA mutations at positions 16224 and 16311 were first independently detected in Munich and Oxford with the use of muscle and bone specimens from the mummy [3]. This result was further confirmed in 2001 by Mark G. Thomas and colleagues [37] in London and, after more than ten years since the first results, by Rollo and colleagues [9, 11] in Camerino, who used specimens from three different portions of the mummy's intestine. We can further add that due to the transfer of the Laboratorio di Archeo-Antropologia Molecolare to a new building in the summer of 2007, the results obtained in Camerino [9, 11] were, in fact, replicated several times in two different ancient-DNA facilities.

The results of the present research strengthen the idea of a good preservation of the remains:

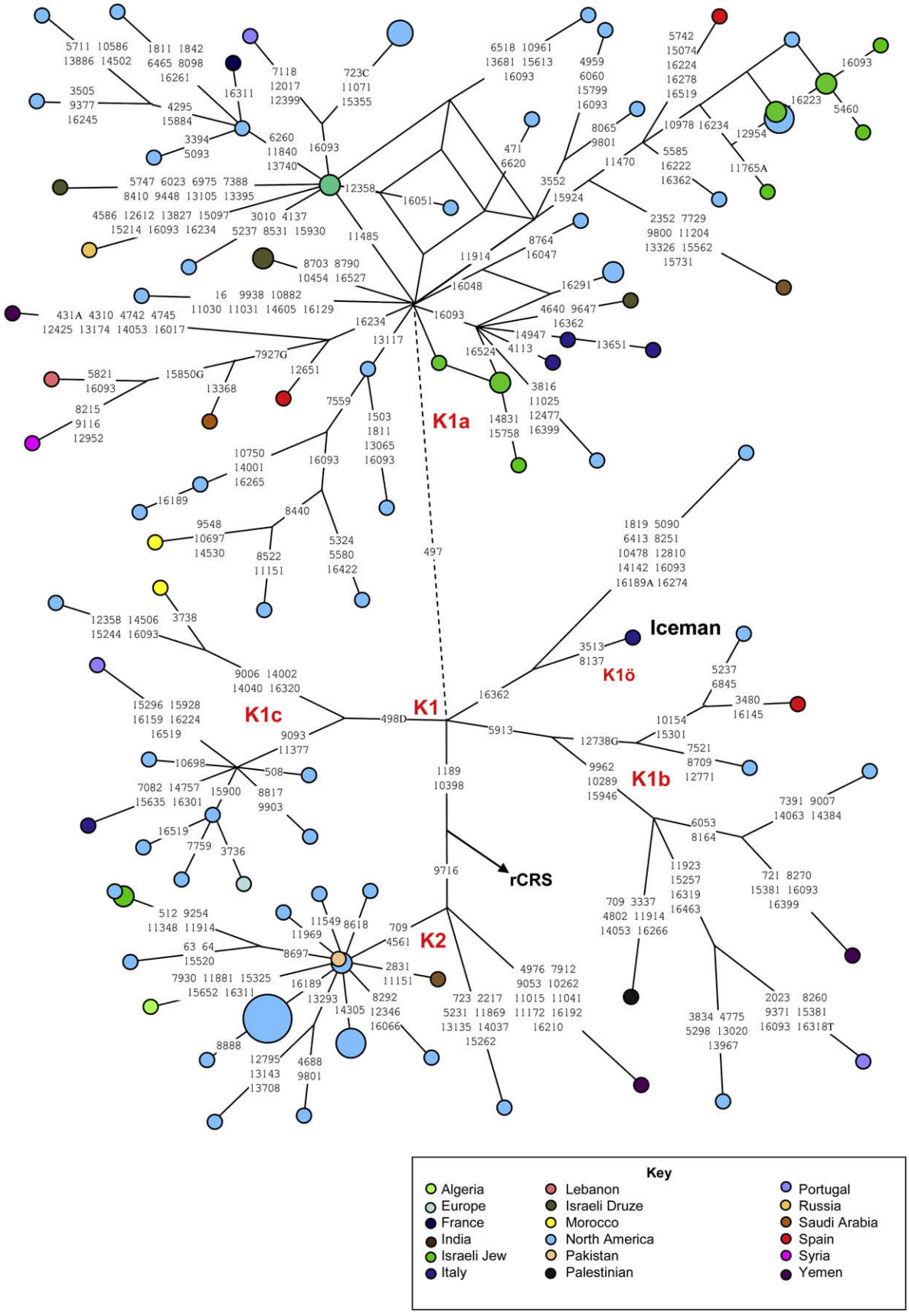
Quantitative PCR assays show a high copy number of the human mtDNA fragments in the intestinal-content specimen. We calculated that, under our reaction conditions, each amplification reaction started from at least 5600 copies of template DNA, as determined for a fragment 162 bp in length. This figure corresponds to approximately 112×10^6 copies per g of sample and is close to recently reported figures for mtDNA copy number in particularly well-preserved Siberian mammoth bones [18].

