An Assessment of the Use of Human Samples in Ancient DNA Studies

M Thomas P Gilbert

HenryWellcome Ancient Biomolecules Centre
Department of Zoology
University of Oxford
Trinity Term, 2003

Marcus Thomas Pius Gilbert
New College

Doctor of Philosophy
Trinity Term, 2003

An Assessment of the Use of Human Samples in Ancient DNA Studies

Abstract

This thesis addresses gaps that exist in the theory and knowledge of ancient DNA (aDNA). Much of the underlying basis of the field has been neglected in the excitement that followed the first aDNA studies. Therefore the results of many studies have been based on untested assumptions about the nature of *post mortem* DNA damage, sample preservation, contamination, and the efficacy of sample decontamination techniques. The validity of such results is questionable if the assumptions prove false.

Hydrolytic *post mortem* DNA damage may modify recovered aDNA sequences. This thesis reports new insights into the biochemical basis of, predisposition of certain sequences and nucleotide positions towards, and subsequent effects of, such damage. Parallels of *post mortem* damage with *in vivo* mutation also enable insights into DNA sequence evolution.

The long-term survival of DNA, and contamination of samples with exogenous DNA are two related problems characteristic to aDNA. The survival of endogenous DNA within bone, teeth and hair samples, the susceptibility of such samples to contamination, and the efficacy of decontamination techniques used to remedy such problems are investigated. The results highlight serious flaws in using bone and teeth as a DNA source. In contrast, the results demonstrate that hair may present a valuable DNA source for future studies.

Numerous studies have reported the retrieval of ancient pathogen DNA from human samples. Analyses of the DNA content within teeth extracted from putative victims of the 2nd plague argue that such studies are at great risk from DNA degradation, and contamination arising due to environmental microorganisms. An extrapolation of these results using basic physical and chemical theory is used to evaluate the potential survival of aDNA in ancient Egyptian remains. This suggests that positive results from such samples are unlikely.

Acknowledgements

Although my undergraduate degree at Oxford was in Biology, I had my heart set on switching into the exhilarating world of banking. I would therefore firstly like to thank my undergraduate tutor at Oriel college, Prof Sarah Randolph, my D.Phil sponsor, Dr Edward Holmes, and my D.Phil supervisor, Prof Alan Cooper for 'showing me the light', teaching me that Biology *can* be both exciting and rewarding, and saving me from that path. In addition, I have to thank Prof Cooper for his many illuminating discussions, his infectious enthusiasm for all things Ancient DNA, his helpful hints on writing technique, and those long hours of sitting through the conception of chapter 2 and 3 with me! In addition I would like to thank Dr Ian Barnes, Dr Beth Shapiro and Dr Michael Bunce for patiently teaching me everything I know about Molecular Biology, and all the other members of the ABC/Molecular Evolution group who provided so much help, ideas, and pub-time – Ross Barnett, Ollie Barnett, Greger Larson, Jen Jackson, Jan Strugnell, Phil Endicott, Shelley Cook, Adriana Grativol, Alexei Drummond – this means all of you.

Others to mention are those from who provided me with so much help throughout my diverse studies, freely giving me data, samples and advice and continually reading drafts of chapters. In particular I would like to mention (in no particular order) Dr Matthew Collins, Dr Hendrik Poinar, Dr Colin Smith, Dr Eske Willerslev, Dr Anders Hansen, Dr Lars Rudbeck, Dr Niels Lynnerup, Dr Carles Lalueza-Fox, Prof Martin Biddle, Birthe Kjølbye-Biddle, Dr Andy Wilson, Dr Des Tobin, Dr Mike Prentice and Prof Gordon Turner-Walker. Furthermore, I am deeply in debt to the Wellcome Trust who funded this D.Phil, on their now sadly extinct Bioarchaeology program. Both the Trust's financial generosity and 'student-support' schemes are excellent, and played a vital role in this project. Similarly, I must thank the Department of Zoology and New College Oxford, for their support. In addition I am very grateful to my examiners who stepped up to the challenge of my thesis – Dr Anthony Boyce of St John's College, Oxford, and Dr Mark Thomas of UCL.

On a more personal level, I would like to thank my other colleagues and friends for their friendship over the last 3 years. I haven't room to mention all of you, but I am grateful to you all. However, after three years of intellectual stimulation and discourse I have to single out one of you – Arnab Chatterjee – Thanks Arnie! In addition a big thank you goes to my family, and particular my wife, Trine Christensen for 3 years of support and encouragement, putting up with my tantrums, encouraging me in my lows, and generally motivating me to produce this thesis. Trine, du er lækker og fantastisk - Og jeg elsker dig!

Contents

List of common abbreviations

$\bullet O_2$	peroxide radical
•OH	hydroxyl radical
16s	b-subunit of the16s mitochondrial ribosomal RNA
^{32}P	Phosphor 32
5-OH-5-MeHyd	5-hydroxy-5-methylhydantoin
5-OH-Hyd	5-hydroxyhydantoin
A	adenine
AAR	Amino acid racemisation
ABC	Henry Wellcome Ancient Biomolecules Centre
aDNA	ancient DNA
Ala	alanine
Asx	aspartine/asparagine
BbRS	*Bos bison* reference sequence
BCE	Before Common Era
bp	base pairs
BSA	bovine serum albumin
BtRS	*Bos taurus* reference sequence
C	cytosine
CE	Common Era
CRS	Cambridge Reference Sequence (Anderson et al. 1981)
CSB1 – 3	mitochondrial HVR2 conserved sequence blocks 1 to 3
D-loop	mitochondrial Displacement-loop
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotide bases
EDTA	ethylene-diaminetetra-acetic acid
ETAS	Extended Termination-associated sequence
G	guanine
GC/MS	gas chromatography/mass spectroscopy
Glx	glycine/glutamine
GuSCN	guanidinium thiocyanate
H	mitochondrial DNA Heavy strand
H_20_2	hydrogen peroxide
HVR1	mitochondrial hypervariable 1 region
HVR2	mitochondrial hypervariable 2 region
hX	hypoxanthine
kya	thousand years ago
L	mitochondrial DNA Light strand
MAP	Maximum aposteriori
MCMC	Markov Chain Monte Carlo
ML	Maximum likelihood
MR	HVR1 middle region (as defined in chapter 3)
mtDNA	mitochondrial DNA
MW	molecular weight
mya	million years ago
OH	mitochondrial origin of H strand replication
OL	mitochondrial origin of L strand replication
OR	HVR1 outer region (as defined in chapter 3)
PCR	polymerase chain reaction

Pk	proteinase K
pla	*Yersinia pestis* plasminogen activator gene
PTB	N-phenacylthiazolium bromide
PTC	3-phenacyl-4,5-dimethylthiazolium chloride
RFLP	Restriction Digest Fragment Polymorphism
rpoB	RNA polymerase b-subunit encoding gene
SBH	St Bartholomew's Hospital
SDS	sodium dodecyl sulphate
Ser	serine
SNP	single nucleotide polymorphisms
T	thymine
TAS	Termination-associated sequence
TE	Trizma®-EDTA
U	uracil
X-gal	5-bromo-4-chloro-3-indoyl-b-D-galactopyranosidase

‘Think of how many religions attempt to validate themselves with prophecy. Think of how many people rely on these prophecies, however vague, however unfulfilled, to support or prop up their beliefs. Yet has there ever been a religion with the prophetic accuracy and reliability of science?

Is this worshipping the altar of science? Is this replacing one faith by another, equally arbitrary? In my view, not at all. The directly observed success of science is the reason I advocate its use. The reason science works so well is partly that built-in error-correcting machinery. There are no forbidden questions in science, no matters too sensitive or delicate to be probed, no sacred truths. That openness to new ideas, combined with the most rigorous, skeptical scrutiny of all ideas, sifts the wheat from the chaff. Diversity and debate are valued. Opinions are encouraged to contend-substantively and in depth.’

‘*The Demon-haunted World*’
Carl Sagan

Chapter 1

Introduction and scope of thesis

1.1 The history of ancient DNA

In 1984 Russ Higuchi and colleagues at Berkeley published a finding that was to revolutionise the scope of molecular biology - that traces of deoxyribonucleic acid (DNA) from a museum specimen of the Quagga, *Equus quagga*, (an Equid believed to have gone extinct in the late 19th century) not only remained in the specimen over 150 years after death of the individual, but could be extracted and sequenced (Higuchi *et al.* 1984). Over the next two years, through investigations into natural and artificially mummified specimens, Svante Pääbo both confirmed that this phenomena was not limited to relatively recent museum specimens, but could apparently be replicated in a range of mummified human samples that dated as far back as several thousand years (Pääbo 1985a; Pääbo 1985b; Pääbo 1986).

Nevertheless, the laborious processes that were required at that time to sequence such DNA (through bacterial cloning) were an effective brake on the development of the field of ancient DNA (aDNA). However, with the development of the Polymerase Chain Reaction (PCR) (Mullis and Faloona 1987; Saiki *et al.* 1988) in the late 1980s (see section 1.3) the field was presented with the ability to rapidly progress.

1.1.1 'Antediluvian' DNA studies

The 'post-PCR' era heralded a wave of publications as numerous research groups tried their hands at aDNA. Soon a series of incredible findings had been published, claiming authentic DNA could be extracted from specimens that were millions of years old, into the realms of what Lindahl (1993b) has labelled 'Antediluvian DNA'. In true 'Jurassic Park' (Crichton 1991) style, the majority of such claims were based on the retrieval of DNA from organisms preserved in amber. Insects such as stingless bees (Cano *et al.* 1992a; Cano *et al.* 1992b), termites (De Salle *et al.* 1992; De Salle

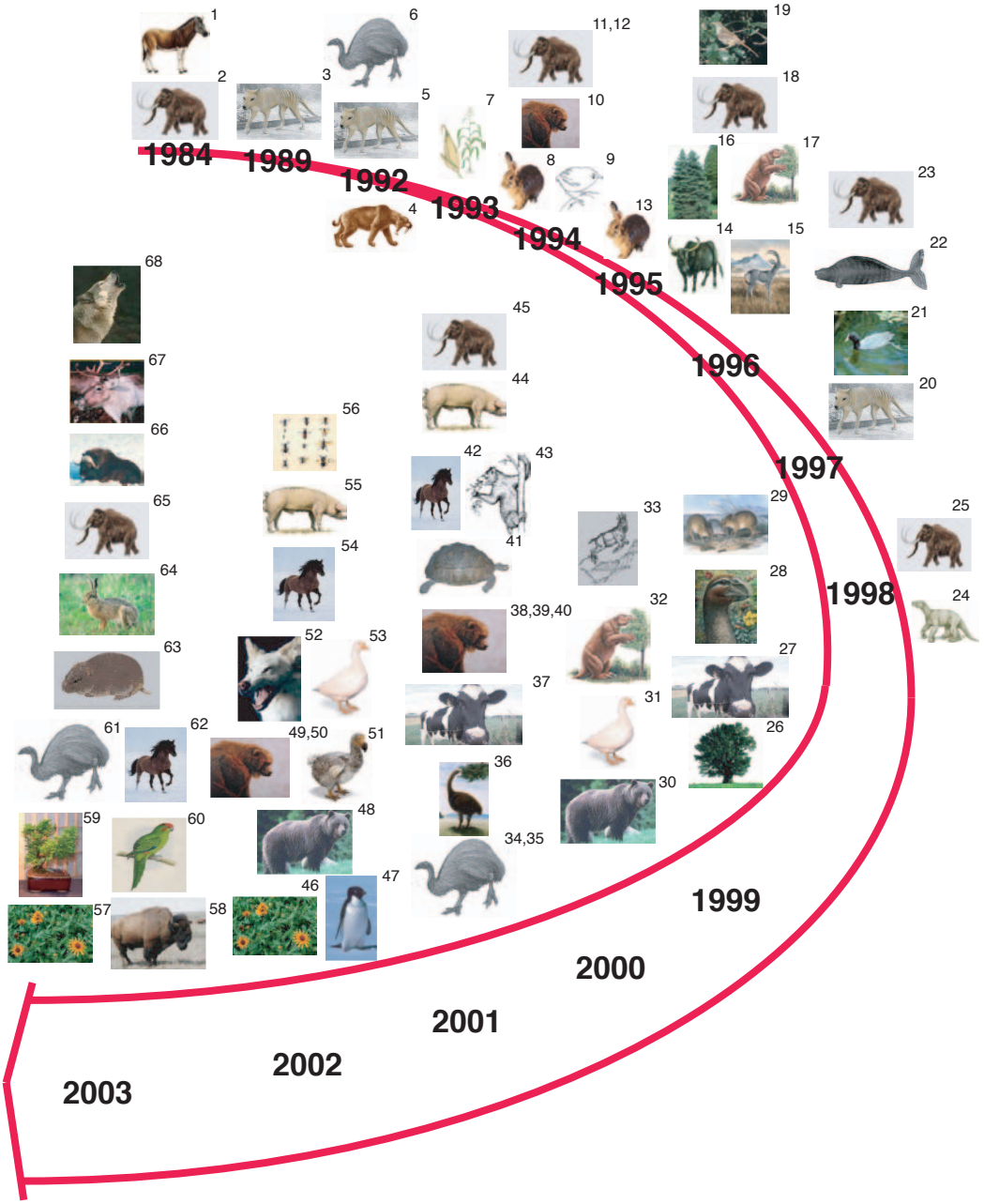
et al. 1993) and wood gnats (De Salle and Grimaldi 1994), as well as plant (Poinar *et al*. 1993) and bacterial (Cano *et al*. 1994) sequences were extracted from Dominican amber dating to the Oligocene (25-35 million years ago (mya)). Older still sources of Lebanese amber-encased weevils, dating to within the Cretaceous (between 120 and 135 mya), reportedly also yielded authentic DNA (Cano *et al*. 1993). DNA retrieval was also not limited to amber. Several sediment-preserved plant remains dating to the Miocene were successfully investigated (Golenberg *et al*. 1990; Golenberg 1991). Then, in 1994 and to international acclaim, Woodward *et al*. reported the most exciting results to date - mitochondrial *cytochrome b* sequences that had apparently been extracted from dinosaur bones dating to over 80mya. When in 1995 two further studies reported dinosaur DNA sequences extracted from a Cretaceous egg (An *et al*. 1995; Li *et al*. 1995) it seemed that the field would truly revolutionise knowledge of the Earth's evolutionary past.

Unfortunately the 'golden days' of aDNA did not last. A critical review of ancient DNA literature through the development of the field highlights that, with two notorious and criticised exceptions that claim the retrieval of 250 mya halobacterial sequences from halite (crystallised salt) (Vreeland *et al*. 2000; Fish *et al*. 2002), few recent studies have succeeded in amplifying DNA from remains older than several hundred thousand years (ky) (*c.f*. Willerslev *et al*. 2003).

1.1.2 Ancient DNA studies

Despite the problems associated with 'antediluvian' DNA (see section 1.5), a wide, and ever-increasing range of aDNA sequences have now been published from a range of animal and plant taxa (Fig 1.1). Tissues examined include artificially or naturally mummified animal remains *c.f*. Higuchi *et al*. 1984; Thomas *et al*. 1989), bone (*c.f*.

Figure 1.1: Examples of some animal and plant ancient DNA studies.



For key see opposite.

Specimen	Organism	Study
1	Quagga	Higuchi *et al*. 1984
2	Woolly mammoth	Higuchi and Wilson 1984
3	Marsupial wolf	Thomas *et al*. 1989
4	Sabre-tooth cat	Janczewski *et al*. 1992
5	Marsupial wolf	Krajewski *et al*. 1992
6	Moa spp.	Cooper *et al*.1992
7	Maize	Goloubinoff *et al*. 1993
8	Rabbit	Hardy *et al*.1994
9	Cave bear	Hänni *et al*. 1994b
10	New Zealand wren	Cooper 1994
11,12	Woolly mammoth	Hagelberg *et al*.1994, Höss *et al*.1994
13	Rabbit	Hardy *et al*.1995
14	Auroch	Bailey *et al*. 1996
15	Blue antelope	Robinson *et al*. 1996
16	Fir spp.	Suyama *et al*. 1996
17	Giant sloth	Höss *et al*.1996
18	South Island Piopio	Christidis *et al*. 1996
19	Woolly mammoth	Yang *et al*. 1996
20	Marsupial wolf	Krajewski *et al*. 1997
21	New Zealand Coot	Trewick 1997
22	Steller's sea cow	Ozawa *et al*. 1997
23	Woolly mammoth	Ozawa *et al*. 1997
24	Shasta ground sloth	Poinar *et al*. 1998
25	Woolly mammoth	Noro *et al*.1998
26	Oak	Dumolin-Lapegue *et al*. 1999
27	Cow	MacHugh *et al*. 1999
28	Moa-nalo	Sorenson *et al*.1999
29	Pig-footed bandicoot	Westerman *et al*. 1999
30	Brown bear	Leonard *et al*. 2000
31	Goose	Barnes *et al*. 2000
32	Shasta ground sloth	Hofreiter *et al*. 2000
33	Balearic antelope-gazelle	Lalueza-Fox *et al*. 2000
34, 35	Moa spp.	Cooper *et al*. 2001, Haddrath *et al*. 2001
36	Elephant bird	Cooper *et al*. 2001
37	Cow	Troy *et al*. 2001
38,39,40	Cave bear	Hofreiter *et al*. 2001a,Kühn *et al*. 2001, Loreille *et al*. 2001
41	Giant tortoise	Austin and Arnold 2001
42	Horse	Vila *et al*. 2001
43	Giant lemur	Montagnon *et al*. 2001
44	Pig	Watanobe *et al*. 2001
45	Woolly mammoth	Greenwood *et al*. 2001
46	13 plant families	Kuch *et al*. 2002
47	Adelie penguins	Lambert *et al*. 2002
48	Brown bear	Barnes *et al*. 2002
49,50	Cave bear	Hofreiter *et al*. 2002, Orlando *et al*. 2002
51	Dodo	Shapiro *et al*. 2002
52	Dog	Leonard *et al*. 2002
53	Goose	Haynes *et al*. 2002
54	Horse	Di Bernardo 2002
55	Pig	Watanobe *et al*. 2002
56	Insect spp.	Junqueira *et al*. 2002
57	48 plant taxa	Willerslev *et al*. 2003
58	Bison spp.	Willerslev *et al*. 2003
59	Japanese cedar	Tani *et al*. 2003
60	New Zealand parakeet	Willerslev *et al*. 2003
61	Moa spp.	Willerslev *et al*. 2003
62	Horse	Willerslev *et al*. 2003
63	Norway lemming	Willerslev *et al*. 2003
64	European hare	Willerslev *et al*. 2003
65	Woolly mammoth	Willerslev *et al*. 2003
66	Musk ox	Willerslev *et al*. 2003
67	Reindeer	Willerslev *et al*. 2003
68	Wolf	Flagstad *et al*. 2003

Cooper *et al*. 1992; Hänni *et al*. 1994b; Hagelberg *et al*. 1994), paleofaeces (Poinar *et al*. 1998; Hofreiter *et al*. 2000), alcohol preserved specimens (Junqueira *et al*. 2002), rodent middens (Küch *et al*. 2002), dried plant remains (Goloubinoff *et al*. 1993; Dumolin-Lapegue *et al*. 1999) and recently, extractions of animal and plant DNA directly from soil samples (Willerslev *et al*. 2003).

1.1.3 Ancient DNA studies on human remains

Due to the considerable anthropological, archaeological, and public interest directed towards human remains (the main subject of this thesis), it is only natural that they have received a similar amount of attention from the aDNA community (Table 1.1). Due to their obvious signs of morphological preservation, many studies utilised mummified tissue as a source of ancient human DNA. Examples include both naturally preserved specimens, for example those preserved in ice (such as the Tyrol man 'Ötzi' (Handt *et al*. 1994b), or through rapid desiccation (for example high-altitude mummies from the Andes (*c.f*. Pääbo 1986; Montiel *et al*. 2001)), as well as various sources of artificially preserved tissue (such as the chemically treated mummies of ancient Egypt (*c.f*. Hänni *et al*. 1994a)). However, mummified remains are a limited resource, thus the majority of human aDNA studies have focused on extracting DNA from two tissues that are much more common in the archaeological record - bone and teeth. Recently, several other sources have also yielded DNA, including paleofaeces (Poinar *et al*. 2001) and hair (Baker *et al*. 2001).

Table 1.1: Some published human ancient DNA studies.

Material	Age[1]	Study	DNA	Sample Origin[2]	Details of Interest
Bone/Teeth	<60,000	Adcock *et al.* 2001	Mitochondrial	Australia	
Bone/Teeth	6,000-5,000	Anzai *et al.* 1999	Nuclear	Japan	HLA typing
Bone/Teeth/Mummy	<5,200	Bachmann *et al.* 2000	Nuclear	Germany/Americas	Autosomal STR
Bone/Teeth	<14,000	Di Benedetto *et al.* 2001	Mitochondrial	Alps	
Bone/Teeth	4,000-3,000	Burger *et al.* 1999	Mitochondrial	Eurasian	
Bone/Teeth	<25,000	Caramelli *et al.* 2003	Mitochondrial	Italy	Cro-Magnon samples
Bone/Teeth	79 CE	Cipollaro *et al.* 1998	Nuclear	Pompeii	Lava preserved, Amelogenin
Bone/Teeth	79 CE	Cipollaro *et al.* 1999	Nuclear	Pompeii	Lava preserved, Y haplotyping
Bone/Teeth	79 CE	Cipollaro *et al.* 1999	Nuclear	Pompeii	Lava preserved, Autosomal
Bone/Teeth	79 CE	Cipollaro *et al.* 1998	Nuclear	Pompeii	Lava preserved, Y haplotyping
Mummified tissue	<2,500	Clisson *et al.* 2002	Mitochondrial	Kazakhstan	Naturally mummified
Bone/Teeth	<2,500	Clisson *et al.* 2002	Nuclear	Kazakhstan	X/Y STR
Bone/Teeth	<12,000	Colson *et al.* 1997	Mitochondrial	Global	
Bone/Teeth	7,000-5,000	Demarchi *et al.* 2001	Mitochondrial	Argentina	
Bone/Teeth	<100	Endicott *et al.* 2003	Mitochondrial	Andaman Islands	
Bone/Teeth	<8,000	Faerman *et al.* 1995	Nuclear	Israel	Amelogenin
Bone/Teeth	<3,700	Fily *et al.* 1998	Mitochondrial	Basque Country	
Bone/Teeth	<3700	Fily *et al.* 1998	Nuclear	Basque Country	Amelogenin
Bone/Teeth	110-140	Gaensslen and Berka 1993	Nuclear	America	X/Y STR
Bone/Teeth	1546-1749 CE	Gerstenberger *et al.* 1999	Nuclear	Germany	X/Y STR
Bone/Teeth	1546-1749 CE	Gerstenberger *et al.* 1999	Nuclear	Germany	Y haplotyping
Bone/Teeth	<2,000	Gilbert *et al.* 2003a	Mitochondrial	N Europe	
Bone/Teeth	<2,000	Gilbert *et al.* 2003b	Mitochondrial	N Europe	
Bone/Teeth	600-1521 CE	Gonzalez-Olivier *et al.* 2001	Mitochondrial	C.America	
Bone/Teeth	750-300	Hagelberg *et al.* 1991	Mitochondrial	England	
Bone/Teeth	5,500-300	Hagelberg and Sykes	Mitochondrial	England	
Mummified tissue	5,000	Handt *et al.* 1994a	Mitochondrial	Tyrol	Natural ice mummy
Mummified tissue	600	Handt *et al.* 1996	Mitochondrial	N. America	
Bone/Teeth	5,500	Hänii *et al.* 1990	Mitochondrial		
Mummified tissue	<7,000	Hänni *et al.* 1994	Mitochondrial	Egypt/Americas	Artificially mummified
Mummified tissue	8,000-7,000	Hauswirth 1994	Mitochondrial	America	Naturally mummified
Bone/Teeth	6,000	Horai *et al.* 1989	Mitochondrial	Japan	

Material	Age (yrs)	Study	DNA	Location	Details of Interest
Bone/Teeth	250-3,000	Hummel *et al*. 2002	Nuclear	Europe	ABO blood typing
Bone/Teeth	17th-18th century	Hummel and Herrmann 1991	Mitochondrial	Norway/Germany	
Mummified Heart	1795 CE	Jehaes *et al*. 2001	Mitochondrial	French	Preserved heart of Louis XVII
Bone/Teeth	350-9,200	Kaestle and Glenn-Smith 2001	Mitochondrial	N.America	
Bone/Teeth	500-1,200	Kalmar *et al*. 2000	Mitochondrial	Hungary	
Bone/Teeth	800-1600 CE	Koman and Tuross 2000	Mitochondrial	N.America	
Bone/Teeth	100,000-30,000	Krings *et al*. 1997	Mitochondrial	Germany	Neanderthal
Bone/Teeth	100,000-30,000	Krings *et al*. 1999	Mitochondrial	Germany	Neanderthal
Bone/Teeth	21,000-14,000	Lalueza *et al*. 1997	Mitochondrial	Patagonia	
Bone/Teeth	670-1680 CE	Lalueza-Fox *et al*. 2001	Mitochondrial	Carribean	
Bone/Teeth	1620 CE	Lalueza-Fox *et al*. 2003	Mitochondrial	Carribean	
Bog remains	7,500	Lawlor *et al*. 1991	Mitochondrial	Europe	
Bone/Teeth	1,340	Malhi and Glenn Smith 2002	Mitochondrial	N.America	
Bone/Teeth	500	Meijer *et al*. 1992	Mitochondrial		
Mummy/Bone/Teeth	Not disclosed	Merriweather *et al*. 1994	Mitochondrial	N.America	Naturally mummified
Bone/Teeth	<7,000	Meyer *et al*. 2000	Nuclear	Germany	Amelogenin
Mummified Umbilical Cord	<44	Minakata *et al*. 2002	Mitochondrial	Japan	Naturally mummified
Mummified Umbilical Cord	<44	Minakata *et al*. 2002	Nuclear	Japan	Autosomal
Mummified tissue	550	Monsalve *et al*. 2002	Mitochondrial	Canada	Natural ice mummy
Bone/Teeth	16th century	Montiel *et al*. 2001	Mitochondrial	Spain	
Bone/Teeth	2,900-600	Moraga *et al*. 2001	Mitochondrial	Chile	
Bone/Teeth	<7,000	Meyer *et al*. 2000	Nuclear	Germany	Amelogenin
Bone/Teeth	<25,800	Oota *et al*. 2001	Mitochondrial	S.E. Asia	
Bone/Teeth	<2,000	Oota *et al*. 1999	Mitochondrial	China	
Bone/Teeth	<2,000	Oota *et al*. 1995	Mitochondrial	Japan	
Bone/Teeth	6,000-300	O'Rourke *et al*. 2000	Mitochondrial	N.America	
Bone/Teeth	29,000	Ovchinnikov *et al*. 2000	Mitochondrial	N. Caucasus	Neanderthal
Mummified tissue	<4,000	Pääbo 1985a	Mitochondrial	Egypt	Artificially mummified
Mummified tissue	2,400	Pääbo 1985b	Mitochondrial	Egypt	Artificially mummified
Mummified tissue	<4,000	Pääbo 1986	Mitochondrial	Egypt/S America	Artificial/Natural/Natural
Bog remains	7,000	Pääbo 1988	Mitochondrial	America	

Material	Age (yrs)	Study	DNA	Location	Details of Interest
Mummified tissue	<5,000	Pääbo *et al*. 1989b	Nuclear	Egypt/Americas	
Bone/Teeth	<800	de Pancorbo *et al*. 1995	Nuclear	Basque Country	Autosomal
Bone/Teeth	<1,600	Parr *et al*. 1996	Mitochondrial	America	
Paleofaeces	2,000	Poinar *et al*. 2001	Mitochondrial	N.America	
Bone/Teeth	450-600 CE	Pusch *et al*. 2002	Nuclear	Germany	
Bone/Teeth	>25,000	Polotoraus *et al*. 2000	Mitochondrial	Russia	Sunghir
Bone/Teeth	<5,000	Ramos *et al*. 1995	Nuclear	Basque Country	Microsatellite repeats
Bone/Teeth	<4,000	Riberio-dos-Santos *et al*. 1996	Mitochondrial	America	
Bone/Teeth	<5,300	Richards *et al*. 1995	Mitochondrial	Europe/Newfoundland	
Bone/Teeth	16-17th century	Rickards *et al*. 2001	Mitochondrial	Sicily	
Bone/Teeth	16-17th century	Rickards *et al*. 2001	Nuclear	Sicily	Amelogenin
Bone/Teeth	<40,000	Schmitz *et al*. 2002	Mitochondrial	Germany	Neanderthal
Bone/Teeth	<3,000	Schultes *et al*. 1999	Nuclear	N.Europe	Y multiplexing
Bone/Teeth	1300 CE	Stone and Stoneking 1998	Mitochondrial	N.America	
Bone/Teeth	1300 CE	Stone *et al*. 1996	Nuclear	America	Amelogenin
Mummified tissue/Bone	1475 CE	Thuesen and Engberg 1990	Nuclear	Greenland	
Bone/Teeth	7-8th century	Vernesi *et al*. 1999	Nuclear	Italy	Amelogenin
Bone/Teeth	2,500	Wang *et al*. 2000	Mitochondrial	China	
Mummified tissue		Wang and Lu 1981	Mitochondrial	China	Naturally mummified
Bone/Teeth	<5,000	Yang *et al*. 1998	Nuclear	China	Autosomal
Bone/Teeth	<800	Zierdt *et al*. 1996	Nuclear	Germany	X/Y STR
Mummified tissue	600-200 BCE	Voevoda *et al*. 2000	Mitochondrial	Altai plateau	

[1]*Age* either in years or by date, as detailed in original study. [2]*Sample origin* as detailed in original study.

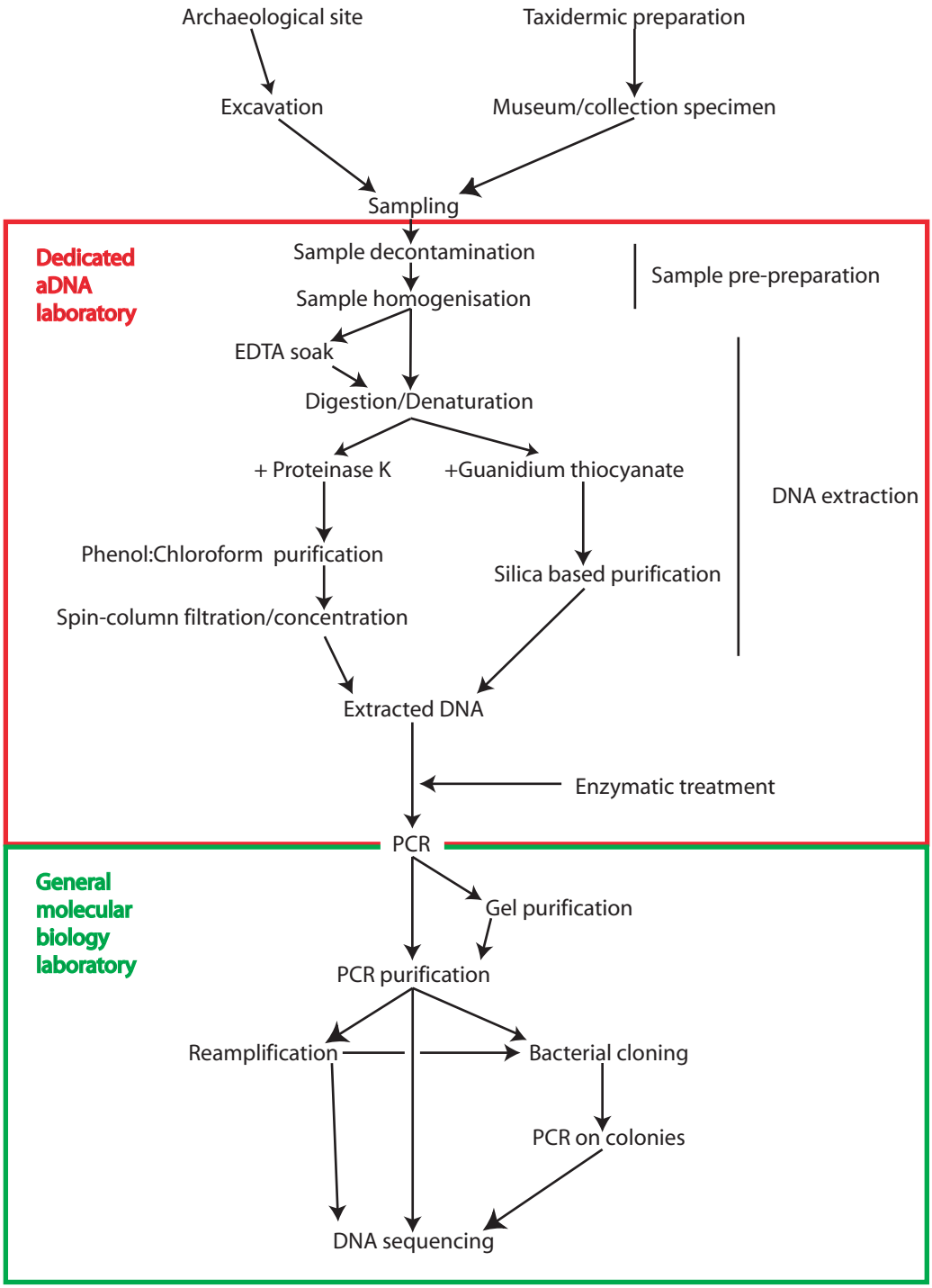
1.2 The extraction of ancient DNA from samples

An understanding of the techniques used in aDNA extraction and amplification is required to comprehend the significance of contamination and damage on aDNA studies, thus the key steps are discussed below.

Techniques employed to extract DNA from ancient specimens vary according to the tissue. Nevertheless the majority of aDNA studies utilise one of two protocols, both which are based around common molecular biology methods (*c.f.* Sambrook *et al.* 1989) and are commonly used in the related field of forensics (O'Rourke *et al.* 2000). Such methods rely on an initial digestion of the tissue to release DNA, with a subsequent purification step using either organic solvents or the DNA binding properties of silica.

Figure 1.2 is a schematic representation of typical DNA extraction, amplification and sequencing protocols. It is important to note that in order to prevent contamination of the sample with modern DNA (see section 1.5.6), all steps prior to amplification should be undertaken in a dedicated aDNA laboratory where no previous DNA amplification or work with modern DNA has been undertaken. It is usual that post sampling a specimen is 'pre-prepared' using various techniques that are aimed at removing any surface contaminants. Following pre-preparation, bones, teeth, and often other tissues as well, are powdered using a freezer mill, microdismembrator or similar homogenisation equipment. Grinding has the effect of increasing surface area, thus aiding digestion. Bones and teeth are then decalcified for up to 72 hours in ethylenediaminetetra-acetic acid (EDTA) (Hagelberg and Clegg 1991; Hagelberg *et al.* 1991; O'Rourke *et al.* 1996; O'Rourke *et al.* 1999). The sample is centrifuged, the supernatant EDTA is removed, and the pellet is re-suspended in a digestion buffer.

Figure 1.2: Steps involved in the extraction and analysis of ancient DNA, using various common techniques.



1.2.1 DNA extractions using Proteinase K and organic solvents

The original tissue digestion protocols utilised for aDNA retrieval contain Proteinase K (Pk), a protease originally derived from fungi that chemically digests proteins (Blin and Stafford 1976). Unlike other proteases its activity is not markedly inhibited by EDTA that may remain from the decalcification step, making it ideal for extractions on bone and teeth. In addition, various detergents or surfactants are added to emulsify the lipids, and/or aid in the denaturation of proteins (Sambrook *et al.* 1989), most commonly sodium dodecyl sulphate (SDS) or sodium lauryl sarkosyl. The addition of 1-2 ml of this extraction buffer to a small amount of powdered bone (0.5-1 gm) is sufficient to digest the sample at moderate temperatures (ca. 50-60ºC). Recently it has also been suggested that, to counter the effects of Maillard products (protein:DNA crosslinks) that may form *post mortem*, the addition of small concentrations of the thiazolium salt N-phenacylthiazolium bromide (PTB) to digestions may be beneficial (Poinar *et al.* 1996).

Post digestion the samples are extracted twice with equal volumes of phenol, or once with phenol and once with a 1:1 mixture of phenol:isoamyl alcohol. Since DNA is polar, it remains dissolved in the aqueous phase, whereas many other compounds commonly found in the tissue remain in the organic phase. Thus removal of the supernatant (or base if very high concentrations of salt or EDTA are present in the extraction) is a rapid and efficient means of DNA isolation and purification. The supernatant is extracted once with either chloroform or a mixture of 24:1 chloroform:isoamyl alcohol. The resulting aqueous phase is carefully removed (as carryover of organic solvents to subsequent extraction and amplification stages may inhibit PCR), and desalted and concentrated using centrifugation-driven dialysis (Horai *et al.* 1989; Pääbo *et al.* 1989) or filtration using micro-concentrators (*c.f.*

Cooper *et al*. 2001a). While this removes macromolecules less than 30,000 molecular weight (MW) that may inhibit subsequent PCR reactions, in some specimens (such as those buried in the soil and paleofaeces) it is common to find some co-extraction of PCR inhibitors along with the DNA. One simple solution to this problem is the incorporation of bovine serum albumin (BSA) (or other molecules that bind to inhibitory chemicals) to PCR reactions (Pääbo *et al*. 1988), which binds to inhibitors thus prevents them from inhibiting the polymerase enzymes.

1.2.2 Silica-based DNA extractions

DNA extractions that exploit the DNA-binding properties of silica (Boom *et al*. 1990; Höss and Pääbo 1993) require the extraction of DNA in high concentrations of salts such as guanidinium thiocyanate (GuSCN), which also has the ability to denature proteins, but also acts as a chaotropic agent that facilitates the binding of DNA to silica particles. After incubation at modest temperatures (50-60ºC) an aliquot of the digestion mix is then added to a silica suspension (that contains additional GuSCN extraction buffer to encourage DNA-silica binding), briefly re-incubated, and washed in various combinations of ethanol and/or acetone. The pellet is finally re-dissolved in a mild Trizma-EDTA (TE) buffer to release the DNA from the silica. This protocol has the advantage that it is less likely to co-extract PCR inhibitors and avoids the use of hazardous chemicals (Cattaneo *et al*. 1997). However, the silica itself is a strong PCR inhibitor, so great care must be taken to remove all silica during the washes. Additionally, due to its extreme affinity for DNA, the GuSCN can easily become contaminated with modern nucleic acids (*c.f*. O'Rourke *et al*. 2000). Nevertheless the technique is popular, and several silica-gel based kits have been developed for use in extracting aDNA. Despite these advantages it has been demonstrated that the classic

Pk-phenol:chloroform techniques produce better yields of aDNA (Cattaneo *et al*. 1997; Yang *et al*. 1998), though this may be outweighed by the lower levels of inhibitors present in silica-extracted aDNA (Cattaneo *et al*. 1997).

1.3 Analysis of ancient DNA by hybridisation and PCR

Post-extraction it is normal for aDNA to be analysed through PCR amplification. Nevertheless, some studies have utilised other techniques, which are also briefly addressed below.

1.3.1 Analyses incorporating PCR

Prior to the development of the PCR reaction, DNA sequencing, amplification and analysis was based around the cleavage of DNA extracts into multiple fragments using restriction enzymes, followed by conventional plasmid cloning in bacteria (*c.f.* Higuchi *et al*. 1984; Pääbo 1985a; Pääbo 1985b; Pääbo 1986). Discrete colonies were then sequenced to examine for the specific 'incorporated' DNA sequence within each bacterial clone. The inefficiency and inadequacies of such a processes are numerous. Firstly, it is estimated that even using a modern mammalian DNA extract, the generation of a complete clone library of the genome, would require the analysis of approximately half a million colonies (O'Rourke *et al*. 2000). With the smaller DNA fragment sizes associated with aDNA (due to DNA degradation, see section 1.5.2), this number would be significantly higher. Secondly it is not possible to easily identify a sequence of interest, instead colonies must be screened for the 'target' sequence through the almost 'random' approach of hybridization analysis. Finally in order to generate accurate DNA sequences, it may be required to re-iterate the process two or more times to achieve the necessary purification of the desired

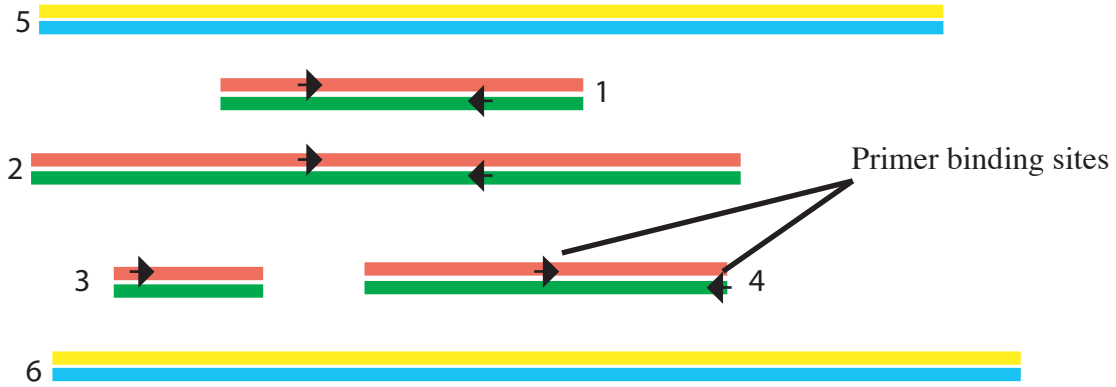
fragment (O'Rourke *et al*. 2000). Thus the development of the PCR reaction (Mullis and Faloona 1987; Saiki *et al*. 1988) provided immediate improvements to DNA amplification and sequence analysis.

PCR is a re-iterative process that relies on the annealing of 'sequence' specific oligonucleotide probes (termed 'primers') to complementary DNA sequences. Typically two primers per reaction are used, each ≈20bp in length and designed so as to bind to the 5' end of the target strand (Fig 1.3). DNA polymerase enzymes, deoxyribonucleotide bases (dNTPs) and various buffers are then added to the mixture of DNA, primers, and usually BSA. Through a cyclical process involving the repeated denaturation of templates, binding of primers to DNA targets (annealing), and strand replication by the enzymes (elongation) (Fig 1.3), the specific DNA sequence can be exponentially amplified. Due to DNA sequence heterogeneity among different organisms, primers can be designed so as to bind to different spectra of related sequences, while specifically 'excluding' others such as bacterial DNA derived from environmental bacteria co-extracted within an aDNA sample, or human DNA present within non-human specimen (Fig 1. 3).

