Characterization of Genetic Miscoding Lesions Caused by Postmortem Damage

M. Thomas P. Gilbert,[1] Anders J. Hansen,[2,*] Eske Willerslev,[2,*] Lars Rudbeck,[3] Ian Barnes,[1,†] Niels Lynnerup,[4] and Alan Cooper[1]

[1]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, Oxford University, Oxford, United Kingdom; and [2]Department of Evolutionary Biology, Zoological Institute, and [3]Research Laboratory and [4]Laboratory of Biological Anthropology, Institute of Forensic Medicine, University of Copenhagen, Copenhagen

The spectrum of postmortem damage in mitochondrial DNA was analyzed in a large data set of cloned sequences from ancient human specimens. The most common forms of damage observed are two complementary groups of transitions, termed "type 1" (adenine→guanine/thymine→cytosine) and "type 2" (cytosine→thymine/guanine→adenine). Single-primer extension PCR and enzymatic digestion with uracil-N-glycosylase confirm that each of these groups of transitions result from a single event, the deamination of adenine to hypoxanthine, and cytosine to uracil, respectively. The predominant form of transition-manifested damage varies by sample, though a marked bias toward type 2 is observed with increasing amounts of damage. The two transition types can be used to identify the original strand, light (L) or heavy (H), on which the initial damage event occurred, and this can increase the number of detected jumping-PCR artifacts by up to 80%. No bias toward H-strand–specific damage events is noted within the hypervariable 1 region of human mitochondria, suggesting the rapid postmortem degradation of the secondary displacement (D-loop) H strand. The data also indicate that, as damage increases within a sample, fewer H strands retain the ability to act as templates for enzymatic amplification. Last, a significant correlation between archaeological site and sample-specific level of DNA damage was detected.

Introduction

DNA decays rapidly after death in biological samples, and the ensuing damage is manifested in many forms. Strand fragmentation is caused by endogenous endonuclease activity (Pääbo 1989) or hydrolytic attacks that lead to the depurination of deoxyribose–adenine (A) or deoxyribose–guanine (G) bonds, rapidly destroying the DNA backbone (Lindahl 1993; Höss et al. 1996; Bada et al. 1999). Much of the DNA is also modified oxidatively via free radicals (Höss et al. 1996). Oxidative damage is most commonly seen as modifications of sugar residues and the pyrimidines cytosine (C) and thymine (T) to hydantoins, as well as baseless sites and intermolecular cross-links (Pääbo 1989), all of which block the activity of PCR enzymes (Höss et al. 1996). However, a small proportion of damage events do not hinder rep-

lication but generate miscoding lesions (Pääbo 1989). These are manifested as base modifications in the amplified sequence, changing the appearance of a DNA template (Fattorini et al. 1999) and potentially generating misleading haplotype analyses (Gilbert et al. 2003 [in this issue]). The few detailed studies of miscoding lesions concur with earlier hypotheses (e.g., see Pääbo 1989; Lindahl 1993; Höss et al. 1996) that the majority of changes arise from the deamination of C to uracil (U), an analogue of T, or the deamination of A to hypoxanthine (HX), an analogue of G (Hansen et al. 2001; Hofreiter et al. 2001). For simplicity, both the chemical event and the phenotype are referred to here simply as C→T or A→G changes. However, because either of the complementary DNA strands can be sequenced after amplification, each of these transitions can produce two observable phenotypes. For example, a C→T degradation may simply be observed as C→T, but, if the complementary strand is sequenced, then it will be read as a G→A transition. Similarly, an A→G degradation may be observed as either A→G or as a T→C transition (Hansen et al. 2001; Hofreiter et al. 2001). Following the nomenclature of Hansen et al. (2001), we term each set of miscoding lesions as "type 1" (A→G/T→C) or "type 2" (C→T/G→A) transitions, respectively.

In the present study, a large data set of previously published cloned sequences of human and nonhuman mitochondrial ancient DNA (aDNA) are analyzed. The

Table 1

Damage Calculations, Archaeological Site of Origin, and Age of Samples Studied

Sample	PCRs[a]	Bases[b]	Damage[c]	d^{d}	Site[e]	Code[f]	Approximate Age (years)
Tg44	1	147	8	.0544	Greenland	5	800
Tg54	1	147	8	.0544	Greenland	5	800
Tg76	1	355	2	.0056	Denmark	1	300
Tg77	3	857	5	.0058	Denmark	1	500
Tg80	1	355	7	.0197	Denmark	1	600
Tg85	1	355	10	.0282	Denmark	1	300
Tg103	1	344	10	.0291	Denmark	1	800
Tg104	1	147	3	.0204	Orkney	6	?
Tg105	1	355	0	0	Denmark	1	800
Tg112	1	147	13	.0884	Greenland	5	800
Tg114	1	355	8	.0225	Denmark	1	300
Tg116	2	502	19	.0378	Denmark	1	600
Tg120	1	355	2	.0056	Denmark	1	300
Tg123	1	147	7	.0476	Denmark	1	500
Tg127	1	147	7	.0476	Denmark	1	300
Tg128	2	502	2	.004	Greenland	5	800
Tg129	2	502	8	.0159	Greenland	5	800
Tg131	1	147	3	.0204	Greenland	5	800
Tg133	1	147	3	.0204	Greenland	5	800
Tg136a	2	502	9	.0179	Repton	7	800
Tg136b	2	502	11	.0219	Repton	7	800
Tg137a	3	294	8	.0272	Repton	7	1,300
Tg137b	3	294	13	.0442	Repton	7	1,200
Tg138	1	147	5	.034	Repton	7	1,100
Tg141	1	147	4	.0272	Repton	7	1,100
Tg142	1	355	2	.0056	Repton	7	1,100
Tg143	1	147	4	.0272	Repton	7	1,100
Tg145	1	355	10	.0282	Repton	7	1,200
Tg146	1	147	5	.034	Repton	7	1,000
Tg148	7	1,029	17	.0165	Repton	7	1,300
Tg149	7	1,029	140	.1361	Repton	7	1,100
Tg192	2	502	3	.006	Southern Britain	4	10,000
Tg196	2	502	18	.0359	Caribbean	3	600
Tg232	1	147	13	.0884	Northern Britain	2	1,800
Tg233	2	147	22	.1497	Northern Britain	2	1,800
Total	61	12,259	409				

NOTE.—Samples studied were those described in the companion article by Gilbert et al. (2003).

[a] Number of independent PCRs amplified per sample.

[b] Total number of bases amplified per sample.

[c] Total number of observed damage events per sample.

[d] Standardized damage calculation (for details, see the "Material and Methods" section).

[e] Archaeological site of sample origin.

[f] Categorical geographical site coding for statistical analysis.

fidelity of the polymerase enzyme used to generate the data is examined by comparing sequencing error rates between ancient and modern extracts, as well as modern contaminants in the ancient extracts. The biochemical causes of postmortem miscoding lesions are investigated by the digestion of samples with the enzyme uracil-N-glycosylase prior to amplification, as well as with single-primer extension PCRs (SP-PCRs) (Hofreiter et al. 2001). The miscoding-lesion data allow analysis of the processes involved in DNA damage and reveal a direct correlation between archaeological sites and the extent and type of damage. The data also show that the ratio of type 2:type 1 transition events differs significantly between samples and is related to the overall level of damage. These findings provide a means to identify which DNA strand was initially damaged and show that there is no strand-specific propensity to hydrolytic damage within the control region, despite the presence of the extra copy of the H strand (the displacement strand, or D-loop) (Wallace et al. 1995). Interestingly, as the amount of damage increases, fewer amplifications are initiated from the H strand. These data also provide a new methodology for improved detection of jumping-PCR events (Pääbo et al. 1990).

