Am. J. Hum. Genet. 59:368–376, 1996

The Retrieval of Ancient Human DNA Sequences

Oliva Handt,[1] Matthias Krings,[1] R. H. Ward,[2] and Svante Pääbo[1]

[1]Zoological Institute, University of Munich, Munich; and [2]Eccles Institute of Human Genetics, University of Utah, Salt Lake City

Summary

DNA was extracted from ~600-year-old human remains found at an archaeological site in the southwestern United States, and mtDNA fragments were amplified by PCR. When these fragments were sequenced directly, multiple sequences seemed to be present. From three representative individuals, DNA fragments of different lengths were quantified and short overlapping amplification products cloned. When amplifications started from <40 molecules, clones contained several different sequences. In contrast, when they were initiated by a few thousand molecules, unambiguous and reproducible results were achieved. These results show that more experimental work than is often applied is necessary to ensure that DNA sequences amplified from ancient human remains are authentic. In particular, quantitation of the numbers of amplifiable molecules is a useful tool to determine the role of contaminating contemporary molecules and PCR errors in amplifications from ancient DNA.

Introduction

In general, inferences about population history are drawn from studies of genetic diversity in contemporary populations. However, the retrieval of ancient DNA from archaeological remains holds the promise to add a temporal component to such studies. With the invention of PCR (Saiki et al. 1985; Mullis and Faloona 1987), it became possible, in principle, to study ancient DNA sequences (Pääbo et al. 1989). However, most of the successes have come from the elucidation of phylogenetic relationships of extinct animals (e.g., Thomas et al. 1989; Cooper et al. 1992; Höss et al. 1996*a*), whereas results in other areas have remained controversial (Sidow et al. 1991; Lindahl 1993) or difficult to authenticate. This is the situation, for example, for the retrieval of DNA sequences from ancient human remains (Richards et al. 1995).

We have analyzed a number of mummified human remains from the southwestern United States in order to elucidate the extent to which they yield endogenous, ancient DNA that can be amplified by PCR. The results show that, even when amplification products are achieved, DNA sequences often cannot be reconstructed reliably. Various experimental approaches to remedy this problem are suggested.

Material and Methods

Samples

Ten mummies from the Ventana Cave (VC), Castle Mountains, Arizona, were investigated (Arizona State Museum, Tucson, burial numbers 3 [0-495], 5, 6 [0-513], 9 [0-499], 11 [0-497], 15A, 16, 24 [0-484], 25, 29). They are attributed to a Hohokam cultural horizon (A.D. 1000–1400) on the basis of archaeological context. From each individual, three to four independent tissue samples were removed, when possible from internal parts of the bodies.

Prevention of Contamination

DNA extractions and setting up of PCR reactions were carried out in a laboratory dedicated to ancient DNA work in which disposable lab coats, face shields, and gloves were used by all personnel. Benches and equipment were frequently treated with bleach and the work areas irradiated by UV light for several hours nightly. Two sets of pipettes, one for PCR reagents and one for DNA extracts, were used with filter tips. All PCR reagents were separated into aliquots. Wherever possible, disposable laboratory ware was used. Nondisposable vessel or glassware was treated with 1 N HCl and rinsed with double-distilled (dd) water before use.

DNA Extraction

Twenty to 70 mg of mummified tissue were cut into small pieces with a sterile scalpel, and nucleic acids were isolated. In initial experiments, proteinase K digestion followed by phenol:chloroform extraction was used. However, since these extracts inhibited the *Taq* polymerase, a silica-based protocol was applied (Höss and Pääbo 1993). In brief, samples were incubated in 1,000 μl 5 M guanidinium thiocyanate, 0.1 M Tris HCl (pH

Received December 14, 1995; accepted for publication May 24, 1996.

Address for correspondence and reprints: Dr. Svante Pääbo, Zoological Institute, University of Munich, P.O. Box 202136, D-80021 Munich, Germany. E-mail: paabo@zi.biologie.uni-muenchen.de

0002-9297/96/5902-0012$02.00

Table 1

Primer Sequences and Lengths of the Amplification Products (Including Primers)

	Primer	Sequence	Length (bp)
Control region	L16055	5′-GAAGCAGATTTGGGTACCAC-3′	123
	H16139	5′-TACTACAGGTGGTCAAGTAT-3′	
Control region	L16131	5′-CACCATGAATATTGTACGGT-3′	126
	H16218	5′-TGTGTGATAGTTGAGGGTTG-3′	
Control region	L16209	5′-CCATGCTTACAAGCAAGT-3′	131
	H16303	5′-TGGCTTTATGTACTATGTAC-3′	
Control region	L16287	5′-CACTAGGATACCAACAAACC-3′	108
	H16356	5′-GTCATCCATGGGGACGAGAA-3′	
Control region	L16347	5′-CGTACATAGCACATTACAGT-3′	102
	H16410	5′-GCGGGATATTGATTTCACGG-3′	
9-bp deletion	L8196	5′-ACAGTTTCATGCCCATCGTC-3′	112[a]
	H8297	5′-ATGCTAAGTTAGCTTTACAG-3′	
*Hae*III site	L635	5′-TGAAAATGTTTAGACGGCCTCACATC-3′	89
	H677	5′-GCATGTGTAATCTTACTAAGAG-3′	
*Alu*I site	L5127	5′-ACTACCGCATTCCTACTACTCA-3′	106
	H5189	5′-GGGTGGATGGAATTAAGGGTGT-3′	
*Hin*cII site	L13257	5′-AATCGTAGCCTTCTCCACTTCA-3′	75
	H13290	5′-GCTAGGTGTGGTTGGTTGATG-3′	

NOTE.—The numbers give the 3′-ends of the primers according to Anderson et al. (1981). L and H refer to the light and the heavy strands, respectively.