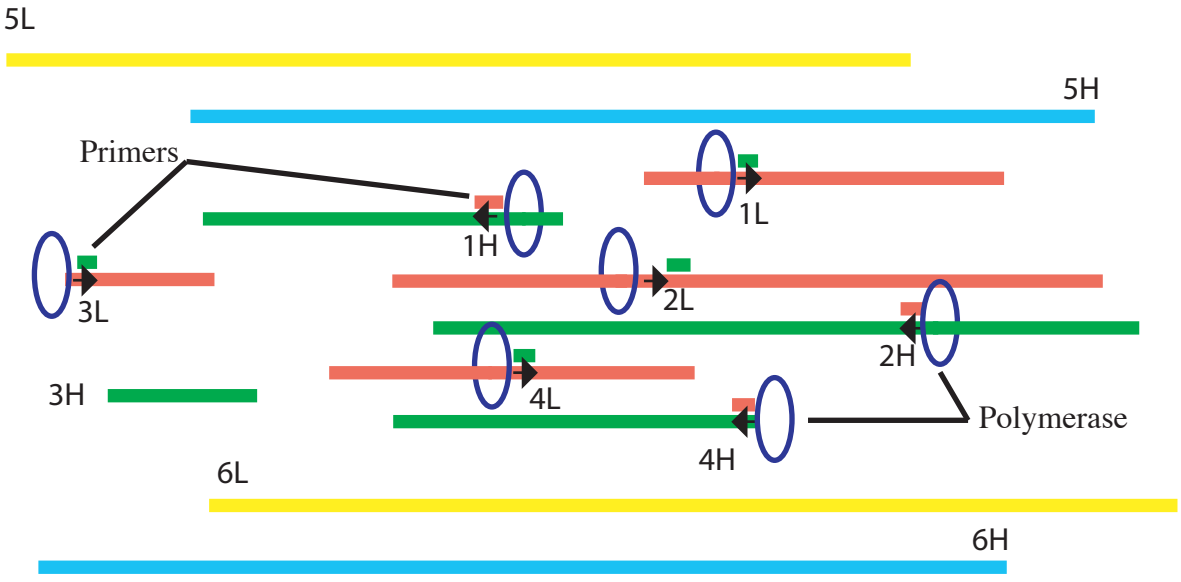
Amplified DNA may be analysed using several techniques. Most studies have sequenced amplified product, using Dye-labeled chain terminators (*c.f.* Ansorge *et al*. 1986; Smith *et al*. 1986; Prober *et al*. 1987). Others however, particularly those that analyse sequence variation among closely related groups of sequences (for example human mitochondrial DNA haplotypes), have used the faster and cheaper techniques of restriction enzyme digests (*c.f.* Merriweather *et al*. 1994).

Theoretically a PCR should result in the sole amplification of the target regions of DNA. However, due to several complications (that include uncertainty in the specificity of primers (see chapter 7), sequence modification due to *post mortem*

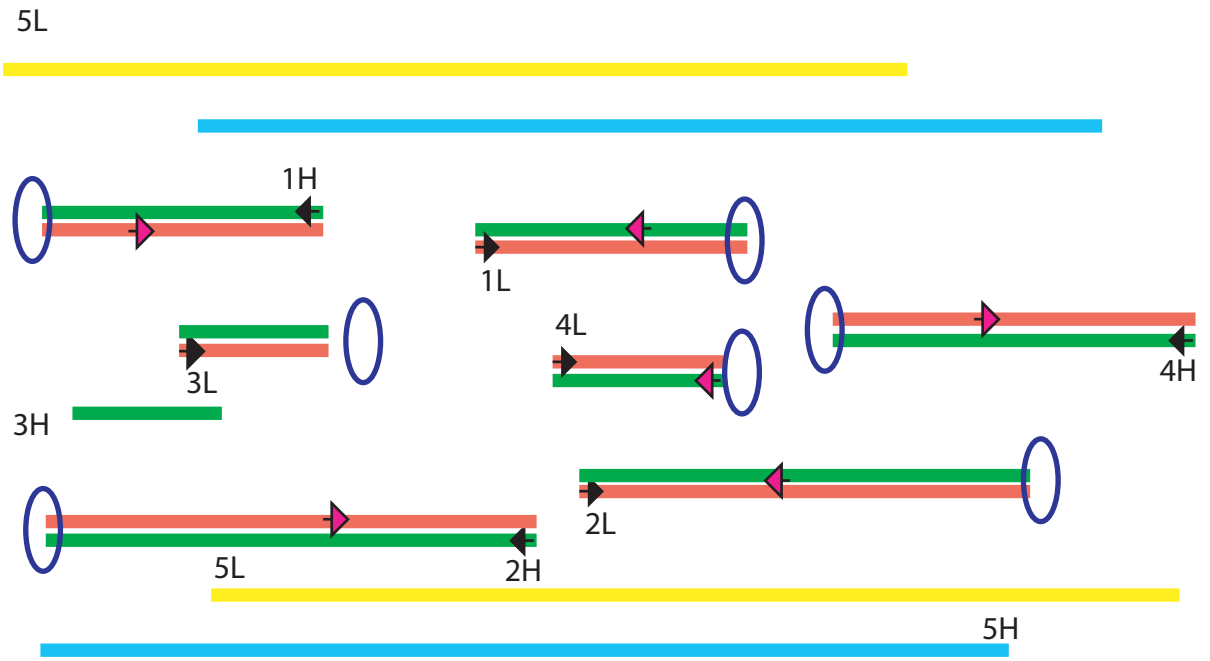
Figure 1.3: Schematic of the Polymerase Chain Reaction (PCR) on a mtDNA template.



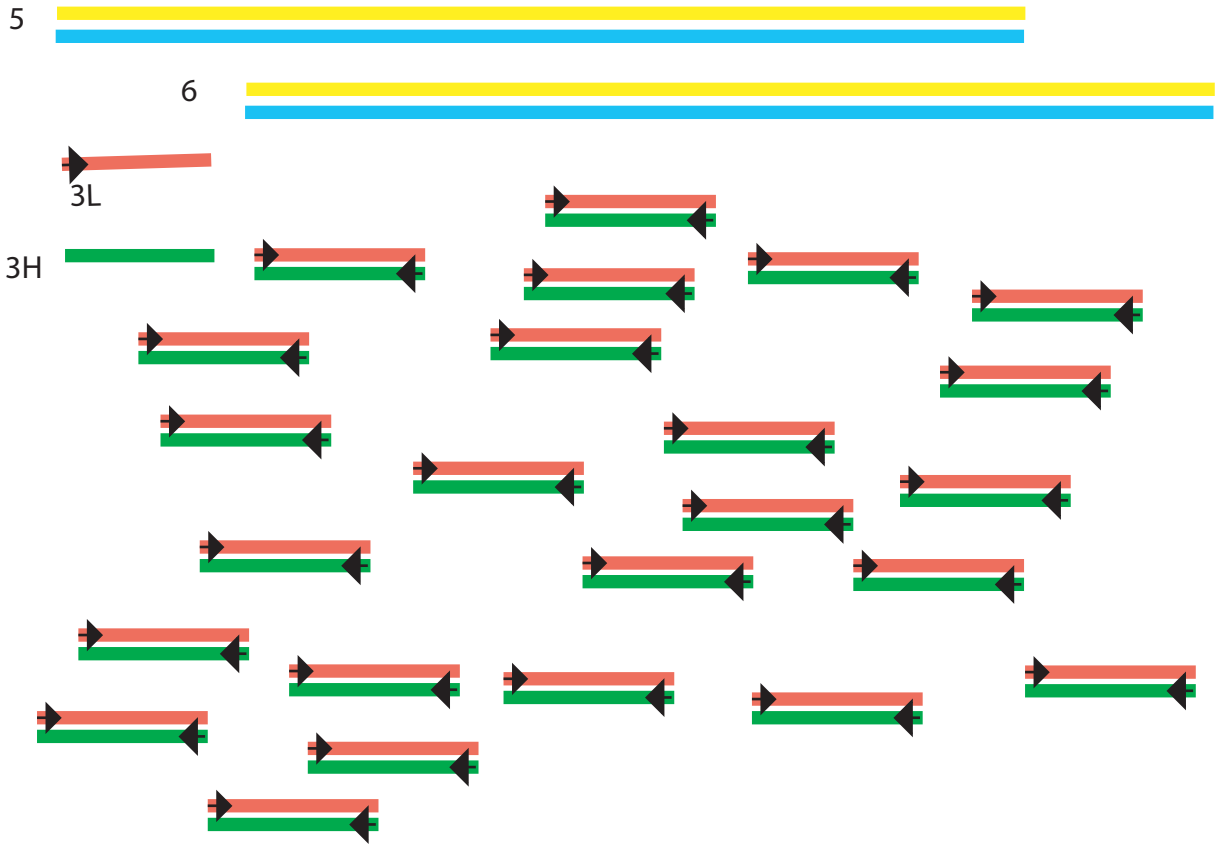
A: Due to DNA degradation and contamination, an ancient DNA extract may contain authentic, but fragmented DNA templates (samples 1-4), as well as contaminant DNA sequences (samples 5 and 6) - in this example unrelated at the genetic level. Strand 3 is fragmented in such a way as to contain only 1 of the two primer binding sites required for PCR. This could also represent a DNA polymerase block due to various forms of DNA damage (as discussed in section1.5).



B: The temperature of the PCR reaction is increased to approximately 94 ^{o}C during the initial denaturation cycle, the template molecules dissasociate into single stranded DNA (here the Light (L) and Heavy (H) strands of the mitochondria. The cycle then cools to approx 56 ^{o}C , oligonucleotide primers anneal to the complementary 'target' on the single stranded DNA. Due to a lack of sequence similarity the primers do not bind to the contaminant DNA sequences (5 and 6). Polymerase enzymes bind to double stranded DNA and elongation commences.



C: During elongation, double stranded molecules are replicated. The damaged H strand of template 3 (3H), with no primer site, is not replicated. The damaged L strand of template 3 (3L), with a primer binding site, is only partially replicated up to the locus of the damage.



D: After several cycles, numerous copies have been generated of undamaged target DNA sequences (1,2,4). The ratio of contaminant to target sequences is significantly reduced. The damaged template (3) has not been replicated. After 35 or 40 cycles of exponential growth, the damaged and contaminant sequences are reduced to an insignificant proportion of the total DNA.

DNA damage, and contamination, PCR reactions often result in the amplification of multiple DNA sequences. Due to the nature of sequence screening methods, the analysis of such amplicons can provide unreadable sequence data, or conflicting base calls. One method of solving such problems is through cloning of PCR products (amplicons/amplimers).

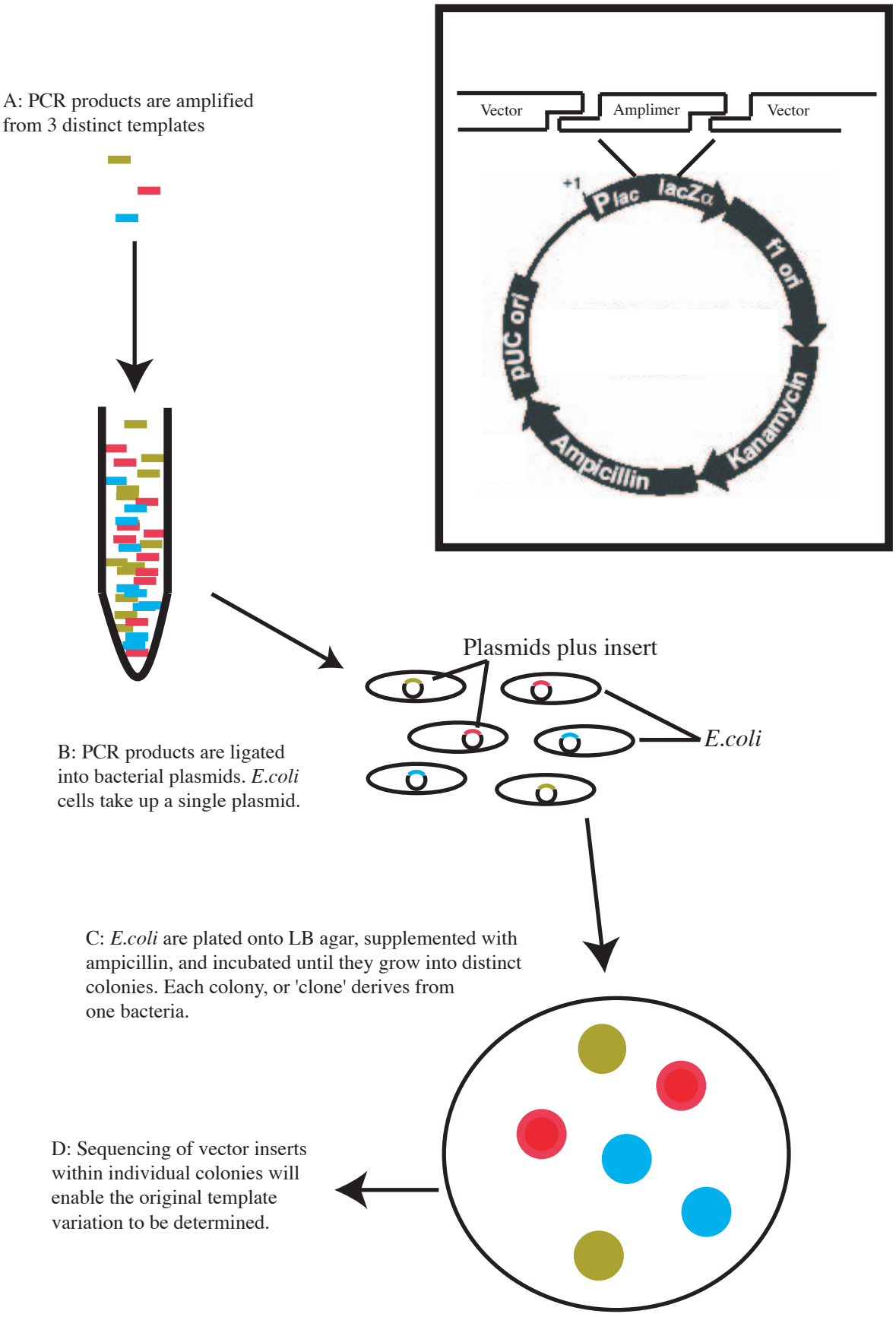
1.3.2 The molecular cloning of ancient DNA

Molecular cloning techniques are used to detect sequence heterogeneity within a single PCR reaction. Commercial cloning kits, such as the 'Topo TA' cloning kit (Invitrogen), rely on the insertion of amplicons into bacterial plasmids, which are subsequently introduced into *Escherichia coli* cells (Fig 1.4). Colonies derived from successfully transformed cells can be identified through blue/white screening due to the disruption of one of a range of subunits of the plasmid's β-galactosidase gene (which metabolises X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) into a blue product). As each clone will have incorporated a single amplicon, by screening numbers of clones post-mortem DNA damage, jumping (recombination between PCR template molecules, Pääbo *et al.*1990) and contamination can be assessed.

1.3.3 DNA-Probe hybridisation

One particular technique touted as an alternative to PCR is hybridisation analysis, either on extracts, or in some cases, directly onto samples prior to DNA extraction. Hybridization relies on the base pairing that occurs between a strand of DNA and its compliment. In practice, a known segment of DNA is labeled (*e.g.* with ^{32}P), making hybridization probes. The target DNA is covalently bound to solid support (*e.g.* a nylon membrane), is bathed in a solution containing the probe, and washed under

Figure 1.4: Molecular cloning of PCR products using TA cloning.



different stringency regimes. Should any target DNA have bound (*i.e.* hybridized) to the probe, this will be detectable either visually or by exposure to film. Theoretically, hundreds of samples can be tested in a single experiment, and, with appropriate choice of probe, the plant or animal species identified (Sofer 1991). Unfortunately however, problems commonly associated with aDNA (such as low yields of DNA, poor resolution of the DNA sequence, DNA damage, and contamination) have demonstrated this technique to be limited in application. In addition, several studies have been strongly criticized (Geigl 2001). In particular, the method does not reveal the actual sequence, only percentage similarities of the target and probe.

1.4 The analysis of aDNA from human remains

O'Rourke *et al.* (2000) have written that '(The) extraction of DNA sequence information from the fossil remains of extinct members of the genus Homo or other ancestral species can perhaps be considered the Holy Grail of biomolecular archaeology'. While preliminary studies on *H.sapiens neanderthalensis* mitochondrial DNA have started to investigate such areas (arguably demonstrating the lack of any large-scale interbreeding between Neanderthals and anatomically modern humans (Krings *et al.* 1997; Krings *et al.*1999; Ovchinnikov *et al.* 2001; Schmitz *et al.* 2002)), DNA retrieved from ancient human specimens has also been used to investigate a wide array of other issues (Table 1.1).

1.4.1 Ancient DNA studies and mitochondrial DNA

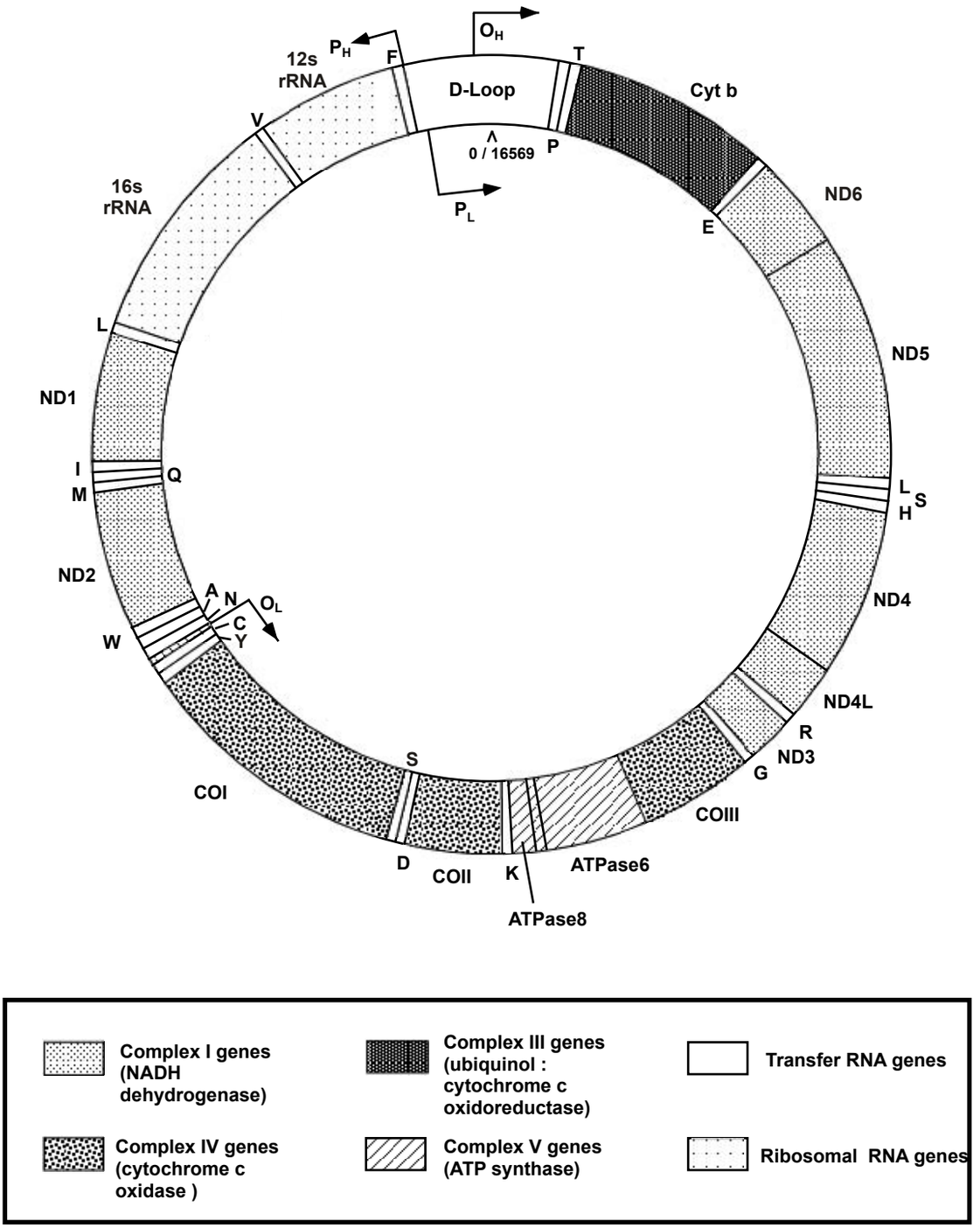
Due to its presence in copy numbers that are estimated to range from between 1000-10,000 times that of single-copy nuclear DNA sequences (Taanman 1999), it is to be expected that at similar rates of DNA degradation, mitochondrial DNA (mtDNA) will

survive for far longer periods than its nuclear counterparts. Thus mtDNA has received the majority of attention in human aDNA studies to date.

Multiple copies of the human mitochondrial genome are located within each mitochondria organelle (Taanman 1999), spanning approximately 16569 bases pairs (bp) in length, and encoding for a small number of products (2 ribosomal RNAs, 22 transfer RNAs and 13 proteins) as well as containing a few non-coding regions (Fig 1.5). The two individual strands of each mitochondrial genome can be differentiated on the basis of guanine (G) and thymine (T) composition, which leads to a differential strand separation in denaturing caesium chloride gradients (Taanman 1999). Although most genetic information is encoded on the G-T rich Heavy (H) strand (Taanman 1999), base positions have conventionally been numbered relative to the Light (L) strand. As the mitochondria in most species is inherited solely through the maternal line, and recombination events appear rare (Moritz *et al.* 1987), it represents a useful tool in DNA based phylogenetic reconstruction, and has thus been used to investigate the phylogeny of human populations (using modern DNA samples) on numerous occasions (*c.f.* Vigilant *et al.* 1991). Within the genome itself, most aDNA phylogenetic studies have focused on an area of limited functional importance, termed the Displacement loop (D-loop) (so named as in most metabolically active vertebrate cells, the region contains a short nucleic acid strand complementary to the L strand, which displaces the H strand, forming the eponymous 'displaced' loop (Taanman 1999)).

Several features of the D-loop explain why it is commonly used in studies of ancient populations (Fig 1.6). Through interspecific sequence comparison Saccone *et al.* (1991) identified three distinct regions within the D-loop, a conserved middle region,

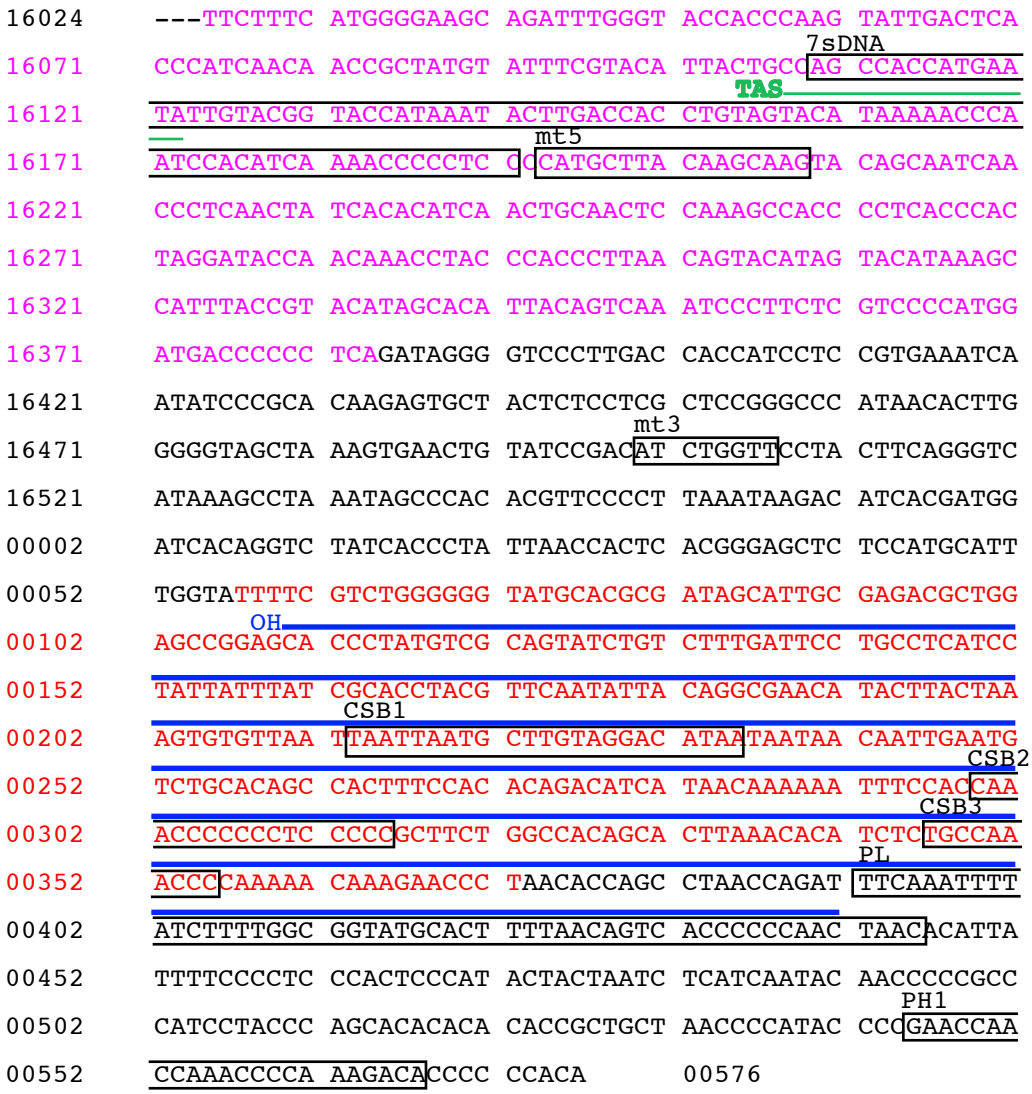
Figure 1.5: Gene organisation of the human mitochondrial genome.



For gene information see opposite. Numbering is with reference to Light strand, with relation to the Cambridge Reference Sequences (CRS) (Anderson et al. 1981). Modified with permission from Marie Lott and the 'Mitomap' database (http://www.mitomap.org/MITOMAP/mitomapgenome).

Position (CRS)	Code	Description
57-372	HVS2	Hypervariable Segment 2
110-441	OH	H-strand origin
213-235	CSB1	Conserved Sequence Block I
299-315	CSB2	Conserved Sequence Block II
346-363	CSB3	Conserved Sequence Block III
392-445	PL	L-strand promoter
545-567	PH1	Major H-strand promoter
577-647	F	tRNA Phenylalanine
645-645	PJ2	Minor H-strand promoter
648-1601	12S	12S rRNA
1602-1670	V	tRNA Valine
1671-3229	16S	16S rRNA
3230-3304	L	tRNA Leucine 1
3307-4262	ND1	NADH dehydrogenase 1
4263-4331	I	tRNA Isoleucine
4329-4400	Q	tRNA Glutamine
4402-4469	M	tRNA Methionine
4470-5511	ND2	NADH dehydrogenase 2
5512-5576	W	tRNA Typtophan
5587-5655	A	tRNA Alanine
5657-5729	N	tRNA Asparagine
5721-5798	OL	L-strand origin
5761-5826	C	tRNA Cysteine
5826-5891	Y	tRNA Tyrosine
5904-7445	COI	Cytochrome c oxidase I
7445-7516	S	tRNA Serine 1
7518-7585	D	tRNA Aspartic acid
7586-8269	COII	Cytochrome c oxidase II
8295-8364	K	tRNA Lysine
8366-8572	ATPase8	ATP synthase 8
8572-9207	ATPase6	ATP synthase 6
9207-9990	COIII	Cytochrome c oxidase IIII
9991-10058	G	tRNA Glycine
10059-10404	ND3	NADH dehydrogenase 3
10405-10469	R	tRNA Arginine
10470-10766	ND4L	NADH dehydrogenase 4L
10760-12137	ND4	NADH dehydrogenase 4
12138-12206	H	tRNA Histidine
12207-12265	S	tRNA Serine2
12266-12336	L	tRNA Leucine 2
12337-14148	ND5	NADH dehydrogenase 5
14149-14673	ND6	NADH dehydrogenase 6
14674-14742	E	tRNA Glutamic acid
14747-15887	Cytb	Cytochrome b
15888-15953	T	tRNA Threonine
15955-16023	P	tRNA Proline
16024-576	D-Loop	Displacement-Loop
16024-16383	HVR1	Hypervariable Segment 1
16106-191	7S	7sDNA
16157-16172	TAS	termination-associated sequence
16194-16208	mt5	control element
16499-16506	mt3	L-strand control element

Figure 1.6: Sequence of the human mitochondrial D-Loop Light strand, demonstrating the location of functionally relevant motifs.



Annotation is with reference to figure 1.5, numbering to the CRS (Anderson *et al.* 1981). HVR1 sequence is indicated in pink, HVR2 in red. Conserved domains are boxed where possible, with the exception of TAS (overscored in green), and OH (overscored in blue).

the central domain, as well as two more variable regions, originally termed the Left and Right domains (now termed Hypervariable Regions 1 (HVR1) and 2 (HVR2) respectively. HVR2 has been identified as probably the most important functional part of the D-loop (Saccone *et al.* 1991), and contains the site of origin of the heavy strand replication (OH), as well as the 2 promoters required for the replication of the Light (PL) and Heavy (PH) DNA strands. In addition, HVR2 contains 3 conserved sequence blocks (CSB1-3) which have been suggested to act as processing signals for the enzymes involved in the generation of RNA primers for H strand replication (Saccone *et al.* 1991).

HVR1 is also of crucial importance to mitochondrial function, due to its controlling role in the synthesis of new mitochondrial DNA. Within this domain lies a triplex structure formed by the arrested replication of the nascent H strand (7sDNA), as well as three further conserved blocks of debated function, the Termination-associated sequence (TAS), and Mt3 and Mt5, putative protein binding sites. (Taanman 1999).

Despite such conserved loci, the interest in the domain lies with the fact that the majority of the positions appear to be unconstrained relative to protein coding sequences grouped in triplet codons, and mutate at rates rapid enough (recent estimates suggest the rate as approximately 0.0043 per generation: Sigurdardottir *et al.* 2000) to allow resolution in studies of the history of large-scale human populations.

As a result of the limits placed on the size of amplifiable fragments due to degradation of aDNA, in order to help provide maximum resolution and differentiation among samples, phylogenetic studies using aDNA have focused on amplifications of the HVR1 (*note – in some papers this region is named Hypervariable Segment 1 - HVS1*). One notable example includes the original report

of Neanderthal mtDNA sequences (Krings *et al.* 1997). Several studies have also focused on the less widely used, HVR2 (or HVS2) region, for example an analysis by Fily *et al.* (1998) on ancient human remains from the Basque country. In addition, a few studies target sequences in the more-conserved coding region of the mitochondria for either sequence (*c.f.* Demarchi *et al.* 2001; Glenn Smith *et al.* 1999; Endicott *et al.* 2003) or Restriction-digest Fragment Length Polymorphism (RFLP) analyses of the sites (*c.f.* Merriweather *et al.* 1994; Montiel *et al.* 2001; Malhi and Glenn Smith 2002; Monsalve *et al.* 2002).

1.4.2 Targeting ancient nuclear DNA

In contrast to the majority of studies on other species, which limit themselves to the task of analysing mtDNA, a large number of human studies have attempted the analysis of nuclear DNA (nuDNA) (Table 1.1). One common use is for the amplification of X and Y chromosome sequences, either in order to 'sex' specimens using systems such as the amelogenin assay (*c.f.* Faerman *et al.* 1995, Fily *et al.* 1998) or to phylogenetically classify samples using various microsatellite repeat and Single Nucleotide Polymorphism (SNP) systems (*c.f.* Zierdt *et al.* 1996; Gaensslen *et al.* 1993). Some studies have even attempted this feat using multiplex reactions, where numerous primer sets are utilised to simultaneously amplify multiple loci in a single PCR reaction (*c.f.* Schultes *et al.* 1999). Various autosomal (non sex-chromosome nuclear DNA) markers have also been targeted, for example Hummel *et al.* (2002) attempt to blood type skeletons using the genetic markers of the ABO blood typing system located on chromosome 9, while Anzai *et al.* (1999) report success in HLA typing various archaeological specimens from Japan.

1.4.3 Pathogen and microorganism aDNA analyses using human remains

The use of degraded human samples in aDNA analyses has not been limited to the amplification of human DNA. It is reasonable to assume that for a period of time *post mortem*, DNA may survive from any microorganisms present in the specimen at death. This not only includes pathogens present at the time of death (either the cause of death or long-term infections) but commensals and other associated microbes. Despite several studies that have reported limited preservation of such DNA, for example the lack of preservation of *Helicobacter pylori* in ethanol-preserved specimens dating to the 18th century (Barnes *et al.* 2000a), over 45 published studies report the successful retrieval of ancient pathogen DNA from samples dating back to over 5,000 years old in humans, and as long as 17,000 years ago in other species (Table 1.2). As well as the usual sources of mummified tissue, bones and teeth, such studies have also examined a range of other tissue samples, including calcified pleura (Donoghue *et al.* 1998), tissue embedded in paraffin (Jackson *et al.* 1998; Basler *et al.* 2001), and formalin-fixed tissue (Taubenberger *et al.* 1997).

1.5 Ancient DNA: Damage, Contamination and Jumping PCR

1.5.1 Early dreams of ancient DNA shattered

Despite the publication of numerous, often high profile, aDNA studies, the field is troubled and its credibility hangs in question. Doubts first arose when attempts to independently replicate the 'successful' amplification of 17-20 mya fossilised plant chloroplast DNA (Golenberg *et al.* 1990) failed. The fossils showed remarkable preservation, including cell structures such as the chloroplasts, thus at first glance seemed ideal specimens. However, an attempt to replicate the results in a second laboratory failed (Sidow *et al.* 1991). The authors report only the presence of high

Table 1.2: Some ancient DNA studies reporting successful pathogen and micro-organism DNA retrieval.

Pathogen	Tissue	Author	Sample Origin[1]	Age[2]
Mycobacterium tuberculosis	Bone	Arriaza *et al*. 1995	Chile	2000BCE -1500 CE
M.tuberculosis	Bone	Braun *et al*. 1998	N. America	11th-15th century
M.tuberculosis	Bone	Baron *et al*. 1996	Germany	<200 years old
M.tuberculosis	Bone	Crubezy *et al*. 1998	Egypt	5400
M.tuberculosis	Bone	Faerman *et al*. 1997b	Lithuania	15th-17th century
M.tuberculosis	Bone	Gernay *et al*. 2001	England	10th century
M.tuberculosis	Bone	Haas *et al*. 2000	Hungary	7/8th - 17th century
M.tuberculosis	Bone	Mays *et al*. 2001	England	10-16th century
M.tuberculosis	Bone	Mays *et al*. 2002	England	10-16th century
M.tuberculosis	Bone	Nuorala *et al*. 1999	Baltic seabed, off Sweden	1679 CE
M.tuberculosis	Bone	Rafi *et al*. 1994	River Jordan	600 CE
M.tuberculosis	Bone[3]	Rothschild *et al*. 2001	N. America	17000 y old.
M.tuberculosis	Bone	Spigelman and Lemma 1993	Britain	Romano Britain to 18C
M.tuberculosis	Bone	Spigelman *et al*. 2003	Britain	Romano Britain to 18th century
M.tuberculosis	Bone	Taylor *et al*. 1996	Britain	1350-1538 CE
M.tuberculosis	Bone	Taylor *et al*. 1999	Britain	1350-1538 CE
M.tuberculosis	Bone	Zink *et al*. 2000a	Egypt	2120-500 BCE
M.tuberculosis	Bone	Zink *et al*. 2003a	Egypt	2050-500 BCE
M.tuberculosis	Bone	Zink *et al*. 2003b	Egypt	3500-500 BCE
M.tuberculosis	Calcified Pleura	Donoghue *et al*. 1998	Negev Desert	600 CE
M.tuberculosis	Mummified tissue	Fletcher *et al*. 2003	Hungary	18th century
M.tuberculosis	Mummified tissue	Fletcher *et al*. 2003	Hungary	18-19th century
M.tuberculosis	Mummified tissue	Nerlich *et al*. 1997	Egypt	1550-1080 BCE
M.tuberculosis	Mummified tissue	Salo *et al*. 1994	Peru	1000 CE
M.leprae	Bone	Taylor *et al*. 2000	Orkney	13th century
Yersinia pestis	Teeth	Drancourt *et al*. 1998	France	1590-1722 CE
Y.pestis	Teeth	Raoult *et al*. 2000	France	14th century
Various gastro intestinal bacteria	Mummified tissue - Bog body	Fricker *et al*. 1997	Britain	300 BCE
Various gastro intestinal bacteria	Mummified tissue	Zink *et al*. 2000	Egypt	1424-1398 BCE
Cornyebacterium spp.	Mummified tissue	Zink *et al*. 2001b	Egypt	1550-1080 BCE
Bacillus anthracis	Parrafin embedded tissue	Jackson *et al*. 1998	Sverdlorsk, former USSR	1979
Treponema pallidum pallidum	Bone	Kolman *et al*. 1999	Easter Island	200

Pathogen	Tissue	Author	Origin[1]	Age[2]
T.p.pallidum	Bone	Kolman *et al.*1999	Easter Island	200
Human T-Cell lymphotropic virus type 1	Mummified tissue	Li *et al*. 1999	S. America	1500
Influenza	Preserved and frozen	Basler *et al*. 2001	N. America	1918 CE
Influenza	Permafrost frozen	Reid *et al*. 1999	N. America	1918 CE
Influenza	Formalin fixed lung	Taubenberger *et al*. 1997	N. America	1918 CE
Phytophthora infestans	Herbarium samples[4]	Ristaino *et al*. 2001	Herbarium	1845-present
P. falciparum	Bone	Sallares *et al*. 2001	Mediterranean	5th century
P. falciparum	Bone	Taylor *et al*. 1997a	England	60 years old
P. falciparum	Mummified tissue	Taylor *et al*. 1997b	England	700
Trypanosoma cruzi	Mummified tissue	Guhl *et al*. 1997	S. America	4000
Trypanosoma cruzi	Mummified tissue	Guhl *et al*. 1999	S. America	4000
Trypanosoma cruzi	Bone	Guhl *et al*. 1999	S. America	4000
Ligula spp.	Formalin fixed tissue[5]	Li *et al*. 2000	China Quinghai-Tibet plateau	< 24
Ascaris spp.	Eggs in copralite	Loreille *et al*. 2001	Belgium	Middle Ages

[1]*Sample Origin* Geographical origin as given in original study. [2]*Age* in years or time period as given in original study. All tissues were human, with the exceptions of [3]Bison, [4]Potato spp., [5]Fish spp.

concentrations of high molecular mass bacterial DNA derived from living bacteria, and suggest that the most likely origin of such DNA was through recent contamination during the excavation process. Pääbo and Wilson (1991) also evince skepticism that a 770 bp sequence can be recovered after 20 million years. Based on laboratory calculations of depurination rates (and subsequent fragmentation) under hydrated conditions (Lindahl and Nyberg 1972), they claim no sequence over 800 bp should exist after 5,000 years, and nothing amplifiable should exist after 4 million years.

In response to these criticisms, Golenberg (1991) claimed the results were replicated by his own laboratory, as well as by other (unpublished) groups. Moreover, several possible explanations as to how DNA may be surviving over several million years are discussed. Firstly dehydration of the leaves may rupture tonoplasts, releasing high concentrations of phenolics. Additionally, any subsequent rehydration by tannic-rich water may have kept the leaves in an excellent state of preservation. Despite such arguments, an overwhelming body of additional evidence suggest that Golenberg *et al*. (1990) did not successfully retrieve authentic sequences. Not only did a further attempt at obtaining sequences fail (Soltis 1995), but analysis of other 'long-lived' biomolecules in the samples (including polysaccharides and proteins (Logan *et al*. 1993)) and amino acid racemisation (Poinar *et al*. 1996) demonstrated a lack of preservation of all biomolecules in the sample.

To compound matters, attempts to both replicate data produced in similar studies, and re-analyses of the original sequence data, have highlighted a worrying trend. These include: the 80-million-year-old Cretaceous 'Dinosaur' DNA sequence of Woodward *et al*. (1994), identified to most likely be contaminant mammalian pseudogene sequences (Allard *et al*. 1995; Collura and Stewart 1995; Hedges and Schweitzer

1995; Henikoff 1995; Van der Kuyl *et al.* 1995; Young *et al.* 1995; Zischler *et al.* 1995b); 18S DNA from Cretaceous dinosaur egg fossils (An *et al.* 1995, Li *et al.* 1995), re-identified as having 85% homology to flowering plants and fungi (Wang *et al.* 1997); and several analyses on amber encased specimens (Howland and Hewitt 1994; Austin *et al.* 1997a; Austin *et al.* 1997b; Walden and Robertson 1997) that fail to yield any DNA at all.

1.5.2 The *post mortem* degradation of DNA

As the field has matured it has become apparent that many of the early claims were over-optimistic, and the scope of aDNA is limited by the *post mortem* degradation of DNA. This rate of decay is so fast, that based on calculated deamination and depurination kinetics for the four nucleotides, it has been estimated that under physiological salt conditions, neutral pH and an ambient temperature of 15ºC, 100,000 years is a likely estimate of the time beyond which DNA will be degraded to below the size of a useful amplifiable fragment (Lindahl 1993a). However it is possible that under certain optimal environmental conditions, such as encasement in ice or permafrost, this period may be extended by two or three times. In warmer environments, the survival limit of amplifiable DNA is much shorter. For example, using observations on the decay rate of chloroplast DNA in papyri (and by inference human DNA in similar conditions), Marota *et al.* (2001) calculate that DNA at Egyptian archaeological sites has a half-life of 19-24 years, which suggests an upper limit of 672 years on the survival of authentic DNA. What is quite clear from this and other studies is that an assessment of the likelihood of DNA retrieval from ancient specimens must take into account the degradation processes that it undergoes.