Table 2

Sample Details, Base-Change Measurements, and Bias Calculations

Species and Sample (Region[a])	AT/GC Ratio[b]	Individual changes[c]												Total[h]	Complementary Groups[d]						Calculations[e]			Enzyme[f]	Source[g]
		A→C	T→G	A→G	T→C	A→T	T→A	C→A	G→T	C→G	G→C	C→T	G→A		A→G T→C	A→T T→A	A→C T→G	C→T G→A	C→G G→C	C→A G→T	(A→G)-(T→C)	(C→T)-(G→A)	Types 2-1		
Homo sapiens:																									
Tg44 (16209–16356)	1.26	0	0	1	0	0	0	1.26	0	1.26	0	6.31	0	9.83	1.0	.0	.0	6.3	1.3	1.3	1.00	6.31	5.31	1	1
Tg54 (16209–16356)	1.26	0	0	1	3	1	0	0	0	0	0	3.78	0	8.78	4.0	1.0	.0	3.8	.0	.0	−2.00	3.78	−.22	1	1
Tg76 (16055–16410)	1.16	0	0	1	0	0	0	0	0	0	0	1.16	0	2.16	1.0	.0	.0	1.2	.0	.0	1.00	1.16	.16	1	1
Tg76 (16209–16356)	1.26	0	1	1	0	0	0	0	0	0	0	1.26	0	2.26	1.0	.0	.0	1.3	.0	.0	1.00	1.26	.26	1	1
Tg77 (16055–16410)	1.16	1	0	2	0	0	0	0	0	0	0	0	1.16	5.16	2.0	.0	2.0	1.2	.0	.0	2.00	−1.16	−.84	1	1
Tg80 (16055–16410)	1.16	0	0	2	0	0	0	0	0	0	0	0	5.8	7.80	2.0	.0	.0	5.8	.0	.0	2.00	−5.80	3.80	1	1
Tg80 (16209–16356)	1.26	0	0	0	0	0	0	0	0	0	0	0	6.31	6.31	.0	.0	.0	6.3	.0	.0	.00	−6.31	6.31	1	1
Tg85 (16209–16356)	1.26	0	0	1	2	0	0	0	0	0	0	6.96	1.16	11.12	3.0	.0	.0	8.1	.0	.0	−1.00	5.80	5.12	1	1
Tg85 (16055–16410)	1.16	0	0	0	0	0	0	0	0	0	0	7.57	1.26	8.83	.0	.0	.0	8.8	.0	.0	.00	6.31	8.83	1	1
Tg103 (16055–16410)	1.16	0	0	2	3	0	0	0	0	0	0	5.8	0	10.80	5.0	.0	.0	5.8	.0	.0	−1.00	5.80	.80	1	1
Tg103 (16209–16356)	1.26	0	0	0	3	0	0	0	0	0	0	0	6.31	9.31	3.0	.0	.0	6.3	.0	.0	−3.00	−6.31	3.31	1	1
Tg104 (16209–16356)	1.26	0	0	0	0	0	0	0	0	0	0	1.26	2.52	3.78	.0	.0	.0	3.8	.0	.0	.00	−1.26	3.78	1	1
Tg112 (16209–16356)	1.26	0	0	2	2	0	0	0	0	1.26	1.26	6.31	2.52	15.35	4.0	.0	.0	8.8	2.5	.0	.00	3.78	4.83	1	1
Tg114 (16055–16410)	1.16	0	0	3	2	0	0	0	0	0	0	2.32	1.16	8.48	5.0	.0	.0	3.5	.0	.0	1.00	1.16	−1.52	1	1
Tg114 (16209–16356)	1.26	0	0	3	0	0	0	0	0	0	0	2.52	1.26	6.78	3.0	.0	.0	3.8	.0	.0	3.00	1.26	.78	1	1
Tg116 (16209–16356)	1.26	0	0	2	1	0	0	0	0	0	0	5.05	5.05	13.09	3.0	.0	.0	10.1	.0	.0	1.00	.00	7.09	1	1
Tg116 (16055–16410)	1.16	0	0	3	2	0	0	0	0	0	0	3.48	0	8.48	5.0	.0	.0	3.5	.0	.0	1.00	3.48	−1.52	1	1
Tg120 (16055–16410)	1.16	0	0	0	1	0	0	0	0	0	0	1.16	0	2.16	1.0	.0	.0	1.2	.0	.0	−1.00	1.16	.16	1	1
Tg120 (16209–16356)	1.26	0	0	0	1	0	0	0	0	0	0	1.26	0	2.26	1.0	.0	.0	1.3	.0	.0	−1.00	1.26	.26	1	1
Tg123 (16209–16356)	1.26	1	0	0	0	0	0	2.52	0	0	0	2.52	5.05	8.57	.0	.0	1.0	7.6	.0	.0	.00	−2.52	7.57	1	1
Tg127 (16209–16356)	1.26	1	0	1	0	0	0	0	0	0	0	3.78	0	8.31	1.0	.0	1.0	3.8	.0	2.5	1.00	3.78	2.78	1	1
Tg128 (16055–16410)	1.16	0	0	1	0	0	0	0	0	0	0	0	1.16	2.16	1.0	.0	.0	1.2	.0	.0	1.00	−1.16	.16	1	1
Tg128 (16209–16356)	1.26	0	0	0	0	0	0	0	0	0	0	0	1.26	1.26	.0	.0	.0	1.3	.0	.0	.00	−1.26	1.26	1	1
Tg129 (16209–16356)	1.26	0	0	1	0	0	0	0	0	0	0	1.26	0	1.26	1.0	.0	.0	1.3	.0	.0	.00	1.26	1.26	1	1
Tg129 (16055–16410)	1.16	0	0	3	2	0	0	0	0	0	0	0	1.16	7.16	5.0	.0	.0	1.2	.0	.0	1.00	−1.16	−3.84	1	1
Tg131 (16209–16356)	1.26	0	0	0	0	0	0	0	0	0	0	3.78	0	3.78	.0	.0	.0	3.8	.0	.0	.00	3.78	3.78	1	1
Tg133 (16209–16356)	1.26	0	0	2	0	1	1	0	0	0	0	0	0	3.00	2.0	2.0	.0	.0	.0	.0	2.00	.00	−2.00	1	1
Tg138 (16209–16356)	1.26	0	0	2	1	0	0	0	0	0	0	0	0	5.00	2.0	.0	.0	.0	.0	.0	2.00	.00	−2.00	1	1
Tg141 (16209–16356)	1.26	1	0	0	3	0	0	0	0	0	0	1.26	0	4.26	3.0	.0	1.0	1.3	.0	.0	−3.00	1.26	−1.74	1	1
Tg142 (16209–16356)	1.26	0	0	0	2	0	0	0	0	0	0	0	0	2.00	2.0	.0	.0	.0	.0	.0	−2.00	.00	−2.00	1	1
Tg142 (16055–16410)	1.16	0	0	0	2	0	0	0	0	0	0	0	0	2.00	2.0	.0	.0	.0	.0	.0	−2.00	.00	−2.00	1	1
Tg143 (16209–16356)	1.26	0	0	2	1	0	0	0	0	0	0	0	1.26	4.26	3.0	.0	.0	1.3	.0	.0	1.00	−1.26	−1.74	1	1
Tg145 (16055–16410)	1.16	0	0	2	2	0	0	0	0	0	0	6.96	0	10.96	4.0	.0	.0	7.0	.0	.0	.00	6.96	2.96	1	1
Tg145 (16209–16356)	1.26	0	0	2	2	0	0	0	0	0	0	7.57	0	11.57	4.0	.0	.0	7.6	.0	.0	.00	7.57	3.57	1	1
Tg146 (16209–16356)	1.26	0	0	0	1	0	0	0	0	0	0	5.05	0	6.05	1.0	.0	.0	5.0	.0	.0	−1.00	5.05	4.05	1	1
Tg148 (16209–16356)	1.26	0	0	6	4	0	0	0	0	0	0	5.05	3.78	18.83	10.0	.0	.0	8.8	.0	.0	2.00	1.26	−1.17	1	1
Tg149 (16209–16356)	1.26	0	0	11	13	1	1	0	0	0	0	138	3.78	168.55	24.0	1.0	.0	142.6	.0	.0	−2.00	134.98	118.55	1	1
Tg192 (16209–16356)	1.26	0	0	1	0	0	0	0	0	0	0	1.26	0	2.26	1.0	.0	.0	1.3	.0	.0	1.00	1.26	.26	1	1
Tg192 (16055–16410)	1.16	0	0	1	0	0	0	0	0	0	0	0	0	1.00	1.0	.0	.0	.0	.0	.0	1.00	.00	−1.00	1	1
Tg196 (16055–16410)	1.16	1	0	4	2	1	0	1.26	0	0	0	10.1	1.26	9.52	6.0	1.0	1.0	1.3	.0	1.3	2.00	−1.26	−4.74	1	1
Tg232 (16209–16356)	1.26	0	0	3	2	1	0	0	0	0	0	10.1	1.26	16.62	3.0	1.0	.0	11.4	1.3	.0	3.00	8.83	8.35	1	1
Tg233 (16209–16356)	1.26	0	0	2	1	0	0	0	0	0	1.26	15.1	8.83	26.97	3.0	.0	.0	24.0	1.3	.0	1.00	6.31	20.97	1	1
Tg233 (16055–16410)	1.16	0	0	1	0	0	0	0	0	0	0	1.16	4.64	6.80	1.0	.0	.0	5.8	.0	.0	1.00	−3.48	4.80	1	1
Tg129 (16055–16410)	1.16	0	0	3	2	0	0	0	0	0	0	0	1.16	7.16	5.0	.0	.0	1.2	.0	.0	1.00	−1.16	−3.84	1	1
Tg136a (16209–16356)	1.26	0	1	1	4	0	0	0	0	0	0	2.52	1.26	9.78	5.0	.0	1.0	3.8	.0	.0	−3.00	1.26	−1.22	1	1
Tg136b (16209–16356)	1.26	0	0	1	2	0	0	0	0	0	0	0	1.26	4.26	3.0	.0	.0	1.3	.0	.0	−1.00	−1.26	−1.74	1	1
Tg136a (16055–16410)	1.16	1	0	1	1	1	0	1.16	0	0	0	0	0	5.16	2.0	1.0	1.0	.0	.0	1.2	.00	.00	−2.00	1	1
Tg136b (16055–16410)	1.16	1	0	1	1	1	0	0	0	0	0	1.26	0	5.26	2.0	1.0	1.0	1.3	.0	.0	.00	1.26	−.74	1	1