Although all the premises indicate that, in the case of the present research, contamination is probably not a problem, we have scrutinized the results to test the reliability of the sequence assemblage and identify possible contaminants. Details are given in the Contamination Analysis paragraph of the Supplemental Data.

Discussion

The discovery of a previously unidentified mtDNA lineage in the Iceman that does not fall within the principal extant subclades of haplogroup K1 (indicated by 115 modern complete haplogroup K genomes) provides additional support for the authenticity of the sequence, given that there are sufficient data to indicate that if such a sequence were still extant in the modern gene pool, it must at least be quite rare and, therefore, unlikely to appear as a contaminant. Indeed, only one other K1* complete genome sequence has been found to date, among a database totalling 2698 complete mtDNA sequences, suggesting a modern frequency of < 1% of haplogroup K lineages, or < 0.05% of modern European lineages as a whole. This is further confirmed by the frequency of K lineages with the 16362 transition in the HVS-I database (doubtless an overestimate for lineages related to the Iceman as a result of recurrences of the fast-evolving 16362 mutation), all of which amount to < 3% of K HVS-I lineages.

This is the oldest complete *H. sapiens* mtDNA genome generated to date. Our mixed sequencing procedure, based on PCR amplification and 454 sequencing of pooled amplification products, is potentially applicable to many other ancient human remains. Indeed, Gilbert et al. [38] recently used a related pyrosequencing approach to describe the analysis of the complete mtDNA genome from a Paleo-Eskimo dated to 3400–4500 years ago. They showed that this individual's mtDNA lineage, belonging to haplogroup D2a1, was distinct from that of both modern Native Americans and Neo-Eskimos. Similarly, the Iceman's mtDNA sequence belongs to a branch of haplogroup K so far unidentified in modern European populations, although whether this is due to changes in central