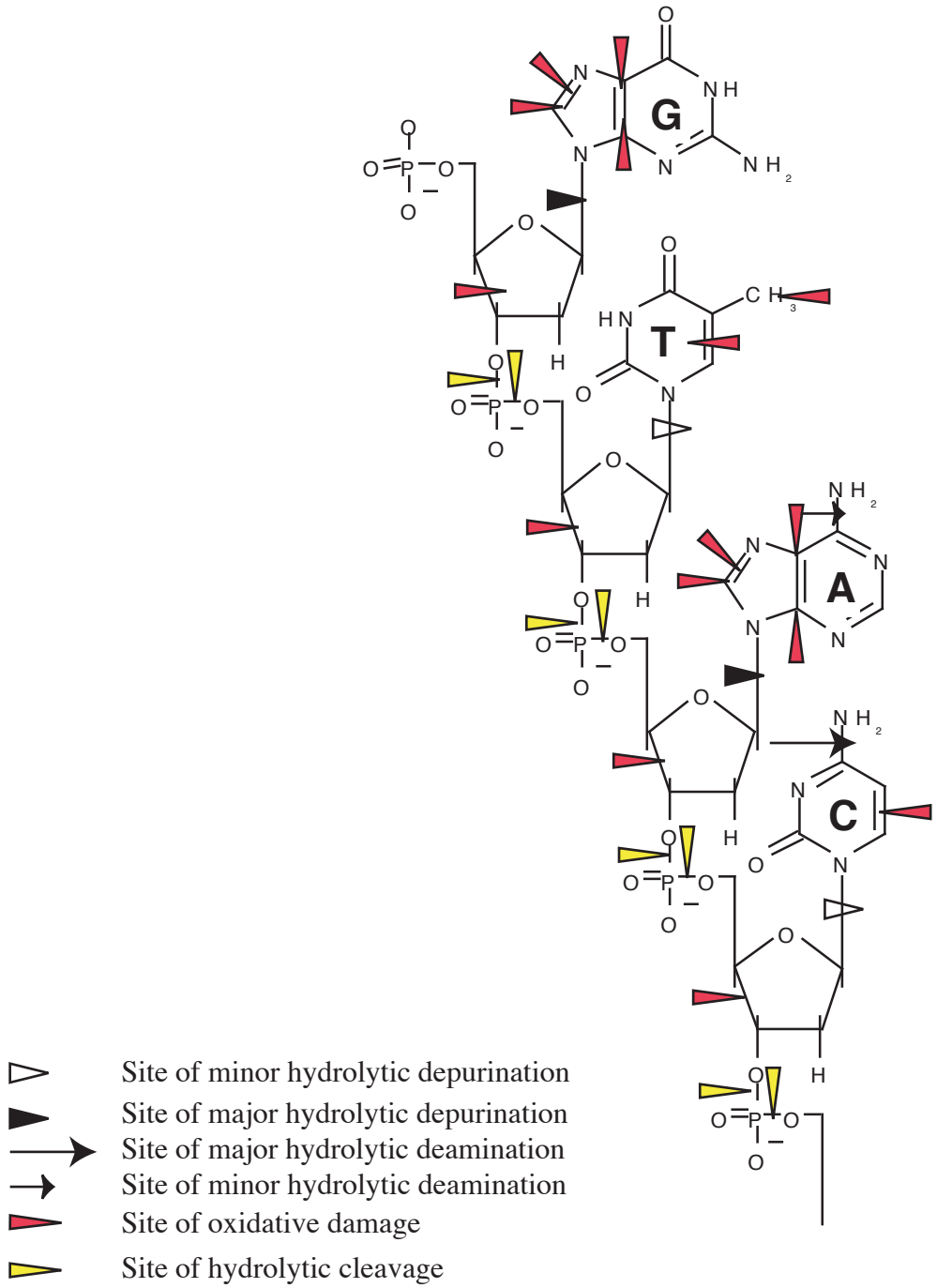
Exceptional circumstances apart, the degradation of endogenous DNA (*i.e.* that belonging to a sample of interest) commences shortly following its death. In humans, within 4-5 minutes of the death cell autolysis commences (Vass 2001). As the cells of the body are deprived of oxygen, carbon dioxide in the blood increases, pH decreases and wastes accumulate which poison the cells. Concomitantly, unchecked cellular enzymes, including lipases, proteases, amylases and nucleases, begin to dissolve the cell from the inside out. Soon the cells rupture, releasing nutrient-rich fluids that encourage the growth of internal and environmental micro-organisms (mainly bacteria, fungi and protozoa) involved in the further putrefaction of the remains. These contribute to further degradation of the DNA as they spread through the corpse. From the molecular biologist's point of view, despite the fact that most human diploid cells contain several billion bases of nuclear DNA, and thousands of copies of mitochondrial DNA, its decay is often so fast that within months, if not weeks, no PCR amplifiable template remains (Lindahl 1993b). Nevertheless, in exceptional circumstances, this degradation can be significantly reduced or halted altogether. Such conditions, believed to destroy/inactivate the nucleases and/or inhibit the action of micro-organisms include rapid desiccation (*e.g.* natural mummies), low temperatures (*e.g.* ice mummies or samples buried in permafrost) and high salt concentrations (Hofreiter *et al.* 2001b) as well as those in which the flesh is quickly removed from the host, thus limiting the substrate that micro-organisms thrive on and subsequent putrefaction (*e.g.* through mild cooking, butchering or very rapid degradation of the flesh).

In such scenarios, slower processes of DNA degradation modify the molecule. While the histone proteins associated with nuclear DNA can be expected to offer some protection from damage, their absence makes mtDNA very susceptible to a range of

biochemical attack (Poinar 2002). These biochemical modifications are believed to be analogous to those seen *in vivo*, and act via both the cross-linking and fragmentation of the molecule's chemical backbone and the alteration of individual nucleotide bases. Although the exact contributions of individual processes to the damage will vary with the direct environment surrounding specimens (*c.f.* Hofreiter *et al.* 2001b) a brief summary of those believed to be important follows.

A seminal study by Svante Pääbo (1989) still provides much of what is known today about DNA degradation. Although his analysis of the DNA retrieved from desiccated mummy samples demonstrated that they contained yields of double stranded DNA, attempts to PCR amplify the DNA ceded only small fragments of product. This lack of amplifiable DNA is a common phenomena of aDNA, and is believed to arise following the cross-linkage of the DNA backbone (inhibiting enzymatic amplification), and through double-strand breaks that arise as a result of sample exposure to various forms of environmental radiation. Further modifications precluding amplification of DNA arise as a result of post-mortem hydrolytic and oxidative damage (Fig 1.7). Diester bonds present on the phosphate sugar backbone of DNA are subject to rapid hydrolytic cleavage resulting in single strand nicks, and are thought to be the most common form of hydrolytic attack (Greer and Zamenof 1962; Shapiro 1981). Other structures within the DNA molecule are also at risk from hydrolytic attack. Deoxyribose sugar lacks the 2'-OH bond that is present in ribose, which in turn weakens the glycosidic bond joining the bases to the sugars. This susceptibility to hydrolytic attack (in particular of the purines, adenine (A) and guanine (G)) may result in depurination into a baseless (apurinic) site via base protonation (Lindahl and Andersson 1972; Lindahl and Nyberg 1972). Such sites rapidly undergo cleavage, producing additional sources of single strand nicks.

Figure 1.7: Potential targets of post mortem DNA damage.



Modified with permission from Lindahl (1993a), Hofreiter *et al*. (2001a) and Poinar (2002)

Pyrimidines (cytosine and thymine) are also susceptible to such reactions, forming apyrimidinic sites, though at rates *in vivo* as much as 100-500 times slower than purines (Lindahl and Nyberg 1972; Lindahl and Karlström 1973; Shaaper *et al.* 1983).

The notion that hydrolytic damage relies to a degree on the presence of water may suggest that archaeological samples that have remained fairly dry are more protected from such damage. However in such situations it is likely that role of oxidative modifications play a significant role (Poinar 2002). As in living systems, a large proportion of oxidative damage is believed to occur through the action of free radicals, such as the hydroxyl radical (•OH), peroxide radical ($•O_2$) and hydrogen peroxide (H_2O_2) (Rogan and Salvo 1992; Lindahl 1993). These may arise from exogenous sources such as ionizing radiation, UV light, as well as cellular processes that arise during bacterial and fungal degradation (Poinar 2002). Free radicals attack the integrity of the DNA molecule by adding to either the carbon C5 or C6 of pyrimidines via their shared double bond, or the C4, C5 and C8 of purines (Fig 1.7). Radicals thus created are unstable and often go through a series of reactions in the presence of oxygen to form other reactive radicals. One such example is thymine, which once attacked by •OH produces thymine peroxyl radicals. As with hydrolytic damage, the net result of the presence of such radicals is further fragmentation of the DNA molecule (Halliwell 1991). The sugar backbone can also be attacked by •OH through removal of hydrogen atoms from any of the five carbons. Oxidative damage is also believed to affect the DNA molecule through the modification of nucleotides to hydantoins. Using Gas Chromatography/Mass Spectrometry (GC/MS), Höss *et al.* (1996) detected the occurrence of two predominant forms, the oxidised pyrimidines 5-hydroxy-5-methylhydantoin (5-OH-5-MeHyd), a major derivative of thymine

following exposure to γ-radiation, and 5-hydroxyhydantoin (5-OH-Hyd). Unlike other forms of oxidative and hydrolytic damage, the authors suggested that these do not prevent DNA amplification through the fragmentation of the DNA but simply block the polymerase during PCR.

1.5.3 DNA damage results in limited templates available for aDNA analyses

To summarise, amplifiable aDNA is in short supply. It would thus seem logical (even if not always practised) that the first question any aDNA researcher should ask is 'Does the sample of interest contain any amplifiable DNA?' The opportunities for a sample to undergo DNA damage are substantial, the extent of each damage type (and thus the resulting modifications) are closely linked to the diagenetic environment of the sample. Without this knowledge it becomes very difficult to predict exactly what damage DNA within a sample has undergone, or whether DNA is even present in the sample at all. However, several general rules of thumb can be outlined:

1) DNA preservation correlates with archaeological site.
2) An estimate of DNA preservation between any two unrelated samples cannot be predicted using a simple age correlate.
3) Biochemical preservation of some other macromolecules (*e.g.* proteins) correlate with DNA survival.

It has been repeatedly demonstrated that unless two samples from an individual archaeological site are considered (and even then a correlation is not always guaranteed) age of sample does not correlate with state of DNA preservation (*c.f.* Pääbo 1989; Höss *et al.* 1996). Such preservation indices have prompted various methods of indirectly assessing the preservation of samples for DNA survival. One

approach, directly takes into account the biochemical limitation of temperature, and has been demonstrated as providing reasonable predictions of DNA survival. The so-called 'thermal age' of the sample, is an age calculation taking into account the preservation temperature of the sample (*c.f.* Smith *et al.* 2001). Other indirect methods of estimating DNA preservation (of variable accuracy) use correlates such as the frequency of water change (Nielsen-Marsh and Hedges 2000), microbial content (Burger *et al.* 1999), or biochemical data such as assays of amino-acid racemisation (Poinar *et al.* 1996; Poinar and Stankiewicz 1999), composition (Bada *et al.* 1999), and levels of hydantoin bases measured by Gas Chromatography/Mass Spectroscopy (GC/MS) (Höss *et al.* 1996).

1.5.4 Chemically-induced DNA damage

While the damage mentioned above largely derives from natural causes, many specimens examined in aDNA studies were originally collected for medical and/or other biological interest over the past few centuries. A large portion of such samples have been fixed to prevent microbial attack, using a range of preservatives (Vachot and Monnerot 1996). The most common preservative, and coincidentally that demonstrated to produce the most adverse effects on DNA, is unbuffered formaldehyde (Douglas and Rogers 1998). Through time formaldehyde oxidises to formic acid, resulting in a pH increase of approximately pH4 to pH2.5 (Vachot and Monnerot 1996), and resulting in many hydrolytic deamination events. Additionally, formylation of nucleic acids results in the production of Schiff bases on free amino groups of nucleotides, thus irreversibly cross-linking the formaldehyde aldehydic groups and DNA (Dubeau *et al.* 1986; Koshiba *et al.* 1993). This process is dynamic and depends principally on the length of time over which tissue is exposed to

formaldehyde (Hamazaki *et al*. 1993; Karlsen et al 1994), the pH of the fixative, and temperature of fixation (Goebel and Simmons 1993; Koshiba *et al*. 1993).

1.5.5 DNA copy number: an important factor

DNA degradation presents several problems for the aDNA field. Firstly, researchers are limited as to the quantities of amplifiable DNA that they can recover. Secondly, due to DNA fragmentation and cross-linking, researchers are rarely able to amplify templates over a few hundred bp in size (*c.f*. Pääbo 1989; Cooper and Poinar 2000; Hofreiter *et al*. 2001b). Several other copy-related issues further complicate matters. For example, the relative abundance of mtDNA over nuDNA suggests that mtDNA will be an easier template to recover than nuDNA. In addition, any pathogen DNA that is present in surviving tissues (*e.g*. bones, teeth, hair) will exist at significantly lower copy numbers than nuclear DNA, and thus should theoretically be harder yet to recover. Therefore, there are natural limitations on what can be achieved in the field – even to recover the complete nuclear genome of recently extinct and well-preserved animals, such as permafrost-preserved mammoths, would take many hundreds of thousands of PCR amplifications. A fifth problem, and that most serious for the credibility of the field is that of sample contamination. The low abundance of DNA extracted from a degraded specimen can easily be 'swamped' by modern DNA from external contaminants of the same or similar DNA sequence. Without proper treatment to remove the environmental DNA, it is a very simple (and unfortunately, common) matter to co-amplify both sources of DNA. Depending on the relative concentrations of host-to-contaminant DNA in the PCR, the target DNA may be either completely swamped, or modified suitably to result in erroneous data. One common cause of such problems identified in the early days of the field is via 'jumping' PCR (Pääbo *et al*. 1990), whereby incomplete amplification of 2 template

strands may result in a process that effectively mimics recombination (Figs 1.8 and 1.9). The chimeric products of such reaction can easily be mistaken for nuclear-mitochondrial copies (numts), or more seriously, a novel aDNA sequence. A classic example of such misinterpretation was reported by Kuznetzov *et al.* (2001), who claim to have identified a new species of buffalo, the Vietnamese 'Linh Duong' specimen, based on 12S rRNA amplified from several unusual museum horn sheaths. In an attempt to reproduce such results, Hassanin (2002) demonstrated that the reported sequence derived from a 'Jumping' chimera of 3 forms of mammal DNA present in the extract.

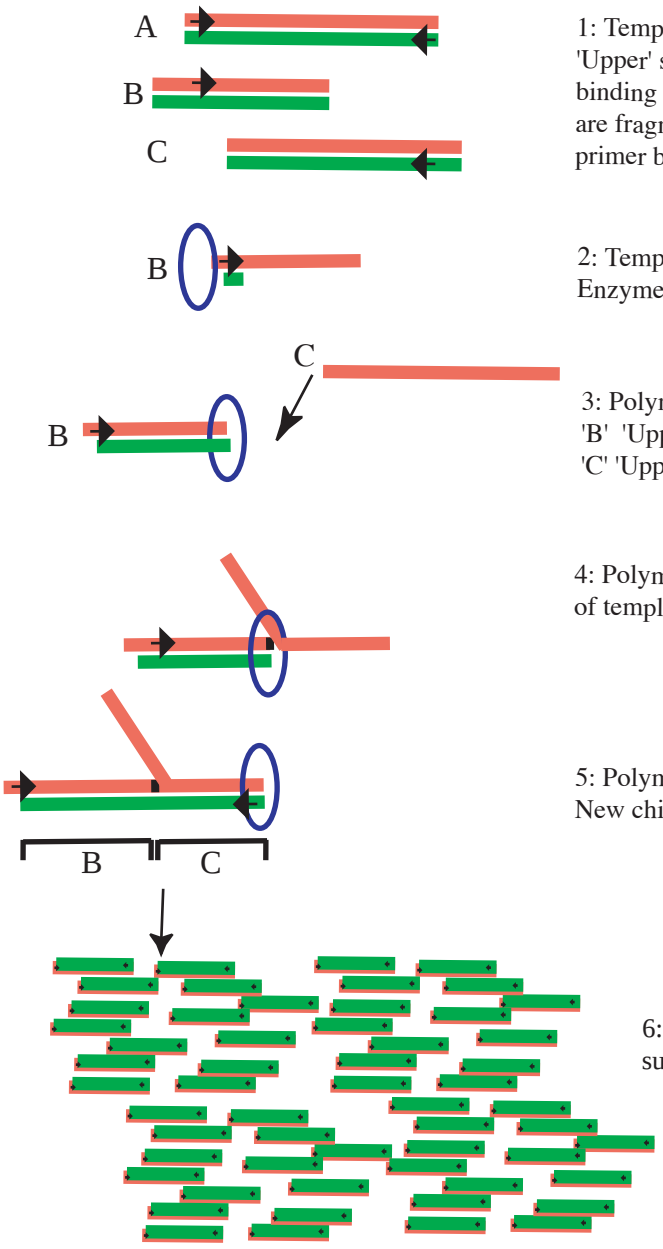
1.5.6 DNA damage renders samples susceptible to contamination

Due to the multi-step processes that are employed to successfully extract the low concentrations of DNA associated with ancient remains, contaminants can enter the system at many stages (Fig 1.10). The most common sources of contaminants include the following:

1) Related specimens. (Cooper 1994) It is possible for tissue or DNA extracted from one specimen to cross-contaminate others extracted in the same environment. In this case, when both sources are co-extracted, PCR primers may bind equally well to the authentic and contaminant sequence (or in the worst case, preferentially to the contaminant), thus co-amplify.

2) Micro-organisms (Sidow *et al.* 1991). Most PCR reactions are based upon the premise that primers will bind to specific loci not present in unrelated organisms, and thus will selectively amplify only DNA of interest. However, due to the vast numbers of genetically different micro-organisms that are estimated to exist (Finlay and Clark 1999; Finlay *et al.* 1999), and that DNA sequences do not exist for the overwhelming

Figure 1.8: 'Jumping' PCR may result in recombination between two damaged target DNA templates, to produce a chimeric sequence.



1: Template 'A' represents an undamaged template, showing 'Upper' strand (red), 'Lower' strand (green) and primer binding sites (→). Templates 'B' and 'C' represent two templates that are fragmented due to damage. Template 'B' lacks the 'Lower' strand primer binding site, template 'C' the 'Upper' strand primer binding site.

2: Template 'B' dissasociates in early stage of PCR cycle. Enzyme and complementary primer bind to 'Upper' strand.

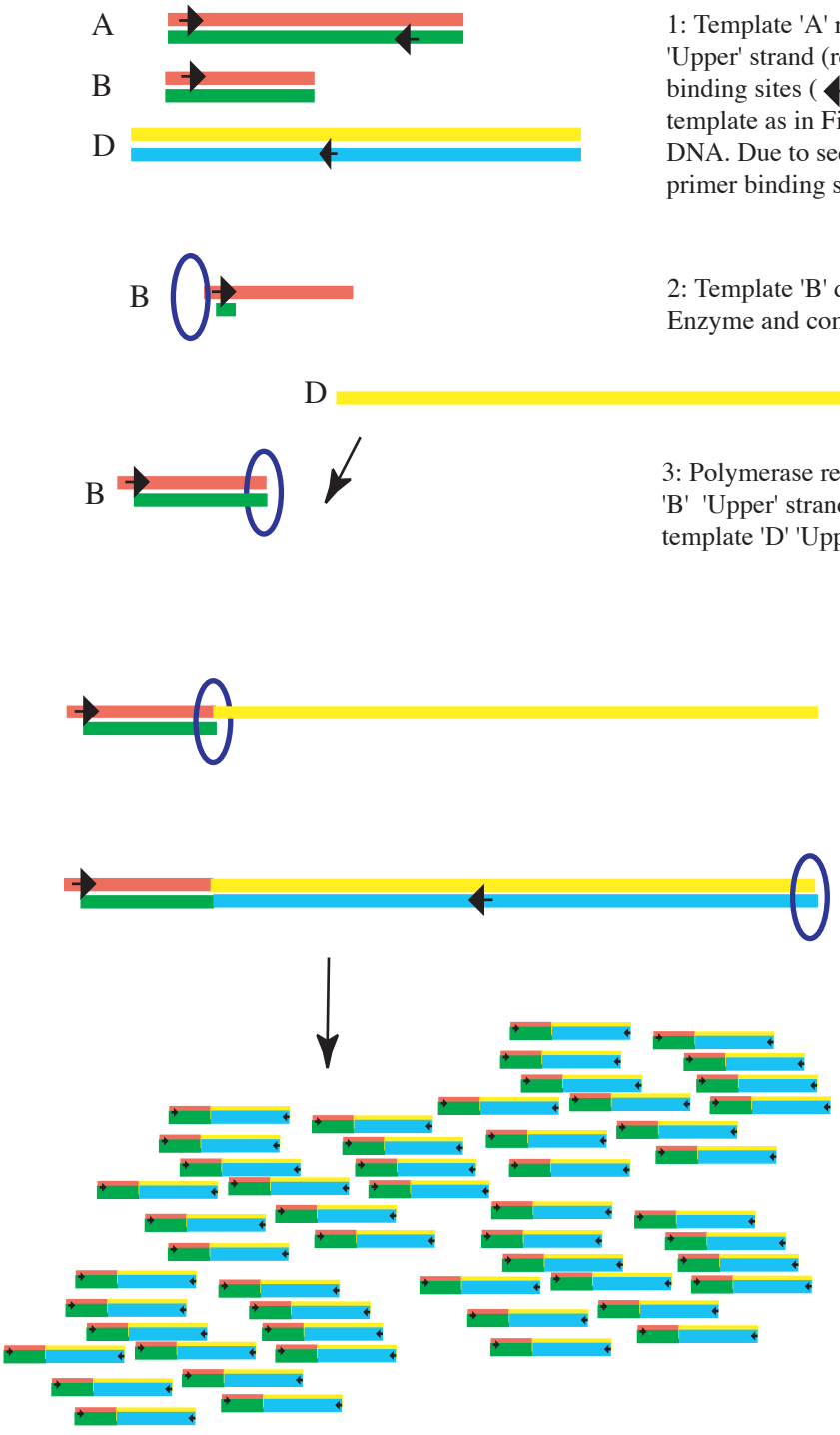
3: Polymerase reaches end of amplifiable template 'B' 'Upper' strand and pauses. Dissasociated template 'C' 'Upper' strand binds to enzyme-DNA complex.

4: Polymerase continues replication along 'Upper' strand of template 'C'

5: Polymerase reaches end of template 'C' 'Upper' strand. New chimera 'BC" is formed.

6: Chimera 'BC" amplified numerous times during subsequent PCR cycles

Figure 1.9: 'Jumping' PCR between damaged target template and contaminant template containing sequence similarity at primer binding site may also form recombinant DNA.



1: Template 'A' represents an undamaged template, showing 'Upper' strand (red), 'Lower' strand (green) and primer binding sites (◄). Template 'B' represents damaged template as in Fig.1.8. Template 'D' represents contaminant DNA. Due to sequence similarity, an 'Upper' strand primer binding site exists in the contaminant

2: Template 'B' dissasociates in early stage of PCR cycle. Enzyme and complementary primer bind to 'Upper' strand.

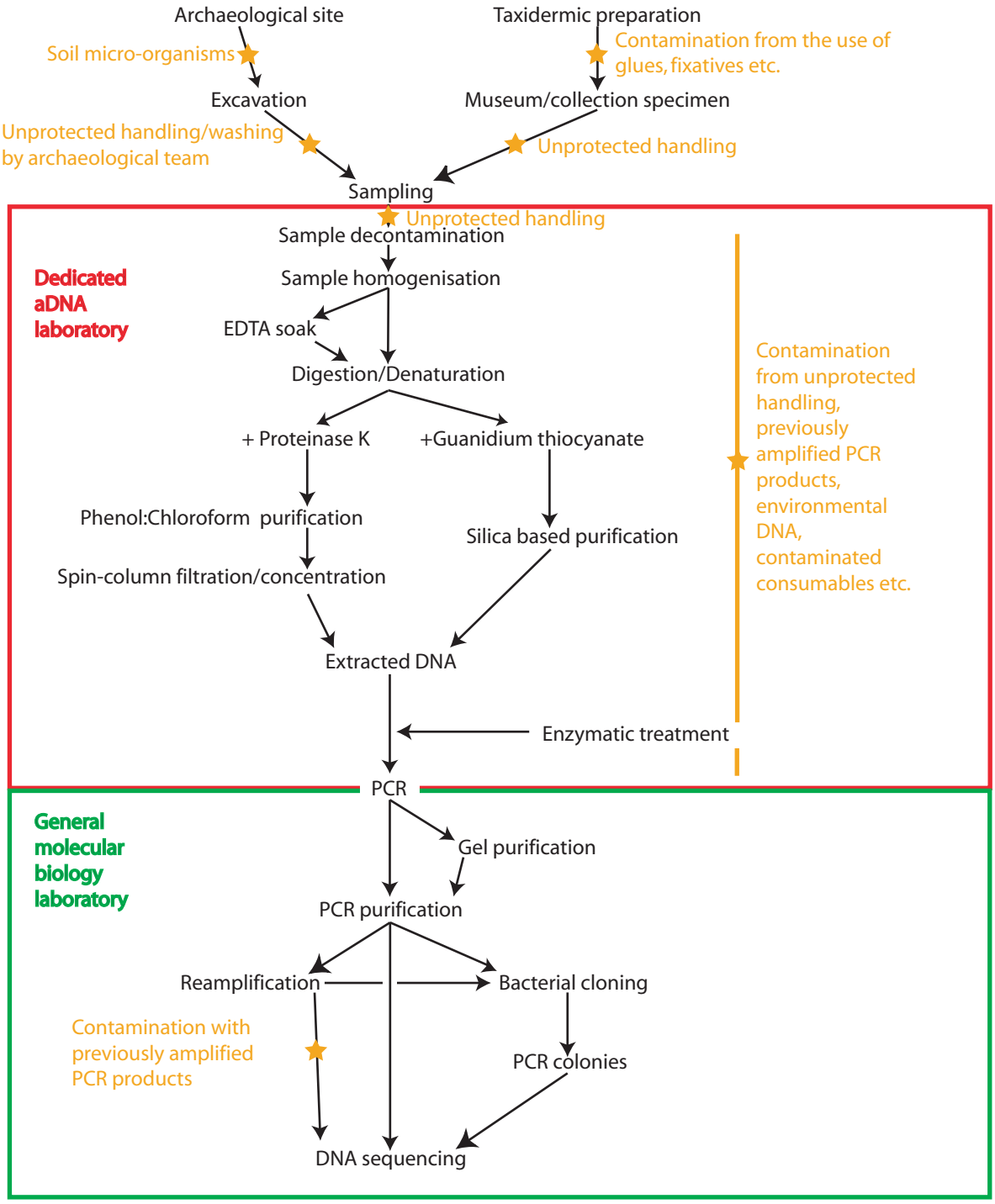
3: Polymerase reaches end of amplifiable template 'B' 'Upper' strand and pauses. Dissasociated contaminant template 'D' 'Upper' strand binds to enzyme-DNA complex.

4: Polymerase continues replication along 'Upper' strand of template 'D'

5: Polymerase reaches end of template 'D' 'Upper' strand. New chimera 'BD' is formed, containing 'Uppert' strand primer binding site.

6: Chimera 'BD' amplified numerous times during subsequent PCR cycles. Final product is recombinant of endogenous and contaminant DNA

Figure 1.10: Steps within the ancient DNA extraction and analysis process where samples are susceptible to contamination.



majority, it is very difficult to rule out the presence of short sequence homologues between the sample of interest and contaminant micro-organisms in the environment.

3) Sample handling (Handt *et al.* 1994a). Of particular importance in the analyses of ancient human remains, but also in any species in which the gene examined is highly conserved (*e.g.* 12S rRNA). Unprotected handling may well coat or impregnate samples with the handler's sweat or skin cells, and thus exogenous DNA.

4) Glue and other treatments (Nicholson *et al.* 2002). Many museum and other 'collection' specimens have been treated with glues and varnishes containing some animal-derived products, and thus provide a contaminating DNA source. This may explain the presence of contaminant DNA in specimens such as the aforementioned 'Linh Duong' buffalo (Hassanin 2002).

5) Previously amplified PCR fragments (Kwok and Higuchi 1989). The exponential nature of the PCR results in huge numbers of amplified products (amplicons) concentrated into tiny volumes. For example, if 100% amplification efficiency is assumed, a 40-cycle reaction that initially starts with just one amplifiable template molecule in a volume of 25µl, will result in a final product of almost $1.1x10^{12}$ amplicons. To put this into context, Kwok and Higuchi (1989) have described this phenomenon in the following manner - if the above reaction was diluted into an Olympic size swimming pool, each µl of the swimming pool would still hold 1 copy of template! As a result of such high concentrations of PCR products, it is not surprising that high levels of amplicons may spread rapidly and contaminate ancient samples with DNA, swamping the authentic DNA. In addition, high levels of contamination can persist in a laboratory environment over long periods of time. For example, Kolman (1999) reports the persistence of goldfish DNA amplicons in a molecular biology laboratory, over one year after the last PCR was amplified.

1.5.7 DNA damage may also lead to artefactual sequence modification

Not only can DNA damage result in low template number and poor amplification, it can also generate 'artefactual' base calls during sequencing. Observed initially by Pääbo (1989) via the molecular cloning of amplified aDNA templates, various forms of DNA degradation manifest themselves as modifications to the actual DNA sequence itself (termed miscoding lesions). Although Pääbo first demonstrated the presence of these modifications using enzymatic treatment of his aDNA samples, it is only recently that the full extent of the modifications, and the biochemistry behind the changes has started to become understood. This is partially a result of the nature of PCR enzymes to incorporate a low frequency of 'incorrect' bases during amplification, ranging from ~$2x10^{-4}$ to <1 x 10^{-5} per nucleotide per cycle (Hansen *et al*. 2001), and masking real changes. Recently, the advent of so called 'Hi-fidelity' enzymes, with much lower misincorporation rates than 'standard' *Taq,* has enabled the identification and characterisation of two very common complementary pairs of base modification. These are all due to transitions, and have been termed Type 1 (thymine to cytosine (T→C); and adenine to guanine (A→G)) and Type 2 (cytosine to thymine (C→T); guanine to adenine (G→A)) (Hansen *et al*. 2001). This issue is dealt with in greater depth in chapter 2.

In vivo and *in vitro* studies have demonstrated that under comparable environmental (*i.e*. temperature and pH) conditions the rate of deamination of cytosine is over twenty times that of adenine (Lindahl 1979; Karran and Lindahl 1980), and thus it is intuitive to expect that cytosine dependent Type 2 transitions will accumulate much faster. However, analyses of *post mortem* samples have demonstrated that the rate of accumulation of the breakdown products, and hence miscoding lesions, is dependent

on the sample environment (Gilbert *et al.* 2003a). For example the common environmental mutagens nitrous acid and bisulfite, will preferentially deaminate cytosine over adenine (Schuster 1960; Lindahl and Nyberg 1974; Lindahl 1979). Cytosine is also more susceptible to heat induced deamination than adenine (Shapiro and Klein 1966; Notari 1967), though following the assumption that cytosine and adenine deamination are associated with similar activation energies, it is predicted that adenine deamination will persist at lower temperatures than cytosine (Karran and Lindahl 1980). The pH of the immediate environment also affects each base differently (Jones *et al.*1966) and adds a further variable to the model.

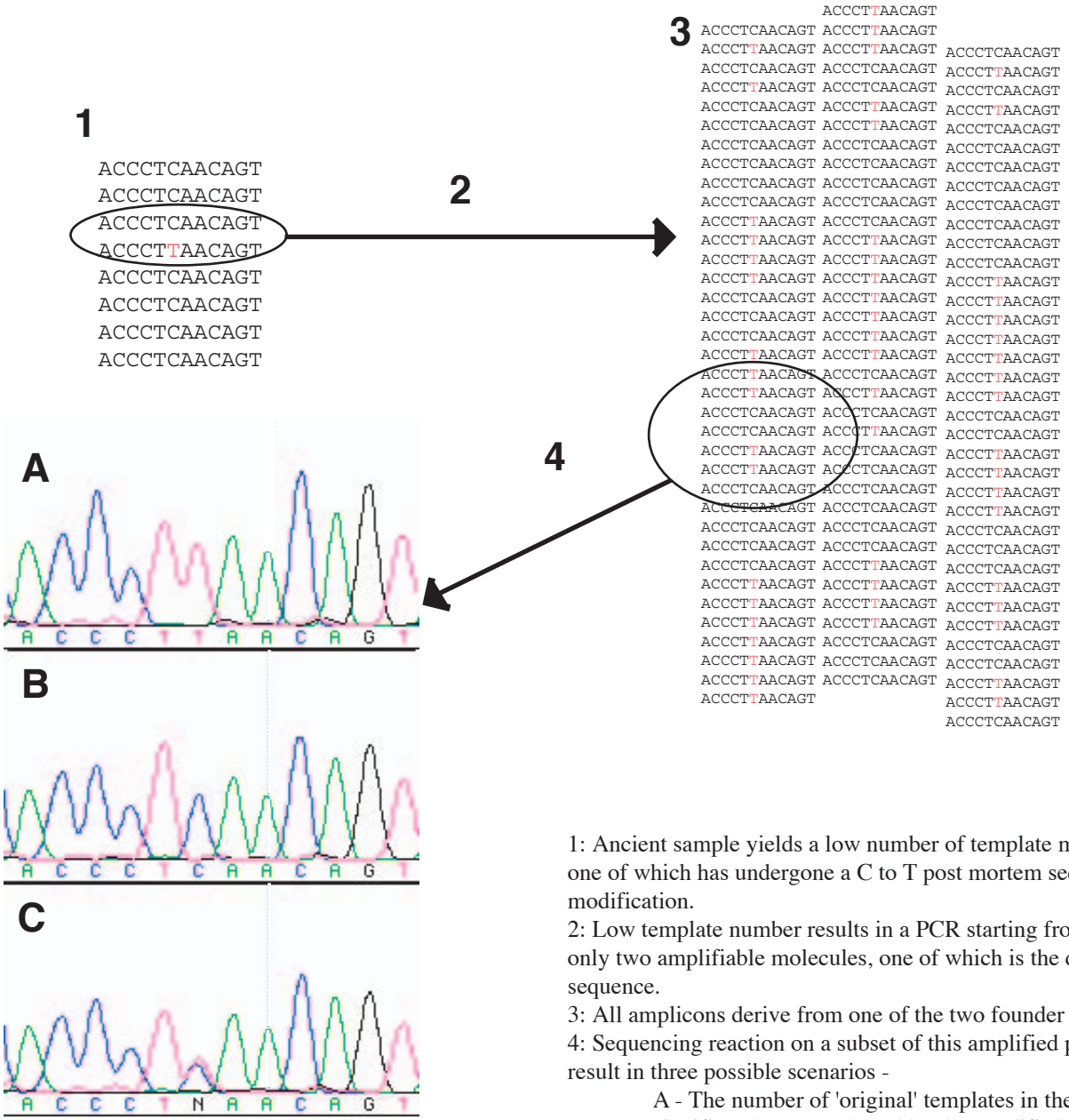
The presence of modified nucleotides in samples containing low template numbers can be manifested as sequence misidentification. As demonstrated in Fig 1.11 a single miscoding lesion on one template molecule in a PCR reaction can, once amplified, result in sequence that is easily misinterpreted.

1.6 Remedies to the pitfalls of aDNA

Various remedies have been suggested to help eliminate the pitfalls that plague aDNA. These can be roughly divided into two strategies, those trying to remove contaminants that may be in the sample prior to analyses, and those trying to prevent further contamination during analysis.

A review of literature published by the majority of the aDNA field leaves the reader under the impression that the removal of contaminants present in samples prior to their collection for aDNA analysis is simple. Additionally it would appear that samples are in fact hard to contaminate, with their morphologically sound appearance (especially in bones and teeth) presenting an efficient barrier to the permeation of contaminants. It is assumed by the majority of the field that current decontamination

Figure 1.11: Modification of sequences through miscoding lesions.



1: Ancient sample yields a low number of template molecules, one of which has undergone a C to T post mortem sequence modification.
2: Low template number results in a PCR starting from only two amplifiable molecules, one of which is the damaged sequence.
3: All amplicons derive from one of the two founder templates.
4: Sequencing reaction on a subset of this amplified product may result in three possible scenarios -

A - The number of 'original' templates in the reaction significantly outnumbered by the 'modified' templates, yielding the incorrect sequence.

B - The number of 'original' templates in the reaction significantly outnumber the 'modified' templates, yielding the correct sequence.

C - A heterozygous phenotype is observed. Base may be called at the analyst's discretion (or reported as heteroplasmy in mitochondrial DNA).

techniques, which focus on cleansing or removing the external layers of samples, are suitable in removing contaminants. Techniques include the treatment of samples (and plastic-ware that may also be contaminated) with products that destroy DNA, such as DNAase (Eshleman and Smith 2001), diluted bleach solutions (Merriweather *et al.* 1994; Richards *et al.* 1995) and/or H_2O_2 (Merriweather *et al.* 1994), exposure to ultra-violet light (λ=254nm) (Oota *et al.* 1995; Zierdt *et al.* 1996) as well as the physical removal of tissue surfaces through techniques such as shot-blasting (Richards *et al.* 1995), sanding (Pusch *et al.* 2000), slicing (Zierdt *et al.* 1996) or even 'exhaustive' brushing with a toothbrush (Gonzalez-Oliver *et al.* 2001).

Therefore, the majority of the field's attention has been turned towards the prevention of further contaminants from entering the samples during the extraction process. Initial methods included the routine utilisation of blank extractions and PCR negative controls. However due to phenomena such as the 'carrier effect' (Cooper 1994; Handt *et al.* 1994), (whereby a few real templates which normally would be adsorbed to the wall of the plastic-ware, and thus not amplify, actually amplify when an extract's DNA (or other molecules such as sugar) 'carry' them into the reaction), such controls have repeatedly been shown to be inefficient in spotting low-level contaminants (Handt *et al.* 1994a; Richards *et al.* 1995; Kolman 1999; Kolman and Tuross 2000).

1.6.1 Authentication criteria to help minimise erroneous results

A list of suggested criteria has thus been published to help limit the effect of sample contamination. This list, originally suggested by Handt *et al.* (1994a), and later updated by Cooper and Poinar (2000) has many similarities with criteria suggested

for use in forensic studies (Carracedo *et al*. 2000) and aims to help minimise some of the field's errors. The list is as follows:

1) Isolation of work areas in order to separate sample preparation from amplified products.

2) Negative control extractions and amplifications to screen for contaminants entering the process at either stage.

3) Appropriate molecular behaviour – due to DNA degradation, aDNA studies should be suspicious of success in amplifying large DNA fragments.

4) Reproducibility – multiple PCR and extractions should yield consistent results.

5) Cloning of products to assess for damage, contamination and 'jumping' PCR.

6) Independent replication – the generation of consistent results by independent research groups.

7) Biochemical preservation – preservation of other biomolecules that correlate with DNA survival (*e.g*. collagen, amino acid racemisation) should indicate good sample preservation.

8) Quantification - by competitive PCR or real-time PCR to give an indication of the number of starting templates in the reaction.

9) Associated remains – Are associated remains equally well preserved, and do they show evidence of contamination?

Some human aDNA studies listed in Table 1.1 do adhere (in the most part) to these guidelines - the initial analyses on Ötzi, the Tyrol Ice Man (Handt *et al*. 1994b), a study on the heart of Louis XVII (Jehaes *et al*. 2001), all four Neanderthal DNA studies (Krings *et al*. 1997; Krings *et al*. 1999; Ovchinnikov *et al*. 2000; Schmitz *et al*. 2002), a study on remains from paleolithic humans excavated from various Alpine locations (di Benedetto *et al*. 2001), a study on a 500 year old Canadian Ice mummy

(Monsalve *et al*. 2002), a study on Amerindian paleofaeces (Poinar *et al*. 2001), a study on relatively recent Andaman Islander samples (Endicott *et al*. 2003), and finally a recent study on Cro-Magnon remains (Caramelli *et al*. 2003). Nevertheless, the majority do not. Similarly, only four ancient pathogen DNA studies also follow the guidelines, all of which claim the retrieval of DNA from members of the *Mycobacterium tuberculosis* complex (Rothschild *et al*. 2001; Zink *et al*. 2001a; Zink *et al*. 2003a; Zink *et al*. 2003b).

Given the numerous problems reported on 'antidiluvian' DNA studies, the guidelines appear at first sensible. Further, various authors highlight similar problems within aDNA studies examining younger material. For example Ovchinnikov and Goodwin (2003) highlight potential problems with DNA sequences purportedly extracted from the famed Russian 'Sunghir' burials (Polotoraus *et al*. 2000). Additionally, given the relatively low levels of successful mitochondrial aDNA amplification on well controlled, well preserved goose bones from northern English Anglo-Saxon burial sites (Barnes *et al*. 2000b; Haynes *et al*. 2002), the much higher amplification success rates from warmer, older sites (such as ancient Egypt and the Mediterranean coast) are curious: using the nuclear marker amelogenin Faerman *et al*. (1995) successfully amplified 18/22 samples examined. The anomalous results do not stop here - various human aDNA studies also report the successful retrieval of nuclear, but not mitochondrial DNA (Fily *et al*. 1998).

In contrast, there are also examples of studies within the field that suggest the guidelines may not be strict enough. Various authors remark on the persistence of contaminant DNA even when strict guidelines are followed (Richards *et al*. 1995; Kolman and Tuross 2000; Schmitz et al 2002).

1.7 The scope of this thesis

Unlike most other scientific fields, where the testing of basic assumptions are *de rigueur* before advanced studies are undertaken, the high profile nature of the field of aDNA has encouraged such testing to be overlooked. Thus the field still both lacks basic knowledge about the degradation and preservation of aDNA, and is rich with untested assumptions about the suitability of aDNA techniques in minimising error. Many of these assumptions underpin the credibility of the field, and include the suitability of different materials as sources of uncontaminated and authentic DNA, as well as the efficacy of sample pre-preparation techniques. In addition the effectiveness of suggested 'aDNA guidelines' (*c.f.* Cooper and Poinar 2000) in preventing the publishing of artificial results is not fully understood.

In this thesis I address some of the gaps in aDNA theory and knowledge through initially investigating in detail the nature of hydrolytic *post mortem* DNA modifications. I then progress to address the relative merits of using bone, teeth and hair as aDNA sources. Finally I extend the investigation away from human aDNA, to investigate the reliability of results that claim the retrieval of authentic pathogen aDNA from human samples.

The structure of this thesis incorporates seven experimental/research chapters, in addition to the introduction and conclusion. The experimental/research chapters have been written as 'stand-alone' scientific papers, and have either already been published (chapter 2 - Gilbert *et al.* 2003a, chapter 3 – Gilbert *et al.* 2003b, chapter 7 – Gilbert *et al.* 2004a), are in press (chapters 8 – Gilbert *et al.* 2004b), or have been submitted for publication (chapters 4,5,6).

The first 3 chapters focus on a fundamental problem in aDNA studies, the nature of *post mortem* damage.

Chapter 2 contains an in-depth study into the phenomenon of damage-driven miscoding lesions, using the largest dataset of cloned ancient human DNA sequences generated to date. The chapter provides new understandings of the mechanisms of such damage. In chapter 3, the discoveries of chapter 2 are extended to investigate the spatial clustering of such damage in human mitochondrial DNA. The findings of this chapter highlight implications of misleading sequence modifications of aDNA sequences. In addition the results provide insights into the mutational processes shaping mitochondrial DNA *in vivo*.

Due to the omnipresent risk that results generated in human aDNA studies may result from contamination, chapter 4 confirmed the results presented in chapter 3 using a dataset of non-human ancient remains. The findings also provide further insights into *in vivo* mtDNA mutation.

Chapters 5 to 8 focus on the reliability of human samples as DNA sources in aDNA analyses.

In chapter 5 an analysis is undertaken into the suitability of bones and teeth as sources of uncontaminated human DNA, using a range of human archaeological samples of varied preservation conditions. The results provide insights into the persistence of contamination in such samples.

In chapter 6, the long-term survival of DNA in hair shafts is explored. The results suggest that such material presents an alternative and more reliable source of DNA for human studies.

In chapter 7, the suitability of human bone and teeth remains for use in ancient pathogen DNA studies is assayed. Although the focus of this chapter is on the retrieval of *Yersinia pestis* DNA, (the putative source of the black death) the results

enable parallels to be drawn with other pathogen aDNA studies, and shed some doubt over previous results.