50

Region[a]	Ratio[b]													Sum										Pol[f]	Source[g]
Tg137a (16209–16356)	1.26	0	0	1	0	5	0	0	0	0	0	2.52	0	8.52	5.0	.0	1.0	.0	2.5	.0	5.00	2.52	−2.48	1	1
Tg137b (16209–16356)	1.26	0	0	0	4	1	0	0	0	0	0	5.05	5.05	15.09	5.0	.0	.0	.0	10.1	.0	3.00	.00	5.09	1	1
Tg191a (16209–16356)	1.26	0	0	1	1	5	0	0	0	0	0	1.26	0	8.26	6.0	1.0	.0	.0	1.3	.0	−4.00	1.26	−4.74	1	1
Tg196 (16209–16356)	1.26	1	0	0	0	0	1	0	0	0	0	3.78	0	6.05	.0	1.0	.0	.0	3.8	1.3	.00	3.78	3.78	1	1
Tg77d (16055–16410)	1.16	1	0	0	2	2	0	0	0	0	1.26	0	0	5.16	2.0	.0	.0	2.0	1.2	.0	2.00	−1.16	−.84	1	1
Tg99i (16209–16356)	1.26	0	0	1	0	7	0	0	0	0	0	11.35	0	19.35	8.0	.0	.0	.0	11.4	.0	6.00	11.35	3.35	1	1
Tg99i (16209–16356)	1.26	0	0	0	1	1	0	0	0	0	0	0	0	5.00	5.0	.0	.0	.0	.0	.0	−3.00	.00	−5.00	1	1
Borgo Nuovo (16055–16410)	1.16	1	0	0	6	4	1	0	0	0	0	24.3	2.32	42.65	14.0	1.0	.0	1.0	26.6	.0	−2.00	22.01	12.65	3	2
Mezzocorona (16055–16410)	1.16	0	0	0	3	8	0	0	0	0	0	45.2	4.63	59.82	8.0	2.0	.0	.0	49.8	.0	−2.00	40.55	41.82	3	2
Villabruna (16055–16410)	1.16	0	0	0	4	5	0	0	0	0	0	30.1	4.63	54.39	9.0	6.0	.0	.0	34.8	1.2	−1.00	25.49	25.76	3	2
Handt (16055–16410)	1.16	1	0	0	1	1	2	2.32	2.32	1.16	0	2.32	0	10.64	2.0	4.0	.0	.0	3.5	3.5	.00	1.16	1.48	3	3
Lalueza (16209–16410)	1.08	0	1	0	0	0	2	1.16	1.16	0	1.08	3.23	0	5.30	.0	.0	.0	1.0	3.2	1.2	.00	3.23	3.23	1	4
Neanderthal2 (16055–16400)	1.18	1	0	0	0	1	0	0	0	0	0	3.53	1.18	5.70	1.0	.0	.0	.0	4.7	1.1	−1.00	2.35	3.70	2	5
Poinar (16131–16218)	1.38	1	0	0	2	2	0	0	0	0	0	5.51	0	10.51	4.0	.0	.0	1.0	5.5	.0	.00	5.51	1.51	2	6
H. sapiens neanderthalis:																									
Neanderthal1 (16055–16400)	1.18	1	0	5	6	10	0	2.35	0	0	0	17.64	0	48.87	16.0	5.0	.0	2.0	23.5	2.4	−4.00	11.76	7.52	2	7
Dinornis giganticus:																									
Dinornis (11120–11958)	1.34	0	0	1	2	1	1	2.36	2.36	0	0	2.36	0	9.72	3.0	2.0	.0	.0	2.4	2.4	1.00	2.36	−.64	1	8
Dinornis12s (1753–2148)	1.1	0	1	0	3	1	0	0	0	0	0	1.1	0	5.10	4.0	.0	.0	.0	1.1	.0	2.00	1.10	−2.90	3	8
DinornisCOI (7807–8325)	1.3	1	0	0	6	0	0	0	0	0	0	1.3	0	8.30	6.0	.0	1.0	.0	1.3	.0	6.00	1.30	−4.70	3	8
DinornisCOII (8861–9349)	1.14	0	0	0	0	1	0	1.14	1.14	0	0	4.56	0	6.70	1.0	.0	.0	.0	4.6	1.1	−1.00	4.56	3.56	3	8
DinornisCR (16733–00441)	1.34	0	0	1	2	7	0	1.34	1.34	0	0	5.34	2.67	18.35	9.0	.0	.0	.0	8.0	1.3	−5.00	2.67	−.99	3	8
DinornisCytb (15303–15783)	1.4	0	0	0	1	0	0	0	0	0	0	2.8	0	3.80	1.0	.0	.0	.0	2.8	.0	1.00	2.80	1.80	3	8
DiornisnND1 (4747–5201)	1.35	0	0	0	2	0	0	0	1.35	0	0	5.4	0	8.75	2.0	.0	.0	.0	5.4	1.4	2.00	5.40	3.40	3	8
Emeus crassus:																									
Emeus16s (3787–4311)	1.33	0	0	1	1	1	0	0	0	0	0	13.3	0	15.30	2.0	.0	.0	.0	13.3	.0	.00	13.30	11.30	3	8
EmeusCOI (7807–8328)	1.33	0	0	0	0	1	0	0	0	0	0	3.99	0	4.99	1.0	.0	.0	.0	4.0	.0	−1.00	2.99	2.99	3	8
EmeusCOII (8320–8807)	1.3	0	0	0	3	3	0	0	1.3	0	1.3	6.5	0	16.40	6.0	.0	1.3	.0	6.5	1.3	.00	6.50	.50	3	8
EmeusCOIII (10161–10743)	1.16	0	0	0	2	3	0	1.16	0	1.16	0	6.96	6.96	22.40	5.0	.0	2.3	.0	13.9	2.3	−1.00	8.92	8.92	3	8
EmeusND4/5 (12788–13200)	1.4	0	0	0	3	2	0	0	0	0	0	7	0	12.00	5.0	.0	.0	.0	7.0	.0	1.00	2.00	2.00	3	8
Mulleronis agilis:																									
Mulleronis12s (1856–2020)	1.15	0	0	0	0	0	0	1.15	1.15	0	0	5.75	0	6.90	.0	.0	.0	.0	5.8	1.2	.00	5.75	5.75	3	8
Ursus arctos:																									
Ursus147 (control region)	1.07	0	0	0	2.28	2.28	0	0	0	0	0	2	0	5.28	2.3	.0	.0	.0	3.0	.0	2.28	1.00	.72	1	9
Ursus221a (control region)	1.07	0	0	0	1.14	1.14	0	0	0	0	0	2	0	6.28	2.3	.0	.0	.0	4.0	.0	.00	.00	1.72	1	9
Ursus222 (control region)	1.07	0	0	1	0	0	1.14	0	0	0	0	3	0	15.28	1.1	.0	1.0	.0	12.0	.0	−6.00	−6.00	10.86	1	9
Ursus223a (control region)	1.07	1.14	0	0	1.14	2.28	0	0	0	0	0	1	0	6.56	3.4	.0	1.1	.0	2.0	.0	−1.14	.00	−1.42	1	9
Ursus221b (12s)	.9	0	0	0	0	1.07	0	0	0	0	0	8	2	11.07	1.1	.0	.0	.0	10.0	.0	−1.14	6.00	8.93	1	1
Ursus223b (12s)	.88	0	0	0	3.2	0	1.07	0	0	0	0	0	0	4.27	3.2	1.1	.0	.0	.0	.0	3.20	.00	−3.20	1	1
U. spelaeus:																									
Ursus47910 (control region)	1.84	0	0	0	0	0	0	0	0	0	1.84	1	0	1.00	1.0	.0	.0	.0	1.0	.0	.00	1.00	1.00	2	10
UrsusCLA 1 (control region)	1.84	0	1	2	1	2	0	1.84	5.52	1.84	0	0	0	16.20	3.0	3.0	.0	1.0	.0	7.4	−1.00	−3.00	−3.00	2	10
UrsusSC11700 1 (control region)	1.84	0	0	1	1	1	2	0	0	0	3.68	0	3.68	8.68	2.0	3.0	.0	.0	3.7	.0	.00	−3.68	1.68	2	10
UrsusSC157001 (control region)	1.84	0	1	0	0	0	0	0	0	0	3.68	3.68	0	3.68	.0	.0	.0	.0	3.7	.0	.00	3.68	3.68	2	10
UrsusSC5300 (control region)	1.84	1	0	0	0	0	0	0	0	0	3.68	3.68	0	5.68	1.0	.0	1.0	.0	3.7	.0	−1.00	3.68	2.68	2	10
UrsusSCL3800 (control region)	1.84	0	0	0	0	0	1	0	0	1.84	0	0	0	2.84	1.0	.0	.0	.0	.0	1.8	.00	.00	.00	2	10
UrsusSCL3500 (control region)	1.84	1	0	0	0	1	0	0	0	1.84	0	1.84	0	6.68	1.0	1.0	.0	1.0	1.8	1.8	1.00	1.84	.84	2	10
UrsusTAB151a (control region)	1.84	1	0	0	0	0	2	0	1.84	0	1.84	1.84	0	10.36	2.0	.0	1.8	.0	1.8	5.5	−2.00	1.84	−.16	2	10
UrsusTAB21a (control region)	1.84	0	0	0	1	0	0	0	0	5.52	0	5.52	1.84	15.88	.0	.0	1.0	.0	7.4	5.5	−1.00	3.68	6.36	2	10

[a] Human regions are numbered with reference to the Cambridge reference sequence (Anderson et al. 1981), ratite regions are numbered with reference to Cooper et al. (2001), and bear regions are located by gene. All cloned regions are mitochondrial.