[a] The size of the product when no deletion is present is 121 bp.

7.4), 0.02 M EDTA, and 1.3% Triton X-100 overnight at room temperature. Subsequently, solid tissue remains were pelleted by centrifugation, supernatants were added to 40 µl silica suspension, and nucleic acids were isolated and eluted in two aliquots of 65 µl ddH_2O at 56°C and stored at −20°C. Two blank extractions containing all reagents but no tissue were included in every set of extractions.

PCR

Hot-start PCR was carried out such that, before the initial denaturation, a barrier of wax (BDH) separated a lower reaction mixture of 20 µl (67 mM Tris [pH 8.8], 2 mM $MgCl_2$, 80 µg BSA, 0.5 µM of each primer, 0.14 mM each of dATP, dCTP, dGTP, dUTP) from an upper mixture of 20 µl (67 mM Tris [pH 8.8], 2 mM $MgCl_2$, 0.75 U *Taq* Polymerase [Perkin-Elmer], and 2–8 µl of DNA extract). Forty cycles of PCR (50 s at 92°C, 50 s at relevant annealing temperature, and 50 s at 72°C) were carried out on a DNA Thermal Cycler (Perkin Elmer Cetus). Table 1 gives the primers used. Each series of PCR included two blank extraction controls and a negative PCR control. Amplification products were analyzed by electrophoresis in 2.8%–5% agarose gels.

Quantitation

To determine the number of template molecules in DNA extracts, a molecule of 308 bp was constructed, encompassing positions 16190 to 16517 in the human mitochondrial control region (Anderson et al. 1981) but with bases 16210 to 16231 (22 bp) deleted (Förster 1994). This molecule was cloned into a TA vector (Invitrogen), and the concentration of the purified construct was determined by optical density at 260/280 nm. A series of fivefold dilution steps down to an estimated number of three construct copies/5 µl was prepared. In order to determine the approximate number of template molecules in an extract, a hot-start PCR was performed with three to six consecutive dilution steps of the standard, to which a constant amount of extract (2 µl for the Ice Man and VC15A and 5 µl for VC16 and VC25) had been added. To quantitate template molecules of different lengths, primer L16209 (table 1) was used with primers H16271 (5′-GTGGGTAGGTTTGTTGGTATCCTA-3′, 81 bp/103 bp), H16303 (table 1, 109 bp/131 bp), or H16410 (table 1, 216 bp/238 bp). Numbers in parentheses refer to lengths of PCR products from the competitor constructs and the mtDNA templates, respectively.

Cloning and Sequencing

Either PCR products were sequenced directly (Bachmann et al. 1990) or 1–4 µl of amplification products were ligated into a vector (TA vector) and cloned in CJ236 (Kunkel 1985). Clones for the control region sequences were screened for the presence of inserts with oligonucleotides H16095 (5′-GTGGCTGGCAGTAATGTACG-3′, 57°C), H16139 (table 1, 51°C), L16228 (5′-

Table 2

Quantitation of mtDNA Fragments in Extracts of VC15A, VC16, Two Samples of VC25 (1 and 3), and the Ice Man

	103 bp	131 bp	189 bp	238 bp
VC15A (.41 mg)	1,000	ND	ND	8
VC16 (.17 mg)	8	5	ND	ND
VC25-1 (.25 mg)	2	0	ND	ND
VC25-3 (.25 mg)	3	0	ND	ND
Ice Man (.56 mg)	1,000	ND	5	ND

NOTE.—The nos. of amplifiable DNA molecules/μl extract are given for products of different lengths. The amounts of tissue that have been added per microliter of extract are given in parentheses. "ND" indicates that quantitations have not been performed.

TACAGCAATCAACCCTCAAC-3′, 53°C), L16313 (5′-CCCTTAACAGTACATAGTAC-3′, 51°C), and H16356 (table 1, 57°C), respectively. Miniprep plasmid purifications were carried out as described by Sambrook et al. (1989) and were sequenced by the dideoxy chain-termination method (Sanger et al. 1977) using Sequenase (USB), α-^{33}P-dATP, and primers T7PBS (5′-AATACGACTCACTATAG-3′) and SP6 (5′-ATTTAGGTGACACTATAG-3′). Reactions were separated on a 6% denaturing polyacrylamide gel and visualized by autoradiography.