K1a
K1
K1c
K1ö
K1b
K2
Iceman
rCRS
497
498D
5711 10586 13886 14502
1811 1842 6465 8098 16261
7118 12017 12399
723C 11071 15355
3505 9377 16245
4295 15884
16311
3394 5093
6260 11840 13740
16093
12358
16051
5747 6023 6975 7388 8410 9448 13105 13395
4586 12612 13827 15097 15214 16093 16234
3010 4137
5237 8531 15930
11485
6518 10961 13681 15613 16093
4959 6060 15799 16093
471 6620
8065 9801
3552
15924
11470
5742 15074 16224 16278 16519
16093
16223
5460
12954
10978 16234
5585 16222 16362
11765A
2352 7729 9800 11204 13326 15562 15731
8764 16047
11914
16048
8703 8790 10454 16527
16 9938 10882 11030 11031 14605 16129
431A 4310 4742 4745 12425 13174 14053 16017
16234
13117
16093
16291
4640 9647 16362
14947
4113
13651
16524
3816 11025 12477 16399
14831 15758
7927G
12651
13368
5821 16093
15850G
8215 9116 12952
7559
1503 1811 13065 16093
10750 14001 16265
16189
8440
9548 10697 14530
8522 11151
5324 5580 16422
1819 5090 6413 8251 10478 12810 14142 16093 16189A 16274
3513 8137
5237 6845
3480 16145
10154 15301
7521 8709 12771
12738G
9962 10289 15946
5913
1189 10398
9716
3738
12358 14506 15244 16093
9006 14002 14040 16320
15296 15928 16159 16224 16519
9093
11377
10698
508
7082 14757 15635 16301
15900
8817 9903
16519
7759
3736
512 9254 11348 11914
63 64 15520
11549 8618 11969
8697
709 4561
2831 11151
7930 11881 15325 15652 16311
16189
13293
14305
8292 12346 16066
8888
12795 13143 13708
4688 9801
723 2217 5231 11869 13135 14037 15262
4976 7912 9053 10262 11015 11041 11172 16192 16210
6053 8164
7391 9007 14063 14384
11923 15257 16319 16463
721 8270 15381 16093 16399
709 3337 4802 11914 14053 16266
3834 4775 5298 13020 13967
2023 8260 9371 15381 16093 16318T
Key
Algeria
Europe
France
India
Israeli Jew
Italy
Lebanon
Israeli Druze
Morocco
North America
Pakistan
Palestinian
Portugal
Russia
Saudi Arabia
Spain
Syria
Yemen

European settlement patterns since the late Neolithic era, random genetic drift, or simply insufficient sampling of modern populations is not yet clear. Nevertheless, both sets of results point to the potential significance of complete-ancient-mtDNA studies in addressing questions concerning the genetic history of human populations that the phylogeography of modern lineages is unable to tackle. Similar considerations on the importance of phylogenies based on complete mtDNA data sets can also be made for animal-domestication studies [39]. Further investigation by means of complete sequencing of modern and ancient central European mtDNA genomes should, in the future, allow researchers to pin down the time and place of the origin of the maternal lineage of the Iceman with considerable precision.

Experimental Procedures

Ancient-DNA Work

We performed all manipulations of ancient-DNA samples at the Laboratorio di Archeo-Antropologia Molecolare/DNA Antico of the Camerino University, in a dedicated facility. This facility is composed of an antechamber—in which the operator wears a full-body sterile suit, gloves, a face screen, and a breathing mask—and a laboratory. Both environments are equipped with UV lights and a positive-pressure air-filtering system providing 99.97% particle elimination and a complete change of air every 10 min. We frequently cleaned all surfaces with bleach. We typed all of the operators for mtDNA: they belonged to haplogroups J, U5, V, and H.

Sample Collection and DNA Extraction

On September 25th, 2000, at the South Tyrol Archaeological Museum in Bolzano, Italy, after the first complete defrosting of the mummy, three specimens of intestinal content, corresponding to ileum (68 mg), colon (58 mg), and rectum (181 mg), were taken under sterile conditions. The specimens were then taken to Camerino and kept at −25°C until use. Through the years, the whole of the ileum and the colon samples and part of the rectum samples (111 mg) were used for DNA investigations [8, 9].

For the present investigation, we extracted DNA from the remaining rectum sample (70 mg). We resuspended the samples in 350 μl of 50 mM Na_2EDTA, 50 mM Tris-HCl (pH 8.0), 1% (weight/volume) sodium dodecyl sulfate (SDS), and 6% (volume/volume) water-saturated phenol. After imbibition (soaking), samples were left overnight at 4°C. We then transferred them into sterile mortars and homogenized them with sterile pestles. During the milling phase, 350 μl of the Tris–EDTA was added to each sample. We collected the homogenates in Eppendorf tubes, taking care to rinse the mortar and pestle with a further 350 μl of extraction medium. Then we extracted the homogenates sequentially with equal volumes of phenol, phenol/chloroform/isoamyl alcohol (25:24:1), and ether. We added to the solution 1:10 volume of 2 M sodium acetate and 2.5 volumes of cold (−20°C) ethanol, and the DNA fraction was precipitated from the final supernatant by centrifugation at 13,500 g for 5 min.