[b] Ratio used to scale damage events occurring on G and C, to allow for the composition bias of each cloned region, calculated as (total A and T bases)/(total C and G bases) per cloned region.

[c] Scaled counts of base modifications per sample subdivided into the 12 hypothetical postmortem base-modification events, represented as "i→j" (with "i" being the original nucleotide and "j" being the modified nucleotide).

[d] Complementary groups are as with "Individual Changes" but with subdivision into the six complementary pairings.

[e] Bias within type 1 transitions, within type 2 transitions, and between types 2 and 1 transitions respectively.

[f] Categorical coding of polymerase enzymes used in studies: 1 = Platinum Taq Hifidelity; 2 = AmpliTaq Gold; 3 = Taq DNA Polymerase.

[g] Original sources of data: 1 = Gilbert et al. (2003); 2 = Di Benedetto et al. (2000); 3 = Handt et al. (1996); 4 = Lalueza-Fox et al. (2001); 5 = Ovchinnikov et al. (2000); 6 = Poinar et al. (2001); 7 = Krings et al. (1997); 8 = Cooper et al. (2001); 9 = Barnes et al. (2002); 10 = Loreille et al. (2001).

[h] Total sum of modified damage per cloned region. Apparent discrepancies are due to rounding.

Material and Methods

The present study uses the large data set of cloned ancient human mitochondrial sequences from the companion article (Gilbert et al. 2003), as well as several other studies of ancient humans and Neanderthals (Handt et al. 1996; Krings et al. 1997; Di Benedetto et al. 2000; Ovchinnikov et al. 2000; Lalueza-Fox et al. 2001; Poinar et al. 2001), bears (Loreille et al. 2001; Barnes et al. 2002), and ratites (Cooper et al. 2001). Full details of the samples and sequences are given in table 1. Base damage and authentication criteria follow the method of Gilbert et al. (2003), and insertions/deletions were removed from all data sets. By convention, all sequences are described in the L-strand orientation.

The majority of the data analyzed were generated using low-error-rate polymerases (e.g., Platinum Hifi [Invitrogen], an enzyme mixture composed of recombinant *Taq* DNA polymerase, *Pyrococcus* spp. GB-D thermostable polymerase, and Platinum *Taq* Antibody), which have been shown to generate very few errors, allowing site variation in cloned sequences to be attributed to miscoding lesions (Willerslev et al. 1999). To test the fidelity of the low-error-rate polymerase, we amplified three modern human samples extracted with a standard phenol:chloroform protocol (Hillis et al. 1996), on two occasions, after the method of Barnes et al. (2002) but without secondary reamplification. Mitochondrial hypervariable region 1 (HVR1) primer pair L16209/H16356 (Handt et al. 1996) were used, and a large number of clones were sequenced ($n = 16$ per PCR). In addition, misincorporation rates for standard *Taq* polymerase that were determined from six data sets were analyzed (Dunning et al. 1988; Saiki et al. 1988; Tindall and Kunkel 1988; Eckert and Kunkel 1990; Sanson et al. 2002). Other enzymes used in the data set are indicated in table 2.

To test whether enzyme-misincorporation rates are modified by the local environment of the ancient extracts, we examined modern contaminants in two sets of aDNA extracts, in detail. Three 18th-century human teeth from Denmark (supplied by N. Lynnerup) were assayed for modern bacterial contaminants, using primers *rpoB,* designed to target the RNA polymerase β-subunit–encoding gene (Drancourt et al. 1998). PCR products were cloned, and the spectrum of damage was examined. Similarly, clones of obvious modern human mtDNA sequences amplified from three ancient Nordic and two Neanderthal teeth (supplied by N. Lynnerup and C. Lalueza-Fox, respectively) were assayed for levels of damage. The primer pairs used were L16055/H16410 and L16209/H16356 (Handt et al. 1996), and amplification and cloning followed the method of Barnes et al. (2002).

The mitochondrial region sequenced for each sample varies within and between the data sets, and the term "cloned region" is therefore used to define each independently amplified area bound by a primer pair. For the largest data set (Gilbert et al. 2003), a measure of DNA damage, *d,* was calculated as $d = D/Lt$, where D is the total number of base changes observed per cloned region, L is the base length of an amplified sequence, and t is the number of independent PCRs amplified. The null hypothesis, H_0, that no significant correlation exists between d and either sample age or archaeological site of origin was tested using the general linear model (GLM) function of the statistical program Minitab.

The most common damage-driven base changes observed in aDNA sequences are the four transitions: C→T, G→A, T→C, and A→G (Hansen et al. 2001). However, because of the complementary nature of DNA, each of these observations can be explained by two possible causative events (Hofreiter et al. 2001). Figure 1 demonstrates this for an observed C→T transition on the L strand. Because of an original damage event, on the L strand, causing C→U, after two stages of replication, we observe the C→T transition on the L strand (fig. 1*A*). However, this phenotype can also occur via an H-strand G→A transition, which, after one PCR cycle, is observed as the C→T transition on the L strand (fig. 1*B*). The same problem applies to each of the transitions, so that each postmortem biochemical change can result in two observed outcomes, depending on which strand is sequenced. Because of this complementarity, Hansen et al. (2001) have termed A→G and T→C changes as "type 1 transitions" (A→G/T→C) and C→T and G→A changes as "type 2 transitions" (C→T/G→A), to indicate the uncertainty about which base was originally damaged.

Although this situation appears intractable, a solution is offered by the possible biochemical pathways by which nucleotide damage can occur. Hofreiter et al. (2001) demonstrated that the damage-driven modification of G to an A analogue is highly unlikely, if not impossible. This excludes the possibility of the events shown in figure 1*B*. It can therefore be argued that any G→A transition that is observed on the L strand and due to damage must have originated as an H-strand C→T modification event, because G→A modification on the L strand is impossible. Conversely, any C→T transition observed on the L strand will actually have originated as an L-strand C→T modification event.

A similar argument can be applied to type 1 damage by assuming that modification of T→C analogues is biochemically unlikely. In this situation, any L-strand T→C modification will actually be due to an H-strand A→G event, whereas L-strand A→G events can be attributed to an original A→G damage event on the L strand. However, in this case, the logic is potentially weakened by in vivo and in vitro studies, of several polymerases, that have shown that a major oxidative derivative of thymine, 5-formyluracil (fU), has the capacity to pair with A, T,
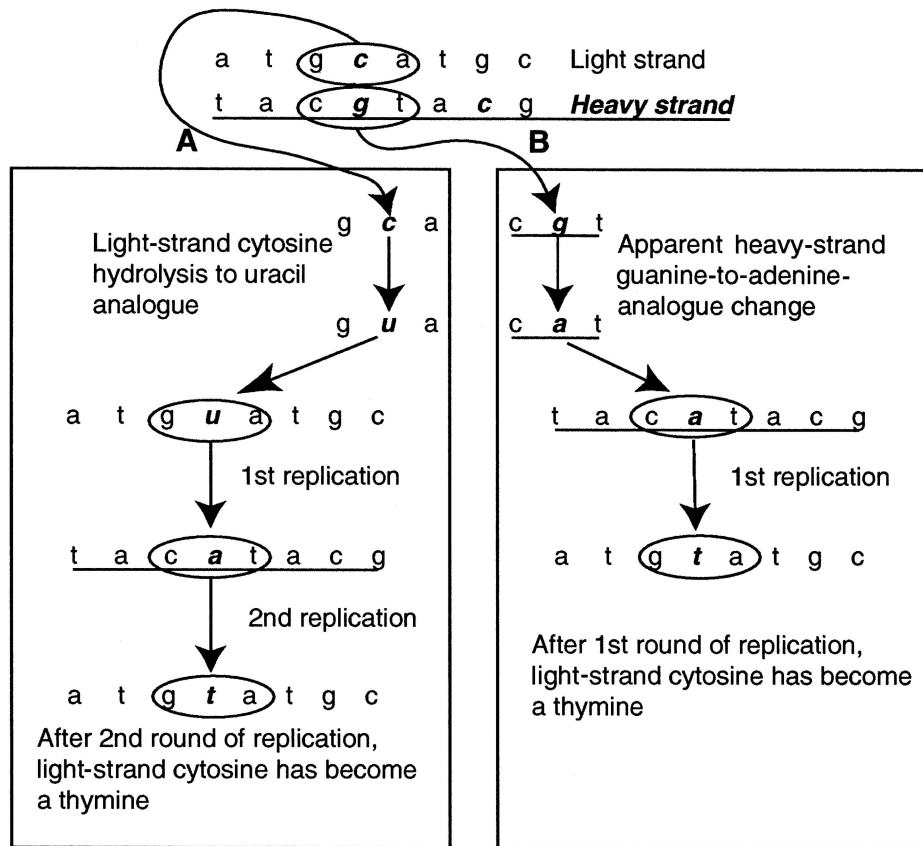
Figure 1 Determination of a strand of origin for postmortem-DNA-damage events by using type 2 (C→T/G→A) transitions as an example. *A*, L-strand C→T transitions after two cycles of amplifications, resulting in a permanent L-strand change. *B*, A theoretical H-strand G→A change, producing the L-strand phenotype of C→T change following one cycle of amplification. However, since a direct G→A postmortem modification is chemically impossible, the example depicted in this panel is not possible. Thus, all C→T changes observed on the L strand must have occurred as L-strand C→T postmortem damage, and all G→A changes on the L strand must have occurred as H-strand C→T postmortem damage.