Restriction Site and Deletion Analyses

Hot-start PCR was carried out with primers amplifying fragments of 75–121 bp (table 1). PCR products were sequenced either directly or from a minimum of three clones. The 9-bp deletion was amplified with primers L8196/H8297 and analyzed by electrophoresis in a 2.8% agarose gel. These products were not sequenced in every case.

Results

Nucleic acids were extracted from samples of 10 mummies from VC, Castle Mountains, Arizona. This site is attributed to the Hohokam culture and dated to 1000–1400 A.D. Mitochondrial control region sequences <250 bp were amplified by PCR from multiple aliquots of the extracts, and the products were analyzed by ethidium-stained agarose gels. One individual (VC15A) showed strong bands in all amplifications attempted, whereas extracts from six individuals yielded much weaker amplification products after 40 cycles of PCR. For three individuals, some amplifications yielded no products, whereas others resulted in very weak bands. Direct sequencing of the amplification products often revealed several positions at which more than one base seemed to exist (not shown). Thus, by a straightforward approach it was impossible to determine unambiguous DNA sequences from these remains.

Further analyses concentrated on the remains of three individuals that represent each of the groups of samples: one individual who consistently yielded strong amplification products (VC15A), one who consistently yielded weak bands (VC16), and a third who only sporadically yielded weak amplification products (VC25).

DNA Quantitation

The numbers of amplifiable human DNA molecules present in the extracts were determined by a competitive PCR assay (table 2). Extracts were quantitated for fragments of length 103 bp as well as for the longest fragment that was possible to amplify. An extract from mummy VC15A contained ~1,000 molecules/μl of extract of the 103-bp fragment. The longest amplification product achieved was 238 bp. However, only ~8 molecules/μl of extract—that is, in the order of single molecules—of these longer molecules were present. An extract of VC16 contained approximately eight copies of the 103-bp fragment per microliter of extract, and the longest product achieved was 131 bp, of which equally few copies were present. From VC25, finally, several attempts were necessary in order to achieve amplifications of the 103-bp fragment, and in the order of single molecules (2–3 molecules/μl) of this length were present in two extracts from different tissue samples. When 8 μl of the extracts were used, amplification of fragments of 131 bp was occasionally possible. However, because of the limited amounts of extract, the larger fragments could not be quantitated for VC25. For comparison, extracts of the so-called Ice Man were analyzed in a similar fashion. Here, the situation is reminiscent of VC15A, in that ~1,000 molecules/μl of extract were present for the 103-bp fragment (fig. 1), whereas ~1 molecule/μl of extract of a 189-bp fragment was present (Handt et al. 1994*b*). Thus, the mummy VC15A as well as the Ice Man contain ~2,000 mitochondrial molecules

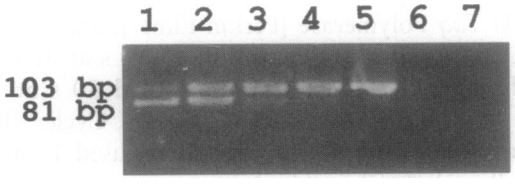

Figure 1 Quantitation of a 103-bp-long mtDNA fragment in an extract from the Ice Man. A dilution series of a competitor template 22 bp shorter was added to a constant amount of extract. Ten percent of the amplification products was separated in a 2%BRL/3%NuSieve agarose gel. The figure shows the amplification products to which 9,500 (lane 1), 1,900 (lane 2), 380 (lane 3), and 76 (lane 4) construct molecules have been added to the PCR. Lanes 5–7 show the results for the extract only, the extraction control, and the mock PCR control.

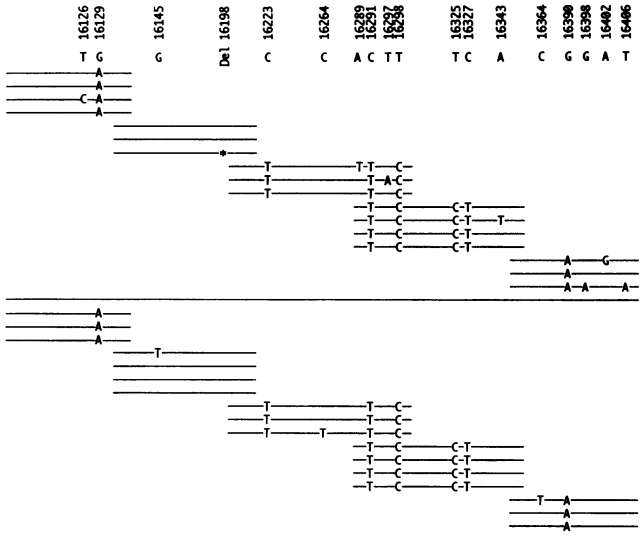

Figure 2 Schematic picture of DNA sequences from cloned PCR products of VC15A. The sequence was determined for five overlapping fragments from two extracts. Results for the individual extracts are separated by a solid line. Only the positions that vary from the reference sequence (Anderson et al. 1981) are given. The asterisks represent deletions.

of a length of ~100 bp/mg of tissue, whereas VC16 and VC25 contain ~10–40 molecules/mg of tissue.