We resuspended the DNA precipitates in 20 μl of sterile distilled water and stored them at −80°C until use. We prepared extraction blanks throughout the procedure.

Primer Design and PCR Amplification

To set up PCR systems for the whole mtDNA genome, we aligned complete mtDNA sequences from nine different European haplogroups (H, V, U, K1, K2, X, W, J, and T). We selected primers located in the most conservative regions in order to keep the number of polymorphic positions in the oligonucleotides to a minimum and we designed them to produce substantially overlapping PCR fragments. The PCR systems were directly tested on the ancient-DNA template, and no positive (i.e., modern DNA) control was used, to minimize the risk of contamination.

We performed DNA amplifications in 50 μl of reaction mix of the following composition: 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 2.5 enzyme units of Taq polymerase (HotStarTaq DNA polymerase, QIAGEN), 200 mM each dNTP, 300 ng each primer, and 1 μl of DNA template (diluted 1:70 to reduce the effect of Taq polymerase inhibitors). We pretreated the reaction mixture with DNase (2 enzyme units for 30 min at room temperature) to eliminate any contaminant DNA and subsequently inactivated the DNase by heating to 95°C for 15 min. The thermal profile was as follows: 1 min at 94°C, 30 s at the relevant annealing temperature, and 1 min at 72°C, with a final extension of 10 min at 72°C. The number of cycles ranged from 40 to 55. To reduce the extent of errors generated by the reaction, we never reamplified a PCR product; only PCR products obtained after a first amplification round were used for the subsequent pyrosequencing step. We set up negative controls throughout the whole procedure, and we performed more than 800 amplifications in total. We checked the amplification products by electrophoresis with 2.5% (w/v) agarose, then purified them by using the High Pure PCR Product purification kit (Roche Molecular Biochemicals) and stored them at −80°C until the subsequent pyrosequencing.

The list of oligonucleotide primer pairs utilized, with the corresponding annealing temperatures, is given in Table S1.

Quantitative PCR

We determined the number of amplifiable mtDNA fragments in the ancient DNA preparation with real-time PCR (Stratagene), using the L16209/H16311 primer-pair system. The primers bind to a 162 bp fragment of HVS-I. We performed the reactions under the conditions described above, using Brilliant SYBR Green QPCR Master Mix in the Mx3000P qPCR system (Stratagene).

We generated a standard curve (correlation coefficient, $R^2 = 0.987$; amplification efficiency, Eff = 104%) by amplifying a serial dilution of a plasmid containing the template sequence (Figure S1).

454 Sequencing

We checked amplification products by gel electrophoresis, diluted them to equal concentrations, then pooled them in equimolar proportion and used them as a substrate for the pyrosequencing reaction. We further spectrophotometrically quantitated and purified the mixture (MinElute PCR Purification Kit, QIAGEN, Hilden, Germany), and we checked for concentration and fragment length using the Agilent Bioanalyzer. Because of the short size of the template, we did not perform any DNA fragmentation as suggested by the Roche GS FLX library preparation protocol for modern DNA. We polished the fragments, using T4 DNA polymerase to generate blunt ends. We then ligated each DNA fragment to specific adaptors, and a double-stranded DNA library was obtained. This library was bound to Sepharose beads for isolation of the ligation products with the correct combination of adaptors. A single-stranded template DNA (sstDNA) library was then obtained. We assessed the quality and quantity of the sstDNA library with the Agilent 2100 Bioanalyzer and the RiboGreen RNA Quantitation Kit (Invitrogen, Carlsbad, CA). In order to obtain a 1 molecule:1 bead ratio, we diluted the sstDNA library, and each sstDNA-library fragment was bound onto a single DNA capture bead and clonally amplified by emulsion PCR performed in microvesicles obtained by a water-in-oil emulsion. Emulsion PCR generated for each bead a clonal population of molecules that were positively selected by an enrichment step that eliminates null beads. We disrupted the emulsion using propanol, ethanol, and buffers provided by 454-Roche emPCR kit, and we recovered beads with an amplified sstDNA fragment for the subsequent pyrosequencing step. We annealed these positive beads with sequencing primer and counted them using the Multisizer 3 Coulter Counter (Beckman Coulter, Fullerton, CA, USA). Finally, we loaded the beads onto a 70 × 75 mm PicoTiterPlate and, using a FLX Genome Sequencer (FLX Roche 454 LifeSciences), we sequenced the ssDNA fragments.