Following the conclusions of chapter 7, and in response to a paper defending the retrieval of ancient pathogen DNA from Egypt, chapter 8 critically reviews the potential for successful retrieval of aDNA from ancient Egypt.

1.8 Collaboration on experimental studies

The majority of aDNA sequence generation was undertaken at the Henry Wellcome Ancient Biomolecules Center (ABC) by the author with the exception of where stated in the thesis as independent replication by colleagues in Denmark – Eske Willerslev, Anders Hansen of Københavns Universitet, and Lars Rudbeck of the Panum Institut, Copenhagen. Various other experimental data were generated, and archaeological samples donated, by colleagues elsewhere. These are as follows:

In chapter 2 and 3, archaeological specimens were provided by Prof Martin Biddle and Birthe Kjølbye-Biddle, of the Winchester Research Unit, Ana Topf of the Department of Archaeology at the University of Durham, Dr Carles Lalueza-Fox of the University of Barcelona, Spain, and Dr Niels Lynnerup of the Panum Institute, Denmark. Extracted bear DNA was provided by Dr Ian Barnes (formerly) of the ABC.

In chapter 4, bison DNA extracts were provided by Beth Shapiro of the ABC. Unpublished bison phylogenetic trees were provided by Dr Alexei Drummond of the ABC.

In chapter 5, archaeological specimens were provided for by Prof Gordon Turner-Walker (then) of the University of Trondheim, Norway, Prof Paul Arthur of the University of Lecce, Italy, Prof Martin Biddle and Birthe Kjølbye-Biddle, Dr Niels

Lynnerup, and the York Archaeological Trust. All biochemical preservation data on the Matera samples was assayed and kindly provided by Dr Matthew Collins, Dr Colin Smith, Dr Christina Nielsen-Marsh, Miranda Jans, Kurt Prangenberg and Kirsty Penkman at NRG, University of Newcastle.

In chapter 6, hair samples were provided by Professors Vyacheslav Mododin and Natalia Polo'smak of the Institute of Archaeology, Novosibirsk, Dr Tamsin O'Connell of RLAHA, University of Oxford, Dr Andrew Wilson of the Department of Archaeological Sciences at the University of Bradford, Prof Paul Lehrmann of The University of Ohio, Professor Alan Cooper and Malcolm Neame of the BBC. Histological preservation information on samples was provided by Dr Andrew Wilson and Dr Desmond Tobin. Isoptope analyses on samples was provided for by Dr Tamsin O'Connell, and Dr Mike Richards of the Department of Archaeological Sciences, University of Bradford. ^{14}C dating on the Bison sample was provided by Dr Tom Higham of RLAHA, University of Oxford.

In chapter 7 teeth samples were provided by Dr Niels Lynnerup, Prof Elisabeth Carniel of the Pasteur Institute, Paris, Prof Bill White of the Museum of London and Dr Mike Prentice of Bart's and the London Medical School. As part of a dual group study into the reliability of ancient plague DNA sequences, a proportion of the samples analysed were undertaken by Dr Mike Prentice, Dr Kahn Voong and Jon Cuccui of Bart's and the London Medical School. These are clearly marked in the chapter.

In chapter 8, estimates of DNA survival with temperature were provided from an unpublished manuscript by Dr Matthew Collins, and Dr Colin Smith. Additionally, unpublished manuscripts detailing the interaction of slaked lime with natron were provided by Prof Jaap Goldsmit of the Academic Medical Centre of Amsterdam.

Helpful discussion on the degradation of ancient DNA, and the archaeology of ancient Egypt were provided by Dr Hendrik Poinar of the Max Plank Institute for Evolutionary Anthropology, Leipzig, and Dr Ian Barnes and Julie Eklund of University College London.

Chapter 2

Characterisation of genetic miscoding lesions caused by *post mortem* damage

Modified from the'American Journal of Human Genetics' publication

'Characterisation of genetic miscoding lesions caused by *post mortem* damage'

Am. J. Hum. Genet. 72:48-61 (2003)

M Thomas P Gilbert[1], Anders J Hansen[2], Eske Willerslev[2], Lars Rudbeck[3], Ian Barnes[1,5], Niels Lynnerup[4], Alan Cooper[1]

[1]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, Oxford University, Oxford, England
[2]Department of Evolutionary Biology, Zoological Institute, University of Copenhagen, Denmark
[3] Research Laboratory, Institute of Forensic Medicine, University of Copenhagen, Denmark
[4]Laboratory of Biological Anthropology, Institute of Forensic Medicine, University of Copenhagen, Denmark
[5]*Current address* Department of Biology, Darwin Building, University College London, London, England

Summary

The spectrum of *post mortem* damage in mitochondrial DNA was analysed in a large dataset 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 (SP-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 towards Type 2 is observed with increasing amounts of damage. The two transition Types can be used to identify the original DNA strand, Light (L) or Heavy (H), on which the initial damage event occurred, and this can increase the number of jumping PCR artefacts detected by up to 80%. No bias toward H strand specific damage events is noted within the hypervariable 1 (HVR1) region of human mitochondria, suggesting the rapid *post mortem* 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. Lastly, a significant correlation between archaeological site and sample-specific level of DNA damage was detected.

2.1 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-A (adenine) or –G (guanine) bonds, rapidly destroying the DNA backbone (Lindahl 1993a; 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 replication, but instead 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 (see chapter 3). The few detailed studies of miscoding lesions concur with earlier hypotheses (Pääbo 1989; Lindahl 1993a; 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.* 2001a). 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 it will be read as a G→A transition. Similarly, an A→G degradation may be observed as either A→G, or a T→C transition (Hansen *et al.* 2001; Hofreiter *et al.* 2001a). 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 this study a large dataset of previously published cloned sequences of human and non-human mitochondrial aDNA are analysed. The 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 *post mortem* miscoding lesions are investigated by digesting samples with the enzyme Uracil-N-Glycosylase (UNG) prior to amplification, as well as with single primer extension PCR reactions (Hofreiter *et al*. 2001a). 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. This data also provides a new methodology for improved detection of PCR jumping events (Pääbo *et al*. 1990).

2.2 Samples, Materials & Methods

2.2.1 Samples

This study uses the large dataset of cloned ancient human mitochondrial sequences published in Gilbert *et al*. (2003b) (chapter 3), 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; Poinar *et al*. 2001; Lalueza-Fox *et al*. 2001), bears (Loreille *et al*. 2000a; Barnes *et al*. 2002) and ratites (Cooper *et al*. 2001a). Full details of the samples and sequences are given in Table 2.1. Base damage, and authentication criteria follow chapter 3), and insertions/deletions were removed from all datasets. By convention, all sequences are described in the L strand orientation.

2.2.2 Sequence generation and enzyme fidelity

The majority of the data analysed was generated using low error-rate polymerases (*i.e*. Platinum Hifi (Invitrogen, UK), an enzyme mixture composed of recombinant *Taq* DNA polymerase, *Pyrococcus spp*. GB-D thermostable polymerase and Platinum *Taq* Antibody) which has 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, three modern human samples extracted with a standard phenol:chloroform protocol (Hillis *et al*. 1996) were amplified on two occasions after Barnes *et al*. (2002), but without secondary re-amplification. Mitochondrial Hypervariable Region 1 (HVR1) primers 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 determined from six data sets were analysed (Saiki *et al*. 1988; Dunning *et al*. 1988;

Table 2.1: Damage calculations, archaeological site of origin and age of samples.

Sample[1]	PCRs[2]	Ampd[3]	Damage[4]	d[5]	Site[6]	Code[7]	Age[8]
Tg44	1	147	8	0.0544	Greenland	5	800
Tg54	1	147	8	0.0544	Greenland	5	800
Tg76	1	355	2	0.0056	Denmark	1	300
Tg77	3	857	5	0.0058	Denmark	1	500
Tg80	1	355	7	0.0197	Denmark	1	600
Tg85	1	355	10	0.0282	Denmark	1	300
Tg103	1	344	10	0.0291	Denmark	1	800
Tg104	1	147	3	0.0204	Orkney	6	1000
Tg105	1	355	0	0	Denmark	1	800
Tg112	1	147	13	0.0884	Greenland	5	800
Tg114	1	355	8	0.0225	Denmark	1	300
Tg116	2	502	19	0.0378	Denmark	1	600
Tg120	1	355	2	0.0056	Denmark	1	300
Tg123	1	147	7	0.0476	Denmark	1	500
Tg127	1	147	7	0.0476	Denmark	1	300
Tg128	2	502	2	0.004	Greenland	5	800
Tg129	2	502	8	0.0159	Greenland	5	800
Tg131	1	147	3	0.0204	Greenland	5	800
Tg133	1	147	3	0.0204	Greenland	5	800
Tg136 a	2	502	9	0.0179	Repton	7	800
Tg136 b	2	502	11	0.0219	Repton	7	800
Tg137 a	3	294	8	0.0272	Repton	7	1300
Tg137 b	3	294	13	0.0442	Repton	7	1200
Tg138	1	147	5	0.034	Repton	7	1100
Tg141	1	147	4	0.0272	Repton	7	1100
Tg142	1	355	2	0.0056	Repton	7	1100
Tg143	1	147	4	0.0272	Repton	7	1100
Tg145	1	355	10	0.0282	Repton	7	1200
Tg146	1	147	5	0.034	Repton	7	1000
Tg148	7	1029	17	0.0165	Repton	7	1300
Tg149	7	1029	140	0.1361	Repton	7	1100
Tg192	2	502	3	0.006	S.Britain	4	10000
Tg196	2	502	18	0.0359	Carribean	3	600
Tg232	1	147	13	0.0884	N.Britain	2	1800
Tg233	2	147	22	0.1497	N.Britain	2	1800
Total	61	12259	409				

[1]*Sample* Author's DNA extraction identification. [2]*PCRs* Independent PCRs amplified per sample. [3]*Ampd* Total bases amplified per sample. [4]*Damage* Total observed damage events per sample. [5]*d* Standardised damage calculation. For details see text. [6]*Site* Archaeological site of sample origin. [7]*Code* Categorical geographical site coding for statistical analysis. [8]*Age* Approximate age of sample (years).

Tindall and Kunkel 1988; Eckert and Kunkel 1990; Sanson *et al*. 2002). Other enzymes used in the dataset are indicated in Table 2.2.

To test whether enzyme misincorporation rates are modified by the local environment of the ancient extracts, modern contaminants in two sets of ancient DNA extracts were examined 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 bacterial RNA polymerase β-subunit-encoding gene (Drancourt *et al*. 1998). PCR products were cloned and the spectrum of damage examined. Similarly, clones of obvious modern human mitochondrial DNA 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 Barnes et al (2002).

2.2.3 Analyses of damage rates

The mitochondrial region sequenced for each sample varies within and between the datasets, and therefore the term 'cloned region' is used to define each independently amplified area bound by a primer pair. For the largest dataset (chapter 3), a measure of DNA damage (*d*) was calculated as *d*=D/Lt, where D=total number of base changes observed per cloned region, L= the base length of an amplified sequence, and t=number of independent PCRs amplified. The null hypothesis, Ho, that no significant correlation exists between *d* and sample age or archaeological site of origin was tested using the General Linear Model (GLM) function of statistical program, Minitab (Minitab Inc).

Table 2.2: Sample details, base change measurements and bias calculations.

				Ratio[2]	Individual changes[3]												Complementary groups[5]						Calculations[6]				
Sample	Species	Region[1]	Clones	AT/GC	A>C	T>G	A>G	T>C	A>T	T>A	C>A	G>T	C>G	G>C	C>T	G>A	Total[4]	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)	Type2-1	E[7]	Source[8]
Tg44	H.sapiens	16209-16356	2	1.26	0	0	1	0	0	0	1.26	0	1.26	0	6.31	0	9.83	1.0	0.0	0.0	6.3	1.3	1.3	1.00	6.31	5.31	1	This study
Tg54	H.sapiens	16209-16356	5	1.26	0	0	1	3	1	0	0	0	0	0	3.78	0	8.78	4.0	1.0	0.0	3.8	0.0	0.0	-2.00	3.78	-0.22	1	This study
Tg76	H.sapiens	16055-16410	5	1.16	0	0	1	0	0	0	0	0	0	0	1.16	0	2.16	1.0	0.0	0.0	1.2	0.0	0.0	1.00	1.16	0.16	1	This study
Tg76	H.sapiens	16209-16356	12	1.26	0	0	1	0	0	0	0	0	0	0	1.26	0	2.26	1.0	0.0	0.0	1.3	0.0	0.0	1.00	1.26	0.26	1	This study
Tg77	H.sapiens	16055-16410	9	1.16	1	1	2	0	0	0	0	0	0	0	0	1.16	5.16	2.0	0.0	2.0	1.2	0.0	0.0	2.00	-1.16	-0.84	1	This study
Tg80	H.sapiens	16055-16410	7	1.16	0	0	2	0	0	0	0	0	0	0	0	5.8	7.80	2.0	0.0	0.0	5.8	0.0	0.0	2.00	-5.80	3.80	1	This study
Tg80	H.sapiens	16209-16356	10	1.26	0	0	0	0	0	0	0	0	0	0	0	6.31	6.31	0.0	0.0	0.0	6.3	0.0	0.0	0.00	-6.31	6.31	1	This study
Tg85	H.sapiens	16209-16356	5	1.26	0	0	1	2	0	0	0	0	0	0	6.96	1.16	11.12	3.0	0.0	0.0	8.1	0.0	0.0	-1.00	5.80	5.12	1	This study
Tg85	H.sapiens	16055-16410	10	1.16	0	0	0	0	0	0	0	0	0	0	7.57	1.26	8.83	0.0	0.0	0.0	8.8	0.0	0.0	0.00	6.31	8.83	1	This study
Tg103	H.sapiens	16055-16410	4	1.16	0	0	2	3	0	0	0	0	0	0	5.8	0	10.80	5.0	0.0	0.0	5.8	0.0	0.0	-1.00	5.80	0.80	1	This study
Tg103	H.sapiens	16209-16356	4	1.26	0	0	0	3	0	0	0	0	0	0	0	6.31	9.31	3.0	0.0	0.0	6.3	0.0	0.0	-3.00	-6.31	3.31	1	This study
Tg104	H.sapiens	16209-16356	15	1.26	0	0	0	0	0	0	0	0	0	0	1.26	2.52	3.78	0.0	0.0	0.0	3.8	0.0	0.0	0.00	-1.26	3.78	1	This study
Tg112	H.sapiens	16209-16356	6	1.26	0	0	2	2	0	0	0	0	1.26	1.26	6.31	2.52	15.35	4.0	0.0	0.0	8.8	2.5	0.0	0.00	3.78	4.83	1	This study
Tg114	H.sapiens	16055-16410	5	1.16	0	0	3	2	0	0	0	0	0	0	2.32	1.16	8.48	5.0	0.0	0.0	3.5	0.0	0.0	1.00	1.16	-1.52	1	This study
Tg114	H.sapiens	16209-16356	11	1.26	0	0	3	0	0	0	0	0	0	0	2.52	1.26	6.78	3.0	0.0	0.0	3.8	0.0	0.0	3.00	1.26	0.78	1	This study
Tg116	H.sapiens	16209-16356	4	1.26	0	0	2	1	0	0	0	0	0	0	5.05	5.05	13.09	3.0	0.0	0.0	10.1	0.0	0.0	1.00	0.00	7.09	1	This study
Tg116	H.sapiens	16055-16410	6	1.16	0	0	3	2	0	0	0	0	0	0	3.48	0	8.48	5.0	0.0	0.0	3.5	0.0	0.0	1.00	3.48	-1.52	1	This study
Tg120	H.sapiens	16055-16410	2	1.16	0	0	0	1	0	0	0	0	0	0	1.16	0	2.16	1.0	0.0	0.0	1.2	0.0	0.0	-1.00	1.16	0.16	1	This study
Tg120	H.sapiens	16209-16356	1	1.26	0	0	0	1	0	0	0	0	0	0	1.26	0	2.26	1.0	0.0	0.0	1.3	0.0	0.0	-1.00	1.26	0.26	1	This study
Tg123	H.sapiens	16209-16356	6	1.26	1	0	0	0	0	0	0	0	0	0	2.52	5.05	8.57	0.0	0.0	1.0	7.6	0.0	0.0	0.00	-2.52	7.57	1	This study
Tg127	H.sapiens	16209-16356	8	1.26	1	0	1	0	0	0	2.52	0	0	0	3.78	0	8.31	1.0	0.0	1.0	3.8	0.0	2.5	1.00	3.78	2.78	1	This study
Tg128	H.sapiens	16055-16410	2	1.16	0	0	1	0	0	0	0	0	0	0	0	1.16	2.16	1.0	0.0	0.0	1.2	0.0	0.0	1.00	-1.16	0.16	1	This study
Tg128	H.sapiens	16209-16356	12	1.26	0	0	0	0	0	0	0	0	0	0	0	1.26	1.26	0.0	0.0	0.0	1.3	0.0	0.0	0.00	-1.26	1.26	1	This study
Tg129	H.sapiens	16209-16356	2	1.26	0	0	0	0	0	0	0	0	0	0	1.26	0	1.26	0.0	0.0	0.0	1.3	0.0	0.0	0.00	1.26	1.26	1	This study
Tg129	H.sapiens	16055-16410	9	1.16	1	0	3	2	0	0	0	0	0	0	0	1.16	7.16	5.0	0.0	1.0	1.2	0.0	0.0	1.00	-1.16	-3.84	1	This study
Tg131	H.sapiens	16209-16356	7	1.26	0	0	0	0	0	0	0	0	0	0	3.78	0	3.78	0.0	0.0	0.0	3.8	0.0	0.0	0.00	3.78	3.78	1	This study
Tg133	H.sapiens	16209-16356	2	1.26	1	0	2	0	0	0	0	0	0	0	0	0	3.00	2.0	0.0	1.0	0.0	0.0	0.0	2.00	0.00	-2.00	1	This study
Tg138	H.sapiens	16209-16356	8	1.26	1	2	1	1	0	0	0	0	0	0	0	0	5.00	2.0	0.0	3.0	0.0	0.0	0.0	0.00	0.00	-2.00	1	This study
Tg141	H.sapiens	16209-16356	6	1.26	0	0	0	3	0	0	0	0	0	0	1.26	0	4.26	3.0	0.0	0.0	1.3	0.0	0.0	-3.00	1.26	-1.74	1	This study

			Ratio[2]		Individual changes[3]												Complementary groups[5]						Calculations[6]				
Sample	Species	Region[1]	Clones	AT/GC	A>C	T>G	A>G	T>C	A>T	T>A	C>A	G>T	C>G	G>C	C>T	G>A	Total[4]	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)	Type2-1	E[7]	Source[8]
Tg142	H.sapiens	16209-16356	7	1.26	0	0	0	2	0	0	0	0	0	0	0	0	2.00	2.0	0.0	0.0	0.0	0.0	0.0	-2.00	0.00	-2.00	1	This study
Tg142	H.sapiens	16055-16410	2	1.16	0	0	0	2	0	0	0	0	0	0	0	0	2.00	2.0	0.0	0.0	0.0	0.0	0.0	-2.00	0.00	-2.00	1	This study
Tg143	H.sapiens	16209-16356	6	1.26	0	0	2	1	0	0	0	0	0	0	0	1.26	4.26	3.0	0.0	0.0	1.3	0.0	0.0	1.00	-1.26	-1.74	1	This study
Tg145	H.sapiens	16055-16410	10	1.16	0	0	2	2	0	0	0	0	0	0	6.96	0	10.96	4.0	0.0	0.0	7.0	0.0	0.0	0.00	6.96	2.96	1	This study
Tg145	H.sapiens	16209-16356	5	1.26	0	0	2	2	0	0	0	0	0	0	7.57	0	11.57	4.0	0.0	0.0	7.6	0.0	0.0	0.00	7.57	3.57	1	This study
Tg146	H.sapiens	16209-16356	6	1.26	0	0	0	1	0	0	0	0	0	0	5.05	0	6.05	1.0	0.0	0.0	5.0	0.0	0.0	-1.00	5.05	4.05	1	This study
Tg148	H.sapiens	16209-16356	80	1.26	0	0	6	4	0	0	0	0	0	0	5.05	3.78	18.83	10.0	0.0	0.0	8.8	0.0	0.0	2.00	1.26	-1.17	1	This study
Tg149	H.sapiens	16209-16356	57	1.26	1	0	11	13	0	1	0	0	0	0	138	3.78	168.55	24.0	1.0	1.0	142.6	0.0	0.0	-2.00	134.98	118.55	1	This study
Tg192	H.sapiens	16209-16356	12	1.26	0	0	1	0	0	0	0	0	0	0	1.26	0	2.26	1.0	0.0	0.0	1.3	0.0	0.0	1.00	1.26	0.26	1	This study
Tg192	H.sapiens	16055-16410	10	1.16	0	0	1	0	0	0	0	0	0	0	0	0	1.00	1.0	0.0	0.0	0.0	0.0	0.0	1.00	0.00	-1.00	1	This study
Tg196	H.sapiens	16055-16410	12	1.16	1	0	4	2	0	0	1.26	0	0	0	0	1.26	9.52	6.0	0.0	1.0	1.3	0.0	1.3	2.00	-1.26	-4.74	1	This study
Tg232	H.sapiens	16209-16356	8	1.26	0	0	3	0	1	0	0	0	1.26	0	10.1	1.26	16.62	3.0	1.0	0.0	11.4	1.3	0.0	3.00	8.83	8.35	1	This study
Tg233	H.sapiens	16209-16356	14	1.26	0	0	2	1	0	0	0	0	0	0	15.1	8.83	26.97	3.0	0.0	0.0	24.0	0.0	0.0	1.00	6.31	20.97	1	This study
Tg233	H.sapiens	16055-16410	4	1.16	0	0	1	0	0	0	0	0	0	0	1.16	4.64	6.80	1.0	0.0	0.0	5.8	0.0	0.0	1.00	-3.48	4.80	1	This study
Tg129	H.sapiens	16055-16410	9	1.16	1	0	3	2	0	0	0	0	0	0	0	1.16	7.16	5.0	0.0	1.0	1.2	0.0	0.0	1.00	-1.16	-3.84	1	This study
Tg136a	H.sapiens	16209-16356	17	1.26	0	1	1	4	0	0	0	0	0	0	2.52	1.26	9.78	5.0	0.0	1.0	3.8	0.0	0.0	-3.00	1.26	-1.22	1	This study
Tg136b	H.sapiens	16209-16356	26	1.26	0	0	1	2	0	0	0	0	0	0	0	1.26	4.26	3.0	0.0	0.0	1.3	0.0	0.0	-1.00	-1.26	-1.74	1	This study
Tg136c	H.sapiens	16055-16410	27	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	0.0	1.2	0.00	0.00	-2.00	1	This study
Tg136d	H.sapiens	16055-16410	38	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	0.0	0.00	1.26	-0.74	1	This study
Tg137a	H.sapiens	16209-16356	4	1.26	0	1	5	0	0	0	0	0	0	0	2.52	0	8.52	5.0	0.0	1.0	2.5	0.0	0.0	5.00	2.52	-2.48	1	This study
Tg137b	H.sapiens	16209-16356	6	1.26	0	0	4	1	0	0	0	0	0	0	5.05	5.05	15.09	5.0	0.0	0.0	10.1	0.0	0.0	3.00	0.00	5.09	1	This study
Tg191a	H.sapiens	16209-16356	5	1.26	0	0	1	5	0	1	0	0	0	0	1.26	0	8.26	6.0	1.0	0.0	1.3	0.0	0.0	-4.00	1.26	-4.74	1	This study
Tg196	H.sapiens	16209-16356	16	1.26	0	0	0	0	1	0	0	0	1.26	0	3.78	0	6.05	0.0	1.0	0.0	3.8	1.3	0.0	0.00	3.78	3.78	1	This study
Tg77d	H.sapiens	16055-16410	16	1.16	1	1	2	0	0	0	0	0	0	0	0	1.16	5.16	2.0	0.0	2.0	1.2	0.0	0.0	2.00	-1.16	-0.84	1	This study
Tg99i	H.sapiens	16209-16356	10	1.26	0	0	7	1	0	0	0	0	0	0	11.35	0	19.35	8.0	0.0	0.0	11.4	0.0	0.0	6.00	11.35	3.35	1	This study
Tg99j	H.sapiens	16209-16356	9	1.26	0	0	1	4	0	0	0	0	0	0	0	0	5.00	5.0	0.0	0.0	0.0	0.0	0.0	-3.00	0.00	-5.00	1	This study
Dinornis	Dinornis giganticus	11120-11958	9	1.34	0	0	2	1	1	1	2.36	0	0	0	2.36	0	9.72	3.0	2.0	0.0	2.4	0.0	2.4	1.00	2.36	-0.64	1	Cooper *et al.* 2001
Dinornis12s	Dinornis giganticus	1753-2148	10	1.1	0	0	3	1	0	0	0	0	0	0	1.1	0	5.10	4.0	0.0	0.0	1.1	0.0	0.0	2.00	1.10	-2.90	3	Cooper *et al.* 2001
DinornisCOI	Dinornis giganticus	7807-8325	17	1.3	0	1	6	0	0	0	0	0	0	0	1.3	0	8.30	6.0	0.0	1.0	1.3	0.0	0.0	6.00	1.30	-4.70	3	Cooper *et al.* 2001
DinornisCOII	Dinornis giganticus	8861-9349	7	1.14	0	0	0	1	0	0	1.14	0	0	0	4.56	0	6.70	1.0	0.0	0.0	4.6	0.0	1.1	-1.00	4.56	3.56	3	Cooper *et al.* 2001
DinornisCR	Dinornis giganticus	16733-00441	2	1.34	0	0	2	7	0	0	1.34	0	0	0	5.34	2.67	18.35	9.0	0.0	0.0	8.0	0.0	1.3	-5.00	2.67	-0.99	3	Cooper *et al.* 2001

			Ratio[2]		Individual changes[3]												Complementary groups[5]						Calculations[6]				
Sample	Species	Region[1]	Clones	AT/GC	A>C	T>G	A>G	T>C	A>T	T>A	C>A	G>T	C>G	G>C	C>T	G>A	Total[4]	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)	Type2-1	E[7]	Source[8]
DinornisCytb	Dinornis giganticus	15303-15783	5	1.4	0	0	1	0	0	0	0	0	0	0	2.8	0	3.80	1.0	0.0	0.0	2.8	0.0	0.0	1.00	2.80	1.80	3	Cooper *et al.* 2001
DiornisnND1	Dinornis giganticus	4747-5201	10	1.35	0	0	2	0	0	0	0	0	0	1.35	5.4	0	8.75	2.0	0.0	0.0	5.4	1.4	0.0	2.00	5.40	3.40	3	Cooper *et al.* 2001
Emeus16s	Emeus crassus	3787-4311	10	1.33	0	0	1	1	0	0	0	0	0	0	13.3	0	15.30	2.0	0.0	0.0	13.3	0.0	0.0	0.00	13.30	11.30	3	Cooper *et al.* 2001
EmeusCOI	Emeus crassus	7807-8328	7	1.33	0	0	0	1	0	0	0	0	0	0	3.99	0	4.99	1.0	0.0	0.0	4.0	0.0	0.0	-1.00	3.99	2.99	3	Cooper *et al.* 2001
EmeusCOII	Emeus crassus	8320-8807	11	1.3	0	0	3	3	0	0	1.3	1.3	1.3	0	6.5	0	16.40	6.0	0.0	0.0	6.5	1.3	2.6	0.00	6.50	0.50	3	Cooper *et al.* 2001
EmeusCOIII	Emeus crassus	10161-10743	19	1.16	0	0	2	3	0	0	1.16	0	1.16	1.16	6.96	6.96	22.40	5.0	0.0	0.0	13.9	2.3	1.2	-1.00	0.00	8.92	3	Cooper *et al.* 2001
EmeusND4/5	Emeus crassus	12788-13200	10	1.4	0	0	3	2	0	0	0	0	0	0	7	0	12.00	5.0	0.0	0.0	7.0	0.0	0.0	1.00	7.00	2.00	3	Cooper *et al.* 2001
Mulleronis12s	Mulleronis agilis	1856-2020	8	1.15	0	0	0	0	0	0	1.15	0	0	0	5.75	0	6.90	0.0	0.0	0.0	5.8	0.0	1.2	0.00	5.75	5.75	3	Cooper *et al.* 2001
Borgo Nuovo	H.sapiens	055-410	129	1.16	1	0	6	8	1	0	0	0	0	0	24.3	2.32	42.65	14.0	1.0	1.0	26.6	0.0	0.0	-2.00	22.01	12.65	3	Di Benedetto *et al.* 2000
Mezzocorona	H.sapiens	055-410	111	1.16	0	0	3	5	2	0	0	0	0	0	45.2	4.63	59.82	8.0	2.0	0.0	49.8	0.0	0.0	-2.00	40.55	41.82	3	Di Benedetto *et al.* 2000
Villabruna	H.sapiens	055-410	161	1.16	0	0	4	5	6	0	2.32	1.16	0	1.16	30.1	4.63	54.39	9.0	6.0	0.0	34.8	1.2	3.5	-1.00	25.49	25.76	3	Di Benedetto *et al.* 2000
Handt	H.sapiens	055-410	34	1.16	0	0	1	1	2	2	0	1.16	0	0	2.32	1.16	10.64	2.0	4.0	0.0	3.5	0.0	1.2	0.00	1.16	1.48	3	Handt *et al.* 1996
Ursus147	Ursus arctos	Control region	5	1.07	0	0	2.28	0	0	0	0	0	0	0	2	1	5.28	2.3	0.0	0.0	3.0	0.0	0.0	2.28	1.00	0.72	1	Barnes *et al.* 2002
Ursus221a	Ursus arctos	Control region	6	1.07	0	0	1.14	1.14	0	0	0	0	0	0	2	2	6.28	2.3	0.0	0.0	4.0	0.0	0.0	0.00	0.00	1.72	1	Barnes *et al.* 2002
Ursus222	Ursus arctos	Control region	15	1.07	0	0	0	1.14	0	1.14	0	0	0	1	3	9	15.28	1.1	1.1	0.0	12.0	1.0	0.0	-1.14	-6.00	10.86	1	Barnes *et al.* 2002
Ursus223a	Ursus arctos	Control region	6	1.07	1.14	0	1.14	2.28	0	0	0	0	0	0	1	1	6.56	3.4	0.0	1.1	2.0	0.0	0.0	-1.14	0.00	-1.42	1	Barnes *et al.* 2002
Neanderthal1	H.s. neanderthalis	16055-16400	123	1.18	1	1	6	10	5	0	0	2.35	0	0	17.64	5.88	48.87	16.0	5.0	2.0	23.5	0.0	2.4	-4.00	11.76	7.52	2	Krings *et al.* 1997
Lalueza	H.sapiens	16209-16410	11	1.08	0	1	0	0	0	0	0	0	1.08	0	3.23	0	5.30	0.0	0.0	1.0	3.2	1.1	0.0	0.00	3.23	3.23	1	Lalueza-Fox *et al.* 2001
Neanderthal2	H.sapiens	16055-16400	6	1.18	0	0	0	1	0	0	0	0	0	0	3.53	1.18	5.70	1.0	0.0	0.0	4.7	0.0	0.0	-1.00	2.35	3.70	2	Ovchinnikov etal.2000
Ursus47910	Ursus spelaeus	Control region	2	1.84	0	0	0	0	0	0	0	0	0	0	1	0	1.00	0.0	0.0	0.0	1.0	0.0	0.0	0.00	1.00	1.00	2	Loreille *et al.* 2000
UrsusCLA 1	Ursus spelaeus	Control region	16	1.84	0	1	1	2	1	2	0	1.84	5.52	1.84	0	0	16.20	3.0	3.0	1.0	0.0	7.4	1.8	-1.00	0.00	-3.00	2	Loreille *et al.* 2000
UrsusSC11700 1	Ursus spelaeus	Control region	9	1.84	0	0	1	1	2	1	0	0	0	0	0	3.68	8.68	2.0	3.0	0.0	3.7	0.0	0.0	0.00	-3.68	1.68	2	Loreille *et al.* 2000
UrsusSC157001	Ursus spelaeus	Control region	4	1.84	0	0	0	0	0	0	0	0	0	0	3.68	0	3.68	0.0	0.0	0.0	3.7	0.0	0.0	0.00	3.68	3.68	2	Loreille *et al.* 2000
UrsusSC5300	Ursus spelaeus	Control region	6	1.84	0	1	0	1	0	0	0	0	0	0	3.68	0	5.68	1.0	0.0	1.0	3.7	0.0	0.0	-1.00	3.68	2.68	2	Loreille *et al.* 2000
UrsusSCL3800	Ursus spelaeus	Control region	3	1.84	0	0	0	0	1	0	0	0	0	1.84	0	0	2.84	0.0	1.0	0.0	0.0	1.8	0.0	0.00	0.00	0.00	2	Loreille *et al.* 2000
UrsusSCL3500	Ursus spelaeus	Control region	7	1.84	0	1	1	0	1	0	0	0	0	1.84	1.84	0	6.68	1.0	1.0	1.0	1.8	1.8	0.0	1.00	1.84	0.84	2	Loreille *et al.* 2000
UrsusTAB151a	Ursus spelaeus	Control region	10	1.84	0	1	0	2	0	0	0	0	1.84	3.68	1.84	0	10.36	2.0	0.0	1.0	1.8	5.5	0.0	-2.00	1.84	-0.16	2	Loreille *et al.* 2000
UrsusTAB21a	Ursus spelaeus	Control region	16	1.84	0	1	0	1	0	1	0	0	5.52	0	5.52	1.84	15.88	1.0	1.0	1.0	7.4	5.5	0.0	-1.00	3.68	6.36	2	Loreille *et al.* 2000
Poinar	H.sapiens	16131-16218	13	1.38	1	0	2	2	0	0	0	0	0	0	5.51	0	10.51	4.0	0.0	1.0	5.5	0.0	0.0	0.00	5.51	1.51	2	Poinar *et al.* 2001
Ursus221b	Ursus arctos	12s	8	0.9	0	0	0	1.07	0	0	0	0	0	0	8	2	11.07	1.1	0.0	0.0	10.0	0.0	0.0	-1.07	6.00	8.93	1	This study
Ursus223b	Ursus arctos	12s	7	0.88	0	0	3.2	0	0	1.07	0	0	0	0	0	0	4.27	3.2	1.1	0.0	0.0	0.0	0.0	3.20	0.00	-3.20	1	This study

All cloned regions are mitochondrial. [1]*Region* Human regions are numbered with reference to the CRS (Anderson *et al*. 1981). Ratite regions are numbered with reference to Cooper et al (2001). Bear regions are located by gene. [2]*Ratio* 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. [3]*Individual changes* Scaled counts of base modifications per sample subdivided into the 12 hypothetical *post mortem* base modification events. These are represented as i>j where i=original nucleotide, j=modified nucleotide. [4]*Total* Total sum of modified damage per cloned region. [5]*Complementary groups* As *Individual changes*, but with subdivision into the six complementary pairings. [6]*Calculations* Bias within type 1 transitions, within type 2 transitions, and between types 2 and 1 transitions respectively. [7]*Enzyme* Categorical coding of polymerase enzymes used in studies, where 1=Platinum Taq Hifidelity, 2=Amplitaq Gold, 3= Taq DNA Polymerase. [8]*Source* Original source of data.

2.2.4 DNA base complementarity and damage

The most common damage-driven base changes observed in ancient DNA sequences are the four transitions: cytosine to thymine (C→T); guanine to adenine (G→A); thymine to cytosine (T→C); and adenine to guanine (A→G) (Hansen *et al*. 2001). However, due to the complementary nature of DNA, each of these observations can be explained by two possible causative events (Hofreiter *et al*. 2001a). Figures 2.1a and 2.1b demonstrate this for an observed C→T transition on the L strand. Due to an original damage event on the L strand causing C→U (uracil, a thymine analogue), after two stages of replication amplified sequences present a C→T transition on the L strand (Fig. 2.1a). However, theoretically this phenotype can also occur via an H strand G→A transition, which after one PCR cycle would generate the same C→T transition on the L strand (Fig. 2.1b). A similar problem applies to each of the transitions, so that each *post mortem* biochemical change can result in two observed outcomes, depending on which strand is sequenced. Due to this complementarity, Hansen *et al*. (2001) have termed A→G and T→C 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.

While this situation appears intractable, a solution is offered by the limited number of possible biochemical pathways by which nucleotide damage can occur. Hofreiter *et al*. (2001a) demonstrated that the damage-driven modification of guanine to an adenine analogue is highly unlikely, if not impossible. This excludes the possibility of the events shown in Fig 2.1b. It can therefore be argued that any G→A transition observed on the L strand due to damage must have originated as a H strand C→T modification event, because G→A modification on the L strand is impossible.

Figure 2.1: Determination of a strand of origin for post mortem DNA damage events using Type 2 (C>T/G>A) transitions as an example.

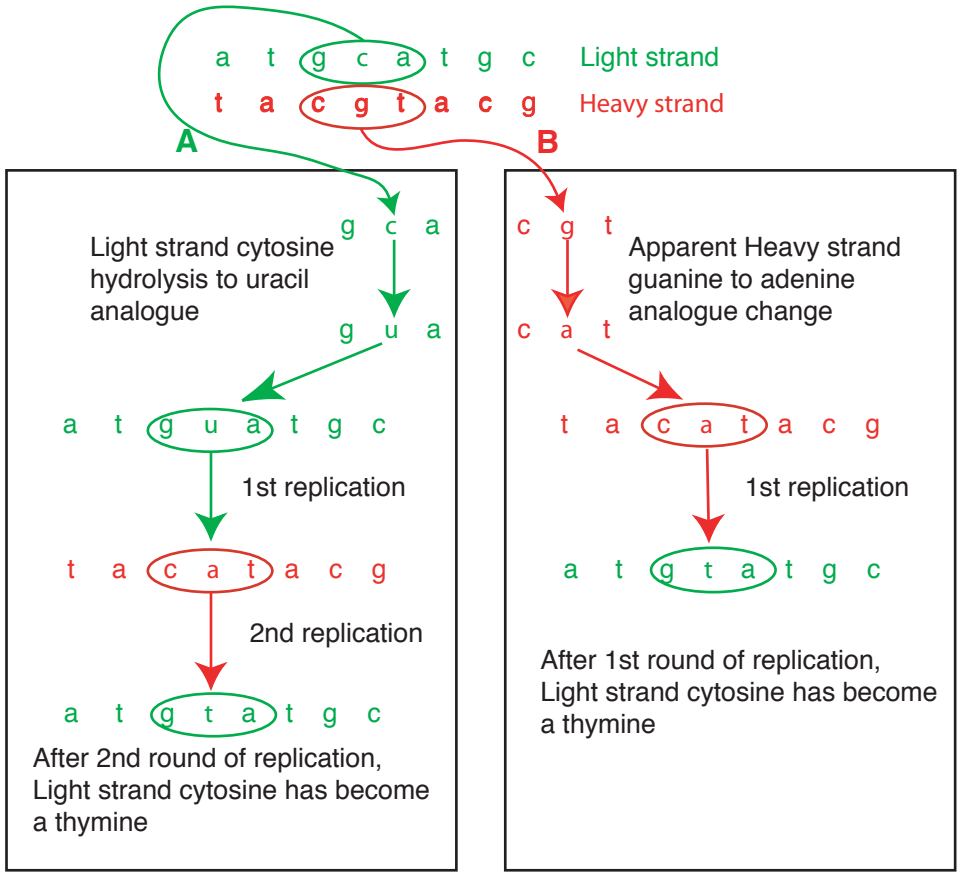


Fig.1a After 2 cycles of amplifications, Light strand C>T transitions will result in a permanent L strand change.
Fig.1b A theoretical Heavy strand G>A change will produce the Light strand phenotype of C>T change following one cycle of amplification. However, as a direct G>A post mortem modification is chemically impossible, Fig.1b is not possible. Thus all C>T changes observed on the L strand must have occurred as L strand C>T post mortem damage, and all G>A changes on the L strand must have occurred as H strand C>T post mortem damage.

A similar argument can be applied to Type 1 damage using the common assumption that the modification of T→C analogues is biochemically unlikely (Hofreiter *et al.* 2001a). In this situation, any L strand T→C modification will actually be due to an H strand A→G event (in fact, adenine deamination to hypoxanthine (HX) a guanine analogue), while 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 which have shown that a major oxidative derivative of thymine, 5-formyluracil (fU), has the capacity to pair with A, T, G or C (Zhang *et al.* 1997; Yoshida *et al.* 1997; Fujikawa *et al.* 1998; Zhang *et al.* 1999). 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). In order to test whether this process is observed in the ancient sequences, a single-primer extension PCR (SP-PCR) was performed to observe the spectrum of *post mortem* damage without the effects of jumping PCR recombining substitutions between H and L strands (Hofreiter *et al.* 2001a). Four samples (Tg129, Tg149, Tg232, Tg233) were each amplified for 25 cycles using either the mitochondrial L strand primer L16209 or the H strand 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 towards the strand initially amplified. Clones were examined for sequences containing both A→G and T→C events which would normally imply a jumping PCR event, but in SP-PCR data would confirm the existence of fU:G or HX:C pairings.

This hypothesised 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 either 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. Lastly, if *in vivo* base modification events are similar to those *post mortem* events, it should be possible to examine modern sequence datasets and determine the strand on which observed transition events originally took place.

2.2.5 Analysis for strand-specific distribution of DNA damage

The distribution of L and H strand damage events in ancient human DNA sequences was examined using the chapter 3 dataset. The total number and location of damaged positions on the L and H strands was determined and compared to the number and distribution expected if there was no strand bias, using a Chi-squared goodness of fit test. The data were scaled to take local base composition into account (generally the under-representation 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.

2.2.6 Analysis of the ratio and distribution of different transition events

To investigate the ratio and distribution of Type 1 (A→G/T→C) and Type 2 (C→T/G→A) damage events in the *post mortem* dataset, the absolute number of each of the twelve 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, T→G) was measured for each cloned region, and scaled for composition bias as above. A second dataset was also created to include the six complementary changes (*e.g.* T→C/A→G etc, Hansen *et al.* 2001; Hofreiter *et al.* 2001a). To examine the variation in the ratio of the Type 2:Type 1 transitions, a value, β, for each region was calculated which was equal to the number of Type 2-

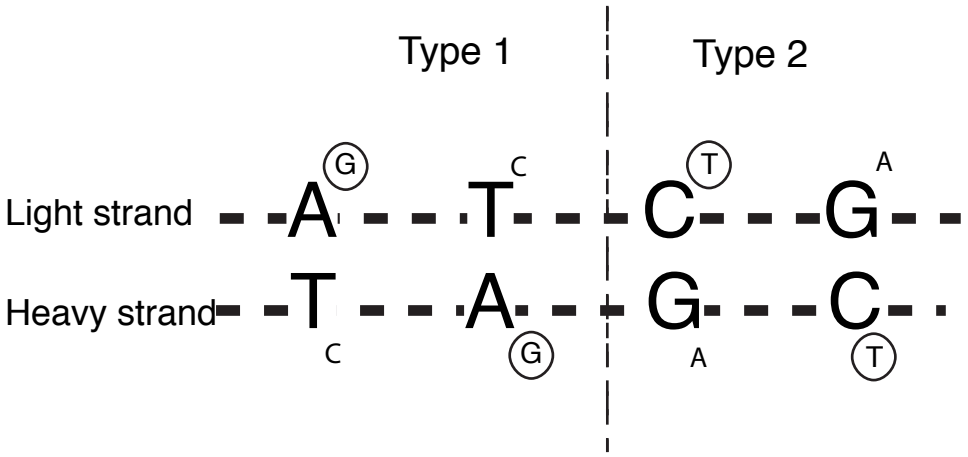
Type1 events. Bias towards either damage event was correlated with the overall extent of template damage, or with different polymerase enzymes, using a GLM.