mtDNA Sequences

In order to determine the nucleotide sequence of the hypervariable region I (HVR I) of the mitochondrial control region (position 16056–16409, Anderson et al. 1981), PCR products encompassing overlapping parts of this region were sequenced directly. Ambiguities, indicating heterogeneity in the amplification products, were often observed (not shown). The PCR products were therefore cloned and multiple clones sequenced. For individual VC15A, the HVR I sequence was determined in five overlapping fragments (fig. 2). When compared to a reference sequence (Anderson et al. 1981), seven substitutions were observed in all clones sequenced (position 16129, G to A; position 16223, C to T; position 16291, C to T; position 16298, T to C; position 16325, T to C; position 16327, C to T; position 16390, G to A). To test the reproducibility of this result, an extract was prepared from a second tissue sample, and the experiments were repeated. Again, substitutions at the same seven positions occurred in all clones. In addition, among the 34 clones sequenced, five transitions, five transversions, and one deletion occurred, each in single clones.

For VC16, the sequence was determined by the same five overlapping fragments (fig. 3). For one extract, substitutions were observed in all clones sequenced at two positions (position 16111, C to T; position 16189, T to C). At three positions, more than one clone, but not all clones, carried changes (position 16183, A to C in three of four clones; after position 16193, insertion of C in two of four clones; position 16217, T to C in three of seven clones). In addition, at 11 positions, single clones carried changes from the reference sequence. An additional extract from another tissue sample was prepared, and three of the five fragments were amplified and cloned. In this case, none of the four positions that carried changes in multiple clones from the first extract showed the same changes. Instead, three other positions carried substitutions in more than one, but not all clones (position 16146, A to G in two of five clones; position 16162, A to G in three of five clones; position 16209, T to C in two of five clones). In addition, eight single substitutions were seen in the 12 clones sequenced. Thus, from these extracts, it was impossible to determine an unambiguous mitochondrial control region sequence.

For individual VC25, the sequence was initially determined in three overlapping fragments, the longest of which was 215 bp (fig. 4, *above*). When these products were cloned and three clones from each amplification sequenced, all clones carried substitutions at five positions (position 16179, C to T; position 16223, C to T; position 16297, T to C; position 16327, C to T; position 16391, G to A), whereas no substitutions occurred in single clones. The sequence was additionally determined from an extract prepared from another tissue sample, this time in five overlapping fragments of a maximum length of 131 bp. When a total of 22 clones of these amplification products were sequenced, none of them

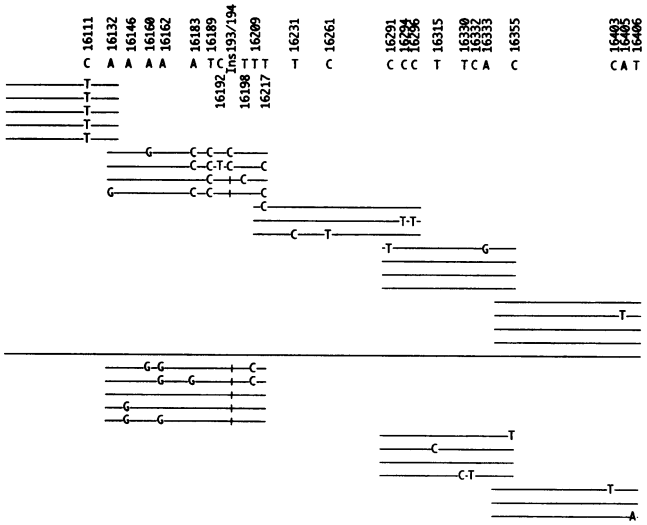

Figure 3 DNA sequences from VC16. Five overlapping fragments were amplified and cloned from one extract, and three fragments were cloned from another extract. The vertical dashes represent a gap in some clones due to an insertion of one nucleotide between position 16,193 and 16,194 in other clones.

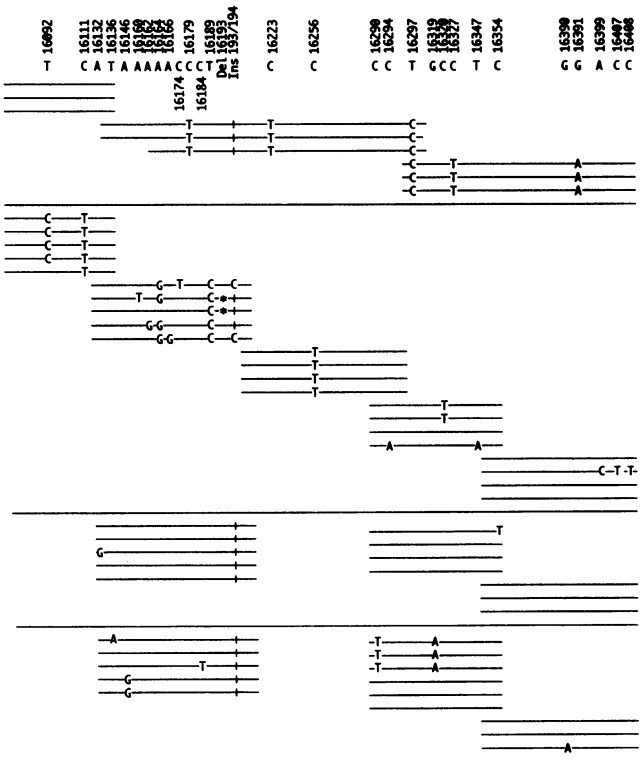

Figure 4 DNA sequences from VC25. Sequences from the first extract was determined for three overlapping fragments, and from the second extract for five overlapping fragments. Sequences for the other two extracts were determined for three fragments.