To assemble and analyze the resulting sequence reads, we used the GS Amplicon Variant Analyzer application (AVA) by Roche and Lasergene 7 Software by DNASTAR [40]. To produce consensus sequences, we applied

Figure 1. Reduced-Median Network of Haplogroup K, Based on 117 Complete mtDNA Sequences, Rooted with the rCRS Sequence

The HVS–II region (nucleotide positions 73–340) was excluded from the network construction for the achievement of a readable network with minimal reticulations. The numbers along the links refer to nucleotide positions and label variants from the rCRS. All nucleotide substitutions are transitions, unless a base change is explicitly indicated, and the suffix D indicates a deletion. The size of each node is proportional to the number of samples. Branch lengths have been distorted to increase legibility. In particular, the broken line representing the 497 link is greatly extended for graphical reasons. The source of the samples is color coded according to the key.

a 95% threshold. A small number of sequences (4%) falling under this threshold were manually processed with BioEdit 7.0.9 [41].

Cloning and Sanger Sequencing

We used both cloning and Sanger sequencing to confirm some mutations found by pyrosequencing and to fill in the pyrosequencing gaps. We purified the PCR products and directly cloned them into plasmids using the pGEM-T Easy Vector System (Promega, Madison, WI), and we isolated recombinant plasmids using a Miniprep kit (Promega, Madison, WI). We assessed insert size and DNA concentration using gel electrophoresis. We performed the DNA sequencing in part with an ABI-Prism 310 automated DNA sequencer and in part with the use of the BMR-Genomics sequencing service (University of Padua). We checked and analyzed all sequences by Sequence Scanner version 1.0 (Applied Biosystems, Foster City, CA), and we aligned them with the program BioEdit v.7.0.9 [41].

Phylogenetic Analysis

We constructed a phylogenetic network of haplogroup K using the Network 4.500 package (http://www.fluxus-technology.com), employing the reduced-median algorithm [21]. We used all haplogroup K complete sequences available in GenBank (115; see Table S6) and included the Iceman's sequence, as well as the rCRS (haplogroup H2) as an outgroup, giving a total of 117 sequences [12, 20, 22, 27, 28, 42–50]. We excluded sequences from Finnilä et al. [13], because these sequences were produced by conformation-sensitive gel electrophoresis (CSGE) and are, therefore, likely to be more error prone than those obtained from DNA sequencing.

To achieve an easily readable network, we excluded the second hypervariable segment (HVS-II, position 73–340) and we downweighted six of the ten fastest hotspot positions [51] (nucleotide positions 16093, 16129, 16189, 16311, 16362 within HVS-I, and 16519) to a fifth and one further to a half (16234). We ignored insertions and deletions (except for the 498 deletion, diagnostic of subhaplogroup K1c) for the network construction.

Supplemental Data

Supplemental Data include Supplemental Results and Discussion, two figures, and six tables and can be found with this article online at http://www.current-biology.com/supplemental/S0960-9822(08)01254-2.