Within Type 2 and Type 1 transitions a similar test was performed to determine which original template strand was damaged, by calculating the product (C→T)-(G→A) for Type 2 and (A→G)-(T→C) for Type 1. For example, SP-PCR amplifications that are initiated from an H strand template will potentially show H-specific components of both Type 2 (G→A) and Type 1 (T→C) damage (Fig 2.2). Conversely, amplifications that derive from L strand templates will contain L-specific components (*i.e.* C→T and A→G). Therefore, if H and L strands are equally represented in a DNA extract, cloned sequences from a single PCR should show approximately equal numbers of sequences containing H- and L-specific components of damage. A significant deviation from this ratio would suggest the extract contains a bias of either H or L strands, or that amplification is initiated preferentially on one strand. Furthermore, mis-coding 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 towards 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 non-zero rates of T→C, which while biochemically unlikely (Lindahl 1993a), have not been experimentally investigated in aDNA.

2.2.7 Analysis for evidence of 'jumping PCR'

The tight correlation between strand-specific Type 1 and Type 2 transitions modifications also provides a means to examine the nature and frequency of PCR jumping artefacts (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

Figure 2.2:Type 1 and type 2 damage induced transitions.

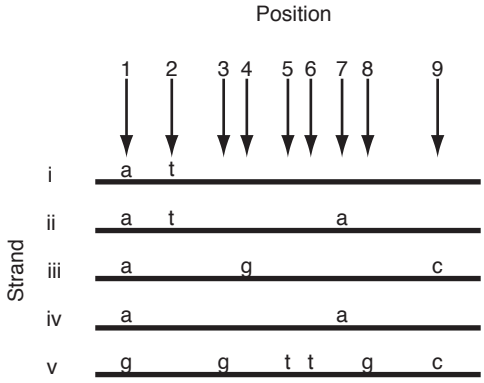


Circled letters represent the principle modifications observed in cloned sequences (e.g. deamination of cytosine to uracil which is read as thymine, or adenine to hypoxanthine, read as guanine). Changes introduced on the complementary strand when the damaged bases are subsequently copied are shown in 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, the type 1 and 2 errors observed are expected to be T>C, and G>A, respectively.

(Fig 2.3). However, when only one genotype is present in a PCR reaction, a jumping event may be 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-specific or L-specific components, but not both. In Table 2.3 the number of clone sequences with such associations, α, is divided into those which can and cannot be explained through jumping PCR (due to the presence of substitutions shared by multiple non-homologous cloned sequences). Sequences that contain both H-specific and L-specific components of damage can only have arisen following a jumping event. These are sub-divided into β, those that contain exclusively Type 1 or 2 transitions, and γ, those that contain Type 1 as well as Type 2 transitions. These are again divided as in α.

A Chi-squared goodness of fit test was performed on the observed distribution of transition modifications among associations α, β and γ, and that expected under randomly paired base changes. Estimates of the number of jumping PCR events within the cloned sequence dataset before and after the identification of associations β and γ are compared to assess jump detection efficiency.

Figure 2.3: Jumping PCR (Pääbo et al. 1990) and the generation of ghost damage events.



Strands i-v represent 5 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, UK). 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 help 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. Assuming 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, but differences at position 2, indicates jumping PCR between the 2 strands. Finally position 9 on strands iii and v represent apparent damage to strand C arising from jumping with the contaminant strand v.

Table 2.3: Samples containing association groups α,β and γ.

Association Group:		α		β		γ	
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
Dinornis	16733-00441	1		1	1j	2	2j
Dinornis	COII					1	
Dinornis	11120-11958	1				1	
Emeus	16s	1					
Emeus	COI					1	
Emeus	COIII		1	1			
Emeus	ND4/ND5	1				1,1j	
Emeus	COII	1					
Emeus	cytb	1				2	
Ursus 221a		1					
Ursus 222		1					
Ursus 223a			1	1			
Ursus 147			1		1		
Ursus 221b				1			
Ursus 3500		2					
Ursus 117001			1				
Ursus 151a			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			
Total associations per group		35		21		33	
Total identified jumps		7	0	7	5	9	6
Total non-identified jumps		22	6	7	2	16	2
Group total jumps		7		12		15	
Group total non-jumps		28		9		18	
Group non-jump:jump ratio		4		0.75		1.2	

The 3 association groups are divided into 2 subgroups each headed X&Y, where X and Y are different base modifications observed together on same cloned sequence. j = direct evidence of jumping as identified following the conventional method (*i.e*. Pääbo *et al*. 1989) where different cloned sequences sharing a proportion of damaged sites are deemed to have arisen due to jumping PCR. These are termed here as 'identified jumps'. Non-identified jumps are those that cannot be identified following the conventional methods.

2.3 Results

2.3.1 Fidelity of polymerase enzyme

The cloned sequences derived from modern samples reveal a very high degree of enzyme fidelity, with only two deletions observed in 14,112 bases. In the 6,451 bases of amplified modern bacterial DNA from the ancient extracts, only one C→A base change was observed, while in over 21,442 bases of human contaminant sequences, only five changes (three A→G and two C→T) were observed. Although these figures are somewhat higher than the enzyme manufacturer's published rate of $2x10^{-6}$, they are vastly lower than the rates observed in our data (Table 2.2). It is also possible that multiple contaminants are causing 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.

2.3.2 The biochemical bases of aDNA transitions

Thirty-two clones deriving 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 *post mortem* damage derived miscoding lesions. Further evidence for a lack of role of fU 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* approximately ten times as much as that of G→A/C→T, G→T/C→A or A→C/T→G. However as seen in Table 2.4, the rates of A→G/T→C occurrence in our data set are only half that of G→A/C→T, and by comparison almost thirty times that of the predicted rates of G→T/C→A and A→C/T→G.

2.3.3 Sample DNA damage

The data gives no indication that sample age correlates with damage (as measured by *d*, p=0.85), though there is significant evidence that archaeological site is important. The H_o that there is no correlation between damage and archaeological site may be rejected (ANOVA F-Value=3.523, p<0.01).

Table 2.2 presents the observed and expected measurements of damage on each of the L and H strands for the HVR1 data set analysed in chapter 3. Although the results suggest that the L strand receives more damaged sites, and more overall damage, proportional to the potential number of bases (A and C) that can change, a Chi-squared goodness of fit test provides no statistical backing for either of these observations (Total hits, p=0.66, Sites hit p=0.55).

The number of each of the six complementary change groups for each cloned region is shown in Table 2.2, while Table 2.4 gives the totals for each dataset (of more than one cloned region) and the averages of the control Taq studies. A bias towards Type 2 events is observable in the data (Table 2.2 and 2.4, and Fig 2.4a), particularly in the complete datasets and agrees 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 (Fig 2.4b, 2.4c) with a few samples displaying a high bias towards Type 1 (*e.g.* TG 99j, TG 129, TG196, Moa12s, MoaCOI) which is characteristic of Taq misincorporation (Hansen *et al.* 2001). Interestingly, these biases correlate poorly with enzyme (p=0.574), though there is a strong positive correlation of Type 2:Type 1 bias with extent of overall damage (p<0.01). The number of modified bases and bias measurements for all 12

Table 2.4: Post mortem damage measurements and bias calculations summarised by study.

		A→C, T→G			A→G, T→C			A→T, T→A			C→A, G→T			C→G, G→C			C→T, G→A			Type1	Type2	Type 2-1
Data set[1]	Cloned Regions[2]	AC	TG	Total	AG	TC	Total	AT	TA	Total	CA	GT	Total	CG	GC	Total	CT	GA	Total	(A→G)-(T→C)	(C→G)-(T→A)	
Control	*			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	365.34	17	209.9	194.3
AC	13	0	1	1	25	20	45	1	1	2	8.4	1.3	9.7	2.5	2.5	5	66.4	9.6	76	5	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	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	12.1	13.1
TG2	2	0	0	0	3.2	1.1	4.5	0	1.1	1.1	0	0	0	0	0	0	8	2	10	2.1	6	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	88	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	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	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

[1]Data set Control=Taq misincorporation measurements, TG1=This thesis (humans), AC=Cooper *et al*. (2001), IB=Barnes *et al*. (2002), OL=Loreille *et al*. (2000), TG2=This thesis (bears), GD=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) [2]Cloned Number of different cloned regions per data set. Other column headings as Table 3. *Presented as percentage. Calculated from following data sets –Saiki *et al*. (1988); Dunning *et al*. (1988); Tindall and Kunkal (1988); Eckert and Kunkel (1990); Sanson *et al*. (2002).

Figures 2.4a-c:Type 1 vs 2 damage.

Fig. 2.4a: Damage per study. All studies demonstrate a Type 2 bias.

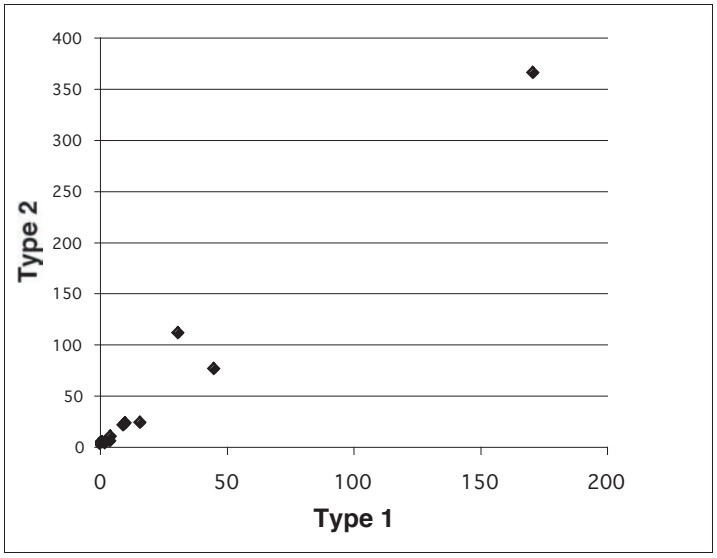
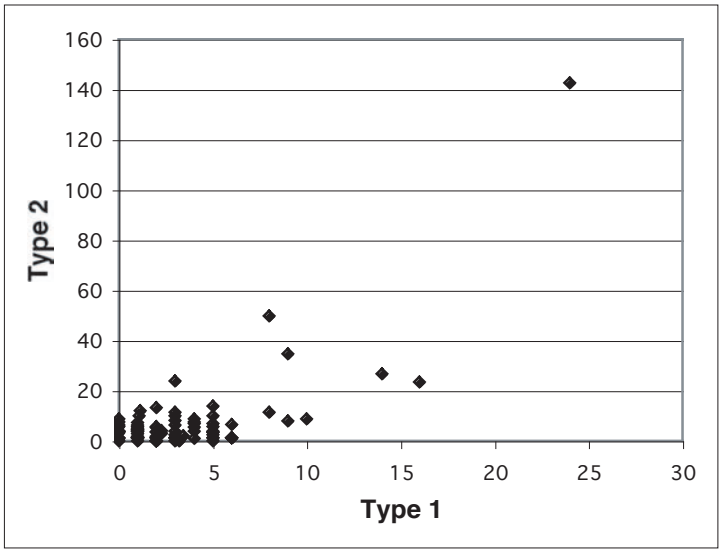
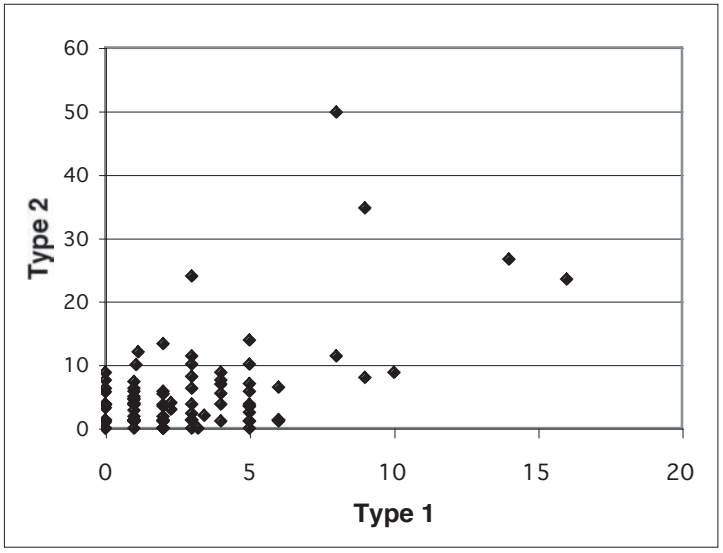


Fig 2.4b: Damage per clone region. Although the Type 2 bias is significant, many samples demonstrate a Type 1 bias.



This is seen clearly in Fig.2.4c in which the outlier is removed to give greater resolution.



possible base changes are also given in Table 2.2 and 2.4. As seen in Fig. 2.5a, data sets as a whole demonstrate a bias towards 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 (Fig. 2.5b, 2.5c). The correlation of bias with enzyme is not significant ($p=0.40$), but bias is highly significant with damage ($p<0.01$). The range of bias within Type 1 damage follows a similar trend, though overall the skew is less pronounced. While there is no significant correlation with enzyme ($p=0.19$), the GLM analysis shows a significant correlation of A→G bias with damage ($p<0.05$). Therefore, for both Type 1 and Type 2 events, there is a trend towards L strand specific damage as the overall extent of damage increases. This finding, combined with the lack of apparent strand bias in damage accumulation, suggests that there is either greater survival of L strands, or at least a prevalence in an amplifiable condition.

2.3.4 'Jumping' PCR

Table 2.3 presents the data on the linkage of damage events in individual clones, and the evidence for jumping PCR artefacts. As expected, pairings are not randomly spread among α, β and γ, and the majority of associated damage events are consistent with a non-jumping origin (Chi-squared goodness of fit test $p=0.14$). Without the data on associated Type 1 and Type 2 changes, 34 obvious jumping artefacts can be seen within the dataset as chimeric sequences. However, the patterns identified in β and γ groups identify another 27 pairings (an increase of 80%) that can be explained through jumping if G→A and T→C transitions are assumed to be chemically impossible.

Figures 2.5a-c: Damage bias within transition types.

Positive Y axis values represent a bias towards Type 1 A>G transitions and Type 2 C>T transitions. Negative values demonstrate a bias towards Type 1 G>A transitions and Type 2 T>C transitions.

Fig.5a Data sets from whole studies. The bias towards Type 2 C>T transitions (indicating damage events on the mitochondrial Light strand) is much stronger than the bias towareds Type 1 A>G transitions. This can be seen clearer in Figs. 5b and 5c.

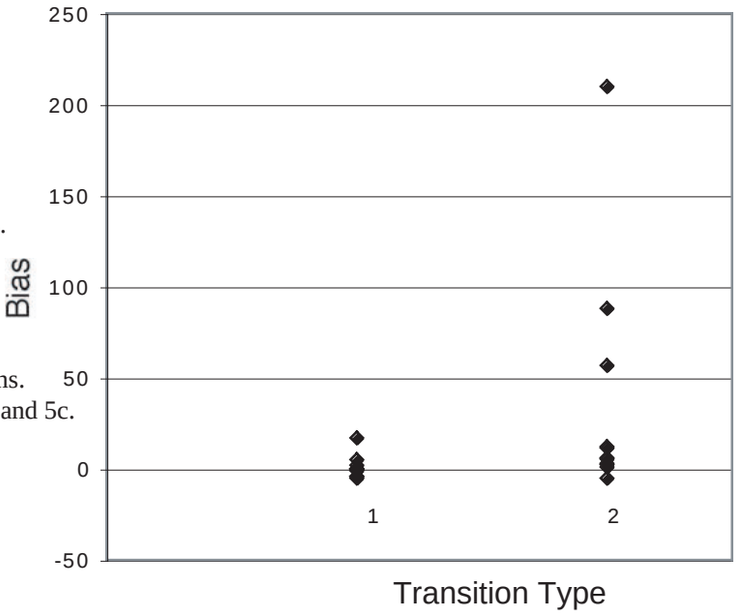


Fig.5b Data from individual cloned regions.

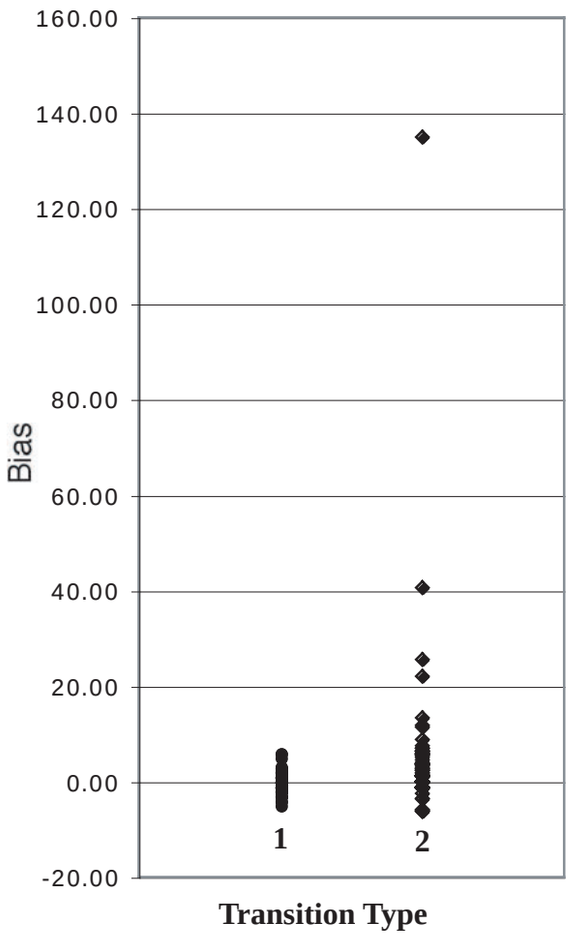
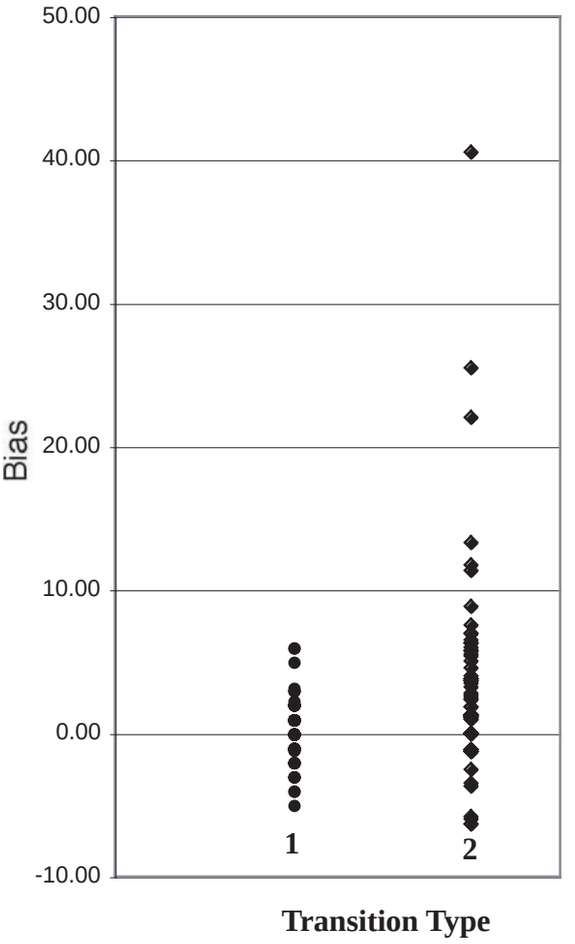


Fig.5c as Fig.5b, but with high values removed to increase overall resolution.



2.4 Discussion

The large dataset 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 mis-incorporation rate observed for the high fidelity enzyme in the control experiments indicates that almost all 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 towards Type 2 over Type 1 transitions increases with the overall extent of damage, in agreement with previous ancient DNA studies (Lindahl 1993a; Willerslev *et al.* 1999; Hofreiter *et al.* 2001a) and the hypothesis of Hansen *et al.* (2001) that Type 1 transitions occur at a slightly slower rate than Type 2. However, the ratio is nowhere near the *in vivo* ratio of 30-50 times (Lindahl 1993a) suggesting that the factors involved in hydrolytic damage vary somewhat between *post mortem* and *in vivo* situations.

The rate of Type 1 damage observed in the dataset is similar to other studies (Hansen *et al.* 2001) and much higher than that reported by Hofreiter *et al.* (2001a), 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 (Amplitaq Gold, ABI, USA) used by Hofreiter *et al.* (2001a). For example, perhaps one of the products in the adenine→HX hydrolytic deamination pathway of Type 1 modifications may hinder the action of Amplitaq Gold relative to Hifi polymerases. However, other ancient DNA studies

that have used Amplitaq Gold (Loreille *et al.* 2001; Krings *et al.* 1997; 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 *post mortem* 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, as 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 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 *post mortem* degradation. Even allowing for this, the apparently reduced rate of overall H strand survival or amplification is unexpected. An increased rate of H strand degradation seems unlikely, as this would also leave the remaining L strands single-stranded, and vulnerable to rapid degradation (Lindahl 1993a). Consequently, there may be some impediment to H strand amplification with increasing damage, perhaps through 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 re-examine mutations in modern human datasets, to determine the strand originally damaged. Such information will facilitate the development of a secondary structure model for the human D-loop, which 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 re-evaluation of the extent of jumping PCR within ancient DNA amplifications. This is an important means to detect haplotypes generated by PCR artefacts (chapter 3), and the high rate detected here is cause for concern about many existing ancient DNA 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 5 is likely to be a considerable under-estimate of the true level, because recombination between homologous template strands (i.e. both H or both L) are not detected. Jumping PCR is believed to be positively correlated with sequence damage (Pääbo *et al.* 1990), and estimates may provide a simple way of comparing the DNA preservation within samples, as well as scrutinising the results for authenticity.

The large dataset of ancient sequences has also permitted the first direct and statistically significant demonstration that the extent of DNA damage within a sample correlates with archaeological site. This has only been indirectly demonstrated previously, using correlates such as the frequency of water change (Nielsen-Marsh and Hedges 2000), temperature (Höss *et al.* 1996; Smith *et al.* 2001), and microbial content (Burger *et al.* 1999), or biochemical data such as amino-acid racemisation (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 use these methods to investigate how environmental factors are effecting 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 1993a). The lack of such a direct correlation in this 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 dessication 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 into the biochemical background, and likelihood of the sequence differences observed between living human groups. Such information is critical for interpreting our recent evolutionary past.

2.5 Acknowledgements

We are indebted to Chris Stringer, Martin Biddle and Birthe Kjølbye-Biddle for samples, and Eddie Holmes, Paul Johnson, Gil McVean, Vincent Macaulay, Svante Pääbo, Mark Stoneking, Ryk Ward and members of the ABC and MPEvA for useful comments. We are also grateful to two anonymous referees for their suggestions and comments. E.W. and A.J.H are grateful to Kim Aaris-Sorensen and Tina B Brandt for help and discussion. M.T.P.G. and A.C. were supported by the Wellcome Trust. E.A. and A.J.H. were financially supported by the VILLUM KANN RASMUSSEN FONDEN, Denmark. L.R. was funded by the Danish Research council for the Humanities. E.W. and A.J.H. contributed equally to the work.

Chapter 3

Distribution patterns of *post mortem* damage in human mitochondrial DNA

Modified from the 'American Journal of Human Genetics' publication

'Distribution patterns of *post mortem* damage in human mitochondrial DNA'

Am. J. Hum. Genet. 72:32-47 (2003)

M Thomas P Gilbert[1], Eske Willerslev[2], Anders J Hansen[2], Ian Barnes[1,5], Lars Rudbeck[3], Niels Lynnerup[4], Alan Cooper[1]

[1]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, Oxford University, South Parks Rd, Oxford OX1 3PS, UK
[2]Department of Evolutionary Biology, Zoological Institute, University of Copenhagen, Denmark
[3]Research Laboratory, Institute of Forensic Medicine, University of Copenhagen, Denmark
[4]Laboratory of Biological Anthropology, Institute of Forensic Medicine, University of Copenhagen, Denmark
[5]*Current address* Department of Biology, Darwin Building, University College London, London, England

Summary

The distribution of *post mortem* damage in mitochondrial DNA retrieved from 37 ancient human DNA samples was analysed by cloning and compared to a selection of published animal data. A relative rate of damage (ρ_v) was calculated for nucleotide positions within the human hypervariable region 1 (HVR1) and cytochrome oxidase III (COIII) genes. A comparison of damaged sites within and between the regions reveals that damage hotspots exist, and that in the HVR1 these correlate with sites known to have high *in vivo* mutation rates. Conversely, HVR1 regions with known structural function such as Mt5 have both lower *in vivo* mutation and *post mortem* damage rates. The *post mortem* data also identify a possibly functional region of HVR1, termed Low-diversity 1 (LD1), through a lack of sequence damage. The amount of *post mortem* damage observed in mitochondrial coding regions was significantly lower than HVR1, and while hotspots were noted these did not correlate with codon position. Finally, a simple method for identifying incorrect archaeological haplogroup designations is introduced, based on the observed spectrum of *post mortem* damage.

3.1 Introduction

The extreme susceptibility of ancient human DNA research to contamination with modern human DNA has caused skepticism about the authenticity of many results (Richards and Sykes 1995; Stoneking 1995; Handt *et al.* 1996, Cooper 1997; Kolman and Tuross 2000; Cooper *et al.* 2001b). Several studies have shown that despite rigorous protocols (*c.f.* Cooper and Poinar 2000), contaminants can still be identified in amplified products (Richards and Sykes 1995; Handt *et al.* 1996; Krings *et al.* 1997; Kolman and Tuross 2000; Hofreiter *et al.* 2001b). Contaminants are usually identified when repeat amplifications yield differing, or multiple, sequences in cloned products. However the effect of *post mortem* damage on endogenous DNA is generally underrated. The majority of *post mortem* DNA damage occurs as double strand breaks and oxidative modification of dinucleotides, which prevent subsequent enzymatic replication (Pääbo 1989; Lindahl 1993a; Höss *et al.* 1996). However at least some minor sequence modifications such as hydrolytic deamination and depurination permit polymerase action, and are manifested as limited amounts of base variation among sequenced clones (Krings *et al.* 1997). Few studies have examined the prevalence of such *post mortem* damage in detail (Handt *et al.* 1996, Kolman and Tuross 2000; Hofreiter *et al.* 2001a; Hansen *et al.* 2001), and it is generally considered that when the initial template number is higher than 1000 copies, *post mortem* damage rates are unlikely to bias results (Handt *et al.* 1996; Krings *et al.* 1997) However when few DNA templates initiate a PCR reaction, the resulting sequences are likely to contain artifacts.

Mitochondrial DNA is generally present in archaeological samples in much greater amounts than nuclear DNA (Greenwood *et al.* 1999), and consequently most ancient genetic analyses have used the well-characterised mitochondrial HVR1 (Oota *et al.*

1995; Oota *et al.* 1999; Ribeiro-dos-Santos *et al.* 1996; Torroni *et al.* 1998, Lalueza-Fox *et al.* 2001). Archaeological HVR1 results can be compared with detailed records of base substitutions within modern populations, and sequences can be located within the commonly used human phylogenetic tree relative to the Cambridge Reference Sequence (CRS) (Anderson *et al.* 1981). Unfortunately, a major limitation with this classification system is that certain haplogroups, especially those of Eurasian origin, are routinely categorised using less than five site changes, and often as few as one. If these key sites are commonly modified by *post mortem* damage, then haplogroup designations may be incorrect. In one of the only studies of this issue, Hofreiter *et al.* (2001a) suggested that *post mortem* damage appeared to be randomly distributed in the control region of the cave bear, *Ursus spelaeus*. However only a small number of samples were studied, and it is also possible that *post mortem* damage may vary according to environmental or taxonomic factors. Within modern human populations, certain nucleotide positions within the HVR1, termed 'sites', appear to mutate at significantly higher rates than others, and these sites have been termed hotspots (Vigilant 1991; Hasegawa and Horai 1991; Hasegawa *et al.* 1993; Wakely 1993; Aris-Brisou and Excoffier 1996; Macauley *et al.* 1997; Excoffier and Yang 1999; Meyer *et al.* 1999; Stoneking 2000; Finnilä *et al.* 2001; Heyer *et al.* 2001). The variation in rates of damage has been suggested to relate to DNA secondary and tertiary structure (Heyer *et al.* 2001), and it is possible that the same sites might also be susceptible to *post mortem* damage. To investigate these issues, human HVR1 and COIII mtDNA was amplified and cloned from a range of archaeological specimens and the human sequences were compared to ancient DNA sequences from brown bear (*Ursus arctos*), domestic cow (*Bos taurus*), pig (*Sus scrofa*), goat (*Capria hiris*) and sheep (*Ovis aries*), as well as published data from other taxa.

3.2 Samples, Materials and Methods

3.2.1 Samples

It is exceedingly difficult to authenticate ancient human sequences, and may even be impossible in certain cases (Cooper 1997). Several authors have demonstrated how even stringent controls can fail to prevent or detect contamination (Handt *et al*. 1996; Kolman and Tuross 2000). Consequently, careful attention was paid to the choice of methods and samples to provide the best chance of limiting contamination. Forty-three whole teeth were obtained from 34 archaeological human skeletons preserved in a range of temporal, geographical and environmental situations (Table 3.1). Teeth were used as a DNA source both because the relatively impervious outer enamel layer provides a degree of protection from contaminating DNA sources (chapter 7), and because teeth have been shown to yield higher amounts of DNA than bone in many environments (Kurosaki *et al*. 1993, Oota *et al*. 1995). Teeth from cow, pig, goat and sheep specimens obtained from the Greenland archaeological site were also analyzed for human DNA as a contamination control. Multiple teeth were taken from five specimens to allow replication of the extraction and amplification procedures. DNA from one of these 5 specimens (tooth EA) was extracted and amplified independently at the Ancient DNA Laboratory, Zoological Institute, Copenhagen using similar techniques, to confirm the spectrum of sequence damage observed was replicable. Two brown bear 12S mtDNA sequences were also analyzed to allow taxonomic comparisons of sequence damage outside the HVR1. Both bear sequences have previously been replicated at UCLA and Oxford (Barnes *et al*. 2002).

Table 3.1: Details of samples studied.

				HVR1[1]				COIII[2]			
Sample	Species	Site	Sample Date	Amps[3]	Clones[4]	UNG[5]	UNG Clones[6]	Amps	Clones	UNG	UNG Clones
Tg44	Homo sapiens	Greenland	1000-1200[α]	1	2	1	8				
Tg54	H.sapiens	Greenland	1000-1200[α]	1	5	1	8				
Tg63*	H.sapiens	Britain	Not known	1	32	1	12				
Tg76	H.sapiens	Denmark	1711[β]	1	5	1	11	1	12	1	12
Tg77[a]	H.sapiens	Denmark	1400-1500[γ]	3	9	1	12				
Tg80	H.sapiens	Denmark	1300-1600[δ]	1	7	1	11	1	10	1	9
Tg85	H.sapiens	Denmark	1711[β]	1	5	1	12				
Tg99.1*[e]	H.sapiens	Greenland	1000-1200[α]	9	60	1	20				
Tg99.2[e]	H.sapiens	Greenland	1000-1200[α]	1	16						
Tg99.3[e]	H.sapiens	Greenland	1000-1200[α]	1	16						
Tg103	H.sapiens	Denmark	1300-1600[δ]	1	4	1	11				
Tg104	H.sapiens	Britain	Not known	1	15	1	16				
Tg105	H.sapiens	Denmark	400-1000[δ]	1	4	1	6				
Tg112	H.sapiens	Greenland	1000-1200[α]	1	6	1	7				
Tg114	H.sapiens	Denmark	1711[β]	1	5	1	12	1	11	1	12
Tg116	H.sapiens	Denmark	700-1000[δ]	2	10	1	12				
Tg120	H.sapiens	Denmark	1711[β]	1	2	1	8				
Tg123	H.sapiens	Denmark	1400-1500[γ]	1	6	1	8				
Tg127[a]	H.sapiens	Denmark	1711[β]	1	8	1	11				
Tg128	H.sapiens	Greenland	1000-1200[α]	2	12						
Tg129	H.sapiens	Greenland	1000-1200[α]	2	9						
Tg131	H.sapiens	Greenland	1000-1200[α]	1	7	1	7				
Tg133	H.sapiens	Greenland	1000-1200[α]	1	2	1	8				
Tg136.1	H.sapiens	Britain	1172-1536	2	17			1	11	1	12
Tg136.2	H.sapiens	Britain	1172-1536	2	26						
Tg137.1[c]	H.sapiens	Britain	675-740	3	27						
Tg137.2[c]	H.sapiens	Britain	675-740	3	38						
Tg138[b]	H.sapiens	Britain	873-874	1	8	1	5				
Tg141	H.sapiens	Britain	873-874	1	6	1	6				
Tg142	H.sapiens	Britain	873-874	1	7	1	9				
Tg143	H.sapiens	Britain	874-876	1	6	1	8				
Tg145	H.sapiens	Britain	750-873	1	5	1	7				
Tg146	H.sapiens	Britain	900-1100	1	6	1	8				
Tg147*	H.sapiens	Britain	873-874	2	9						
Tg148[c]	H.sapiens	Britain	675-740	7	80	1	21	3	33	1	12
Tg149[b]	H.sapiens	Britain	873-874	7	57	1	12	2	21	1	10
Tg191*	H.sapiens	Britain	Not known	2	8						
Tg192	H.sapiens	Britain	Not known	2	10	1	11	1	12	1	12
Tg196	H.sapiens	Not known	Not known	2	12						
Tg232[d]	H.sapiens	Britain	Not known	1	8	1	8				
Tg233[d]	H.sapiens	Britain	Not known	2	14						
Tg200*[e]	H.sapiens	Greenland	1000-1200[α]	1	15	2	19				
Ea[e]	H.sapiens	Greenland	1000-1200[α]	1							
Cow	Bos taurus	Greenland	1000-1200[α]	1‡							

				HVR1				COIII†			
Sample	Species	Site	Sample Date	Amps	Clones	UNG	UNG Clones	Amps	Clones	UNG	UNG Clones
Sheep	Ovis aries	Greenland	1000-1200$^{\alpha}$	1‡							
Pig	Sus scrofa	Greenland	1000-1200$^{\alpha}$	1‡							
Goat	Capra hircus	Greenland	1000-1200$^{\alpha}$	1‡							
IB221	Ursus arctos	Alaska	Not known					1	12		
IB223	U.arctos	Alaska	Not known					1	12		
			Total	83	606	31	314	12	134	7	79

Greenland samples have been identified as Viking (Lynnerup 1998).*Sample rejected due to signs of contamination. ‡ No PCR product. †12s gene amplified in bear samples. [a-e] Multiple teeth originating from one of five skeletons (labelled a-e). Samples dated as follows; [α]Radiocarbon and archaeological analysis (Arneborg *et al*. 1999), [β]Historical records, Copenhagen City Museum, Denmark, [γ]Archaeological analysis, Aalborg Museum, Denmark, [δ]Archaeological analysis, Panum Institute, Denmark. Column headings are as follows. [1]*Hvr1* Amplifications over Hypervariable region I. [2]*CoIII* Amplifications over Cytochrome oxidase subunit 3. [3]*Amps* Number of untreated (ie no UNG) PCRs on sample. [4]*Clones* Total number of non-UNG clones sequenced per sample. [5]*UNG* Number of PCRs with UNG treatment on sample. [6]UNG *clones* Total number of UNG clones sequenced on sample.

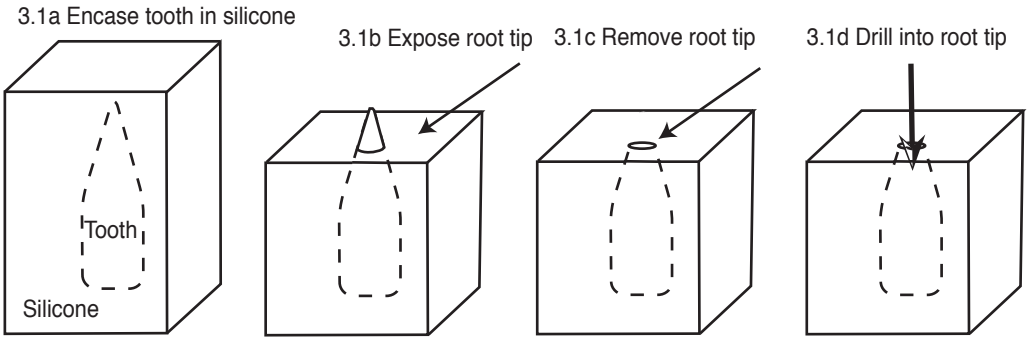
3.2.2 Extraction

All DNA preparation and extraction methods followed strict ancient DNA specific requirements (Cooper and Poinar 2000), and were performed in a dedicated ancient DNA facility, in a building physically isolated from any molecular biology research. Full body suits, breathing masks and face shields were used, and gloves were frequently changed. DNA was extracted from teeth using a new technique that significantly reduces external contamination (chapter 7) (Fig 3.1). Whole teeth were initially washed in 50% bleach for five minutes, followed by exposure to 254nm wavelength UV light for ten minutes on either side. Teeth were fully encased, upside-down, in RTV-11 liquid silicone rubber (Tiranti Ltd, UK), and left overnight while the matrix hardened. The silicone encasing the top 5-10 mm of the root was removed with a horizontal cut, and the root removed flush with the silicon using a single-use carborundum cutting-disc. Dental pulp was powdered and removed from the pulp chamber using a dental drill bit, and digested and extracted using a phenol:chloroform DNA extraction protocol (after Barnes *et al.* 2002), with 0.2M PTB (N-phenacylthiazolium bromide; Poinar *et al.* 1998). PTB has been shown to increase DNA yield in ancient DNA extractions, putatively by cleaving cross-linked proteins and DNA formed by condensation reactions (Vasan *et al.* 1996; Poinar *et al.* 1998; Bada 1999). The bear DNA extractions have previously been described (Barnes *et al.* 2002).

3.2.3 PCR amplification

PCR amplifications used a high-fidelity polymerase (Platinum Taq Hi-Fidelity, Invitrogen UK) after Cooper *et al.* (2001a) to reduce the polymerase error rate, and increase amplification efficiency(Willerslev *et al.* 1999; Hansen *et al.*2001;chapter 2).

Figure 3.1: Dental pulp extraction technique.



Full details of the mitochondrial regions amplified are given in Table 3.1, while primers and annealing temperatures are given in Table 3.2. All samples were amplified and cloned at least twice. PCR products were purified via precipitation using Microclean (Microzone, UK), and cloned using the Topo TA cloning system (Invitrogen, UK). Colonies were used to initiate PCR re-amplifications with vector M13R and T7 primers, purified as before, and sequenced on an ABI 377 or 3700 using the ABI Big Dye 3 PRISM kit (ABI Inc, USA). Blank control extractions were performed at a ratio of 1:5 samples, and negative control amplifications were performed at a ratio of 1:1. No positive controls were used at any step in the amplification process.

3.2.4 UNG treatment

A 5μl aliquot of each human DNA extract was treated with 1 unit of *Escherichia coli* uracil-N-glycosylase (UNG) (Sigma, UK) to excise uracil caused by hydrolytic deamination of cytosine (Dinner *et al.* 2001). UNG reduces sequence artifacts that arise through this common form of *post mortem* damage, which results in an apparent C→T/G→A mutation (Pääbo 1989; Hofreiter *et al.* 2001a; chapter 2). Following UNG treatment, extracts containing multiple sources of DNA can be identified as those in which cloned sequences still share either no consensus sequence, or retain sporadic C→T/G→A base changes that may previously have been attributed to *post mortem* damage, but which now may be recognized as authentic base differences between a contaminant and the sample's endogenous sequence.

Table 3.2: Details of primer used in study.

Primer	Sequence	Annealing Temp (°C)
L09486	5' ttc gca gga ttt ttc tga gc	56
H09652	5'tgg tga gct cag gtg att ga	56
12se	Barnes *et al.* 2002	
12sf3	Barnes *et al.* 2002	
L16055	Handt *et al.* 1996	
H16410	Handt *et al.* 1996	
L16209	Handt *et al.* 1996	
H16356	Handt *et al.* 1996	

Conditions for previously reported primers were identical to the references.

3.2.5 Damage spectrum in ancient specimens

To examine the robustness of the results, samples from a range of sites/time periods were examined (Table 3.1). Previously published cloned ancient DNA sequences of cave bear (Loreille *et al*. 2001a) and brown bear (Barnes *et al*. 2002) were obtained (see Table 3.3). Other reported ancient human DNA clones cannot be included because of an inability to determine the extent of contamination in those studies.

3.2.6 Contamination

Contaminating DNA is a major concern because it can imitate sequence damage, and also permits jumping PCR between endogenous and contaminant strands (Pääbo 1989). Jumping events will increase the apparent number of damaged sites in cloned sequences by introducing positions that differ between the contaminant and authentic DNA. Low-level contamination is often either not observed or reported, but this insidious problem would be manifested in clone sequences as sporadic, apparently *post mortem,* damage at phylogenetically variable sites. To avoid this problem, we analyze specimens representing a variety of mitochondrial haplogroups, so that common contaminants can be detected, and use at least two separate PCR amplifications and UNG treatment (which involves a further PCR replication) to confirm that extracts contain only one source of ancient sequence. It could be argued that an old, and therefore damaged contaminant, could generate a solitary consensus sequence from a tooth, and that UNG treatment would confirm the observed sequence heterogeneity had arisen from *post mortem* damage (as opposed to several contaminating sequences). Although this situation appears to be rare, such results would not affect the analyses as DNA damage is the only issue under investigation, whether from ancient endogenous or contaminant sources.

Table 3.3: Regional specific damage rates per sample.