G, or C (Yoshida et al. 1997; Zhang et al. 1997, 1999; Fujikawa et al. 1998). The pairing of fU:G, producing a T→C modification, has also been demonstrated as the most common of these mispairings (Ånensen et al. 2001). To test whether this process is observed in the ancient sequences, we performed an SP-PCR, to observe the spectrum of postmortem damage without the effects of jumping-PCR recombining substitutions between H and L strands (Hofreiter et al. 2001). Four samples (Tg129, Tg149, Tg232, and Tg233) were each amplified for 25 cycles, using either the mitochondrial L-strand primer L16209 or the H-strand primer H16356 (Handt et al. 1996), followed by 45 cycles of amplification with both primers. The initial, single-stranded phase enriches one of the DNA strands (H or L) prior to conventional PCR amplification, heavily biasing the final product toward the strand initially amplified. Clones were examined for sequences containing both A→G and T→C events, which normally would imply a jumping-PCR event but in

SP-PCR data would confirm the existence of fU:G or HX:C pairings.

This hypothesized skewed distribution of damage provides a method to discriminate the strand on which the original transition-inducing damage event occurred. Furthermore, because each cloned sequence originates from a single H- or L-strand template (but not both), it is possible to contrast the spectrum of damage events occurring on either DNA strand. Last, if in vivo base-modification events are similar to those postmortem events, it should be possible to examine modern sequence data sets and determine the strand on which any observed transition events originally occurred.

The distribution of L- and H-strand damage events in human aDNA sequences was examined using the Gilbert et al. (2003) data set. The total number and location of damaged positions on the L and H strands was determined and was compared to the number and distri-
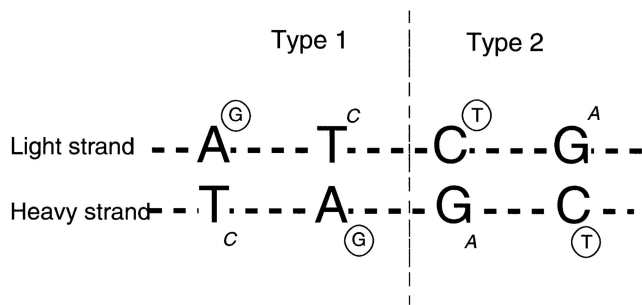
Figure 2 Type 1 and type 2 damage–induced transitions. Circled letters represent the principle modifications observed in cloned sequences (e.g., deamination of C→U [read as T] or A→HX [read as G]). Changes introduced on the complementary strand when the damaged bases are subsequently copied are shown in italics. By convention, sequences are referred to in the L-strand orientation. Therefore, if an amplified sequence was initiated from an original H-strand template, then the type 1 and type 2 errors observed are expected to be T→C and G→A, respectively.

bution expected if there was no strand bias, using a χ^2 goodness-of-fit test. The data were scaled to take local base composition into account (generally the underrepresentation of C and G) by multiplying the number of base changes originating on a C and G by the ratio of (C + G):(A + T) across the cloned region.

To investigate the ratio and distribution of type 1 (A→G/T→C) and type 2 (C→T/G→A) damage events in the postmortem data set, we measured the absolute number of each of the 12 possible base changes (A→C, A→G, A→T, C→A, C→G, C→T, G→A, G→C, G→T, T→A, T→C, and T→G) for each cloned region, and this was scaled for composition bias (as described above). A second data set was also created, to include the six complementary changes (e.g., T→C/A→G, etc.) (Hansen et al. 2001; Hofreiter et al. 2001). For the examination of variation in the ratio of the type 2:type 1 transitions, a value, β, for each region was calculated and was equal to the number of type 2–type 1 events. The bias toward either damage event was correlated with the overall extent of template damage or with different polymerase enzymes, by using a GLM.

Within type 2 and type 1 transitions, a similar test was performed to determine which original template strand was damaged, by calculating the products (C→T)-(G→A) for type 2 and (A→G)-(T→C) for type 1. For example, amplifications that are initiated from an H-strand template will potentially show H-strand–specific components of both type 2 (G→A) and type 1 (T→C) damage (fig. 2). Conversely, amplifications that derive from L-strand templates will contain L-strand–specific components (i.e., C→T and A→G). Therefore, if H and L strands are equally represented in a DNA extract, then cloned

sequences from a single PCR should show approximately equal numbers of sequences containing H- and L-strand–specific components of damage. A significant deviation from this ratio would suggest that the extract contains a bias of either H or L strands or that amplification is initiated preferentially on one strand. Furthermore, miscoding lesions observed in any single cloned sequence can be used to examine whether all type 1 damage arises from A→G transitions. This is because biases toward H- or L-strand–specific events should be mirrored in both type 1 and type 2 changes. Deviations from this correlation would provide evidence for nonzero rates of T→C, which, although biochemically unlikely (Lindahl 1993), have not been experimentally investigated in aDNA.

The tight correlation between strand-specific type 1 and type 2 transition modifications also provides a means to examine the nature and frequency of jumping-PCR artifacts (Pääbo et al. 1990), in which templates recombine during amplification. Previously, jumping events have been identified when substitutions from different genotypes appear on the same amplified strand, in a chimeric sequence (fig. 3). However, when only one genotype is present in a PCR, a jumping event may be
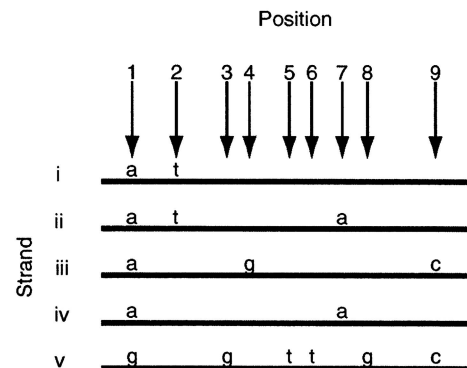
Figure 3 Jumping PCR (Pääbo et al. 1990). Strands i–v represent five sequences obtained from the cloned product of an individual PCR based on one extraction, using a low-error-rate enzyme such as Platinum Taq Hifidelity (Invitrogen). Positions 1–9 represent nucleotide positions that differ between strands, with the altered nucleotide marked above the strand. The shared adenine (a) base on strands i–iv at position 1 helps determine that they derive from one source (though not template molecule) of DNA, with other differences arising due to hydrolytic damage and jumping PCR. Positions 2, 4, and 7 on strands i–iv are base changes resulting from DNA damage. Differences in strand v at positions 1, 3, 5, 6, 8, and 9 identify it as a contaminant. Under the assumption that transitions at identical positions are rare, the shared thymine (t) at position 2 indicates that strands i and ii derived from one template molecule with damage at position 2. The shared adenine (a) base at position 7 on strands ii and iv, in contrast to differences at position 2, indicates jumping PCR between the two strands. Finally, position 9 on strands iii and v represents apparent damage to strand C, arising from jumping with the contaminant strand v.

Table 3

Samples Containing Association Groups *A, B,* and *C*

| | RESULTS FROM ASSOCIATION GROUP | | | | | |
| | *A* | | *B* | | *C* | |
SAMPLE (REGION)	C→T, A→G	G→A, T→C	C→T, G→A	A→G, T→C	C→T, T→C	A→G, G→A
DinornisCR (16733–00441)	1		1	1j	2	2j
DinornisCOII (COII)					1	
Dinornis (11120–11958)	1				1	
Emeus16s (16s)	1					
EmeusCOI (COI)					1	
EmeusCOIII (COIII)		1	1			
EmeusND4/5 (ND4/ND5)	1				1, 1j	
EmeusCOII (COII)	1					
EmeusCytb (cytb)	1				2	
Ursus221a	1					
Ursus222	1					
Ursus223a		1	1			
Ursus147		1		1		
Ursus221b			1			
Ursus3500	2					
Ursus117001		1				
Ursus151a		1	1j			
Tg233 (HVRI)	1					
Tg232 (HVRI)	1					1j
Tg116 (HVRI)	2j				1j	
Tg114 (HVRI)	1j		1j		1j	1j
Tg103 (HVRI)	2			2j		
Tg54 (HVRI)	1j					
Tg145 (HVRI)	1					
Tg44 (HVRI)	1					
Tg149 (HVRI)	3j, 2		2j		3j, 3	2j, 1
Tg137a (HVRI)	1					
Tg112 (HVRI)	1					
Tg105 (COIII)	1					
Tg63 (COIII)	1		1			
Tg148 (HVRI)		1			1	1
Tg136a (HVRI)			1j		1	
Tg147 (HVRI)					1j, 2	
Tg85 (HVRI)			1		1j	
Tg93 (HVRI)					1, 1j	
Tg148 (COIII)			1	1j		
Tg143 (HVRI)				1		
Tg129 (HVRI)				1j		
Tg123 (HVRI)			2j			

| | RESULTS FROM ASSOCIATION GROUP | | | | | |
| | *A* | | *B* | | *C* | |
INDIVIDUAL TOTALS	C→T, A→G	G→A, T→C	C→T, G→A	A→G, T→C	C→T, T→C	A→G, G→A
Identified jumps	7	0	7	5	9	6
Nonidentified jumps	22	6	7	2	16	2

| | RESULTS FROM ASSOCIATION GROUP | | |
GROUP TOTALS	*A*	*B*	*C*
Associations per group	35	21	33
Jumps	7	12	15
Nonjumps	28	9	18
Nonjump:jump ratio	4	.75	1.2