Acknowledgements

We thank Donata Luiselli of the University of Bologna for providing mtDNA typing for the personnel of the Laboratorio di Archeo-Antropologia Molecolare/DNA Antico; Eduard Egarter Vigl of the Regional Hospital of Bolzano and the South Tyrol Archaeological Museum, Bolzano, for providing the mummy specimens; and three anonymous reviewers for their useful comments and suggestions. This research was supported by an Eli Lilly donation to F.R.

Received: July 7, 2008
Revised: August 22, 2008
Accepted: September 3, 2008
Published online: October 30, 2008

References

1. Spindler, K. (1996). Iceman's last weeks. In Human mummies, *Volume 3*, K. Spindler, H. Wilfing, E. Rastbichler-Zissernig, D. Zur Nedden, and H. Nothdurfter, eds. (New York: Springer-Verlag Wein), pp. 249–263.
2. Von Lippert, A., Gostner, P., Egarter-Vigl, E., and Pernter, P. (2007). Vom leben und Sterben des Otztaler Gletschermannes. Germania *85*, 1–21.
3. Handt, O., Richards, M., Trommsdorff, M., Kilger, C., Simanainen, J., Georgiev, O., Bauer, K., Stone, A., Hedges, R., Schaffner, W., et al. (1994). Molecular genetic analyses of the Tyrolean Ice Man. Science *264*, 1775–1778.
4. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., et al. (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457–465.
5. Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L., and Wallace, D.C. (1996). Classification of European mtDNAs from an analysis of three European populations. Genetics *144*, 1835–1850.
6. Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B., and Torroni, A. (1999). The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. Am. J. Hum. Genet. *64*, 232–249.
7. Stone, R. (2000). Ice Man Warms Up for European Scientists. Science *289*, 2253–2254.
8. Rollo, F., Ubaldi, M., Ermini, L., and Marota, I. (2002). Otzi's last meals: DNA analysis of the intestinal content of the Neolithic glacier mummy from the Alps. Proc. Natl. Acad. Sci. USA *99*, 12594–12599.
9. Rollo, F., Ermini, L., Luciani, S., Marota, I., Olivieri, C., and Luiselli, D. (2006). Fine characterization of the Iceman's mtDNA haplogroup. Am. J. Phys. Anthropol. *130*, 557–564.
10. Endicott, P., Sanchez, J.J., Metspalu, E., Behar, D.M., and Kivisild, T. (2007). The unresolved location of Otzi's mtDNA within haplogroup K. Am. J. Phys. Anthropol. *132*, 590–591.
11. Rollo, F., Ermini, L., Luciani, S., Marota, I., Olivieri, C., and Luiselli, D. (2007). The Resolved Location of Ötzi's mtDNA Within Haplogroup K: A Reply to Endicott et al. Am. J. Phys. Anthropol. *132*, 590–593.
12. Behar, D.M., Metspalu, E., Kivisild, T., Achilli, A., Hadid, Y., Tzur, S., Pereira, L., Amorim, A., Quintana-Murci, L., Majamaa, K., et al. (2006). The matrilineal ancestry of Ashkenazi Jewry: Portrait of a recent founder event. Am. J. Hum. Genet. *78*, 487–497.