Sample[1]	Species	Region[2]	Damage Rate[3]	Saturation[4]	Source
Tg76	*H.sapiens*	Hvr1	1.35	0.40	
		CoIII	0.79	0.50	
Tg80	*H.sapiens*	Hvr1	2.82	0.86	
		CoIII	0.79	0.20	
Tg114	*H.sapiens*	Hvr1	4.96	0.80	
		CoIII	3.97	0.36	
Tg136	*H.sapiens*	Hvr1	4.81	0.41	
		CoIII	0.00	0.41	
Tg148	*H.sapiens*	Hvr1	0.68	0.18	
		CoIII	2.91	0.30	
Tg149	*H.sapiens*	Hvr1	5.01	0.88	
		CoIII	4.76	0.58	
Tg192	*H.sapiens*	Hvr1	0.82	0.40	
		CoIII	0.00	0.00	
Ea	*H.sapiens*	Hvr1	0.91	1.00	
		CoIII	1.59	0.60	
TAB15	*U.spelaeus*	CR	3.4	1.0	Loreille *et al.* 2001
		Cytb	0.5	0.9	Loreille *et al.* 2001
IB221	*U.arctos*	CR	6.0	1.0	Barnes *et al.* 2002
		12s	9.0	0.8	
IB223	*U.arctos*	CR	5.4	0.7	Barnes *et al.* 2002
		12s	3.3	0.7	

Where no source is indicated the data is from this study. [1]Authors' DNA extraction identification. [2]*Regions* Fragments from the following genes/non-coding regions; CoIII=Cytochrome oxidase subunit 3, CR=Control Region, Cytb=Apocytochrome b, Hvr1=Hypervariable region I, 12s=12s Ribosomal DNA. [3]*Damage Rate* Sample specific calculation of damage level, only comparable within regions from one sample. [4]*Saturation* Proportion of clones that vary from each other within one cloned region.

3.2.7 Determination of site-specific *post mortem* damage rates

Bases between CRS positions 16209-16356 (here termed Middle Region or MR) contain the majority of phylogenetically variable sites, and were consequently subject to more amplifications than the outer regions (OR), between 16055-16208 and 16357-16410. Four samples which produced multiple sequences in every set of cloned product were presumed to be contaminated, and removed from the analysis. Cloned sequences from the remaining samples were aligned manually with the CRS using the program SEQUENCE NAVIGATOR v 1.0.1 (Applied Biosystems Inc. USA). The consensus sequence for each specimen was determined from the sequences shared between all clones, including the UNG treated amplifications. The remaining intra-clone base differences formed the *post mortem* damage dataset. The few base insertions and deletions (indels) observed were excluded from the analysis.

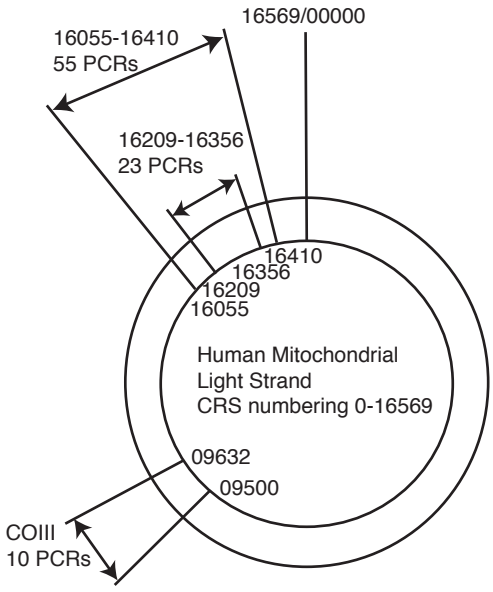
It is difficult to calculate site-specific HVR1 *post mortem* damage-rates for a number of reasons. A major problem is that damaged sites which occur on several clones within a set derived from one PCR are likely to have arisen from the same ancestral template molecule, and therefore the observed number is dependent on the number of clones sequenced. Similarly, the proportion of undamaged starting templates is difficult to assess because of their replication advantages over damaged templates during the early amplification cycles. Furthermore, within a set of clones it is common to see jumping PCR spread damaged sites between daughter amplified strands, generating a few damaged sites at identical nucleotide positions between quite different sequences. Lastly, the different number of PCR amplifications and clones for each fragment (MR and OR) complicate calculations. Consequently, an approximate relative rate of *post mortem* damage (ρ_v) was calculated for each site after Hofreiter *et al*. (2001), modified to account for the dual-fragment amplification

of the HVR1 (Fig 3.2). A modified relative rate of *post mortem* damage was calculated as $\rho_v=\mu_v/\sigma_v$ where v = a specific site with reference to the CRS, μ_v = hits observed at a specific site across all sequences analyzed and σ_v = Total number of amplifications for each specific site. This calculation is biased as it does not take into account the numbers of clones examined at each site (which varies by PCR), or the problems associated with assessing the number of undamaged starting templates. Some previous studies (*c.f.* Cooper *et al.* 2001a) have assumed that the low starting template number implies that all undamaged strands effectively arose from one template. The denominator then becomes the number of different clones (as opposed to PCRs). However, our data demonstrate that different genetic regions exhibit heterogeneity in the rates of *post mortem* damage, and therefore an area with larger amounts of background *post mortem* damage will have a higher denominator in the rate calculation, lowering the apparent relative rates of damage for sites in this region. Consequently it is difficult to compare *post mortem* damage rates between different regions of HVR1, or between HVR1 and COIII, and such calculations should be seen as illustrative, not absolute. It is possible to compare *post mortem* damage estimates within a single region of amplification (MR or OR) however, as all sites have been amplified/cloned the same number of times.

3.2.8 Hotspots

A null hypothesis (H_o) that *post mortem* damaged sites were randomly distributed across HVR1 and COIII was tested through comparison with the expected Poisson distribution (Aris-Brisou and Excoffier 1996; Heyer *et al.* 2001). Heyer *et al.* (2001) have calculated an expected distribution of substitutions in modern human mitochondrial samples and tested this against empirical data. It is a simple matter to

Figure 3.2: Human mitochondrial regions amplified.



modify the test for this dataset, although because of the bias described above (see section 3.2.7), the HVR1 analysis was performed on the MR and OR separately. The probability of sequence sites having exactly X substitutions, P(X), is

$$P(X)=e^{-\lambda}\lambda^{x}/X!$$

in which case Poisson parameter λ refers to the observed density of mutations. To estimate the expected count, LP(X), of sequence positions for each category, given the random distribution λ and the length L of the sequence, we multiply the Poisson probability P(X) by the number of sequence sites L, such that

$$LP(X)=L(e^{-\lambda}\lambda^{x}/X!)$$

A Chi-squared goodness of fit test can then be applied to the observed and expected results to determine whether the H_o of *post mortem* damage being randomly distributed within each of the regions can be rejected.

3.2.9 Relative rate comparison with modern studies

To determine whether there is a correlation between sites with elevated substitution rates in modern populations and sites with high *post mortem* damage rates, estimates of site specific rates of mutation were obtained from recent comprehensive studies (Meyer *et al*. 1999; Excoffier and Yang 1999). Different rate calculation methods were used in each study, preventing direct comparison, so the results were standardized onto a quartile scale of 1-4 (1=lowest, 4=highest), excluding invariant

sites. Sites with consistently high mutation rates (3 or 4) in both modern studies, were contrasted with those in the *post mortem* dataset.

3.2.10 Base structure

The originally damaged mtDNA strand (light (L) or heavy (H)) was determined after Gilbert *et al*. (2003a) (chapter 2), using the prevalence of hydrolytic deamination of cytosine to deoxycytidine residues (leading to an observable C→T or G→A change; C→T/G→A), which are by far the most commonly observed forms of *post mortem* damage (Hansen *et al*. 2001; Hofreiter *et al*. 2001). The base composition around the most mutable and damaged HVR1 sites (those with a standardized mutation/damage rate of 3 or 4) were examined for patterns within a window of 3,5,7,9 or 11 bp around the hotspot. Principal component analysis was performed on the five datasets, and the H_o of the first and second eigenvectors having no correlation with levels of *post mortem* damage were tested using a general linear model.

3.2.11 Functional analysis

The rate of *post mortem* damage, and variation in modern datasets, was analyzed in three short regions whose function is known within OR; 7sDNA, Mt5 and TAS (positions 16106-16191, 16194-16208 and 16157-16172 respectively, Doda *et al*. 1981, Ohno *et al*. 1991). An additional area of low variation observed in the *post mortem* dataset, Low-Damage Region 1 (LDR1), between positions 16365-16395 was also examined. An H_o that the percentage of variable bases did not differ significantly along the OR was tested using a Student's unpaired 2-sample t and a Mann-Whitney test. For the COIII *post mortem* dataset, the H_o tested was that *post mortem* damage (measured as total hits per codon position, 'Hits'), and the number of

variable sites at each codon position (Sites) were evenly distributed amongst different codon positions, by comparing the observed and expected levels with a Chi-squared goodness of fit test.

3.2.12 Regional specificity

Post mortem damage rates in HVR1 and COIII can be compared if it is assumed that amplifications from the same DNA extract will a) start from the same template number and mix of *post mortem* damage and undamaged strands, and b) will amplify under similar conditions. A comparison of relative rates can be made using the number of clones screened for HVR1 and COIII, assuming the number of undamaged sequences results from the same relative number of undamaged templates. The calculation of comparable rate, P, is P=H/LP, where H=total hits, L=total length of sequence and P is the number of PCR amplifications. A saturation value, S, was also calculated for each sample, representing the number of different sequences within the clones. A value of S=1 indicates that all sequences are different, and in such cases it is probable that further clones would yield different products, and therefore that *post mortem* damage for the individual has been underestimated. The H_o that no difference exists in the proportion of sites that show damage between HVR1 and COIII for each sample was tested using a Student's 1-sample t-test and a Wilcoxon test.

3.2.13 Haplogroup analysis

To analyse the possible phylogenetic effects of damage-induced sequence artefacts, a Eurasian mitochondrial tree based predominantly on HVR1 sequences (Richards and Macaulay 2000) was modified to show which branches were defined by sites with high *post mortem* damage rates. This comparison requires that the strand on which

the *post mortem* damage occurred (Light or Heavy) is identified, to allow determination of the direction in which the haplogroup assignments will shift on the phylogenetic tree. In order to standardise the rates of *post mortem* damage between the OR and MR, the OR rates were multiplied by the ratio of the number of PCRs over the MR and OR.

3.3 Results

3.3.1 Data authenticity, *post mortem* damage and UNG

Four of the original 43 human teeth samples appeared to be contaminated (Tg99.1, Tg147, Tg191, Tg200) as cloned sequences from a single PCR amplification differed significantly and/or repeat amplifications yielded different products. Tg99.1 contained at least 6 sequences, while Tg147, Tg191 and Tg200 contained 2 different sequences each. All other samples possessed the same consensus sequence in the separate amplification and cloning experiments. Seven of the 34 sampled individuals shared the relatively common haplogroup (V) of the researcher undertaking the experimental analysis (MTPG) (M. Gilbert, unpublished data). However the sporadic base changes among the clones suggest these sequences are ancient. All other samples used for the analysis (including that sent for independent replication), displayed similar *post mortem* hydrolytic damage. The four animal teeth examined from two of the archaeological sites repeatedly yielded no human DNA, and tests using 'spiked' PCR (*c.f.* Hofreiter et al. 2001a) revealed that this did not result from PCR inhibition.

A consistent, single haplotype may easily be obtained when a sample with no endogenous DNA is contaminated by a single modern source, or even several sources if one is more concentrated than the others (Cooper 1997). However, cloned sequences of all the *post mortem* samples show a pattern of a shared consensus sequence with many scattered singleton substitutions, particularly C→T/G→A, characteristic of aDNA (Handt *et al.* 1996; Krings *et al.* 1997; Hansen *et al.* 2001; chapter 2). In samples treated with UNG no C→T/G→A differences from the CRS were observed unless shared between all clones (therefore deemed to be part of the original sequence). This strongly suggests the sequences derive from one ancient

source (either endogenous, or old, damaged contaminant) and are therefore suitable for this study. The UNG results, together with the very low rate of polymerase misincorporation observed in modern contaminants from both modern and ancient DNA extractions (chapter 2), indicate that the observed sequence heterogeneity in ancient sequences is not a result of enzyme misincorporation.

3.3.2 Rates of *post mortem* damage

The spatial variation of rates of *post mortem* damage per site (calculated as absolute number of hits (A_n) for COIII, and A_n as well as the more comparable relative rate (ρ_n) of *post mortem* damage for HVR1) can be seen in Figures 3.3a-c. The variation initially appears to be over-dispersed with certain sites over represented such as 16110, 16204, 16223, 16270, 16298, 16325 in the HVR1 and 9540, 9545 and 9570 in the COIII. The H_o that *post mortem* damage is randomly distributed can be strongly rejected (Figs 3.4 a-c) (MR χ^2 p-value of $p<0.01$, OR $p=0.02$, COIII $p<0.01$), demonstrating that sites in both regions are not being damaged at random.

3.3.3 Structural analysis

The majority of sites (24/30) with a high rate of *post mortem* damage have also been observed as hyper-mutable in the modern data sets (Table 3.4). The exceptions (16110. 16131, 16144, 16204, 16242 and 16325) are of special interest and the surrounding base composition was examined for a possible cause. The first and second eigenvectors of the Principal Component Analysis explained the majority of the results in all cases, and there was no evidence to reject the H_o that base composition around individual sites plays no role in determining a propensity for *post*

Figure 3.3a-c: Damage variation across HVR1 and COIII.

Fig 3.3a:Absolute damage measured in hits/site for data sets generated by each PCR primer set across HVR1 (16055-16410).

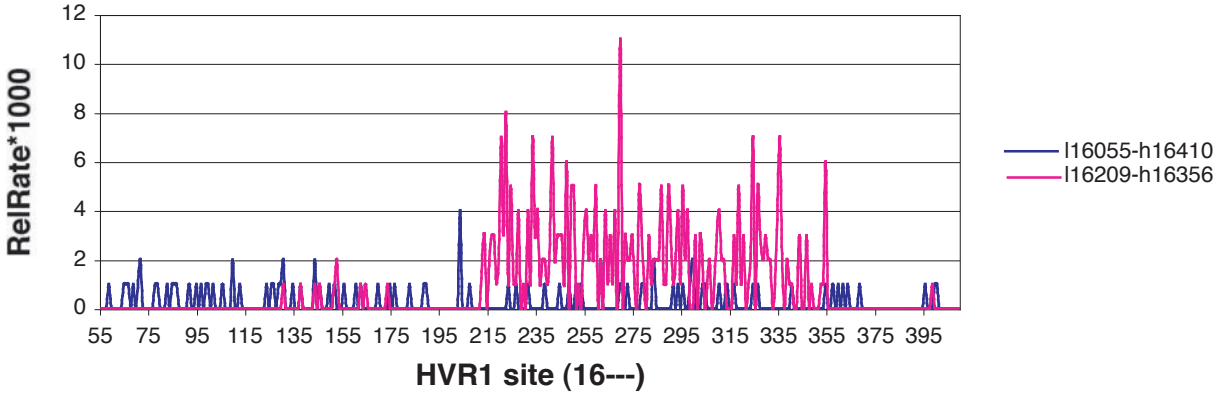


Fig 3.3b:Relative rate variation(*100) across HVR1, calculated as in text. Four areas of interest are marked onto the graph.

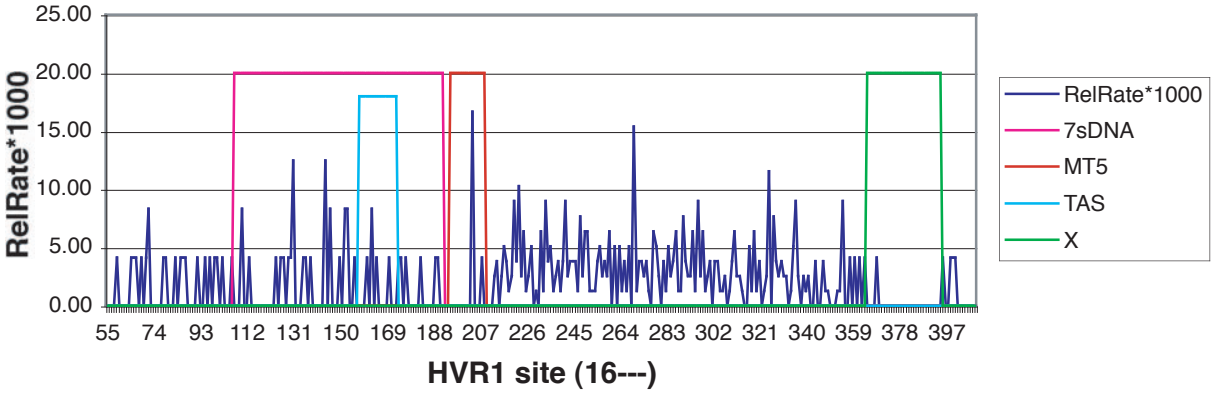


Fig 3.3c: Absolute damage across region of COIII (09500-09632). Numbering is with reference to the CRS (Anderson *et al.* 1981).

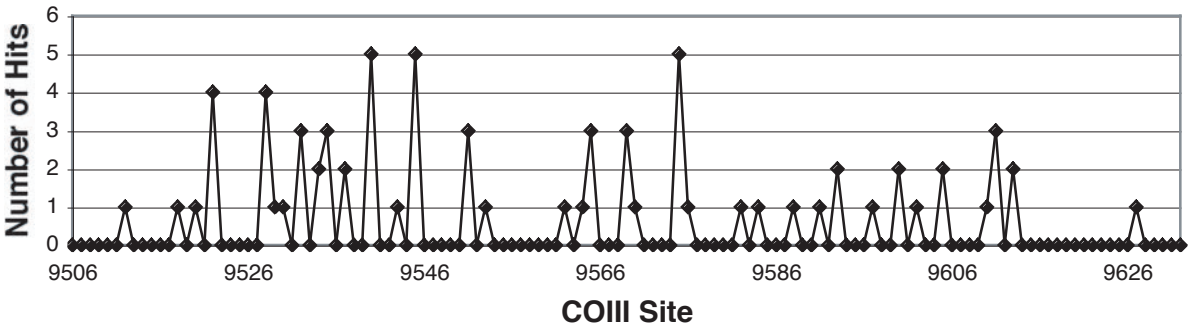
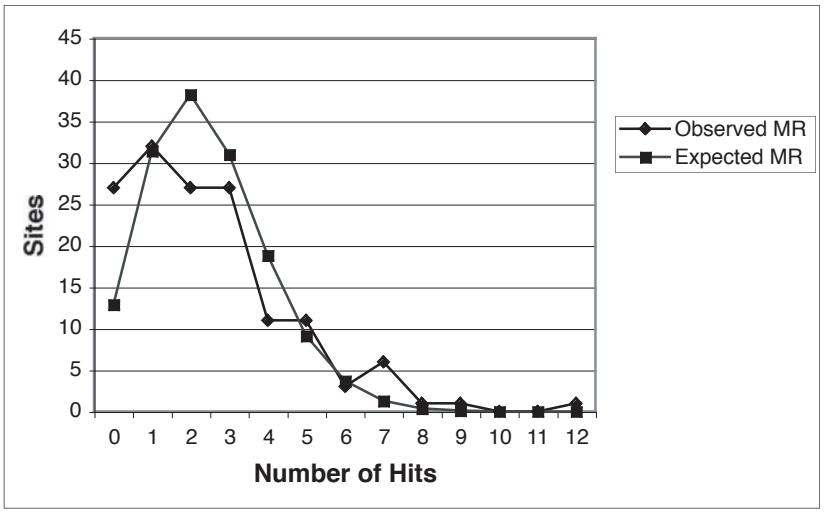
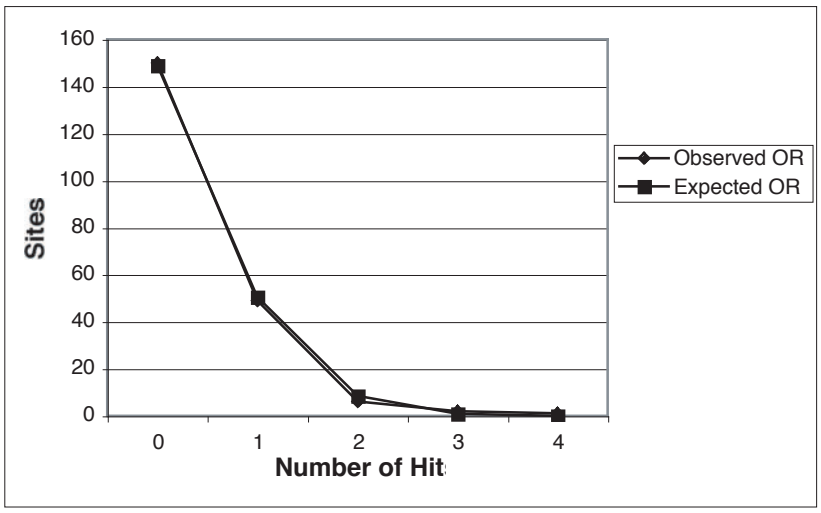


Figure 3.4a-c: Observed and expected random distribution of absolute damage rates for HVR1 MR (3a) HVR1 OR (3b) and COIII (3c). For full details see text.





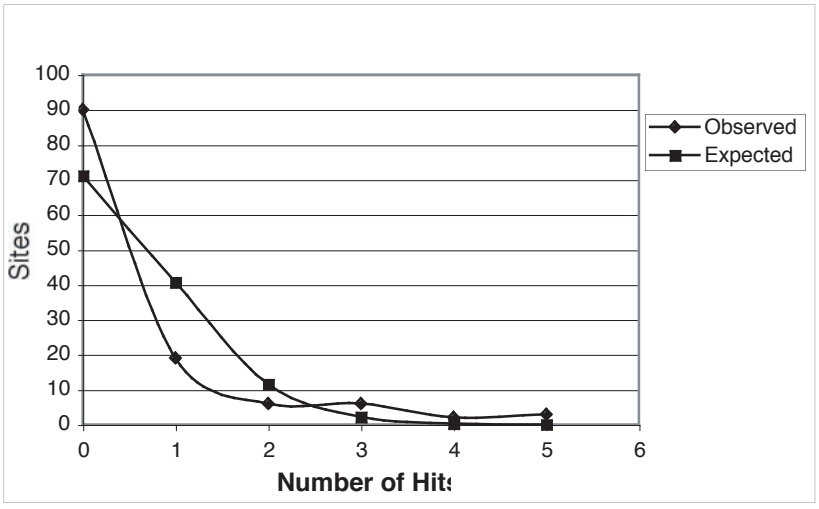


Table 3.4: Standardised mutation and damage rates.

HVR1 Base Position	E99	M99	TG03
93	4	3	4
126	0	4	3
129[a]	4	4	1
148	0	3	1
163	2	3	3
172[a]	4	3	1
182	4	0	1
183	4	3	2
187	0	3	1
189[a]	4	4	1
192[a]	4	3	1
209	4	0	1
219	0	3	1
223	4	4	4
230	0	4	1
234	3	0	2
265	4	2	1
270	4	3	4
274	0	3	2
278	4	4	2
290	2	2	3
291	4	2	1
293[a]	4	4	1
294	4	4	2
298	0	2	4
304	4	0	1
309	4	4	1
311	4	4	2
319	0	3	2
320	3	2	1
327	0	2	3
343	3	2	1
355	3	2	2
362[a]	4	4	1

Site-specific in vivo mutation rates taken from two previous studies (Excoffier and Yang 1999 [E99]; Meyer *et al*. 1999 [M99] were standardized into quartiles and were compared with the standardized postmortem-damage rates from the present study [TG03].[a]Seven sites where major disagreement is observed between rates of occurrence of modern mutations and ancient damage.

mortem damage (results not shown). However 2 of the 3 HVR1 regions with known function appear to have a reduced rate of *post mortem* damage (Fig 3.3b). A fourth region of low *post mortem* damage 'LDR1' is also apparent between 16365-16395 of HVR1. Two-sample t-tests and non-parametric Mann-Whitney tests show the percentage of *post mortem* damaged sites within LDR1 is significantly different from the average rate across the OR at $p<0.05$, and across Mt5 at $p<0.1$ (Table 3.5). Within COIII, there is no evidence to suggest that there is a bias of *post mortem* damage towards codon positions. (Sites χ^2 p=0.47, Hits χ^2 p= 0.14).

3.3.4 Regional and species specificity

Table 3.3 presents sample-specific rates of *post mortem* damage, P, and saturation, S, for each sample with both HVR1 and COIII data. Six out of eight human samples and two out of three bears demonstrate higher *post mortem* damage rates in HVR1 than COIII/Cytb with HVR1 regions showing 1.1 to 6.8 times the *post mortem* damage than corresponding COIII/Cytb regions. Of the three exceptions, two (bear IB 221 and human EA) demonstrate HVR1 saturation values of 1, indicating the *post mortem* damage is underestimated. When these two samples are removed, the H_o of an equal amount of *post mortem* damage in HVR1 and coding regions can be rejected (One-way t-test p=0.033, Wilcoxon p=0.038).

3.3.5 Haplogroup analysis

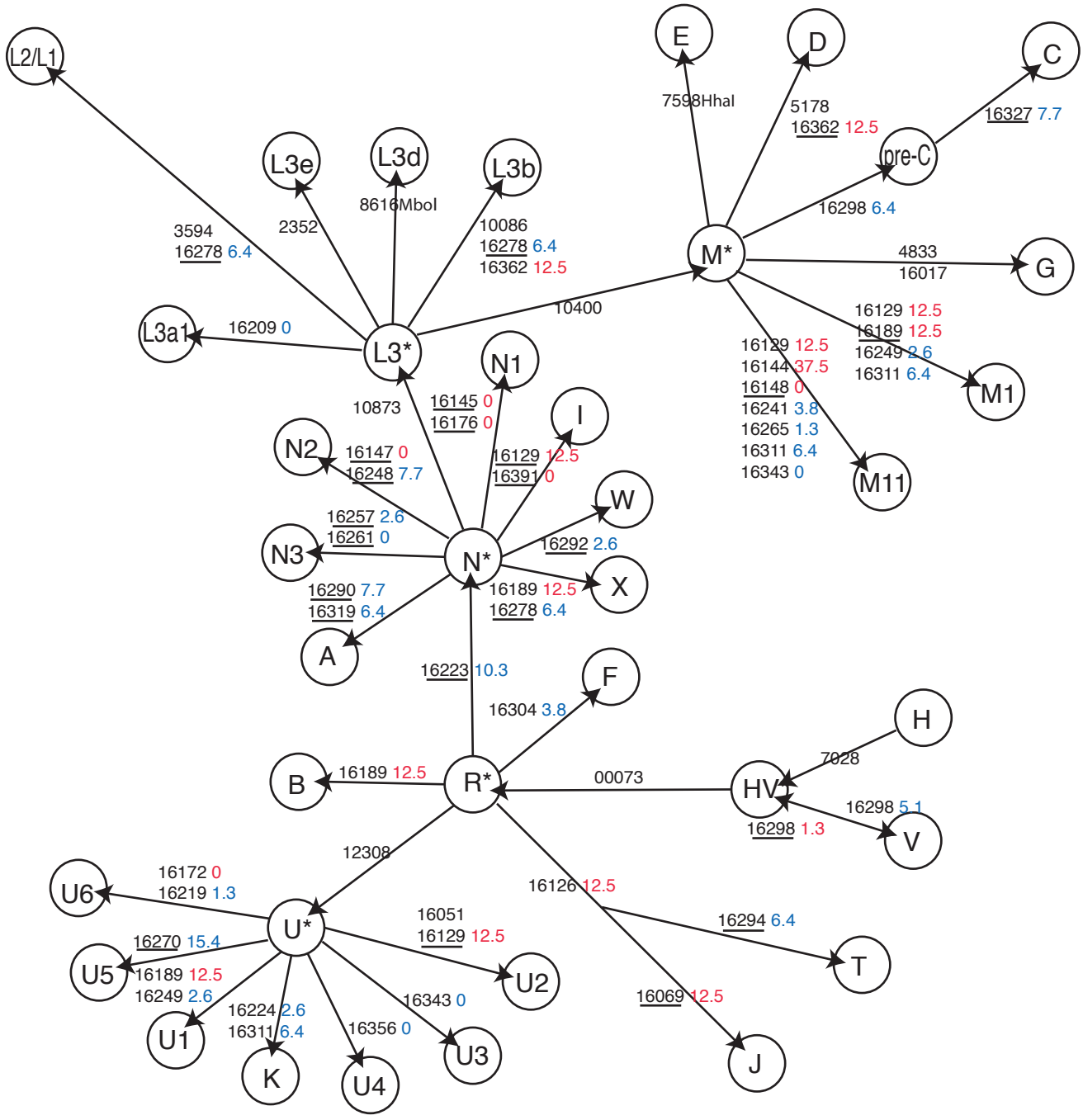
Figure 3.5 shows the relationship between phylogenetic changes and *post mortem* damage rates on each branch of the Eurasian HVR1 mitochondrial tree. The analysis is limited because the majority of our samples are genetically similar north-western Europeans (Anglo-Saxons and Scandinavians/Vikings). Contemporary indigenous

Table 3.5: Statistical tests on absolute number of sites changing across the Outer Regions (OR), and 4 sub-regions.

				2 Sample t		Mann-Whitney	
Region[1]	Bases[2]	Changes[3]	Average[4]	t-statistic	P-value	W statistic	P-value
OR	208	58	0.28				
7sDNA	86	21	0.24	-0.93	0.82	12435	*
TAS	16	3	0.19	0.27	0.40	20112	0.39
MT5	15	1	0.07	1.62	0.06	20396	0.10
LD1	31	1	0.03	6.07	0.00	19291.5	0.00

[1]*Region* Sub-region of mitochondrial HVR1. *Bases* Size of the region in base pairs. [3]*Changes* Number of damaged sites. [4]*Average* Average damage per site. Remaining columns provide statistical calculations from 2-sample t-tests and Mann-Whitney nonparametric test. *No P-value returned by Minitab.

Figure 3.5: Modification of mitochondrial HVR1 haplogroup tree (with permission, Richards and Macaulay 2000) to demonstrate potential for haplogroup mis-identification.



From a starting sample of Haplogroup H or V, arrows indicate the direction along the tree a sample could appear to move due to hydrolytic deamination (or real singleton substitution) of C>T or A>G. Site numbers are with reference to the CRS (Anderson *et al.* 1981). Relative rates are calculated for the MR (blue) and OR (red). HVR1 sites where modifications would arise from Type 2 (C>T/G>A) transitions (those most frequently observed among post-mortem miscoding lesions, and which would undergo cleavage when treated with UNG), are underlined, representing 48% of the available sites. Note that this would include site 16298 when a haplogroup V sample was assessed, but not when a haplogroup H sample was assessed. For full details see text.

European populations reach up to 60% haplogroup H and 40% haplogroup V (Torroni *et al*. 1998), and this is mirrored in our findings (M Gilbert, unpublished data). Consequently, the *post mortem* damage in HVR1 is mostly constrained to non-H or non-V substitutions, and unidirectional changes are almost exclusively observed, (for example, *post mortem* damage at 16223 will nearly always be constrained to changes from the haplogroup H or V cytosine to a 5-hydroxyuracil).

However, where the opportunity for back mutations exists, such as 16298 in haplogroup V (changing back to the CRS state), this is observed in only one clone out of the seven specimens. This scenario is marked on the tree. The data permitted a rough estimation of which haplogroups are most likely to arise through *post mortem* damage starting from the CRS. The limiting factors are naturally the number of sites that are needed to change (two sites are less likely to change than one) as well as the site with lowest rate of *post mortem* damage in a haplogroup. For example haplogroup I differs from the CRS at sites 16129, 16223 and 16391, which have relative rates of *post mortem* damage of 12.5, 10.3 and 0 respectively. Although 2 of these sites are highly susceptible to *post mortem* damage, 16391 effectively prevents the misidentification of haplogroup I.

3.4 Discussion

The sequence results are consistent with derivation from a series of specimens containing single ancient DNA sources. The cloning and UNG experiments, along with the repeat extractions and amplifications, suggest that mtDNA heteroplasmy (Gocke *et al*. 1998), nuclear copies of mitochondrial genes (numts), or polymerase misincorporations (Hansen *et al*. 2001) are not contributing to the observed variation. The results clearly demonstrate that *post mortem* DNA damage is not distributed randomly across the HVR1 or COIII.

Although HVR1 contains no functional genes, modern population studies show it does not evolve in a random manner, as sequence variation is concentrated at a few sites. Relative rates of mutations have been calculated using a range of methods (Hasegawa *et al*. 1993; Wakely 1993; Meyer *et al*. 1999; Excoffier and Yang1999; Heyer *et al*. 2001) which are not directly comparable, although striking similarities are apparent even after standardisation has reduced resolution (Table 3.4). Of the 34 sites that can be compared for *post mortem*-damage and *in vivo* mutation rates, six show very similar rates in all three studies, 23 show similar rates in this and at least one of the other studies, and only seven (sites 16129, 16172, 16189, 16192, 16293, 16309 and 16362) completely disagree (i.e. are not observed to mutate *in vivo* but experience fast *post mortem* damage or *vica versa*). However, at least eleven sites from the two modern studies also exhibit different rates from each other. Thus these findings may relate to sampling stochasticity or the standardization approach. If so, further sampling may provide more accurate estimates of mutation rates at these sites. Other explanations for different rates include the possibility that some form of *in vivo* protection for these sites may be removed/degraded after death.

Structural constraints may also explain the correlation between the *post mortem* damage and mutation rates of mitochondrial DNA in modern studies. For example, if DNA secondary structural conformation predisposed particular sites to hydrolytic attack, then these could appear both hyper-mutable in modern populations, and common targets of *post mortem* damage. Secondary structural models of the human HVR1 may provide a useful insight, but are currently insufficiently detailed. Meyer *et al*. (1999) remark on the correlation of HVR1 site-specific mutation rates with the three known features of structural interest. 7sDNA is a short fragment of H strand DNA which provides the D-loop with its characteristic triple stranded feature, extending from a trinucleotide stop codon at 16104 to at least 00110 (Doda *et al*. 1981; Meyer *et al*. 1999). The stop codon itself exhibits neither *post mortem* damage nor high mutation rates in this, or the data from previous studies (Meyer *et al*. 1999; Excoffier and Yang 1999) and the 7sDNA itself appears to mutate and receive *post mortem* damage at average rates. The other two structural regions, TAS, the Termination Associated Sequence, and the putative control element Mt5, have low observed *in vivo* mutation rates, which have previously been suggested to result from functional constraints (Meyer *et al*. 1999). TAS is located upstream of the trinucleotide stop codon, and interacts with sequence-specific binding factors (Doda *et al*. 1981; Wallace *et al*. 1995), while Mt5 is postulated to be a protein binding site (Meyer *et al*. 1999). Both TAS and Mt5 appear to display lower than average *post mortem* damage rates, although this cannot be statistically validated at $p<0.05$. Certain proteins have been shown to survive for much longer time scales than DNA (Bada *et al*. 1999), and it is possible that the low amount of *post mortem* damage observed in the Mt5 region (Fig 3.3) is related to the continued binding and protection offered by the putative control element protein.

The results also identify a previously unrecognised region with significantly reduced levels of *post mortem* damage, LDR1. HVR1 mutation rate studies (Meyer *et al.* 1999; Excoffier and Yang 1999) do not include all of this region, but do characterise low rates of *in vivo* mutation in the 5' segment. The absence of *post mortem* damage over such a large region (and over 600 clones) is unusual and suggests that the DNA sequence may be protected from hydrolytic damage in some way. One possibility would be a protein-DNA association, similar to that postulated for Mt5 (Meyer *et al.* 1999), and if further studies verify this observation, LDR1 may be the first structural feature identified by aDNA. The hypothesis that there is a correlation between function and *post mortem* damage rate should be tested further by examining ancient sequences of other CR regions with known binding sites such as the main regulatory elements of transcription and regulation, or the origin of heavy strand replication (Chang *et al.* 1985; Wallace *et al.* 1995).

One of the most interesting findings of this study is the observation in both humans and bear species that *post mortem* damage rates are higher in HVR1 than coding regions. Three samples did not show this pattern, two of which (one human and one bear) can be excluded due to high 'saturation' values of clone sequence diversity within the HVR1 region, indicating the true level of *post mortem* damage is underestimated. The third anomaly, human sample Tg148, cannot be excluded on this basis and it is possible that varying amounts of template in the original PCR reactions may be responsible. Finnilä *et al.* (2001) note that *in vivo* mutation rates of mitochondrial 3rd codon positions are also lower than those of the HVR1. A plausible explanation is that the secondary structure conformation of the HVR1 promotes increased rates of both *in vivo* mutation and *post mortem* damage, whereas in certain coding regions this is lacking, or there has been some selection to constrain mutation

rates. Structural models of the human HVR1 may provide a useful test of this hypothesis.

It is interesting that the *post mortem* damage results do not show the selection against 1st and 2nd codon position mutations seen in coding region sequences of modern populations. The ancient sequences may actually reflect the real mutation rates of COIII sites which are normally masked by the effects of selection. This pattern shows no evidence for codon bias, but does shows hotspots for *post mortem* damage like HVR1 sequences.

3.4.1 Problems with studies on human aDNA

Current human aDNA research focuses on the identification of mtDNA haplogroups through either restriction digests (*c.f.* Merriweather *et al.* 1994; Kaestle and Glenn Smith 2001) or sequence analysis (*c.f.* Hänni *et al.* 1994; Krings *et al.* 1997; Stone and Stoneking 1998; Krings *et al.* 1999; Adcock *et al.* 2001), and sexing individuals through systems such as amelogenin (*c.f.* Faerman *et al.* 1995; Fily *et al.* 1998; Vernesi *et al.* 1999). The sexing and restriction-digest studies lack sufficient resolution to allow assessment of the role of either *post mortem* damage or contamination, and so we can only analyse research that uses sequence data to define, or even determine, new haplogroups such as studies of Neanderthal or archaic human samples (Krings *et al.* 1997; Krings *et al.* 1999; Ovchinnikov *et al.* 2000; Adcock *et al.* 2001).

The data in the current study has been generated from a small set of north-western European samples, predominantly of haplogroups H and V. Consequently, the mtDNA tree (Fig 3.5) has been modified to demonstrate the potential haplogroup alteration caused by damage to these haplogroups only. However Figure 3.5

demonstrates the ease with which phylogenetic misidentification can occur if *post mortem* damage is not detected in sequences from any haplogroup. For example, it is not inconceivable that an Amerindian haplogroup A sequence could result from a haplogroup H Viking sample. We suggest that *post mortem* damage may explain many unusual results obtained from ancient human remains where appropriate techniques have not been followed (*e.g.* the 60,000 year old 'Mungo man' sequences, Adcock *et al.* 2001). *Post mortem* damage will also complicate population genetic analyses of ancient humans, and detailed cloning will be needed to avoid overestimation of heterogeneity and population expansion sizes (Lundström *et al.* 1992; Aris-Brisou and Excoffier 1996).

For ancient human DNA research to progress it will be necessary to rigorously enforce aDNA guidelines (Cooper and Poinar 2000). As real-time PCR becomes more routinely available the initial template copy number should be quantified in samples, and detailed cloning experiments should become a requirement. Samples with low template number should be replicated and cloned at least once since even highly damaged sites such as 16270 are unlikely to predominate in 2 independent PCRs. However the simplest procedure appears to be a comparison of sequences before and after enzymatic treatment with UNG (Pääbo 1989; Hofreiter *et al.* 2001a), EndoIV (Pääbo 1989), AAG (Lau *et al.* 2000), or T4 with Pol 1 (Pusch *et al.* 1998; Di Bernardo *et al.* 2002) to reduce the number of templates with *post mortem* damaged bases, and the resulting possibility of mis-identification. Such treatment is expected to reduce the starting template copy number which provides a check for authenticity, but will also limit the number of samples suitable for study. A variety of other non-genetic factors should also be considered such as depositional environment (Nielsen-Marsh and Hedges 2000; Höss *et al.* 1996) and microbial content (Burger *et*

al. 1999), amino-acid racemisation (Poinar *et al*. 1996; Poinar and Stankiewicz 1999) and composition (Bada 1999), and gas chromatography/mass spectroscopy (GC/MS) measured levels of hydantoins (Höss *et al*. 1996). These issues are now well known, and the heavy burden of proof associated with aDNA research demands that subsequent studies address them.

3.5 Acknowledgements

We are indebted to Chris Stringer, Martin Biddle and Birthe Kjølbye-Biddle for samples, and Paul Bradshaw and the BBC who funded initial studies as part of 'Blood of the Vikings' production. We would also like to thank Ziheng Yang and Arndt von Haesler for data provided, and Vincent Macaulay for permission to modify the tree used in Figure 5. M.T.P.G. is grateful to Eddie Holmes, Paul Johnson, Gil McVean, Svante Pääbo, Mark Stoneking, Ryk Ward and members of the ABC and MPEvA for useful comments. E.W. and A.J.H are grateful to Kim Aaris-Sorensen and Tina B Brand for samples, help and discussion. M.T.P.G. and A.C. were supported by the Wellcome Trust. E.A. and A.J.H. were financially supported by the VILLUM KANN RASMUSSEN FONDEN, Denmark. L.R. was funded by the Danish Research council for the Humanities. E.W. and A.J.H. contributed equally to the work. Two anonymous referees provided helpful comments on the manuscript.

Chapter 4

Post mortem DNA damage hotspots in Bison (*Bison bison*) provide supporting evidence for mutational hotspots in human mitochondria

Modified from the manuscript submitted to 'Molecular Biology and Evolution'

M Thomas P Gilbert[1], Beth A Shapiro[1], Alexei Drummond[1,2], Alan Cooper[1]

[1]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, University of Oxford, South Parks Rd, Oxford OX1 3PS, United Kingdom
[2]Department of Statistics, University of Oxford, South Parks Rd, Oxford OX1 3TG, United Kingdom

Summary

Chapter 3 demonstrated that the distribution of *post mortem* DNA damage was not random, but rather mutations were concentrated on hotspots within the human Hypervariable Region. Furthermore, the hotspots correlate with sites that are known to have elevated mutation rates *in vivo*. In this chapter the results of chapter 3 are reproduced in a non-human data set composed of mitochondrial control region DNA extracted, amplified and sequenced from ancient bison (*Bison bison*) specimens. The data identify correlating damage and mutation 'hotspots' in bison and modern bovids. In addition, the analyses highlight a 14bp sequence within the control region, the Mt5 putative protein-binding locus, which exhibits both reduced DNA damage rates in bison, and extraordinary sequence conservation among eutherian mammals. These findings support the hypotheses discussed in chapter 3, that a) hotspots for mitochondrial DNA damage and mutation exist, and b) structural elements of mitochondrial DNA confer a degree of protection from this damage.

4.1 Introduction

Certain nucleotide positions (hereafter termed 'sites') within the human mitochondrial genome have been reported to mutate at rates that are significantly higher than average, and have been described as mutational 'hotspots' (Vigilant 1989; Hasegawa and Horai 1991; Hasegawa *et al.* 1993; Van de Peer 1993; Wakely 1993; Bendall *et al.* 1996; Aris-Brisou and Excoffier 1996; Macaulay *et al.* 1997; Mumm *et al.* 1997; Parsons et al 1997; Excoffier and Yang 1999; Meyer, *et al.* 1999; Sigurðardóttir *et al.* 2000; Stoneking 2000; Van de Peer *et al.* 2000; Finnilä *et al.* 2001; Heyer *et al.* 2001; Pesole and Saccone 2001; Forster *et al.* 2002, Malyarchuk *et al.* 2002; Howell *et al.* 2003; Meyer and von Haeseler 2003). Although some of these studies have identified hotspots through the direct comparison of closely related (*e.g.* familial) DNA sequences (*c.f.* Bendall *et al.* 1996; Mumm *et al.* 1997; Parsons *et al.* 1997; Sigurðardottir *et al.* 2000; Heyer *et al.* 2001; Stoneking 2000; Finnilä *et al.* 2001; Forster *et al.* 2002), such studies have limited resolution due to the small size of data sets, and as mitochondrial mutation rates are too slow to clearly distinguish different sites. For example, in a study of background radiation-induced mutations in over 980 humans, only 22 mutations were observed over the complete Hypervariable Region 1(HVR1) and 2 regions (Forster *et al.* 2002). In order to draw conclusions from larger data sets, most studies have estimated individual site mutation rates using phylogenetic reconstructions of human mitochondrial sequences. One weakness with this method is that the accuracy of hotspot designation is directly related to the accuracy of the phylogenetic reconstruction used in representing the true phylogeny. Consequently it is not unusual for different studies to identify hotspots in conflicting sites. An example is human site 16325, identified as mutating slowly in some studies (*c.f.* Excoffier and Yang 1999; Meyer *et al.* 1999), but above average when analysed with other techniques (Bandelt *et al.* 2002).