NOTE.—The three association groups are each divided into two subgroups named according to the two different base modifications observed together on same cloned sequence. A "j" after the numeral denotes direct evidence of jumping as identified following the conventional method (e.g., Pääbo 1989), in which different cloned sequences sharing a proportion of damaged sites are deemed to have arisen because of jumping PCR. These are termed here as "identified jumps"; nonidentified jumps are those that cannot be identified following the conventional methods.

difficult to observe, and it is likely that the frequency of such events has been considerably underestimated. Without jumping events between templates, cloned sequences that contain both type 2 and type 1 damage should exhibit either H- or L-strand–specific components, but not both. In table 3, the number of clone sequences with such associations, *A,* is divided into those which can and cannot be explained through jumping PCR (because of the presence of substitutions shared by multiple nonhomologous cloned sequences). Sequences that contain both H- and L-strand–specific components of damage can have arisen only after a jumping event. These are subdivided into *B,* those that contain exclusively type 1 or type 2 transitions, and *C,* those that contain type 1 as well as type 2 transitions. These are again divided, as with *A.*

A χ^2 goodness-of-fit test was performed on the observed distribution of transition modifications among associations *A, B,* and *C* and on that expected under randomly paired base changes. Estimates of the number of jumping-PCR events within the cloned sequence data set before and after the identification of associations *B* and *C* were compared, to assess jump-detection efficiency.

Results

The cloned sequences derived from modern samples reveal a very high degree of enzyme fidelity, with only two deletions observed in the 14,112 bases determined. In the 6,451 bases of amplified modern bacterial DNA from the ancient extracts, only one C→A base change was observed, whereas, in >21,442 bases of human contaminant sequences, only five changes (three A→G and two C→T changes) were observed. Although these figures are somewhat higher than the enzyme manufacturer's published rate of 2×10^{-6}, they are vastly lower than the rates observed in our data (table 2). It is also possible that the presence of multiple contaminants causes an overestimation of the rate for the modern human DNA. Consequently, any alteration of enzyme fidelity by aDNA extracts seems insufficient to have an impact on the data.

Thirty-two clones derived from SP-PCR were examined, and in no cases were A→G and T→C modifications (or C→T and G→A modifications) seen on the same sequence. This provides evidence that oxidation of T to fU does not play a major role in postmortem-damage–derived miscoding lesions. Further evidence that fU lacks a role can be inferred from the spectrum of damage observed. Ånensen et al. (2001) report that fU-mediated mispairings of A→G/T→C occur in vivo ~10 times as often as those of G→A/C→T, G→T/C→A, or A→C/T→G. However, as seen in table 4, the rates of A→G/T→C occurrence in our data set are only half those of G→A/C→T and by comparison are almost 30 times the predicted rates of G→T/C→A and A→C/T→G.

The data give no indication that sample age correlates with damage (as measured by d, $P = .85$), although there is significant evidence that archaeological site is important. The H_0 that there is no correlation between damage and archaeological site may be rejected, with $P < .01$.

Table 5 presents the observed and expected measurements of damage on each of the L and H strands for the HVR1 data set analyzed in the companion article by Gilbert et al. (2003). Although the results suggest that the L strand receives more damaged sites, as well as more overall damage, proportional to the potential number of bases (A and C) that can change, a χ^2 goodness-of-fit test provides no statistical backing for either of these observations (for the total number of damage events on the L and H strands, $P = .66$; for the number of different base positions that are seen to modify on the L and H strands, $P = .55$).

The number of each of the six complementary change groups for each cloned region is shown in table 2, and table 4 gives the totals for each data set (of more than one cloned region) and gives the averages of the control *Taq* studies. A bias toward type 2 events is observable in the data (tables 2 and 4 and fig. 4*A*), particularly in the complete data sets, and is in agreement with previous studies on the spectrum of damage in ancient samples (Hansen et al. 2001; Hofreiter et al. 2001). However, within individual cloned regions, the biases range considerably (figs. 4*B* and 4*C*), with a few samples (e.g., Tg99j, Tg129, Tg196, Moa12s, and MoaCOI) displaying a high bias toward type 1, which is characteristic of *Taq* misincorporation (Hansen et al. 2001). Interestingly, these biases correlate poorly with enzyme ($P = .574$), although there is a strong positive correlation between type 2:type 1 bias and the extent of overall damage ($P < .00$).

The number of modified bases and bias measurements for all 12 possible base changes are also given in tables 2 and 4. As seen in figure 5*A*, data sets, as a whole, demonstrate a bias toward C→T change within type 2 damage. However, such a bias is less apparent within type 1 damage. When individual cloned regions are examined, it is apparent that, although the strength and direction of type 2 biases vary, overall there is a C→T trend (figs. 5*B* and 5*C*). The correlation between bias and enzyme is not significant ($P = .40$), but bias is highly significant with damage ($P < .00$). The range of bias within type 1 damage follows a similar trend, although overall the skew is less pronounced. Although there is no significant correlation with enzyme ($P = .19$), the GLM analysis shows a significant correlation between A→G bias and damage ($P < .05$). Therefore, for both type 1 and type 2 events, there is a trend toward L-strand–specific damage as the overall extent of damage increases. This finding, combined with the lack of apparent strand bias in damage accumulation, sug-

Table 4

Postmortem-Damage Measurements and Bias Calculations Summarized by Study

DATA SET[a] (No. of Cloned Regions)	A→C, T→G			A→G, T→C			A→T, T→A			C→A, G→T			C→G, G→C			C→T, G→A			TYPE 1, (A→G)-(T→C)	TYPE 2, (C→T)-(G→A)	TYPES 2-1
	AC	TG	Total	AG	TC	Total	AT	TA	Total	CA	GT	Total	CG	GC	Total	CT	GA	Total			
Control[b]			1.6			62.6			10.2			7.4			3.9			14.3			-48.3
TG1 (55)	12	6	18	94	77	171	5	3	8	6.14	0	6.14	5	1.3	6.3	288	77.7	365.7	17.0	209.9	194.3
AC (13)	0	1	1	2.5	20	45	1	1	2	8.4	1.3	9.7	2.5	2.5	5	66.4	9.6	76	5.0	56.8	31
IB (4)	1.1	0	1.1	4.6	4.6	9.2	0	1.1	1.1	0	0	0	0	1	1	8	13	21	.0	−5.0	11.8
OL (9)	0	5	5	3	7	10	5	4	9	0	1.8	1.8	12.9	9.2	22.1	17.6	5.5	23.1	−4.0	12.1	13.1
TG2 (2)	0	0	0	3.2	1.1	4.3	0	1.1	1.1	0	0	0	0	0	0	8	2	10	2.1	6.0	5.7
GD (3)	1	0	1	13	18	31	9	0	9	2.3	1.2	3.5	0	1.2	1.2	99.6	11.6	111.2	−5.0	88.0	80.2
IO (1)	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	3.5	1.2	4.7	−1.0	2.3	3.7
CL (1)	0	1	1	0	0	0	0	0	0	0	0	0	1.1	0	1.1	3.2	0	3.2	.0	3.2	3.2
HP (1)	1	0	1	2	2	4	0	0	0	0	0	0	0	0	0	5.5	0	5.5	.0	5.5	1.5
MK (1)	1	1	2	6	10	16	5	0	5	0	2.4	2.4	0	0	0	17.6	5.9	23.5	−4.0	11.7	7.5
OH (1)	0	0	0	1	1	2	2	2	4	0	1.2	1.2	0	0	0	2.3	1.2	3.5	.0	1.1	1.5

[a] TG1 = Gilbert et al. (2003), humans; AC = Cooper et al. (2001); IB = Barnes et al. (2002); OL = Loreille et al. (2001); TG2 = Gilbert et al. (2003), bears; D = Di Benedetto et al. (2000); IO = Ovchinnikov et al. (2000); CL = Lalueza-Fox et al. (2001); HP = Poinar et al. (2001); MK = Krings et al. (1997); OH = Handt et al. (1996).

[b] *Taq* misincorporation measurements, presented as a percentage and calculated from the following data sets: Saiki et al. (1998), Dunning et al. (1988), Tindall and Kunkal (1988), Eckert and Kunkel (1990), and Sanson et al. (2002).

gests either that there is greater survival of L strands or, at least, that there is a prevalence in an amplifiable condition.