13. Finnila, S., Lehtonen, M.S., and Majamaa, K. (2001). Phylogenetic network for European mtDNA. Am. J. Hum. Genet. *68*, 1475–1484.
14. Herrnstadt, C., Elson, J.L., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C., Ghosh, S.S., Olefsky, J.M., Beal, M.F., Davis, R.E., and Howell, N. (2002). Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am. J. Hum. Genet. *70*, 1152–1171.
15. Rieder, M.J., Taylor, S.L., Tobe, V.O., and Nickerson, D.A. (1998). Automating the identification of DNA variations using quality-based fluorescence re-sequencing: Analysis of the human mitochondrial genome. Nucleic Acids Res. *26*, 967–973.
16. Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M., and Paabo, S. (2006). Analysis of one million base pairs of Neanderthal DNA. Nature *444*, 330–336.
17. Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Paabo, S., Pritchard, J.K., and Rubin, E.M. (2006). Sequencing and analysis of Neanderthal genomic DNA. Science *314*, 1113–1118.
18. Poinar, H.N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R.D., Buigues, B., Tikhonov, A., Huson, D.H., Tomsho, L.P., Auch, A., et al. (2006). Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. Science *311*, 392–394.
19. Gilbert, M.T., Tomsho, L.P., Rendulic, S., Packard, M., Drautz, D.I., Sher, A., Tikhonov, A., Dalen, L., Kuznetsova, T., Kosintsev, P., et al. (2007). Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. Science *317*, 1927–1930.
20. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. *23*, 147.
21. Bandelt, H.J., Forster, P., Sykes, B.C., and Richards, M.B. (1995). Mitochondrial portraits of human populations using median networks. Genetics *141*, 743–753.
22. Palanichamy, M.G., Sun, C., Agrawal, S., Bandelt, H.J., Kong, Q.P., Khan, F., Wang, C.Y., Chaudhuri, T.K., Palla, V., and Zhang, Y.P. (2004). Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: Implications for the peopling of South Asia. Am. J. Hum. Genet. *75*, 966–978.
23. Meyer, M., Stenzel, U., Myles, S., Prufer, K., and Hofreiter, M. (2007). Targeted high-throughput sequencing of tagged nucleic acid samples. Nucleic Acids Res. *35*, e97.
24. Bandelt, H.J., Lahermo, P., Richards, M., and Macaulay, V. (2001). Detecting errors in mtDNA data by phylogenetic analysis. Int. J. Legal Med. *115*, 64–69.
25. Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., Sica, L., Mouguiama-Daouda, P., Comas, D., Tzur, S., et al. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. Proc. Natl. Acad. Sci. USA *105*, 1596–1601.
26. Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., Scozzari, R., Cruciani, F., Behar, D.M., Dugoujon, J.M., et al. (2006). The

mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science *314*, 1767–1770.
27. Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S., Magri, C., Scozzari, R., Babudri, N., Santachiara-Benerecetti, A.S., et al. (2005). Saami and Berbers–an unexpected mitochondrial DNA link. Am. J. Hum. Genet. *76*, 883–886.
28. Maca-Meyer, N., Gonzalez, A.M., Larruga, J.M., Flores, C., and Cabrera, V.M. (2001). Major genomic mitochondrial lineages delineate early human expansions. BMC Genet. *2*, 13.
29. Bandelt, H.J., Kong, Q.P., Richards, M., and Macaulay, V. (2006). Estimation of mutation rates and coalescence times: Some caveats. In Human mitochondrial DNA and evolution of Homo sapiens, H.J. Bandelt, V. Macaulay, and M. Richards, eds. (Berlin: Springer-Verlag), pp. 227–268.
30. Melchior, L., Kivisild, T., Lynnerup, N., and Dissing, J. (2008). Evidence of authentic DNA from Danish Viking Age skeletons untouched by humans for 1,000 years. PLoS ONE *3*, e2214.
31. Pääbo, S. (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. USA *86*, 1939–1943.
32. Caramelli, D., Lalueza-Fox, C., Vernesi, C., Lari, M., Casoli, A., Mallegni, F., Chiarelli, B., Dupanloup, I., Bertranpetit, J., Barbujani, G., and Bertorelle, G. (2003). Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. Proc. Natl. Acad. Sci. USA *100*, 6593–6597.
33. Bandelt, H.J. (2005). Mosaics of ancient mitochondrial DNA: Positive indicators of nonauthenticity. Eur. J. Hum. Genet. *13*, 1106–1112.
34. Gaber, O., Kunzel, K.H., Maurer, H., and Platzer, W. (1992). Konservierung und Lagerung der Gletschermumie. In Der Mann Im Eis, *Volume 1*, W.P. Frank Höpfel and Konrad Spindler, eds. (Innsbruck, Austria: Universität Innsbruck), pp. 92–99.
35. Cano, R.J., Tiefenbrunner, F., Ubaldi, M., Del Cueto, C., Luciani, S., Cox, T., Orkand, P., Kunzel, K.H., and Rollo, F. (2000). Sequence analysis of bacterial DNA in the colon and stomach of the Tyrolean Iceman. Am. J. Phys. Anthropol. *112*, 297–309.
36. Poinar, H.N., Hoss, M., Bada, J.L., and Pääbo, S. (1996). Amino acid racemization and the preservation of ancient DNA. Science *272*, 864–866.
37. Thomas, M.G., Jones, A.L., Osman, F., Weale, M.E., Egarter-Vigl, E., and Pramstaller, P.P. (2001). Y-chromosome variation in the Iceman and modern Europeans. In The Iceman, Sud-Tyrol Archaeological Museum: Bolzano.
38. Gilbert, M.T., Kivisild, T., Gronnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Gotherstrom, A., Campos, P.F., et al. (2008). Paleo-Eskimo mtDNA Genome Reveals Matrilineal Discontinuity in Greenland. Science *320*, 1787–1789.
39. Achilli, A., Olivieri, A., Pellecchia, M., Uboldi, C., Colli, L., Al-Zahery, N., Accetturo, M., Pala, M., Kashani, B.H., Perego, U.A., et al. (2008). Mitochondrial genomes of extinct aurochs survive in domestic cattle. Curr. Biol. *18*, R157–R158.
40. Burland, T.G. (2000). DNASTAR's Lasergene sequence analysis software. Methods Mol. Biol. *132*, 71–91.
41. Hall, T. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. *41*, 95–98.
42. Coble, M.D., Just, R.S., O'Callaghan, J.E., Letmanyi, I.H., Peterson, C.T., Irwin, J.A., and Parsons, T.J. (2004). Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int. J. Legal Med. *118*, 137–146.
43. Fraumene, C., Belle, E.M., Castri, L., Sanna, S., Mancosu, G., Cosso, M., Marras, F., Barbujani, G., Pirastu, M., and Angius, A. (2006). High resolution analysis and phylogenetic network construction using complete mtDNA sequences in sardinian genetic isolates. Mol. Biol. Evol. *23*, 2101–2111.
44. Gonzalez, A.M., Garcia, O., Larruga, J.M., and Cabrera, V.M. (2006). The mitochondrial lineage U8a reveals a Paleolithic settlement in the Basque country. BMC Genomics *7*, 124.
45. Greenspan, B. (2007). Direct Submission, NCBI http://www.ncbi.nlm.nih.gov/.
46. Hartmann, A., Thieme, M., Nanduri, L.K., Stempfl, T., Moehle, C., Kivisild, T., and Oefner, P.J. (2008). Validation of Microarray-Based Sequencing of 93 Worldwide Mitochondrial Genomes. Hum Mutation, in press (NCBI http://www.ncbi.nlm.nih.gov/).
47. Hurst, W. (2006). Direct Submission, NCBI http://www.ncbi.nlm.nih.gov/.
48. Just, R.S., Diegoli, T.M., Saunier, J.L., Irwin, J.I., and Parsons, T.J. (2008). Complete mitochondrial genome sequences for 265 African American and U.S. 'Hispanic' individuals. Forensic Sci Int Genet *2*, e45–e48.
49. 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.
50. Pereira, L., Goncalves, J., Franco-Duarte, R., Silva, J., Rocha, T., Arnold, C., Richards, M., and Macaulay, V. (2007). No evidence for an mtDNA role in sperm motility: Data from complete sequencing of asthenozoospermic males. Mol. Biol. Evol. *24*, 868–874.
51. Salas, A., Bandelt, H.J., Macaulay, V., and Richards, M.B. (2007). Phylogeographic investigations: The role of trees in forensic genetics. Forensic Sci. Int. *168*, 1–13.