carried any of the five substitutions previously found. Rather, three other positions showed substitutions in all clones (position 16111, C to T; position 16189, T to C; position 16256, C to T), and five positions carried substitutions that occurred in more than one, but not all clones (position 16092, 4 clones C, 1 clone T; position 16164, 4 clones G, 1 clone A; position 16193, deletion in two of five clones; insertion of a C after position 16193 in two of five clones; position 16320, 2 clones T, 2 clones C). In addition, nine substitutions occurred in single clones. In an attempt to clarify the situation, further PCR products of a length ≤126 bp were cloned and sequenced from two additional extracts. From these, a total of 26 clones were sequenced and found to differ from the two extracts analyzed previously, as well as from each other, and within one of the extracts. Consequently, it was not possible to determine an unambiguous mitochondrial control region sequence for individual VC25.

Restriction Sites and Deletion

Three restriction sites and one region carrying a 9-bp repeated sequence exhibiting length polymorphism in several populations (see, e.g., Schurr et al. 1990; Vigilant 1990; Redd et al. 1995) were amplified from two to three extracts from each of the mummies. These polymorphic positions define four clusters of related mitochondrial lineages that dominate the mitochondrial diversity in the Americas (Wallace et al. 1985; Schurr et al. 1990; Torroni et al. 1992; Horai et al. 1993). The amplification products were analyzed by direct sequencing and sequencing of cloned products, and, in some cases, the presence of the deletion was analyzed solely by agarose gel electrophoresis. The results (table 3) show that VC15A lacks the *Hin*cII site at position 13259, the *Hae*III site at position 663, and the 9-bp deletion but carries the *Alu*I site at position 5176. This combination of polymorphisms is typical of group C lineages in the Americas (Torroni et al. 1992). The extracts of VC16 and VC25 are similar to each other in that they lack the *Hae*III site but carry the *Hin*cII and *Alu*I sites. However, when amplifications of the region with the 9-bp repeat were carried out for these extracts, some extracts showed the deletion, whereas others, from the same individuals, did not. In addition, when the amplification products were directly sequenced, they could be shown to contain both deleted and nondeleted molecules (not shown). Thus, VC16 and VC25 are heterogeneous when mitochondrial markers outside the control region are investigated.

Discussion

Success Rate of Amplification

The mummies from VC are macroscopically excellently preserved (Cockburn and Cockburn 1980), and in this respect are similar to, for example, many Egyptian mummies. However, the experience from 69 bone samples, 38 soft tissue samples, and 3 teeth from Egypt (M. Krings, unpublished results), as well as from 36 bones, 6 coproliths, and 3 soft tissue samples from animal remains in the New World (Höss et al. 1996*a*), shows that only a small fraction of such remains yield amplification products. Thus, in our experience, most archaeological samples do not contain DNA molecules that are amplifiable with current techniques. Obviously, that presents a limitation to studies of ancient human populations.

Table 3

Restriction Site Data Collected for Various Tissue Samples of Mummies VC15A (Two Extracts), VC16 (Two Extracts), and VC25 (Three Extracts)

	*Hae*III Site (nt 663)	*Hin*cII Site (nt 13259)	*Alu*I Site (nt 5176)	9-bp Deletion
VC15A	−	−	+	−
VC16	−	+	+	+/−
VC25	−	+	+	+/−

Table 4

Compilation of the Mean Pairwise Sequence Differences (Mean) and Their SDs within the Clusters A, B, C, and D, and between These Clusters and VC15A

Haplogroup	Sample Size	Within Clusters		Comparison with VC15A	
		Mean	SD	Mean	SD
A	19	3.15	1.78	9.68	1.29
B	11	2.80	1.41	9.00	1.18
C	9	3.17	1.21	3.33	.87
D	10	2.82	1.50	6.50	1.18

NOTE.—Sequence data are from Torroni et al. (1993). The sequence segment considered is between positions 16,056 and 16,370.

Potentially even more detrimental, however, are erroneous results in the cases where amplification products are seen. At least four sources exist for such false-positive results and ambiguous sequences. They are, in what we believe to be a decreasing order of importance: first, contamination by contemporary DNA sequences (Handt et al. 1994*a*); second, PCR errors due to misincorporation of nucleotides (Pääbo et al. 1990; Goloubinoff et al. 1993); third, inadvertent amplification of nuclear insertions of mitochondrial sequences (Collura and Stewart 1995; van der Kuyl et al. 1995; Zischler et al. 1995); and, fourth, mtDNA heteroplasmy, which results in the determination of ambiguous sequences (Gill et al. 1994; Bendall et al. 1995; Comas et al. 1995). Below, we discuss the reliability of the sequences determined from the VC remains and use these results to discuss these four sources of erroneous, positive results.