In addition to these issues, the existence of hotspots *per se* has also been questioned. Hagelberg (2003) raises mitochondrial recombination as a preferable explanation for homoplasies observed on phylogenetic trees that would otherwise be attributed to recurrent mutations.

In a previous study, Gilbert et al (2003b) (chapter 3) have investigated the distribution of *post mortem* DNA sequence modifications ('damage') in ancient human DNA. The results of this study, (that *post mortem* DNA sequence modifications are not distributed randomly across the human HVR1 and COIII mitochondrial regions), support the existence of *post mortem* damage hotspots. The data also demonstrates a correlation between *in vivo* mutational and *post mortem* damage hotspots. This suggests that the processes involved in *post mortem* sequence modification are analogous to those driving *in vivo* sequence mutations, and that DNA secondary and tertiary structure may predispose or protect certain sites from being damaged (Meyer *et al*. 1999; chapter 3).

However, as with all human ancient DNA (aDNA) studies the authenticity of the chapter 3 dataset is open to criticism due to the potential for contamination with modern or previously amplified human DNA fragments. Contaminated samples are of concern because the resulting spectrum of base variation among amplified and cloned products would both imitate sequence damage (chapter 3) and also permit 'jumping PCR' between endogenous and contaminant strands (Pääbo *et al*. 1990). Such 'jumping PCR' will increase the recorded number of damaged sites in cloned sequences by introducing positions that differ between the contaminant and authentic DNA.

The existence of *post mortem* DNA damage hotspots can be confirmed through an analysis on a DNA damage dataset derived from non-human specimens (and therefore much less susceptible to sample contamination). Shapiro *et al*. (in preparation) have extracted aDNA from over 350 ancient bison (*Bison bison*) specimens. As at least three

published studies have reported the presence of hypervariable mutational sites in bovids (Troy *et al*. 2001; Cymbron *et al*. 1999; Wu *et al*. 2000), bison present an ideal data-set to study the distributions of *post mortem* damage. Cloned PCR products of mitochondrial control region of 81 ancient bison are generated and analysed for *post mortem* damage hotspots. Damage hotspots are contrasted to putative *in vivo* mutational hotspots in bison identified from two sources: homoplasies on phylogenetic trees generated from a large subset of Shapiro *et al*.'s data, and sites observed as hotspots in three studies on modern bovid DNA. In addition, a specific locus that we believe is conserved among the majority of eutherian mammals, the Mt5 element (Ohno *et al*. 1991), is investigated to gain a deeper insight into the pressures acting on mutation and damage in the HVR1.

4.2 Materials and Methods

A small proportion of *post mortem* DNA damage events do not result in DNA fragmentation, and instead generate miscoding lesions (Pääbo 1989). These are manifested as base modifications in the amplified sequence, changing the appearance of a DNA template, and represent the basis of this analysis. The few detailed studies of miscoding lesions concur that, as *in vivo*, the majority of changes arise from the deamination of cytosine (C) to uracil, an analogue of thymine (T) (Hofreiter *et al.* 2001a; Gilbert *et al.* 2003a (henceforth referred to as chapter 2), or the deamination of adenine (A) to hypoxanthine, an analogue of G (chapter 2). For simplicity, both the chemical event and the phenotype are referred to hereafter 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 it will be read as a G→A transition. Similarly, an A→G degradation may be observed as either A→G, or a T→C transition (Hansen *et al.* 2001; Hofreiter *et al.* 2001a). 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 addition to these, a small proportion of miscoding lesions may be derived through oxidatively derived tranversions.

4.2.1 DNA sequences generated

Eighty-one PCR products were amplified and cloned from 61 ancient bison individuals, between *Bison bison* Reference Sequence (BbRS) (this study) positions 24-653 (Table 4.1). For comparison with the *Bos Taurus* Reference Sequence (BtRS) (Anderson *et al.*

1982) refer to Table 4.1. Details of the different products are shown Table 4.2. Full details of the samples and primers are to be presented elsewhere (Shapiro *et al*. in preparation). PCR amplifications used a high-fidelity polymerase (Platinum Taq Hi-Fidelity, Invitrogen UK) after Cooper *et al*. (2001a) to reduce the polymerase error rate, and increase amplification efficiency (Willerslev *et al*. 1999; Cooper *et al*. 2001a; Hansen *et al*. 2001; chapter 2). The use of such enzymes enables sequence miscoding lesions to be confidently attributed to *post mortem* DNA damage as opposed to errors introduced during enzymatic amplification. PCR products were purified via precipitation using Microclean (Microzone, UK), and cloned using the Topo TA cloning system (Invitrogen, UK). Colonies were used to initiate PCR re-amplifications with vector M13R and T7 primers (Invitrogen, UK), purified as before, and sequenced on an ABI 3700 using the ABI Big Dye V3.1 PRISM kit (ABI Inc, USA).

4.2.2 Determination of site-specific *post mortem* damage rates in bison

Cloned sequences were aligned with the BbRS manually using the program SEQMAN 4.0 (DNAstar INC). The consensus sequence for each specimen was determined from the sequences shared between all clones. The remaining intra-clone base differences formed the *post mortem* damage dataset. The few base insertions and deletions (indels) observed were excluded from the analysis.

As outlined in chapter 3, it is difficult to calculate site-specific *post mortem* damage-rates for a number of reasons. A major problem is that damaged sites which occur on several clones within a set derived from one PCR are likely to have arisen from the same ancestral template molecule, thus should not be counted twice. Secondly, the proportion of undamaged starting templates is difficult to assess because damaged templates may be replicated slower during the early amplification cycles. Furthermore, within a set of

Table 4.1: *Bos bison* Reference Sequence (BbRS).

```
1     CTACAGTCTC ACCGTCAACC CCCAAAGCTG AAGTTCTATT TAAACTATTC
51    CCTGAACGCT ATTAATATAG TTCCATAAAT GCAAAGAGCC TCACCAGTAT
101   TAAATTTACT AAAAATTCCA ATAACTCAAC ACAAATTTTG TACTCTAACC
151   AAATATTACA AACACCACTA GCTAACGTCA CTCACCCCCA AAATGCATTA
201   CCCAAATGGG GGGGACGTAC ATAATATTAA TGTAATAAAA ACATATTATG
251   TATATAGTAC ATTAAATTAT ATGCCCCATG CATATAAGCA AGTACTTAAT
301   CCCTATTGAT AGTACATAGT ACATAAAGTT ATTAATTGTA CATAGCACAT
351   TATGTCAAAT CTACCCTTGA CAACATGCAT ATCCCTTCCA TTAGATCACG
401   AGCTTAACTA CCATGCCGCG TGAAACCAGC AACCCGCTAG GCAGGGACCC
451   CTCTTCTCGC TCCGGGCCCA TGAATTGTGG GGGTCGCTAT TTAATGAACT
501   TTATCAGACA TCTGGTTCTT TCTTCAGGGC CATCTCATCT AAAATTGTCC
551   ATTCTTTCCT CTTAAATAAG ACATCTCGAT GGACTAATGG CTAATCAGCC
601   CATGCTCACA CATAACTGTG CTGTCATACA TTTGGTATTT TTTTATTTTG
651   GGGGATGCTT GGACTCAGCT ATGGCC
```

The *Bos bison* Reference Sequence used in this study corresponds to *Bos taurus* Reference Sequence (BtRS) Light strand positions 15761-16338-00052 (Gen bank J01394: Anderson *et al.*1982). Over this region the bison reference sequences differs from the BtRS by 20 deletions and 1 insertion, as follows: 15916/15917 AA deletion, 15921-15925 inclusive GCCCA deletion, 15927 A deletion, 15931-15933 inclusive AGA deletion, 15957 A deletion, 16058-16059 T insertion, 16121/16122 GT deletion, 16124-16126 inclusive TAT deletion, 16128 T deletion, 16133 T deletion, and16143 A deletion.

Table 4.2: Number of independent PCR amplification events over different sections of the Bison Control Region.

BbRS[1]	Number of Independent PCRs
21-37	8
38-66	15
67-92	16
93-166	19
167-184	17
185-261	23
262-272	20
273-289	7
290-292	6
293-308	8
309-341	10
342-349	9
350-416	6
417-419	17
420-434	30
435-496	29
497-560	24
561-616	22
617-642	16
643-657	3

Due to the different, and often overlapping, primer sets used to amplify the bison DNA, some nucleotide sites within the sequence have been amplified on more occasions than others. This complicates DNA damage rate calculations, and must be taken into account (see text). The above table highlights the number of times that each site in the analysed sequence has been amplified. [1]*BbRS* Sites numbered with relation to *Bos bison* Reference Sequence (BbRS).

clones it is common to see jumping PCR spread damaged sites between daughter amplified strands, generating a few damaged sites at identical nucleotide sites between quite different sequences (*c.f.* Hofreiter *et al.* 2001a). Lastly, the variation in number of PCR amplifications and clones sequenced for different regions of the bison sequence (Table 4.2) complicates calculations and makes it difficult to directly compare rates between regions (see chapter 3). Consequently, an approximate relative rate of *post mortem* damage (ρ_v) that takes into account the above issues was calculated for each site after chapter 3.

In this calculation, $\rho_v=\mu_v/\sigma_v$ where v = a specific site with reference to the reference sequence, μ_v = hits observed at a specific site across all sequences analyzed and σ_v = Total number of amplifications for each specific site. Although 9 primer sets were used for the amplification, producing 20 values of σ_v, two broadly similar groups can be determined to simplify matters. σ_v is similarly large over the DNA sequence at the extremities of the HVR1 (between sites 24-272 and 417-679), and smaller for the central region between sites 273-416. Thus following chapter 3 the former regions are designated 'OR' (outer region), the latter 'MR' (middle region), and subsequent analyses are performed independently on these two datasets.

4.2.3 Hotspots

The null hypothesis (H_o) that *post mortem* damaged sites were randomly distributed across each region (OR and MR) was tested through comparison with the modified expected Poisson distribution of Heyer *et al.* (2001) utilised in chapter 3. The probability of sequence sites having exactly X substitutions, P(X), is

$$P(X)=e^{-\lambda}\lambda^{x}/X!$$

in which case Poisson parameter λ refers to the observed density of mutations. To estimate the expected count, LP(X), of sequence sites for each category, given the random distribution λ and the length L of the sequence, we multiply the Poisson probability P(X) by the number of sequence sites L, such that

$$LP(X)=L(e^{-\lambda}\lambda^{x}/X!)$$

A Chi-squared goodness of fit test was applied to the observed and expected results to determine whether the data rejects the hypothesis that *post mortem* damage is randomly distributed.

4.2.4 *In vivo* hotspots

In vivo site mutation rates were calculated as the average mutation rates of two maximum likelihood (ML) and two maximum *aposteriori* (MAP) mitochondrial genealogies (Shapiro *et al.* unpublished data) to provide a comparison with the *post mortem* damage spectrum. The sample genealogies were generated from 172 bison sequences using Markov chain Monte Carlo (MCMC), and the pairs of trees represent the MAP (generated using BEAST v1.1, Drummond and Rambaut 2003) and ML (generated using PAUP*, V4.0b10, Swofford 2002) estimates from two independent replicate MCMC analyses. Using the software MacClade (Sinauer Associates, Mass, USA), the minimum number of mutations at each site, given each of the four fixed trees, was calculated. This maximum parsimony estimate is necessarily a lower limit of the number of changes at a given site, and represents a conservative estimate of rate variation among sites, as fast evolving sites will typically have experienced many

reversions in a large genealogy that will not be detected by parsimony reconstruction. The presence of mutation hotspots was examined using the above method. Sites identified as hypervariable in three previous studies of *Bos taurus* mitochondrial DNA were compared with the damage and mutation data. As the methods and datasets used to generate the data were very different between the previous studies, the quantitative results are not directly comparable and sites were simply designated hotspot or not-hotspot status. For example, Wu *et al.* (2000) use a familial method of real time observations, whereas Cymbron *et al.* (1999) and Troy *et al.* (2001) use phylogenetic methods.

4.2.5 HVR1 sequence conservation

The inter-specific sequence variation within a conserved mitochondrial element was investigated in order to provide further insights into the possibility of structural protection of DNA from transition-inducing damage. Studies on HVR1 sequence conservation have previously focussed on the 60bp ETAS (extended Terminator Associated Sequence – TAS) elements (Sbisà *et al.* 1997). However, Ohno *et al.* (1991) have demonstrated a more conserved region of HVR sequence, both within humans, and between human, mice, rat, cow and pig DNA. This area, the Mt5 putative protein-binding site (positioned between Cambridge Reference Sequence (CRS: Anderson *et al.* 1981) sites 16194-16208 inclusive) has also been observed to receive *post mortem* damage at unusually low rates (chapter 3). An alignment of mammal mtDNA sequences was generated by eye in order to identify what we believe to be Mt5 analogues in 33 species, representing most of the eutherian orders, as well as several marsupials and a monotreme. Although examined, alignments of avian (Randi and Lucchini 1998) and teleost (Lee *et al.* 1995) mitochondrial sequences did not contain any close homology.

4.2.6 Source of modern Mt5 sequences analysed

The sequences used are as follows:

Homo sapiens (Human) CRS (Anderson *et al*. 1981); *Pan troglodytes* (Chimpanzee1) X93335 (Arnason *et al*. 1996b); *Pan troglodytes* (Chimpanzee 2) NC_001643 (Horai *et al*. 1995); *Pongo pygmaeus abelii* (Sumatran Orangutan) NC_002083 (Xu and Arnason 1996); *Pongo pygmaeus* (Orangutan) NC_001646 (Horai *et al*. 1995); *Hylobates lar* (White haired Gibbon) NC_002082 (Arnason *et al*. 1996a); *Bison bison* (Shapiro and Cooper, unpublished data); *Bos taurus* (Cow) J01394 (Anderson *et al*. 1982); *Sus scrofa* (Pig) AJ002189 (Ursing and Arnason 1998); *Equus asinus* (Donkey) NC_001788 (Xu *et al*. 1996); *Equus caballus* (Horse) NC_001640 (Xu and Arnason, 1994); *Ceratotherium simum* (White Rhino) NC_001808 (Xu and Arnason 1997); *Rhinoceros unicornis* (Indian Rhino) NC_001779 (Xu and Arnason 1997); *Halichoerus grypus* (Grey seal) X72004 (Arnason *et al*. 1993); *Phoca vitulina* (Harbour seal) NC_001325 (Arnason and Johnsson, 1992); *Canis familiaris* (Domestic dog) NC_002008 (Kim *et al*. 1998); *Felis catus* (Domestic cat) NC_001700 (Lopez *et al*. 1996); *Manis tetradactyla* (Long-tailed pangolin) AJ421454 (Arnason *et al*. 2002); *Artibeus jamaicensis* (Jamaican fruit-eating bat) NC_002009 (Pumo *et al*. 1998); *Dugong dugon* (Dugong) NC_003314 (Arnason *et al*. 2002); *Loxodonta africana* (African Savannah Elephant) NC_000934 (Hauf *et al*. 2000); *Orycteropus afer* (Aardvark) NC_002078 (Arnason *et al*. 1999); *Oryctolagus cuniculus* (Rabbit) NC_001913 (Gissi *et al*. 1998); *Rattus norvegicus* (Norway Rat) NC_001665 (Gadaleta *et al*. 1989); *Mus musculus* (House Mouse) NC_001569 (Slott Jr *et al*. 1983); *Erinaceus europaeus* (Western European Hedgehog) NC_002080 (Krettek *et al*. 1995); *Tamandua tetradactyla* (Southern Tamandua) AJ421450 (Arnason *et al*. 2002); *Dasypus novemcinctus* (Nine-banded Armadillo) NC_001821 (Arnason *et al*.

1997); *Balaenoptera musculus* (Blue Whale) NC_001601 (Arnason and Gullenberg, 1993); *Macropus robustus* (Wallaroo) NC_001794 (Janke *et al*. 1997); *Isododon macrourus* (Northern Brown Bandicoot) NC_002746 (Phillips *et al*. 2001); *Didelphis virginiana* (North American Possum) NC_001610 (Janke *et al*. 1994); *Ornithorhynchus anatinus* (Platypus) NC_000891 (Janke *et al*. 1996).

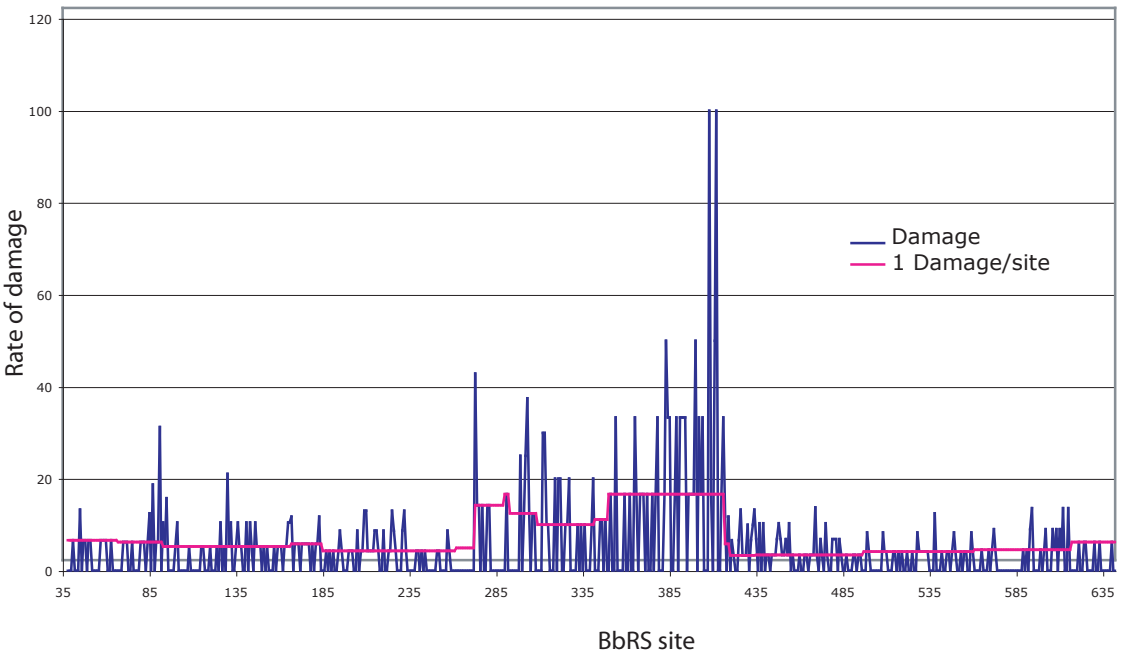
Using a recent mammalian phylogentic tree (Murphy *et al*. 2001), the minimum number of mutations at each site within the Mt5 sequence, and its surrounding bases were calculated for comparison with the mutation and damage rates estimated using the bison data sets.

4.3 Results

679 bison clones were analysed, from 81 initial PCR reactions, producing 135,221 bases of mitochondrial DNA Light (L) strand data. The spectra of observed damage types within the clones of the bison samples is 153/411(37.2%) Type 1 transitions, 244/411(59.4%) Type 2 transitions and only 14 (3.4%) other mutations. This overall bias towards transitions, and specifically towards the more biochemically likely Type 2 transitions (Lindahl 1979; Karran and Lindahl 1980) is expected and has been seen in several other aDNA studies (*c.f.* Hofreiter *et al.* 2001a; Hansen *et al.* 2001; chapter 2). No two samples yielded the same DNA sequence, thus intra-sample contamination is unlikely. Full details of the sequences will be published elsewhere (Shapiro *et al. in preparation*). The variation in *post mortem* damage relative rate across the bison HVR1 is shown in Figure 4.1. Similarly, the variation in modern bison mutation rates is shown in Figure 4.2. The different bison phylogenetic trees that were analysed to generate the estimates of modern mutational rates per site gave almost identical results. The *post mortem* damage distribution initially appears to be over-dispersed with certain sites over represented, notably BbRS sites 273, 303, 312,313, 383, 400, 408, 411 and 412. A random distribution of *post mortem* damage can be strongly rejected for the MR (χ^2 p-value of $p < 0.00$), though not for the OR ($p = 0.20$), demonstrating that at least in the MR sites are not being damaged at random. Similarly, the random distribution of homoplasies in the modern data set can also be rejected at $p<0.01$.

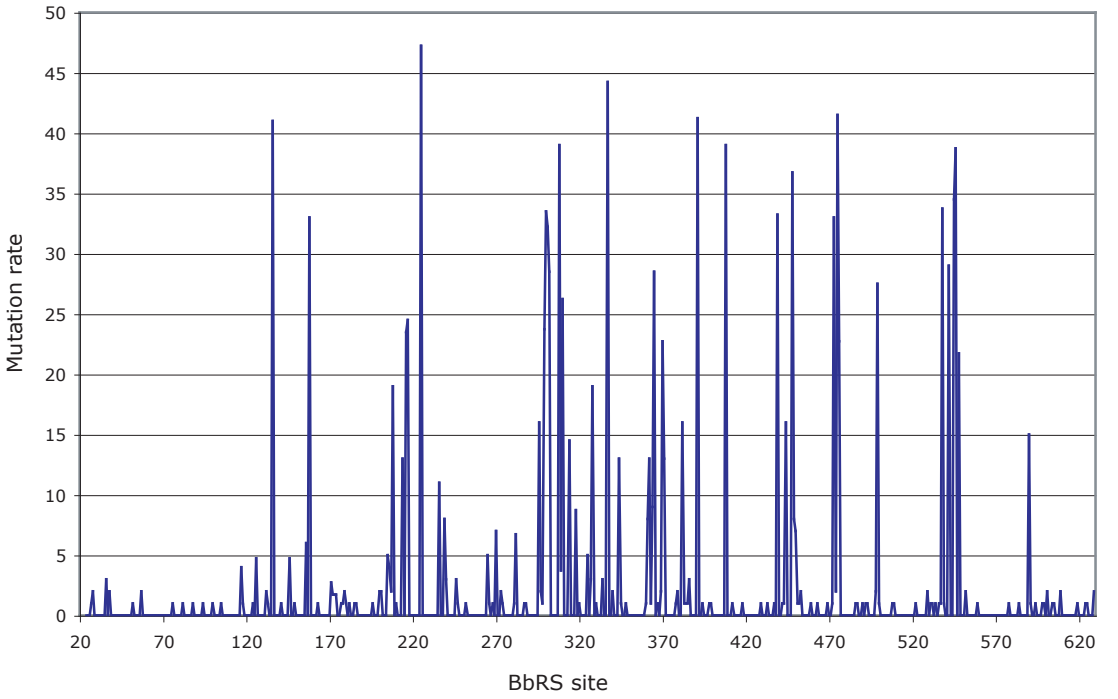
Although the different ranges of measurements of damage (0-48 homoplasies) and mutation rates (0-6 transitions) make direct comparison difficult, the majority of sites within the bison control region show similarly low rates of *post mortem* damage and *in vivo* mutation rates. However, as in the previous study on human mtDNA, some discrepancies exist. In particular, at least 15/144 sites within the MR exhibit obviously

Figure 4.1: Variation in relative damage rates across the Bison Control Region BbRS sites 24-629.



Relative damage rates calculated as number of damage events / number of individual amplifications at each specific site. To provide an idea of the resolution at each site (which varies due to the differing number of PCR amplifications used to cover the Control Region), the rate corresponding to 1 damage event at each site is indicated with the pink line. Site numbering is with reference to the *Bos bison* Reference sequence (BbRS), as in text.

Figure 4.2: Average mutation rate across Bison Control Region.



Mutation rate calculated as the average mutation rates of two pairs of mitochondrial genealogies (for details see text). Site numbering is as above.

conflicting rates, where sites with high values (measured as over half of the maximum rate observed) in one dataset were associated with rates of zero in the other dataset. It should be noted that this disagreement lies not only between the damage and mutational data sets, but also between the sites of mutational hotspots identified here and previously (Table 4.3). For full details of site mutation and damage rates refer to Appendix 1.

With the notable exception of the blue whale, tamadua and armadillo sequences examined (possibly due to alignment error), sequences remarkably similar to the human Mt5 could be identified among all eutherian mammals, and similarities could be seen in the marsupial and monotreme sequences (Fig 4.3). When sequences are compared with phylogenetic position, 2 clear areas of homoplasies can be identified, CRS sites 16197/16198 and 16206/207 (Table 4.4). Interestingly, both the rates of mutation and damage across the complete Mt5 region are much lower than the amount of inter-specific homoplasies observed. Finally we note that using this alignment, the region of conserved sequence (the original criteria used to define the limits of Mt5 in humans) (Ohno *et al.* 1991) can be extended by several bases, to incorporate sites 16193-16211.

Table 4.3 *In vivo* mutation and *post mortem* DNA damage rates at select sites within the 'Middle Region' (MR) of the Bison Control Region.

Site[1]	Mutation[2]	Damage[3]	*Bos taurus*[4]
273	2	3	JW00
293	0	0	TC99
300	34	0	TC99, CT01
301	32	1	TC99, JW00, CT01
308	39	0	TC99, JW00, CT01
309	4	0	TC99, CT01
318	9	0	TC99
320	1	0	TC99
321	0	2	TC99
326	0	1	TC99, CT01
331	0	0	TC99
337	44	0	TC99, CT01
340	2	1	TC99
347	0	0	TC99
361	8	0	TC99
365	29	2	TC99, CT01
373	0	0	CT01
380	0	0	JW00
383	1	3	TC99, CT01
384	1	2	TC99
386	3	0	TC99, JW00
390	0	0	TC99
408	39	6	TC99

[1]*Site Bos bison* Reference Sequence (BbRS) position [2]*Mutation* In vivo mutation rate calculated as average number of homoplasies on 2 pairs of phylogenetic trees [3]*Damage* Observed number of damage events [4]*Bos taurus* Sites identified as hotspots in 3 studies on *Bos taurus*. Studies are as follows TC99: Cymbron *et al*. (1999), JW00: Wu *et al*. (2000), CT01: Troy *et al*. (2001).

Figure 4.3: Aligned mammalian mitochondrial sequences believed to be homologous to the human Mt5 element.

```
Human          16181 AAAACCCCCTCCC--CATGCTTA-CAAG-CAAGTAC--AGCAATCAAC
Chimpanzee1    15603 ......TT.A...--.....-....-..C.C..--.A........
Chimpanzee2    15596 ......TT.A.T.--.....-....-..C.C..--.A........
S. Orangutan   15639 TCG...T..A.A.--.CC.....-....-.....--CC.CCCATG.
Orangutan      15629 CC....T..AAA.--.CC.....-.....GG--GAAGC.TT.A
Gibbon         15598 T.C.AA.GAC...AA.....-....-....C..--CAGC.CATCT
Cow            16013 ...TTATATG...--.....A..-T...-.....--TTA.TC.CTA
Bison            264 ...TTATATG...--.....A..-T...-.....--TTA.TC.CTA
Pig            15599 ..TTG.TAT....--.....A..-T....-..T....--.TATTATT.T
Donkey         15590 ...TTGTT.A...--....AA..AT...-..T....--.TA.TATT.T
Horse          15595 ...TTGT.TG...--....AA..AT...-..T....--.TA.TATC.T
White Rhino    15582 .C.TTGTTTA...--.....A..-T....-..T....--.TT.TATT.T
Indian Rhino   15570 ...TTGTTTG...--.....A..-T....-..T....--.TT.TATT.T
Grey seal      16486 T..TGGTTTG...--.....A..-T....-..T....--.TG..ATGGT
Harbour seal   16515 T..TGGTTTG...--.....A..-T....-..T....--.TATGAAGTG
Domestic dog   15552 T..TGGTTTG...--.....A..-T....-..T....--.TA.TATT.T
Domestic Cat   16538 ..TTG.TAG....--....AA..TT...-..T....--..G.G.TT.T
Pangolin       15549 T..TGATTAA...--....AA..TT...-..T....--.TT.CATT.T
Fruit Bat      15500 ..T.TAATG...T--....AA..TT...-.....A--.TT..GTT.A
Dugong         15527 .C.CTA.TTA...--.....A..-T....-.C....--G.T.GGATT.
A. Elephant    15534 .C.TTATTTA...--.........-T...-.....--T.TTTAACTA
Aardvark       15566 .C.TTAT.TG...--.........-T...-.....--.TA..ATCCA
Rabbit         15599 ..TT..T.A....--....AA..AT...-.T....--.TT.C.GCTT
Norway Rat     15455 ...TTATTT.....--..A..A..-T....-..T...A--TAT.TATCTA
House Mouse    15473 ...CTATTT.....--..A..A..-T....-.T....--.TT..ATC.A
Hedgehog       15567 T.TTTATGTAATT--.TA..A..-T....-..T....--.TT..A.TCT
Tamandua       15511 C.TTTATTTA...--....AA..-TTCT-T....--.TATTATT.T
Armadillo      16579 .C.TATAATA.AT--A.CC.ACC-...A-..TA...--TTTC.CT..A
Blue Whale       525 TTCTTAA.AAA.T--T.CA.A.G-....-.GTC...--.T.CCGGTGA
Wallaroo       15754 ......TTTA...--.TA..A..-T..A-.....--.AAT.A.CTA
Bandicoot      15640 ..TTTATTTA...--.TA..A..-T....-.-TA...--..T.CAT..A
N.Am Possum    16217 .C.TAATTTATAT--T..AAGATAT..CG..TA...--GTGT..ACG.
Platypus        1753 TT..TATTACATA--T..ATA..TT...-.....--.TT.TATGTA
```

The position of the first nucleotide of each sequence, with respect to the specific sequences used for the alignment is included in the alignment.

Table 4.4: Mutation and damage rates at the *Mt5* locus and neighbouring sites.

CRS[1]	BbRS[2]	Parsimony[3]	Damage[4]	Mutation[5]
193	277	0	1	0
194	278	2	0	0
195	279	3	0	0
196	280	0	1	0
197	281	4	1	1
198	282	3	0	7
199	283	0	0	0
200	284	0	0	0
201	285	1	0	0
202	286	0	0	0
203	287	0	0	1
204	288	0	0	1
205	289	0	0	0
206	290	3	0	0
207	291	5	1	0
208	292	0	0	0
209	293	2	0	0
210	294	1	0	0
211	295	3	0	0

[1]*CRS* Site position with relation to the Cambridge Reference Sequence (CRS) (Anderson *et al*. 1981). The human Mt5 locus is highlighted in red. [2]*BbRS* Site position with relation to the *Bos bison* Reference Sequence (BbRS). [3]*Parsimony* Number of mutations at site on parsimony tree of examined mammalian orders. [4]*Damage* Number of observed *post mortem* damage events among cloned ancient sequences at site. [5]*Mutation* Number of observed mutations among modern DNA sequences at site.

4.4 Discussion

The spectra of damage within cloned ancient bison sequences is consistent with previous studies on aDNA damage (*c.f.* Hofreiter *et al.* 2001a; chapter 2). In addition, the lack of evidence for inter-sample contamination suggests the data accurately represent *post mortem* damage variation among the control region of bison. As in previous studies (chapter 3) strong evidence is found for *post mortem* damage hotspots in at least part of the bison control region (MR). Although there is no statistical evidence to support the existence of such hotspots in the OR, there are also sites within the region that appear to receive a disproportionate amount of the damage, such as BbRS sites 91 and 130 (see Appendix 1).

In agreement with the previous study (chapter 3) the site-specific mutation and damage rates at most sites correlate. This suggests that previous data of *post mortem* damage hotspots obtained using human mtDNA do not result from contamination. In turn, the existence of human *post mortem* mitochondrial damage hotspots supports the notion of *in vivo* mitochondrial mutational hotspots. Such hotspots are a more likely explanation than mitochondrial recombination (*e.g.* Hagelberg 2003) for homoplasies observed on human mitochondrial DNA phylogenetic trees. The results also suggest that the secondary structure of mitochondrial DNA somehow influences the propensity of certain nucleotides to be damaged (Meyer *et al.* 1999; chapter 3).

Also in agreement with chapter 3, some sites are found with contradictory mutation and damage rates. Interestingly, these differences are not only seen between the *post mortem* and mutational data set generated from bison in this study, but also between this data-set and hotspots identified in the previous studies on bovid sequences (Cymbron *et al.* 1999; Wu *et al.* 2000; Troy *et al.* 2001). It is possible that a proportion of these inconsistencies have resulted through sampling stochasticity. However, as contradictions are also

common among studies examining modern human rates (Bandelt *et al*. 2001; Bandelt *et al*. 2002), it appears that a large proportion of the disagreement may be explained through the phylogenetic models used in the estimation of mutation rates (*c.f.* Meyer and von Haeseler 2003).

An additional explanation for the discrepancy may lie with small differences among the interactions that lead to mutation *in vivo* and miscoding lesions *post mortem*. While most sequence conservation may arise through a structural limitation of damage, a number of mammalian HVR1 elements have been identified that potentially perform a functional role. These include the two ETAS elements (Sbisà *et al*. 1997) as well as the Mt5 element (Ohno *et al*. 1991). If *in vivo* mutation within such elements results from selection as well as structural constraints (as can be expected of functional sequences), differences between the mutational and damage data may occur.

Such differences are not observed at the bison homologue of the human Mt5 element. Mt5 has been proposed as the binding site for an Mt5-responsive factor produced in response to viral infection, due to sequence similarity with the 5' flanking region of a nuclear interferon-α class 2 gene (Ohno *et al.* 1991). A simple explanation for the observed lack of variation is selection for the binding site *in vivo* and continued protection offered by the bound protein *post mortem*. The relative proximity of Mt5 to the origins of mitochondrial replication, its overall conservation between eutherian species, and the flexibility of some binding proteins recognition sequence suggest this may be possible. Nevertheless, the validity of this explanation relies on the identification of the sequence as a protein-binding site, that the protein survives *post mortem*, and that it can protect the DNA from hydrolytic deamination.

It has been predicted that Mt5 sites that vary inter-specifically should also exhibit elevated mutation and damage rates (chapter 3). The data do not demonstrate the

proposed correlation, although the region is conserved between species. This could be explained through the binding of a protein that confers enough sequence protection to limit the overall ability of the sequence to mutate *in vivo*, but also contains enough structural variation between species to allow for the small amount of sequence variation observed. Furthermore, the loss of requirement for the protein, or a significant modification of the protein's structure, could explain why several unrelated mammals do not appear to contain the sequence homologue.

An alternative explanation to protein-DNA binding conferring *post mortem* protection at Mt5 is direct structural protection conferred through DNA secondary sequence conformation. This would suggest that selection may act upon the surrounding HVR1 DNA sequence to mutate towards/remain at a base composition that confers a protective secondary structural conformation on the Mt5 region. Selection on DNA to confer a protective effect through base composition has been proposed to explain other phenomena, including the mitigation of oxidative damage through cathodic protection of essential chromosomal domains (Heller 2000). For dissolved DNA duplexes in homogenous solutions, holes (electron vacancies) injected by oxidants are transferred by hopping between CG base pairs. The injected holes subsequently oxidize a remote G-site, particularly those comprising sequences of multiple GC base pairs (*c.f.* Heller 2000). In this case the selection includes the composition of the sequence to allow both the transfer of the holes and the final damage event.

The implication of hypotheses that suggest non-neutral evolution in supposedly-neutral sequences (such as the mitochondrial Control Region) are important with regard to the importance place on selection pressures in DNA-based phylogenetic reconstruction. Therefore further investigation into aDNA hotspots, and the characterisation of the nature of conserved elements (such as Mt5) may be warranted.

4.5 Acknowledgements

The authors would like to thank Hans-Jurgen Bandelt and Mike Bunce for useful discussion. M.T.P.G. and A.C. are supported by the Wellcome Trust. B.S. is supported by the Rhodes Trust. A.D. is supported by NERC.

Chapter 5

The suitability of bones and teeth as sources of DNA in ancient human DNA analyses.

Modified from the version in preparation for submission to

'Proceedings of the Royal Society of London, Series B'

M Thomas P Gilbert[1], Lars Rudbeck[2], Eske Willerslev[3], Anders J Hansen[3], Colin Smith[4,5], Kirsty Penkman [5], Kurt Prangenberg [5,6], Christina Nielsen-Marsh[5], Miranda Jans[5], Paul Arthur[7], Niels Lynnerup[8], Gordon Turner-Walker[9,10], Martin Biddle[11], Birthe Kjølbye-Biddle[11], Alan Cooper[1], Matthew Collins[5]

[1]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, University of Oxford, South Parks Rd, Oxford OX1 3PS, United Kingdom
[2]Research Laboratory, Institute of Forensic Medicine, University of Copenhagen, Denmark
[3]Department of Evolutionary Biology, Zoological Institute, University of Copenhagen, Denmark
[4]Current address : Department of Palaeobiology, Museo Nacional de Ciencias Naturales (CSIC),C/José Gutiérrez Abascal. 228006, Madrid, Spain
[5]Fossil Fuels and Environmental Geochemistry, (Postgraduate Institute); NRG, Drummond Building, University of Newcastle, Newcastle upon Tyne NE1 7RU, United Kingdom
[6]Institut für Geowissenschaften, Sigwartstrasse 10, 72076 Tübingen, Germany
[7] Dipartimento di Beni Culturali,Via D. Birago, 64,University of Lecce,73100 Lecce, Italy
[8]Laboratory of Biological Anthropology, Institute of Forensic Medicine, University of Copenhagen, Denmark
[9]Institutt for arkeologi og kulturhistorie, NTNU, Vitenskapsmuseet, 7491 Trondheim, Norway
[10]Current address - Institute for Cultural Heritage Conservation, National Yunlin University of Science and Technology, 123 University Road Sec.3, Touliu, Yunlin 640, Taiwan
[11]Hertford College, University of Oxford, Oxford OX1 3BW, United Kingdom

Summary

The majority of ancient DNA studies on human specimens have utilised teeth and bone as a source of genetic material. In addition to the abundance of such samples, they are perceived to be a source of uncontaminated endogenous DNA. In this paper the levels of endogenous contamination* (*i.e.* present within the sample prior to sampling for the DNA analysis) are assessed within a range of European archaeological human bone and teeth samples. A series of experiments are also undertaken to artificially contaminate samples that have previously been assessed for biochemical and histological preservation. The findings demonstrate two important issues: a) although teeth are more resilient to contamination than bone, both are readily contaminated through handling or washing, and b) once contaminated in this way, both are difficult (if not impossible) to decontaminate. The specific biochemical characteristics of individual samples reveal that contamination is directly correlated to sample preservation, and in particular, levels of microbial attack and related increases in sample porosity. The implication is that the vast majority of the known archaeological record has been deeply contaminated.

* The impetus behind this work was a series of incomprehensible aDNA results on human bone from Repton, Derbyshire, England. See Biddle and Kjolbye-Biddle, 2001, 45-96.

5.1 Introduction

Bones and teeth are normally the longest lasting physical evidence of human or animal presence at an archaeological site are, and also the most widely used sources of samples for ancient DNA (aDNA) studies. Post-hoc explanations of their suitability as a source of ancient DNA have identified retarded rates of decomposition, arising from adsorption of DNA to hydroxyapatite (Lindahl 1993a), low water content (Hummel and Hermann 1994), 'mummification' of individual cells (Bell *et al.*, 2002), and physical exclusion of microbes and external contaminants (Hummel and Hermann 1994). Recently, an awareness of sample handling as a source of contamination has led researchers to investigate teeth as an aDNA source. One hypothetical benefit is protection conferred by enamel (*c.f.* Oota *et al.* 1995). Additionally, although histological studies identify higher numbers of DNA containing cells per unit area of bone than teeth (Currey *et al.* 2002), several studies have reported better DNA yields in teeth than bone (Kurosaki *et al.* 1993; Oota *et al.* 1995; Meyer *et al.* 2000; di Benedetto *et al.* 2001; Montiel *et al.* 2001).

Richards *et al.* (1995) have argued convincingly that contamination, not DNA preservation, is the greatest problem facing the field, although the two are evidently linked – sparse, damaged endogenous DNA is less likely to be amplified than modern contamination. Although it is known that teeth and bone may become contaminated prior to aDNA extraction (*c.f.* Richards *et al.* 1995), current techniques used to decontaminate specimens - the application of bleach, exposure to UV light, and grinding or shot-blasting - reflect a belief that contamination is concentrated in the outer surface of the material. Protocols designed to limit contamination stress the prevention of contact between samples and previously amplified DNA (amplicons) (*c.f.* Handt *et al.* 1994a; Cooper and Poinar 2000). Nevertheless, even when strict

protocols are followed contaminants are frequently observed. For example human DNA has been reported from cave bear (Hofreiter *et al*. 2001b), 500-year-old pig samples (Richards *et al*. 1995) and 109 out of 168 relatively recent fox teeth (Wandeler *et al*. 2003). More seriously, several studies report significant numbers of human remains contaminated with multiple human sequences (*c.f*. Handt *et al*. 1996; Kolman and Tuross 2000). Obviously, decontamination methods are not 100% efficient, and contamination remains a serious threat to the validity of ancient DNA studies, particularly on human templates.

In this study a detailed examination of sample contamination is undertaken. A series of artificial contamination experiments are utilised to directly compare contamination within bones and teeth, and to assess a new method of dentine extraction from teeth. Levels of endogenous contamination (*i.e*. human contaminant DNA sequences present in samples prior to this study) are assessed within a range of European archaeological specimens. A subset of samples that have previously been assayed for 18 measures of biochemical and physical preservation are used to correlate preservation and contamination. The results are used to identify techniques for pre-selecting uncontaminated samples.

5.2 Materials and Methods

5.2.1 Samples

Human bone and teeth samples were examined from five archaeological sites.

5.2.1.1 Trondheim

Sixty-six teeth and five femur samples were taken from 33 different individuals (Table 5.1) excavated at Trondheim, Norway, in the late 1970s (Turner-Walker *et al.* 2002). Samples were washed post-excavation and have subsequently been directly handled by several individuals. Turner-Walker *et al.* (2002) used these samples as reference material for porosimetry analysis due to the unusually good preservation (in particular lack of microbial attack and high collagen content).