Table 3 presents the data on the linkage of damage events in individual clones and the evidence for jumping-PCR artifacts. As expected, pairings are not randomly spread among *A, B,* and *C,* and the majority of associated damage events are consistent with a non-jumping origin (χ^2 goodness-of-fit test $P = .14$). Without the data on associated type 1 and type 2 changes, 34 obvious jumping artifacts can be seen within the data set as chimeric sequences. However, the patterns identified in *B* and *C* groups identify another 27 pairings (an increase of 80%) that can only be explained through jumping (if G→A and T→C transitions are assumed to be chemically impossible).

Discussion

The large data set of ancient sequences has provided new insights into the spectrum and distribution of DNA damage, at both the nucleotide and DNA-strand level. The very low misincorporation rate observed for the high-fidelity enzyme in the control experiments indicates that almost all of the base modifications observed in the ancient sequences are likely to be attributable to damage of the original template strand. The data seem robust and are independent of enzyme or region of DNA. The few departures from the general pattern (e.g., a higher rate of type 1 changes) occur in samples with relatively low levels of damage and are likely to result from sampling stochasticity.

The observed bias toward type 2 over type 1 transitions increases with the overall extent of damage, in agreement with previous aDNA studies (Willerslev et al. 1999; Hofreiter et al. 2001) and the hypothesis, of Hansen et al. (2001), that type 1 transitions occur at a slightly slower rate than those of type 2. However, the ratio is nowhere near the in vivo ratio of 30–50 times (Lindahl 1993), suggesting that the factors involved in hydrolytic damage vary somewhat between postmortem and in vivo situations.

The rate of type 1 damage observed in the data set is similar to other studies (Hansen et al. 2001) and is much higher than that reported by Hofreiter et al. (2001), who recorded a level even lower than the very rare C→A/G→T transversion modifications. This discrepancy is hard to explain, but may be related to the chemically modified *Taq* polymerase (Ampli*Taq* Gold; ABI) used by Hofreiter et al. (2001). For example, perhaps one of the products in the A→HX hydrolytic deamination pathway of type 1 modifications may hinder the action of Ampli*Taq* Gold relative to Hifi polymerases. However, other aDNA studies that have used Ampli*Taq* Gold (Krings et al. 1997; Loreille et al. 2001; Poinar et al. 2001) do not report such a low type 1 rate.

It is significant that the two types of transition modifications can be used to identify the original template strand ancestral to any individual PCR amplicon, because this provides a means to investigate the mode of DNA survival after death. In the postmortem data, it is apparent that, with an increase in the extent of damage, there is a parallel increase in the number of amplicons originating from L-strand templates. This is interesting, since there appears to be no overall strand-specific bias in transition damage, once base composition is taken into account. Furthermore, in living cells, the eponymous displaced 7S strand of the D-loop implies that there should be two copies of the H-strand template for every L strand in this region. The transition data indicate that this extra H strand is not available for amplification, and it is possible that the exposed position contributes to rapid postmortem degradation. Even when we allow for this, the apparently

Table 5

H- and L-Strand Damage Measurements at Positions 16209–16356

			HITS			SITES		
STRAND	C/A[a]	C/A:G/T[b]	Total Hits[c]	Hit Ratio[d]	Expected Hits[e]	Sites Hit[f]	Hit Ratio[g]	Expected Hits[h]
L	110	2.9	119	3.1	116.6	95	3.3	92.1
H	38		38		40.4	29		31.9
Total	148		157		157	124		124

NOTE.—Calculated from the Gilbert et al. (2003) data set.

[a] Total C and adenosine bases on the L strand of the Cambridge reference sequence (Anderson et al. 1981) at positions 16209–16356.

[b] Ratio of C/A:G/T bases on the L strand of the Cambridge reference sequence.

[c] Total number of damage events on the L and H strands.

[d] Ratio of L:H total hits.

[e] Expected number of hits on each strand, using the C/A:G/T ratio if damage is nonrandomly distributed.

[f] Number of different base positions that are seen to modify on the L and H strands.

[g] Ratio of L:H sites hits.

[h] Expected number of sites hit on each strand, using the C/A:G/T ratio if damage is nonrandomly distributed.
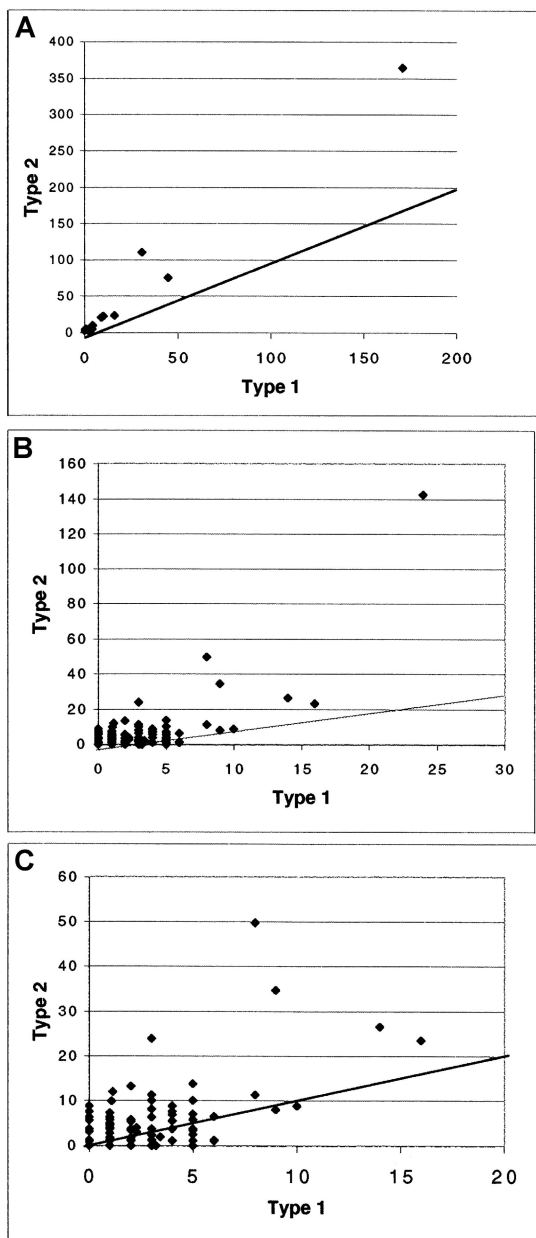
Figure 4 Type 1 versus type 2 damage. *A,* Damage per study. All studies demonstrate a type 2 bias. *B,* Damage per clone region. Although the type 2 bias is significant, many samples demonstrate a type 1 bias. This is seen clearly in panel C, in which the outlier is removed, to give greater resolution.

damage forms such as hydantoins, which block replication but leave the DNA structurally sound (Höss et al. 1996). It will be important to apply the methods developed here to the reexamination of mutations in modern human data sets, to determine the strand originally damaged. Such information will facilitate the development of a secondary-structure model for the human D-loop, a development that is an important requirement for the accurate interpretation of sequence evolution within modern humans.

The ability to identify the original template strand has allowed a reevaluation of the extent of jumping PCR within aDNA amplifications. This is an important means to detect haplotypes generated by PCR artifacts (Gilbert et al. 2003), and the high rate detected here is cause for concern about many existing aDNA studies. Jumping permits recombination between damaged templates, contaminants, nuclear copies, and the endogenous sequence, potentially generating a wide range of sequences. The number of jumping events suggested in table 3 is likely to be a considerable underestimate of the true level, because recombination between homologous template strands (i.e., both H or both L) is not detected. Jumping PCR is believed to be positively correlated with sequence damage (Pääbo et al. 1990), and rate estimates may therefore provide a simple way of comparing the DNA preservation within samples, as well as a way of scrutinizing the results for authenticity.