DNA Sequence Reliability

Since a large proportion of mitochondrial sequences indigenous to the New World fall into four clusters of related lineages (Wallace et al. 1985; Schurr et al. 1990; Torroni et al. 1992; Horai et al. 1993), it can be estimated whether sequences determined from ancient remains are likely to be of Native American origin. As modern contaminations often stem from non–Native American sources, it is possible to distinguish sequences that are likely to be endogenous from those that are likely to stem from contaminations.

When the mitochondrial control region sequence determined from the two extracts of VC15A (fig. 2) is compared with the four sequence clusters among contemporary Native Americans, the mean pairwise difference to cluster C is 3.33 ± 0.87 (table 4). While this difference is two to three times lower than the difference to the other three clusters, it falls well within the mean difference for sequences within cluster C (3.17 ± 1.21). In agreement with this information, the analyses of restriction sites and the 9-bp deletion assign the mitochondrial sequence of this individual to cluster C (table 3). In contrast, the VC15A sequence differs by 8.3–11.4 substitutions from sequences from other areas of the World (not shown).

For VC16, the predominant sequence present in clones from the first extract (16111 T, 16183 C, 16189 C) shows 3.45 substitutions to cluster B sequences and 5.90–7.78 substitutions to the other clusters. Furthermore, it differs by 4.52–4.91 substitutions from European populations and by 9.02 substitutions from African sequences. This information is not incompatible with a Native American origin of this sequence. However, the clones are not unambiguous, since two additional positions carry substitutions that occur in more than one, but not all clones. Similarly, whereas some amplifications revealed the presence of the 9-bp deletion typical of cluster B sequences, others failed to do so. Finally, a second tissue sample did not allow the sequence to be reproduced (fig. 3). Thus, from this individual several sequences have been amplified, and it is impossible to determine which, if any, of them is of ancient origin.

From individual VC25, the first extract yielded a sequence that differs by 4.78 substitutions from cluster C. Contemporary cluster C sequences differ by an average of 3.17 substitutions from each other. While the distances to European sequences are larger (6.86–7.04), this extract did not reveal the expected loss of the *Hin*cII site at position 13259. Therefore, this sequence does not clearly fall into one of the Native American clusters. Furthermore, other extracts yielded different sequences (fig. 4).

Because of several observations, we regard the sequence determined from VC15A to be of putatively ancient origin. First, the cloned amplification products reveal one sequence to be present in all clones, with other changes occurring only in individual clones (fig. 2). Second, this sequence can be reproduced from a second tissue sample (fig. 2). Third, substantial amounts of amplifiable molecules are present (table 2). Fourth, the analysis of restriction sites (table 3) is compatible with the control region sequence in assigning this mitochondrial lineage to Native American cluster C (table 4). In contrast, the sequences determined from VC16 and VC25 are neither unambiguous nor reproducible. In our experience and that of others (Richards et al. 1995), the latter situation is common when amplification products from ancient human remains are analyzed in sufficient detail. Below, we discuss four possible sources of such ambiguous results.

Contamination

Contaminations from contemporary DNA are of two types (Navidi et al. 1992). The first type will affect all

amplifications to the same extent and are due to, for example, contemporary DNA present in reagents used in tissue extractions or PCR. These contaminants are generally easily detected by controls, such as amplifications from mock extractions and amplifications where no template DNA is added. Of more concern are contaminations of the second type, which affect individual extractions or amplifications. They may stem, for example, from handling of archaeological specimens prior to sampling or from laboratory aerosols containing human DNA or amplification products. The main way to identify this type of contamination is to repeat experiments from two or more independent samples from each individual studied.

In the case of VC16 and VC25, independent extracts yielded different sequences, while extraction controls and PCR controls were negative. Thus at least some of these sequences must be due to contamination(s) of the second type. Unless such contaminations are caused by PCR products, they can be assumed to be of relatively limited quantity. It is therefore of importance that for VC15A, where a putatively ancient sequence could be determined, the numbers of amplifiable template molecules were about two orders of magnitude larger than for VC16 and VC25, from which amplifications started from ~10 molecules. Therefore, the contaminations were caused by few molecules in the latter cases. It is likely that sporadic contaminating molecules are present also in the extracts of VC15A. However, there they will be outnumbered by the endogenous molecules.

Thus, a quantitation of template molecules is helpful in determining the possible role and sources of contamination in an extract. In extracts that contain large amounts of short mitochondrial molecules and drastically fewer longer molecules, only contaminations of short PCR products encompassing the amplified region are likely to cause unambiguous false sequences. Consequently, barring that possibility, reproducible DNA sequences of putatively ancient origin can generally be determined from the short molecules (Handt et al. 1994*b*). Furthermore, if both mitochondrial and nuclear sequences are amplified (Béraud-Colomb et al. 1995), a quantitation of both will reveal whether the mtDNA, as expected, occurs in larger amounts. If that is not the case, one or the other, or both, is likely to represent a contamination.

PCR Errors

If direct sequencing reactions result in unambiguous sequences, or if multiple clones are sequenced and found to have the same sequence, it is generally assumed that PCR errors are not responsible for the substitutions seen (Pääbo and Wilson 1988). However, ancient DNA is often extensively damaged (Pääbo 1989; Höss et al. 1996*b*) and may contain base modifications that are miscoding. If amplifications start from few or even single DNA strands carrying miscoding lesions, all molecules in the resulting amplification product will carry an incorrect base. Furthermore, in amplifications that start from few initial template molecules, errors occurring during the first cycles of the PCR will be present in a large proportion of the final products. Therefore, a quantitation of the numbers of template molecules from which the PCR reactions starts are of importance in order to rule out that nucleotide misincorporations during the early cycles of the PCR may cause incorrect sequences.

In the case of VC15A, amplifications of length 103 bp start from ~2,000 copies. Therefore, in this case, there is no reason to assume that errors during the early cycles of the PCR at any one position would influence a large proportion of the molecules in the final amplification product. In agreement with this reasoning, all substitutions observed among the 34 clones sequenced either are present in all clones and are thus assumed to be present in the original template DNA, or, alternatively, occur in only one clone and are likely due to misincorporations during the PCR. In contrast, for VC16 and VC25, amplifications start from <40 template molecules. In both cases, several substitutions are observed in more than one but not all template molecules. While these results could be due to heteroplasmy (see below), the fact that the amplification starts from so few molecules makes contamination—or, alternatively, misincorporations during the early cycles of PCR—a much likelier source of these substitutions.

To distinguish the two possibilities of contamination and misincorporation, it is instructive to compare the distribution across the sequence of substitutions that occur in single clones, which are likely to be due to polymerase errors, to the distribution of those that occur in multiple but not all clones. Among the 123 different clones sequenced, 41 substitutions are observed in single clones. Of these, 56% (23) fall at positions that are variable among 807 human mitochondrial control region sequences from various populations (not shown). Since 43% of the positions (167 of 387) in the sequence are variable, this proportion is in agreement with the assumption that these substitutions are distributed randomly and are caused by polymerase errors. In contrast, 90% of substitutions (26 of 29) observed in more than one but not all clones fall at variable positions. Thus, the majority of these substitutions are likely due to contamination of the specimens and/or laboratory reagents with contemporary human DNA.

Nuclear mtDNA Insertions

Recently, nuclear insertions of mtDNA have been identified as a source of purportedly ancient sequences (Collura and Stewart 1995; van der Kuyl et al. 1995; Zischler et al. 1995), and it has been demonstrated that

such sequences are numerous within the nuclear genome (Zullo et al. 1991). However, since such sequences are less numerous than the mitochondrial genomes in the cell, they will, in general, not represent a problem. An exception is if a nuclear sequence will be amplified preferentially over the mitochondrial copies. Although the primers used in amplifications of human mtDNA are generally chosen such that they bind to areas of little or no sequence variation, the high sequence diversity of, for example, the mitochondrial control region makes it possible that a primer might be affected by a substitution close to its 3′ end in a particular individual. In such a case, a nuclear insertion that does not carry this substitution might be amplified preferentially. Since such a problem will be limited to a particular primer, it can be detected or excluded by the use of different primer pairs that amplify partially overlapping sequences. For example, in VC15A, substitutions at positions 16291 and 16298 occur in two different amplification products. They are therefore unlikely to stem from a nuclear insertion. Furthermore, if quantitations are performed for all amplifications, a single segment that would stem from a nuclear insertion would be detected by its relatively lower abundance. Novel techniques that allow the quantitation of template molecules without the use of internal standards might make this objective feasible in the near future.

Heteroplasmy

An additional source of heterogeneity in mitochondrial sequences determined from amplification products is heteroplasmy, that is, the occurrence of more than one mitochondrial type within an individual. In humans, this occurrence is extremely rare, even if its rate might be somewhat underestimated (Comas et al. 1995). For example, in the case of the remains of the last tsar, sequence heterogeneity in the amplification product has been attributed to heteroplasmy (Gill et al. 1994), and this attribution has been verified by the analysis of remains of related individuals (Ivanov et al. 1996). However, heteroplasmy as a cause for sequence heterogeneity in ancient remains should only be considered after contamination and PCR errors have been excluded.

Conclusion

The results presented above are in our experience typical when amplifications are performed from ancient human remains. They demonstrate that greater experimental efforts than are generally made are necessary to demonstrate the authenticity of amplification products from such sources. Specifically, the reproduction of ancient human DNA sequences is necessary from at least two independent extracts, preferably performed in different laboratories. Furthermore, the cloning of amplification products is necessary whenever any sign of sequence heterogeneity occurs in direct sequence reactions. Finally, when a quantitation of the number of amplifiable DNA molecules shows that an extract containing few template molecules, sporadic contamination, or PCR errors may contribute a substantial fraction, or all, of the final amplification product. A minimum of 100–1,000 molecules per amplification may be necessary to obviate such problems.