The large data set of ancient sequences has also permitted the first direct and statistically significant demonstration that the extent of DNA damage within a sample is correlated with archaeological site. Previously, this has been demonstrated only indirectly, by using correlates such as the frequency of water change (Nielsen-Marsh 2000), temperature (Höss et al. 1996; Smith et al. 2001), and microbial content (Burger et al. 1999) or by using biochemical data such as amino acid racemization (Poinar et al. 1996; Poinar and Stankiewicz 1999), composition (Bada et al. 1999), and levels of DNA-damaged bases (Höss et al. 1996). It will be important to investigate a range of archaeological sites and to use these methods to investigate how environmental factors affect DNA survival. For example, in vitro experiments predict that, in a constant environment, the DNA damage will correlate with age (Pääbo and Wilson 1991; Lindahl 1993). The lack of such a direct correlation in the present study is likely to be related to the temporal and geographical heterogeneity of specimens within individual sites, and factors such as the rate and extent of decomposition or desiccation before burial (Pääbo 1989). The discovery that the damage spectrum allows a detailed investigation of DNA-degradation processes provides a means to further investigate the role of such archaeological parameters. Furthermore, the data provide important insights
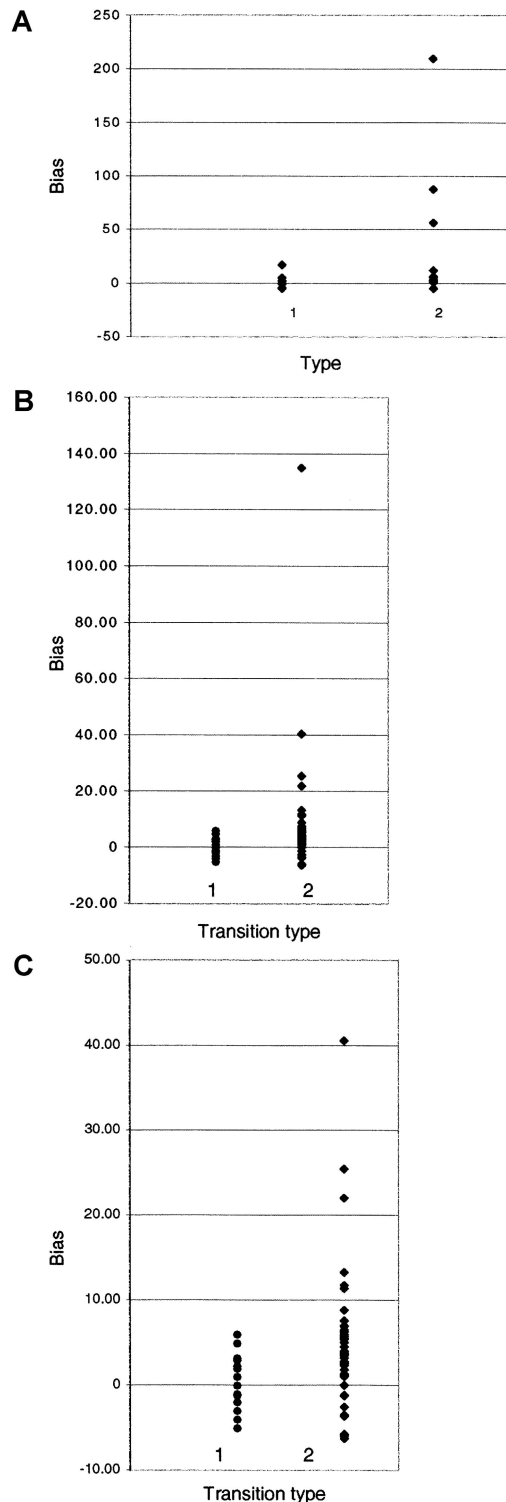
reduced rate of overall H-strand survival or amplification is unexpected. The former is unlikely, because an increased rate of H-strand degradation would also leave the remaining L strands single-stranded and vulnerable to rapid degradation (Lindahl 1993). Consequently, there may be some impediment to H-strand amplification with increasing damage, perhaps through

Figure 5 Damage bias within transition types. Positive *Y*-axis values represent a bias toward type 1 A→G transitions and type 2 C→T transitions. Negative values demonstrate a bias toward type 1 G→C transitions and type 2 T→C transitions. *A,* Data sets from whole studies. *B,* Data from individual cloned regions. *C,* Same as panel B but with high values removed to increase overall resolution.

into the biochemical background and likelihood of the sequence differences observed between living human groups. Such information is critical in understanding our recent evolutionary past.

Acknowledgments

References

Anderson S, Bankier A, Arrell B, de Bruijn M, Coulson A, Drouin J, Eperon I, Nierlich D, Roe B, Sanger F, Schreier P, Smith A, Staden R, Young I (1981) Sequence and organisation of the human mitochondrial genome. Nature 290:457–465

Ånensen H, Provan F, Lian A, Reinertsen S-H, Ueno Y, Matsuda A, Seeberg E, Bjelland S (2001) Mutations induced by 5-formyl-2′-deoxyuridine in *Escherichia coli* include base substitutions that can arise from mispairs of 5-formyluracil with guanine, cytosine and thymine. Mutat Res 476:99–107

Bada J, Wang X, Hamilton H (1999) Preservation of key biomolecules in the fossil record: current knowledge and future challenges. Philos Trans R Soc Lond B Biol Sci 354:77–87

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. Science 295:2267–2270

Burger J, Hummel S, Herrmann B, Henke W (1999) DNA preservation: a microsatellite-DNA study on ancient skeletal remains. Electophoresis 20:1722–1728

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature 409: 704–707

Di Benedetto G, Nasidze IS, Stenico M, Nigro L, Krings M, Lanziger M, Vigilant L, Stoneking M, Pääbo S, Barbujani G (2000) Mitochondrial DNA sequences in prehistoric human remains from the Alps. Eur J Hum Genet 8:669–677

Drancourt M, Aboudharam G, Signoli M, Dutour O, Raoult D (1998) Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. Proc Natl Acad Sci USA 95:12637–12640

Dunning AM, Talmud P, Humphries SE (1988) Errors in the polymerase chain reaction. Nucleic Acids Res 16:10393

Eckert KA, Kunkel TA (1990) The fidelity of DNA polymerase used in the polymerase chain reaction. In: McPherson MJ, Quirke P, Taylor GR (eds) PCR: a practical approach. Oxford University Press, Oxford, UK, pp 225–244

Fattorini P, Ciofuli R, Cossutta F, Giulianini P, Edomi P, Furlanut M, Previdere C (1999) Fidelity of polymerase chain reaction–direct sequencing analysis of damaged forensic samples. Electrophoresis 20:3349–3357

Fujikawa K, Kamiya H, Kasai H (1998) The mutations induced by oxidatively damaged nucleotides, 5-formyl-dUTP and 5-hydroxy-dCTP, in *Escherichia coli.* Nucleic Acids Res 26:4582–4587

Gilbert MTP, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, Cooper A (2003) Distribution patterns of postmortem damage in human mitochondrial DNA. Am J Hum Genet 72:32–47 (in this issue)

Handt O, Krings M, Ward R, Pääbo S (1996) The retrieval of ancient human DNA sequences. Am J Hum Genet 59:368–376

Hansen A, Willerslev E, Wiuf C, Mourier T, Arctander P (2001) Statistical evidence for miscoding lesions in ancient DNA templates. Mol Biol Evol 18:262–265

Hillis DM, Mable BK, Larson A, Davis SK, Zimmer EA (1996) Nucleic acids. IV. Sequencing and cloning. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics. Sinauer Associates, Sunderland, MA, pp 321–381

Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Res 29:4693–4799

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

Krings M, Stone A, Schmitz R, Krainitzki H, Stoneking M, Pääbo S (1997) Neanderthal DNA sequences and the origin of modern humans. Cell 90:19–30

Lalueza-Fox C, Luna Calderón F, Calafell F, Morera B, Bertranpetit J (2001) mtDNA from extinct Tainos and the peopling of the Caribbean. Ann Hum Genet 65:137–151

Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362:709–715

Loreille O, Orlando L, Patou-Mathis M, Philippe M, Taberlet P, Hänni C (2001) Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus,* and brown bear, *Ursus arctos,* lineages. Curr Biol 11:200–203

Nielsen-Marsh C (2000) Patterns of diagenesis in bone. I. Effects of site environments. J Archeol Sci 27:1139–1150

Ovchinnikov I, Götherström A, Romanova G, Kharitonov V, Lidén K, Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404:490–493

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

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

Pääbo S, Wilson AC (1991) Miocene DNA sequences—a dream come true? Curr Biol 1:45–46

Poinar H, Höss M, Bada J, Pääbo S (1996) Amino acid racemization and the preservation of ancient DNA. Science 272:864–866

Poinar HN, Kuch M, Sobolik KD, Barnes I, Stankiewicz AB, Kuder T, Spaulding WG, Bryant VM, Cooper A, Pääbo S (2001) A molecular analysis of dietary diversity for three archaic Native Americans. Proc Natl Acad Sci USA 98:4317–4322

Poinar H, Stankiewicz B (1999) Protein preservation and DNA retrieval from ancient tissues. Proc Natl Acad Sci USA 96:8426–8431

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. Science 239:487–491

Sanson GFO, Kawashita SY, Brunstein A, Briones MRS (2002) Experimental phylogeny of neutrally evolving DNA sequences generated by a bifurcate series of nested polymerase chain reactions. Mol Biol Evol 19:170–178

Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, Collins MJ (2001) Neanderthal DNA: not just old but old and cold? Nature 410:771–772

Tindall KR, Kunkel TA (1988) Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry 27:6008–6013

Wallace DC, Lott MT, Brown MD, Huoponen K, Torroni A (1995) Report of the committee on human mitochondrial DNA. In: Cuticchia AJ (ed) Human gene mapping 1995: a compendium. Johns Hopkins University Press, Baltimore, pp 910–954

Willerslev E, Hansen AJ, Christensen B, Steffensen JP, Arctander P (1999) Diversity of Holocene life forms in fossil glacier ice. Proc Natl Acad Sci USA 96:8017–8021

Yoshida M, Makino K, Morita H, Terato H, Ohyama Y, Ide H (1997) Substrate and mispairing properties of 5-formyl-2′-deoxyuridine 5′-triphosphate assessed by *in vitro* DNA polymerase reactions. Nucleic Acids Res 25:1570–1577

Zhang Q-M, Sugiyama H, Miyabe I, Matsuda S, Kino K, Saito I, Yonei S (1999) Replication *in vitro* and cleavage by restriction endonuclease of 5-formyluracil- and 5-hydroxymethyluracil-containing oligonucleotides. Int J Radiat Biol 75:59–65

Zhang Q-M, Sugiyama H, Miyabe I, Matsuda S, Saito I, Yonei S (1997) Replication of DNA templates containing 5-formyluracil, a major oxidative lesion of thymine in DNA. Nucleic Acids Res 25:3969–3973