Acknowledgments

We thank Arndt von Haeseler, Matthias Höss, Hendrik Poinar, Antti Sajantila, and Hans Zischler for fruitful discussions, and Walter H. Birkby for samples. This work was supported by the Deutsche Forschungsgemeinschaft (Pa 452/1-2) and the Boehringer Ingelheim Fonds.

References

Anderson S, Bankier AT, Barrel BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457–465

Bachmann B, Lüke W, Hunsmann G (1990) Improvement of PCR amplified DNA sequencing with the aid of detergents. Nucleic Acids Res 18:1309

Bendall KE, Sykes BC (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 57:248–256

Béraud-Colomb E, Roubin R, Martin J, Maroc N, Gardeisen A, Trabuchet G, Goossens M (1995) Human β-globin gene polymorphisms characterized in DNA extracted from ancient bones 12,000 years old. Am J Hum Genet 57:1267–1274

Cockburn A, Cockburn E (1980) Mummies, disease and ancient cultures. Cambridge University Press, Cambridge

Collura RV, Stewart CB (1995) Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. Nature 378:485–488

Comas D, Pääbo S, Bertranpetit J (1995) Heteroplasmy in the control region of human mitochondrial DNA. Genome Res 5:89–90

Cooper A, Mourer-Chauviré C, Chambers GK, von Haeseler A, Wilson AC, Pääbo S (1992) Independent origins of New Zealand moas and kiwis. Proc Natl Acad Sci USA 89:8741–8744

Förster E (1994) An improved general method to generate internal standards for competitive PCR. Biotechniques 16: 18–20

Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, et al (1994) Identification of the remains of the Romanov family by DNA analysis. Nat Genet 6:130–135

Goloubinoff P, Pääbo S, Wilson A (1993) Evolution of maize inferred from sequence diversity of an Adh2 gene segment from archaeological specimens. Proc Natl Acad Sci USA 90: 1997–2001

Handt O, Höss M, Krings M, Pääbo S (1994*a*) Ancient

DNA—methodological challenges. Experientia 50:524–529

Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, et al (1994*b*) Molecular genetic analyses of the Tyrolean Ice Man. Science 264:1775–1778

Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonada S, Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Mol Biol Evol 10:23–47

Höss M, Dilling A, Currant A, Pääbo S (1996*a*) Phylogeny of the extinct ground sloth Mylodon darwinii. Proc Natl Acad Sci USA 93:181–185

Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S (1996*b*) DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res 24:1304–1307

Höss M, Pääbo S (1993) DNA extraction from Pleistocene bones by a silica-based purification method. Nucleic Acids Res 21:3913–3914

Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parson TJ (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of the Tsar Nicholas II. Nat Genet 12:417–420

Kunkel T (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci USA 82: 488

Lindahl T (1993) Recovery of antediluvian DNA. Nature 365: 700

Mullis KB, Faloona F (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155:335–350

Navidi W, Arnheim N, Waterman MS (1992) A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. Am J Hum Genet 50:347–359

Pääbo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci USA 86:1939–1943

Pääbo S, Higuchi RG, Wilson AC (1989) Ancient DNA and the polymerase chain reaction. J Biol Chem 264:9709–9712

Pääbo S, Irwin DM, Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic amplification. J Biol Chem 265:4718–4721

Pääbo S, Wilson AC (1988) Polymerase chain reaction reveals cloning artefacts. Nature 334:387–388

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/tRNA^{Lys} intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

Richards MB, Sykes BC, Hedges, REM (1995) Authenticating DNA extracted from ancient skeletal remains. J Archaeol Sci 22:291–299

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia. Science 230:1350–1354

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schurr TG, Ballinger SW, Gan Y-Y, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, et al (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613–623

Sidow A, Wilson AC, Pääbo S (1991) Bacterial DNA in Clarkia fossils. Phil Trans R Soc Lond B 333:429–433

Thomas RH, Schaffner W, Wilson AC, Pääbo S (1989) DNA phylogeny of the extinct marsupial wolf. Nature 340:465–467

Torroni A, Schurr TG, Yang CC, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, et al (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162

Torroni A, Sukernik RI, Schurr T, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, et al (1993) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53:591–608

Van der Kuyl A, Kuiken CL, Dekker JT, Perizonius WRK, Goudsmit J (1995) Nuclear counterparts of the cytoplasmic mitochondrial 12S rRNA gene: a problem of ancient DNA and molecular phylogenies. J Mol Evol 40:652–657

Vigilant L (1990) Control region sequences from African populations and the evolution of human mitochondrial DNA. PhD thesis, University of California, Berkeley

Wallace DC, Garrison K, Knowler WC (1985) Dramatic founder effects in Amerindian mitochondrial DNAs. Am J Phys Anthropol 68:149–155

Zischler H, Höss M, Handt O, von Haeseler A, van der Kuyl AC, Goudsmit J, Pääbo S (1995) Detecting dinosaur DNA. Science 268:1192–1193

Zullo S, Sieu LL, Slightom JL, Hadler HI, Eisenstadt JM (1991) Mitochondrial D-loop sequences are integrated in the rat nuclear genome. J Mol Biol 221:1223–1235