5.2.1.2 Matera

Twenty-four teeth and nine femur samples were taken from twelve individuals excavated at the church of Santa Lucia alle Malve, Matera, Italy (Bruno 2001) (Table 5.2). In a study that investigated more than 200 bones from 43 sites across Europe, this site was identified as a site where human bones were either well preserved or the predominant mode of diagenesis was microbially mediated (Kars and Kars 2002, Smith *et al.* 2002, Jans *et al.* 2002; Table 5.2, Appendix 2). Additionally, as part of this study the extent of aspartic acid (Asx) racemization was determined in one tooth from each individual using the extraction method of Poinar *et al.* (1996) and the analytical method of Kaufman and Manley (1998). Although these samples have been handled since excavation, they have not been washed. All Matera

Table 5.1: Trondheim samples - morphological preservation, presence of non-endogenous human DNA contamination, and treatment regimes.

Sample*	Skeleton†	Tissue	Details‡	Silicone§	Treatment¶	Amplify⇓	Sequence match**	Contamination††
Tg292	686	Tooth	Sealed	Yes	K	No	n/a	No
Tg330	686	Tooth	Sealed	Yes	E	Yes	n/a	No
Tg302	754	Tooth	Sealed	Yes	K	Yes	No	Yes
Tg352	754	Tooth	Cavity	Yes	E	Yes	No	No
Tg321	199(6)8/8 023	Tooth	Worn Crown	Yes	K	Yes	Yes	No
Tg354	199(6)8/8 023	Tooth	Sealed	Yes	C	Yes	Yes	No
Tg307	1996/8 193	Tooth	Cavity	Yes	C	No	n/a	No
Tg313	1996/8 193	Tooth	Worn Crown	Yes	K	Yes	n/a	No
Tg304	1996/8 456	Tooth	Cavity	Yes	B	No	n/a	No
Tg340	1996/8 456	Tooth	Worn Crown	Yes	K	Yes	n/a	Yes
Tg326	1996/8 466	Tooth	Sealed	Yes	K	Yes	n/a	No
Tg347	1996/8 466	Tooth	Worn Crown	No	A	No	n/a	No
Tg311	1996/8(9) 383	Tooth	Sealed	Yes	K	No	n/a	No
Tg344	1996/8(9) 383	Tooth	Sealed	No	A	Yes	n/a	No
Tg341	5/S017	Tooth	Worn Crown	Yes	K	Yes	No	Yes
Tg353	5/S017	Tooth	Worn Crown	No	A	Yes	No	Yes
Tg373	AM68	Femur	n/a	No	J	Yes	Yes	No
Tg379	AM68	Tooth	Worn Crown	Yes	K	No	Yes	No
Tg384	AM68	Tooth	Sealed	No	J	Yes	Yes	No
Tg317	AS10	Tooth	Worn Crown	Yes	K	Yes	Yes	No
Tg358	AS10	Tooth	Worn Crown	No	D	Yes	Yes	No
Tg309	AS159	Tooth	Worn Crown	Yes	C	No	n/a	No
Tg318	AS159	Tooth	Worn Crown	Yes	K	Yes	n/a	No
Tg298	AV36	Tooth	Sealed	Yes	K	Yes	Yes	Yes
Tg336	AV36	Tooth	Worn Crown	Yes	E	Yes	Yes	No
Tg300	N805679	Tooth	Sealed	Yes	K	No	n/a	No
Tg338	N805679	Tooth	Worn Crown	Yes	E	No	n/a	No
Tg314	N80662	Tooth	Worn Crown	Yes	K	No	n/a	No
Tg345	N80662	Tooth	Worn Crown	No	D	Yes	n/a	Yes
Tg306	N82400	Tooth	Sealed	Yes	C	No	n/a	No
Tg312	N82400	Tooth	Sealed	Yes	K	Yes	n/a	Yes
Tg328	SK169	Tooth	Sealed	Yes	B	Yes	n/a	Yes
Tg369	SK192	Femur	n/a	No	F	Yes	Yes	No
Tg375	SK192	Tooth	Worn Crown	Yes	K	No	Yes	Yes
Tg380	SK192	Tooth	Sealed	No	F	Yes	Yes	No
Tg370	SK280	Femur	n/a	No	G	Yes	No	Yes
Tg376	SK280	Tooth	Worn Crown	Yes	K	Yes	No	No
Tg381	SK280	Tooth	Sealed	No	G	Yes	No	Yes
Tg371	SK352	Femur	n/a	No	H	Yes	No	No
Tg377	SK352	Tooth	Worn Crown	Yes	K	Yes	No	Yes
Tg382	SK352	Tooth	Worn Crown	Yes	H	Yes	No	Yes
Tg372	SO11	Femur	n/a	No	I	Yes	Yes	No
Tg378	SO11	Tooth	Worn Crown	Yes	K	Yes	Yes	Yes
Tg383	SO11	Tooth	Sealed	No	I	No	Yes	No
Tg310	SO38	Tooth	Worn Crown	Yes	C	No	n/a	No

Sample	Skeleton	Tissue	Details	Silicone	Treatment	Amplify	Sequence match	Contamination
Tg323	SO38	Tooth	Worn Crown	Yes	K	Yes	n/a	Yes
Tg316	SO51	Tooth	Worn Crown	Yes	K	Yes	Yes	No
Tg356	SO51	Tooth	Worn Crown	No	D	Yes	Yes	No
Tg322	SO54	Tooth	Worn Crown	Yes	K	Yes	No	Yes
Tg359	SO54	Tooth	Sealed	No	A	Yes	No	No
Tg293	SO68	Tooth	Sealed	Yes	K	Yes	No	No
Tg331	SO68	Tooth	Worn Crown	Yes	E	Yes	No	No
Tg301	SO69	Tooth	Sealed	Yes	K	Yes	No	Yes
Tg350	SO69	Tooth	Sealed	Yes	E	Yes	No	Yes
Tg305	T11792	Tooth	Sealed	Yes	B	No	n/a	No
Tg351	T11792	Tooth	Sealed	Yes	K	Yes	n/a	Yes
Tg325	T19564L	Tooth	Sealed	Yes	K	Yes	Yes	No
Tg346	T19564L	Tooth	Worn Crown	No	D	Yes	Yes	No
Tg319	T19564M	Tooth	Sealed	Yes	K	Yes	No	No
Tg349	T19564M	Tooth	Cavity	No	C	Yes	No	Yes
Tg296	T19564N	Tooth	Sealed	Yes	K	No	n/a	No
Tg334	T19564N	Tooth	Sealed	Yes	E	Yes	n/a	No
Tg342	T20769a	Tooth	Worn Crown	Yes	K	Yes	n/a	No
Tg329	T21887	Tooth	Sealed	Yes	B	Yes	No	Yes
Tg339	T21887	Tooth	Worn Crown	Yes	K	Yes	No	Yes
Tg297	T22317a	Tooth	Sealed	Yes	K	Yes	n/a	Yes
Tg335	T22317a	Tooth	Worn Crown	Yes	E	Yes	n/a	No
Tg327	T22317b	Tooth	Sealed	Yes	K	Yes	n/a	No
Tg361	T22317b	Tooth	Sealed	No	D	No	n/a	No
Tg294	T2703	Tooth	Sealed	Yes	K	Yes	No	Yes
Tg333	T2703	Tooth	Sealed	Yes	E	Yes	No	Yes

[*]*Sample* Authors' DNA extraction identification details. [†]*Skeleton* Original sample identification details. [‡]*Details* Morphological preservation of sample: 'Sealed' (teeth demonstrate no observable structural flaws), 'Worn Crown' (enamel on the crown is worn through to the dentine layer), 'Cavities' (teeth contain cavities). [§]*Silicone* Encasement of tooth with silicone prior to dentine extraction. [¶]*Treatment* Artificial contamination regime (*e.g.* handling, washing etc) undergone by sample prior to dentine extraction to investigate the susceptibility of differentially preserved samples to contamination (for full details see text). [⇓]*Amplify* Presence of amplifiable human mtDNA within DNA extract of sample. [**]*Sequence match* Whether different DNA sequences amplified from different samples extracted from the same skeleton match. [††]*Contamination* Presence of contaminant human DNA sequences within sample prior to this investigation as evidenced through the molecular cloning of PCR amplified DNA sequences.

Table 5.2: Matera samples - selected measures of biochemical and micro-morphological preservation indices of bone samples corresponding to each tooth, and presence of endogenous and artificial levels of contamination within bones and teeth.

Sample	Skeleton*	Tissue	Contam†	ΦX174‡	Histology§	Wedl¶	Collagen⇓	C:P**	Total Hg††	Hg‡‡	Asx§§
Tg458	SLM-3	Tooth	N	N/A	5	0	20.9	0.68	0.0853	0.021	0.09
Tg460	SLM-3	Tooth	N	N/A	5	0	20.9	0.68	0.0853	0.021	0.08
Tg425	SLM-4	Tooth	N	N/A	5	0	20.8	1.03	0.1535	0.0339	0.04
Tg453	SLM-4	Tooth	N	N/A	5	0	20.8	1.03	0.1535	0.0339	0.05
Tg447	SLM-5	Tooth	N	N/A	4	0	20.4	0.85	0.237	0.097	0.05
Tg448	SLM-5	Tooth	N	N/A	4	0	20.4	0.85	0.237	0.097	0.05
Tg449	SLM-6	Tooth	N	N/A	2	0	15.2	0.28	0.3117	0.1941	0.06
Tg450	SLM-6	Tooth	N	N/A	2	0	15.2	0.28	0.3117	0.1941	0.06
Tg438	SLM-11	Tooth	N	N/A	4	1	22.2	0.66	0.123	0.0652	0.05
Tg455	SLM-11	Tooth	?	N/A	4	1	22.2	0.66	0.123	0.0652	0.05
Tg446	SLM-7	Tooth	Y	N/A	3	1	13.1	0.29	0.2679	0.163	0.07
Tg461	SLM-7	Tooth	Y	N/A	3	1	13.1	0.29	0.2679	0.163	0.07
Tg423	SLM-8	Tooth	Y	N/A	2	1	9.5	0.21	0.3749	0.266	0.08
Tg435	SLM-8	Tooth	?	N/A	2	1	9.5	0.21	0.3749	0.266	0.06
Tg439	SLM-10	Tooth	Y	N/A	3	1	13.1	0.34	0.2211	0.1592	0.06
Tg426	SLM-10	Tooth	?	N/A	3	1	13.1	0.34	0.2211	0.1592	0.07
Tg454	SLM-12	Tooth	Y	N/A	2	0	13.9	0.23	0.3337	0.223	0.06
Tg452	SLM-14	Tooth	Y	N/A	2	1	7.8	0.31	0.3868	0.2401	0.07
Tg443	SLM-14	Tooth	?	N/A	2	1	7.8	0.31	0.3868	0.2401	0.06
Tg445	SLM-13	Tooth	Y	N/A	1	1	9	0.31	0.2285	0.1666	0.06
Tg457	SLM-1	Tooth	?	N/A	4	1	20.2	1.14	0.1874	0.055	0.06
Tg456	SLM-1	Tooth	?	N/A	4	1	20.2	1.14	0.1874	0.055	0.07
Tg451	SLM-9	Tooth	?	N/A	5	0	21.2	0.53	0.0826	0.032	0.06
Tg436	SLM-9	Tooth	?	N/A	5	0	21.2	0.53	0.0826	0.032	0.06
Tg497	SLM-5	Femur	N	Y	4	0	20.4	0.85	0.237	0.097	n/a
Tg498	SLM-6	Femur	Y	N	2	0	15.2	0.28	0.3117	0.1941	n/a
Tg499	SLM-7	Femur	Y	Y	3	1	13.1	0.29	0.2679	0.163	n/a
Tg500	SLM-8	Femur	Y	Y	2	1	9.5	0.21	0.3749	0.266	n/a
Tg501	SLM-9	Femur	Y	Y	5	0	21.2	0.53	0.0826	0.032	n/a
Tg502	SLM-10	Femur	Y	Y	3	1	13.1	0.34	0.2211	0.1592	n/a
Tg503	SLM-11	Femur	N	Y	4	1	22.2	0.66	0.123	0.0652	n/a
Tg504	SLM-12	Femur	Y	Y	2	0	13.9	0.23	0.3337	0.223	n/a
Tg505	SLM-14	Femur	Y	Y	2	1	7.8	0.31	0.3868	0.2401	n/a

Skeleton* Archaeological specimen. *†Contam* Presence of endogenous contamination. *‡ΦX174* Presence of ΦX174 contaminant DNA in DNA extract after sample decontamination. §,¶,⇓,,††,‡‡ Measures of biochemical preservation of samples, taken from Kars and Kars (2002). In brief: *§Histology* Histology of bones from each individual where 0=minimum, 5=maximum preservation, *¶Wedl* Evidence for presence (1) or absence (0) or Wedl fungal tunnelling, *⇓Collagen* Percentage collagen remaining in specimen, ***C:P* Carbonate to phosphate ratio, *††Total Hg* Total pore volume as measured by mercury porosimetry, *‡‡Hg* Pore volume in the pore range 100nm-6μm as measured by mercury porosimetry, *§§Asx* Combined aspartic acid and asparagine amino acid D:L enantiomer ratios. The 0.034% racemization induced by hydrolysis (6M HCl, 6 hours @ 110°C; Poinar et al. 1996) was subtracted from all the reported data.

samples have been stored together, and are thought to have undergone similar amounts of human handling post-excavation.

5.2.1.3 York, Copenhagen and Aalborg

Eleven teeth were extracted from ten Viking individuals that were excavated at York by the York Archaeological Trust (Table 5.3). Twenty-six teeth were extracted from six individuals that were excavated in 1991 from an 18th century plague pit in Copenhagen (Ringboel Bitsch 1991) (Table 5.3). Nine teeth were extracted from one individual that was excavated from a 16th century cemetery in Aalborg, Denmark (unpublished record of Aalborg museum, Denmark) (Table 5.3). York, Copenhagen and Aalborg samples have not been subject to detailed biochemical screening, but all are less well preserved than the Matera and Trondheim material, and all have been washed and extensively handled.

A series of experiments was undertaken to a) quantify the levels of human contaminant DNA in samples prior to this investigation (endogenous contamination), and to b) examine the susceptibility of samples to further contamination by exogenous DNA - both human and naked DNA from the HaeIII digested bacteriophage ΦX174 that is commonly used as a molecular weight standard.

5.2.2 Ancient DNA extraction and amplification

DNA was extracted from samples following strict ancient DNA protocols in order to prevent sample contamination with previously amplified DNA (*c.f.* Cooper and Poinar 2000; chapter 3). One extraction blank (extraction with no tissue) was used for every four samples in order to monitor contaminants entering during the DNA

Table 5.3: Details of York, Copenhagen and Aalborg samples

Sample*	Details	Skeleton†	Tissue	Contamination‡	Sample	Details	Skeleton	Tissue	Contamination
Tg265	York viking	1841	Tooth	No	Tg214	Copenhagen	AO	Tooth	n/a
Tg271	York viking	1877	Tooth	Yes	Tg216	Copenhagen	AO	Tooth	n/a
Tg272	York viking	1877	Tooth	Yes	Tg217	Copenhagen	AO	Tooth	n/a
Tg277	York viking	1907	Tooth	Yes	Tg219	Copenhagen	AO	Tooth	n/a
Tg264	York viking	2123	Tooth	No	Tg220	Copenhagen	AO	Tooth	n/a
Tg276	York viking	2159	Tooth	No	Tg87	Copenhagen	AO	Tooth	n/a
Tg269	York viking	2408	Tooth	Yes	Tg240	Copenhagen	BD	Tooth	n/a
Tg267	York viking	6078	Tooth	No	Tg241	Copenhagen	BD	Tooth	n/a
Tg278	York viking	6097	Tooth	Yes	Tg242	Copenhagen	BD	Tooth	n/a
Tg270	York viking	6202	Tooth	Yes	Tg243	Copenhagen	BD	Tooth	n/a
Tg266	York viking	6417	Tooth	Yes	Tg244	Copenhagen	BD	Tooth	n/a
LR1a	Aalborg	AHM1086	Tooth	n/a	Tg247	Copenhagen	BD	Tooth	n/a
LR1b	Aalborg	AHM1086	Tooth	n/a	Tg248	Copenhagen	BD	Tooth	n/a
LR1c	Aalborg	AHM1086	Tooth	n/a	Tg249	Copenhagen	BD	Tooth	n/a
LR1d	Aalborg	AHM1086	Tooth	n/a	Tg250	Copenhagen	BD	Tooth	n/a
LR1e	Aalborg	AHM1086	Tooth	n/a	EA1	Copenhagen	AV1	Tooth	No
LR2a	Aalborg	AHM1086	Tooth	n/a	EA2	Copenhagen	AV1	Tooth	No
LR2b	Aalborg	AHM1086	Tooth	n/a	EA3	Copenhagen	CV	Tooth	Yes
LR2c	Aalborg	AHM1086	Tooth	n/a	EA4	Copenhagen	CV	Tooth	Yes
LR2d	Aalborg	AHM1086	Tooth	n/a	EA5	Copenhagen	CM	Tooth	No
Tg210	Copenhagen	AO	Tooth	n/a	EA6	Copenhagen	CM	Tooth	No
Tg211	Copenhagen	AO	Tooth	n/a	EA7	Copenhagen	CP	Tooth	Yes
Tg213	Copenhagen	AO	Tooth	n/a	EA8	Copenhagen	CP	Tooth	Yes

**Sample* Authors' DNA extraction identification details. †*Skeleton* Original sample identification details.
‡*Contamination* Presence of human contaminant DNA sequences within sample prior to this investigation as evidenced through the molecular cloning of PCR amplified DNA sequences.

extraction. DNA extractions from bone used 0.2 g bone powder, collected as in Barnes *et al*. 2002. Unless otherwise stated, all DNA was extracted from teeth following chapter 3. All PCRs were performed on each sample at least twice, using the polymerase enzyme Platinum Taq Hi-Fidelity (Invitrogen), and incorporating one PCR blank (*i.e*. containing no template DNA) to every 3 samples. PCRs for human mitochondrial DNA (mtDNA) used primers L16209-H16356 (Handt 1996) following chapter 3. DNA extracts that did not yield PCR products were screened for the presence of PCR inhibitors (a common phenomena in aDNA studies, Pääbo *et al*. 1988) through spiking PCRs containing amplifiable DNA, and monitoring any reduction in PCR success (following chapter 3). All amplified human PCR products were cloned. A minimum of twelve colonies were sequenced per cloned PCR product following chapter 3. Contaminated samples were identified as described in chapter 3.

The amount of ΦX174 DNA in samples that had been deliberately contaminated was determined using PCR. Primers ΦX1F (5'ctg ccg ttt tgg att taa cc) and ΦX1R (5'ttt gaa tgt tga cgg gat ga) amplify a product of 207 bp and primers ΦX3F (5'cat gac ctt tcc cat ctt gg) and ΦX3R (5'caa tgg aga aag acg gag ag) amplify a product of 129 bp. PCR conditions for both primer sets were: Enzyme activation 94ºCx1min30, 40 cycles of (94ºCx45 secs denaturation, 56ºCx45secs annealing, 68ºCx1min30 extension), final extension 68Cx10mins. The primers had previously been optimised on ΦX174 DNA in a separate laboratory. A proportion of the amplified ΦX174 products were cloned and sequenced to confirm the identification.

5.2.3 Assessment of bone preservation correlates with contamination

The detailed record of bone preservation of the Matera samples, and their similar treatment since excavation presents an ideal opportunity to investigate correlations

between contamination and preservation indices. Nine femur samples were taken (Table 5.2) and divided into two halves. DNA was directly extracted from one half to detect existing contamination. The second half of each sample was thoroughly handled to ensure maximum skin contact, then washed in room temperature tap water to simulate the cleaning methods commonly used at archaeological sites. Handling was undertaken by a volunteer (A.C.) with a mitochondrial haplotype that is unlikely to be found among European samples (HVR1 sequence differs from the Cambridge Reference Sequence (CRS; Anderson *et al.* 1982) by transitions as positions 16223, 16241, 16265, 16311 and 16343). The tap water was additionally spiked with 10^{-12} g/l *HaeIII* digested ΦX174 DNA (Sigma-Aldrich) to enable an investigation into the effect concentration of DNA has on sample contamination. To investigate whether water absorption might facilitate the penetration of DNA, samples were washed for 5 minutes, with the exception of samples Tg498 and Tg501 (extracted from skeletons SLM-6 and SLM–9) that were left to soak for 12 hours. Samples were subsequently left to air dry in a sterile environment, washed through immersion in 50% bleach solution for 10 minutes, rinsed in distilled water, and exposed to λ=254nm UV light for 10minutes. DNA was extracted, PCR amplified, cloned and sequenced as above. ΦX174 DNA was also amplified using PCR as above.

5.2.4 Assessment of endogenous contamination within teeth

DNA was extracted from 85 teeth from the three burial sites (49 Trondheim (Table 5.1), 24 Matera (Table 5.2), and 11 York (Table 5.3), PCR amplified, and cloned to detect contamination in the samples prior this study.

5.2.5 Assessment of teeth dentine extraction methods

Previous tests have demonstrated that carryover contamination between extractions (chapter 7*)* can be reduced by encasing teeth in silicone-rubber prior to drilling dentine. The results suggest that DNA may enter the extraction process *via* contact between the exterior of the tooth and the researcher or laboratory environment (see chapter 7). To investigate the phenomenon further, comparative extractions were performed on nine pairs of teeth taken from each of two Copenhagen skulls, samples AO and BD. This phenomenon was independently investigated by L.R. in Copenhagen using nine teeth, from Aalborg skeletal sample AHM2481x1057A (Table 5.3). Pairs of teeth from each individual were soaked for 5 minutes in water containing ΦX174 DNA at a concentration of 10^{-9}, 10^{-12}, 10^{-15}, 10^{-18} and 0g DNA ml^{-1*}. Teeth were then left to air dry, and subsequently cleaned as above. One of each pair of teeth was extracted using the silicone method of chapter 3, where dentine is removed using a dental drill. The second was prepared using a method that involves direct glove-tooth contact, whereby teeth were longitudinally fractured and dental pulp was scraped from the cavity using a sterile dental pick (Drancourt *et al.* 1998). ΦX174 DNA then was amplified as above to examine the persistence of contaminant DNA.

The potential benefit of the 'silicone' extraction technique, and the effect of direct handling of samples was further investigated in a second experiment. Twenty-eight pairs of Trondheim teeth (Table 5.1) were separated into three categories of external morphology – 'Sealed', 'Worn Crown' and 'Cavities' (for details see Table 5.1). One of each pair was assayed for human DNA contamination prior to the sampling, as detailed above. The second tooth of each pair was subject to one of five treatments -

[*] For comparison, Allen *et al.* (1998) have estimated that the mass of DNA within a single human mitochondrion is approximately $3x10^{-15}$ g.

groups (A-E) - that consisted of different investigations into contamination and extraction methodologies (5.1). The treatment groups were as follows:

A – Washed (as above) by volunteer S.T. (mtDNA sequence 16238 w/r to CRS). Extracted without silicone.

B – Washed (as above) by volunteer S.T. Extracted with silicone.

C – Handled thoroughly by volunteers S.T. and A.C. Extracted with silicone.

D – Handled thoroughly by volunteers S.T. and A.C. Extracted without silicone.

E – Handled and washed thoroughly by volunteers M.Y. (mtDNA sequence 16217, 16224, 16294, 16296, 16304 w/r to CRS) and M.N.K. (mtDNA sequence 16241, 16311 w/r to CRS). Extracted with silicone.

All samples were PCR amplified, cloned, sequenced, and analysed for persistence of the human contaminant sequences as above. Treatments C and D were independently replicated in Denmark at the Zoological Institute by E.W. and A.J.H. using 4 sets of Copenhagen teeth (samples AV1, CV, CM, CP) (Table 5.3), identical conditions, and handling volunteer R.J.V.

5.2.6 Direct comparison of the susceptibility of teeth and bones to contamination

A direct examination of the susceptibility of tooth and bone to contamination was undertaken as follows. One femur and two teeth samples were taken from five Trondheim individuals (Table 5.1). One tooth from each individual was extracted in the silicone method as a control (treatment K). The remaining tooth and femur sample from each individual were soaked in water containing 10^{-9}, 10^{-12}, 10^{-15}, 10^{-18}, 10^{-21} g/ml ΦX174 for 10 minutes (treatments F-J respectively), before being allowed to dry, cleaned and extracted as above (without using the 'silicone' method). Post-extraction

the samples were screened for human DNA and the persistence of ΦX174 DNA as above.

5.3 Results

All extraction and PCR blanks were consistently negative throughout the study, indicating the results are unlikely to derive from contaminants in the extraction or PCR processes.

5.3.1 Preservation and contamination of Matera teeth and bones

Nine Matera teeth samples contained only one DNA sequence, and appear to be uncontaminated, while seven teeth extracts contained multiple sequences and were identified as contaminated (Table 5.2). A further eight teeth could not be accurately identified as contaminated, as it could not be resolved whether observed sequence variation was due to damage-driven miscoding lesions or the presence of multiple DNA sources. The remaining four samples did not yield an amplification product. There are several clear patterns between DNA contamination in *teeth* and the preservation state determined for *bone* samples from the same individuals. All uncontaminated samples (except for teeth from skeleton SLM-6) lack evidence of significant organic degradation, but have high mineral carbonate:phosphate ratios, indicative of secondary carbonate precipitation. Similarly, they all (except teeth from skeleton SLM-11) lack evidence of fungal attack (identified through a lack of Wedl tunneling – Hackett 1976). Most interesting, the presence of human contaminant DNA in teeth correlates with the increased porosity in the bones of the same individual - the pores having a diameter indicative of microbial alteration ("m" porosity, 600 nm – 1.5 nm; Turner-Walker *et al.* 2002). For full biochemical preservation data refer to Appendix 2.

Despite the rigorous bleach and UV cleaning methods, all Matera bone samples contained multiple DNA sequences, with the exception of samples Tg497 and Tg503,

extracted from skeletons SLM-5 and SLM-11. Preservation data (*c.f.* Smith *et al.* 2002) on both these samples identify them as exhibiting similar values for histology, porosity, collagen content and mineralogy as modern bone, with no evidence of microbial alteration (Table 5.2). Perhaps the most intriguing aspect of these samples is their very high carbonate:phosphate ratios in the mineral, indicating the presence of secondary carbonate. All Matera bone samples (including Tg497 and Tg503) that were artificially contaminated with 10^{-12} g/l ΦX174 DNA (≈ 300 mitochondria in a liter) could not be decontaminated successfully using bleach washes and exposure to UV light – ΦX174 DNA was amplified from all extracts with both primer sets (Table 5.2). Curiously, despite the fact that the majority of bones were contaminated with human DNA prior to this study, the handling and washing by volunteer A.C. did not lead to further contamination. Amino acid racemization (AAR) values tended to be lower in uncontaminated samples (with the exception of skeleton SLM-03, although all values were tightly ranged (0.07-0.13). However even in this sample there was no evidence from DL ratios of Ala, Glx or Ser of measurable (peptidoglycan derived) microbial biomass (*c.f.* Pedersen *et al.* 2001; data not shown). The bone from skeleton SLM-03 had the lowest porosity of any bone from Matera, so it is possible that the elevated AAR value is due to the lower rate of leaching of soluble (highly racemized gelatin), as speculated by Collins *et al.* (1999).

5.3.2 Endogenous contamination within other teeth samples

Seventeen of the 66 Trondheim teeth did not produce amplifiable DNA (Table 5.1), and spiked-PCR tests confirmed that this was not a result of inhibition. These samples appear to be free from both ancient and modern human DNA. Of the 49 remaining teeth, 42 were identified as not contaminated (extract contained only one DNA

sequence, consistent among multiple teeth from the same skeleton). The correlation of contamination with macro-morphological preservation of the whole teeth, is unclear. Of the 32 samples exhibiting good morphological condition (Sealed), 20 contained only one DNA source or did not yield a PCR product, and were presumed not contaminated. Of the 30 samples with worn enamel (Worn Crown), 19 had single sequences or did not yield a DNA product. Of the remaining three teeth containing cavities (Cavities), only 1 contained multiple DNA sequences. Although it seems logical that samples with compromised enamel should be more susceptible to contamination, there is no statistical evidence to support this hypothesis (Chi-squared contingency test on similarity, $p>0.99$). Seven of the 11 less-well-preserved York samples and four of the eight Copenhagen teeth were also contaminated (Table 5.3) (Chi-squared contingency test on similarity, $p>0.53$).

5.3.3 Comparison of teeth dentine extraction methods

The concentration of ΦX174 DNA required to permanently contaminate teeth was much higher in teeth where dentine was extracted using the silicone method (10^{-9} g/l) than where the method of Drancourt *et al*. (1998) was used (Table 5.4), and corresponds to an equivalent of $3x10^5$ mitochondria/liter.

5.3.4 Persistence of human contaminant DNA in artificially handled and washed teeth

The results of the experiment in which teeth were deliberately handled and washed are similar to those obtained for the Matera bones, and are equally difficult to explain. Most of the DNA extracts from the 28 Trondheim teeth that were handled in Oxford

Table 5.4: The suitability of two different dentine-removal techniques in minimizing the carryover of artificial contaminant ΦX174 DNA into DNA extracts.

ΦX174[†]	Skeleton[*]					
	AO		BD		AHM1086	
	Si[‡]	Sc[§]	Si	Sc	Si	Sc
0	N/N	N/N	N/N	N/N	N/N	N/N
10^{-21}	N/N	N/N	N/N	N/N	N/N	N/N
10^{-18}	N/N	N/N	N/N	N/N	N/N	N/N
10^{-15}	N/N	N/N	N/N	N/N	N/N	N/N
10^{-12}	N/N	Y/Y	N/N	Y/Y	N/N	Y/Y
10^{-9}	Y/Y	Y/Y	Y/Y	Y/Y	N/N	Y/N

[*]*Skeleton* Archaeological sample from which teeth were extracted. [†]*ΦX174* Concentration (g/l) of contaminant ΦX174 DNA in which sample was washed prior to cleansing with bleach and UV light (λ=254nm). [‡]*Si* Tooth encased in silicone prior to the extraction of dentine using a dental drill (following chapter 3). [§]*Sc* Tooth not encased in silicone, but split longitudinally and dentine removed through scraping pulp cavity (following Drancourt *et al*. 1998). The persistence of contamination is measured through the presence (Y) or absence (N) of an amplifiable ΦX174 PCR product using two ΦX174–specific primer sets (F1/R1) / (F3/R3).

contained multiple human sequences, but in no case was this DNA from the person who handled and washed the teeth (Table 5.1). Attempts to independently replicate this experiment in Copenhagen by E.W. and A.J.H. produced identical results. Thus, although samples are obviously contaminated with human DNA, the processes involved are obviously more complex than those tested here.

5.3.5 Direct comparison of the susceptibility of bones and teeth to contamination

A contaminant load of 10^{-9} g/l ΦX174 DNA was necessary to contaminate the Trondheim teeth extracted without the silicone technique (Table 5.5). Corresponding femur samples were contaminated at 1000-fold lower levels of ΦX174 DNA (10^{-12} g/l ≈ 300 mitochondria/liter). Therefore, in agreement with previous hypotheses (*c.f.* Drancourt *et al.* 1998) teeth are less prone to DNA contamination than bones in comparable conditions. This observation was also indirectly observed by comparing the presence of endogenous human DNA contamination among bones and teeth from the same skeleton.

Table 5.5: Presence (Y) or Absence (N) of amplifiable ΦX174 DNA within different DNA extracts of Trondheim teeth and femur specimens artificially contaminated with varying levels of ΦX174 DNA.

ΦX*	+Silicone†	-Silicone‡	Femur
10^{-9}	N	Y	Y
10^{-12}	N	N	Y
10^{-15}	N	N	N
10^{-18}	N	N	N
10^{-21}	N	N	N

**ΦX* Concentration of ΦX174 DNA used to contaminate sample prior to cleansing using standard decontamination techniques. †*+Silicone* Tooth where dentine was extracted using silicone rubber encasing technique. ‡*-Silicone* Tooth where dentine was extracted without silicone rubber encasing technique.

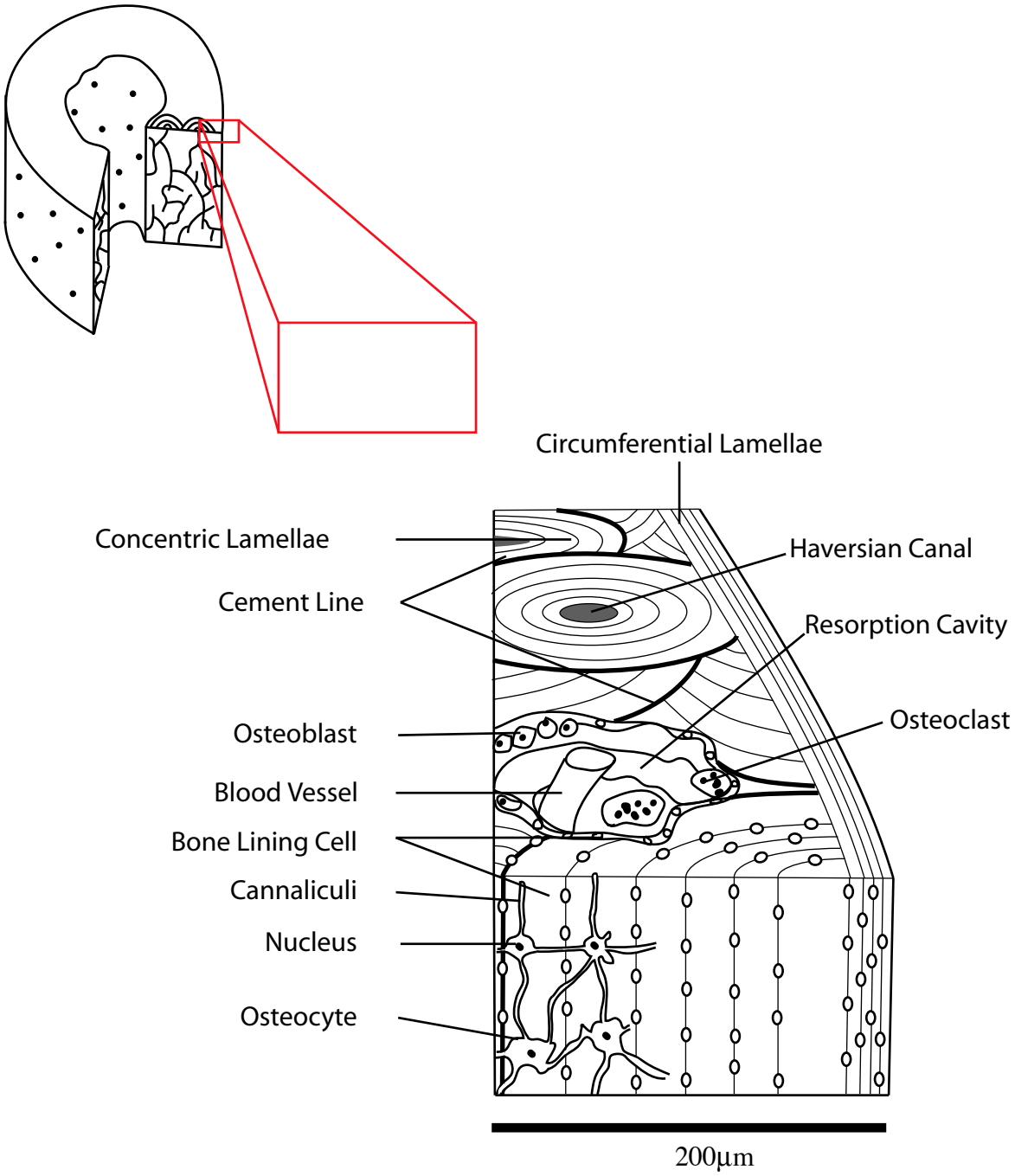
5.4 Discussion

The results of this study are naturally site, sample, and treatment-specific. Nevertheless they highlight the following points. Contamination is widespread among ancient human teeth and bones, and both are easy to contaminate. The susceptibility to contamination correlates directly with sample morphological preservation, and once contaminated samples are difficult (if not impossible) to decontaminate using conventional techniques. However, certain dentin extraction techniques (*e.g.* the silicone rubber method investigated) have advantages in minimizing secondary contamination during analysis.

The degree and persistence of contamination must be largely attributed to the ultra-structural organization of bones and teeth. Mammalian bone is essentially mineralised Type 1 collagen containing a number of specialised cell bodies. The structure of dentin (the bone analogue within teeth) is almost identical, though containing tubular extensions of the odontoblasts (the cells that deposit the dentine) as opposed to the embedded cell bodies of bone, and partly enclosed by enamel, the most highly mineralised structure in the vertebrate body which lacks both cells and cellular debris (Fincham *et al.* 1999).

Although the DNA in living bone is located within several different types of cell (Fig 5.1), the *post mortem* degradation of the cells (and thus the DNA within the cells) is likely to be variable. Osteocytes, which represent 90% of all bone cells, are embedded in the bone matrix, at a density of 90,000-30,000 mm^{-3} (Mullender *et al.* 1996), and connect to neighboring osteocytes *via* canaliculi of approximately 200 nm in diameter (Cooper *et al.* 1966). Thus osteocyte isolation, protection and quantity suggests that they are the most likely source of surviving DNA *post mortem*. However, whether osteocytes represent a contamination free haven is dependent on

Fig 5.1: Histological structure of compact bone.

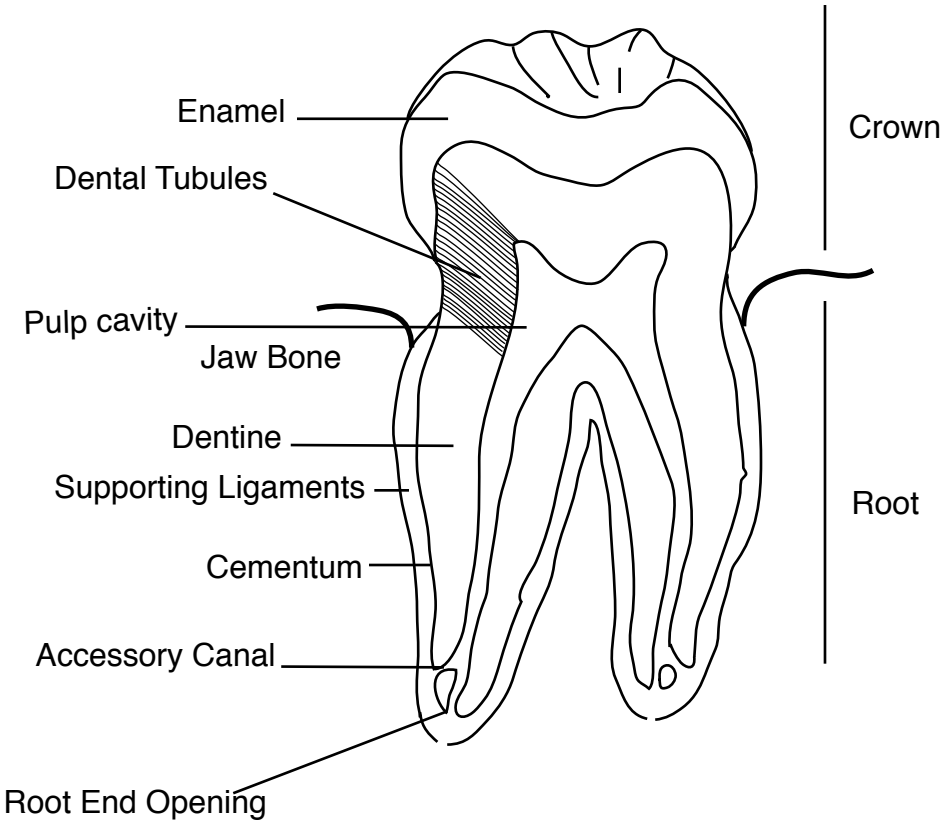


the degree of permeability. Human Haversian canals are generally not more than 0.22mm apart, and connect to osteocytes via the canaliculi syncitium to ensure that no living osteocyte is more than a fraction of a mm from a capillary (Vaughan 1970). Mercury porosimetry demonstrates that the minimum total inter-connected porosity is never less than 8% of bone volume, the majority of which is Haversian, and 2% of which is attributed to canaliculi / osteocyte lacuni. Porosity increases *post mortem*, from both decomposition of the collagen, and microbial destruction of the collagen/apatite composite (Turner-Walker *et al*. 2002).

The results of this study highlight a correlation between endogenous DNA contamination and porosity; the absence of contaminant correlates with low overall sample porosity, and in particular that characteristic of microbial attack (Turner-Walker *et al*. 2002).

The micro-morphological structure of teeth (Fig 5.2) also provides several insights into how DNA may contaminate teeth. Teeth are formed with the migration of ameloblasts (laying enamel) and odontoblasts (laying dentin) out from two initially apposed layers (Currey 2002). The exposed ameloblasts vanish with the eruption of the teeth, leaving no DNA or collagen in the enamel (Smith *et al*. 1993). Enamel is therefore ideally suited to preventing the permeation of contaminants from the tooth crown into the tooth. However the enamel can play no protective role once teeth are washed or the root is directly handled (as demonstrated here by the contamination of the morphologically well-preserved teeth from Trondheim). Although the composition of human dentine is similar to bone it contains no cell bodies (thus no DNA). Instead, the bodies of the odontoblasts are located around the surface of the pulp cavity, trailing living 'dentinal processes' back into the dentine (Currey 2002). The pulp cavity also contains living tissue, both vascularised and well innervated, and

Fig 5.2: Histological structure of a human tooth.



Dental tubules are shown larger than actual size.

connected to the rest of the body through the root canal. This cavity likely represents the only sources of DNA in teeth. In young 'unrooted' teeth, the open root canal provides an obvious pathway for contamination to enter the tooth cavity.

As human teeth develop the root canal gradually seals, rendering them 'rooted'. It is this lack of Haversian canals, and the 'cement layer' surrounding 'rooted' teeth that has lead to suggestions that teeth represent a contamination free niche (Drancourt *et al.* 1998). However, as demonstrated by the widespread contamination of samples in this study, this is patently not the case. Based on the observations that fewer of the well-preserved Trondheim teeth are contaminated than the less well-preserved Matera and Copenhagen samples, and that in bones porosity correlates with contamination, it is plausible that *post mortem* linked increases in porosity are a key factor in tooth contamination

5.4.1 Sample predisposition to contamination

The correlation of various sample preservation indices with contamination provides an important aid to identifying suitable specimens for analysis. Such indices would save considerable time, while limiting destructive sampling, but may help critically assess previous ancient human DNA results. Overall these will be of most use if they help provide a taphonomic context in which DNA preservation and contamination can be understood.

Porosity, in particular that caused by microbial alteration of the bone, appears to be the most helpful measure. Whilst this is not the only factor (*e.g.* accelerated destruction of the organic phase as observed at Apigliano, Italy (Smith *et al.* 2002) will presumably destroy DNA but not lead to microbial porosity), techniques such as microscopy or mercury porosimetry which assay this (*c.f.* Turner-Walker *et al.* 2002)

will provide a useful guide to the contamination load of a sample. An alternative technique now routinely used on ancient samples prior to DNA analysis that may also be of use is amino acid racemization analyses (Poinar *et al.* 1996; Poinar *et al.* 1999) (though see Collins *et al.* 1999) though whether a significant correlation exists between racemization and porosity requires further investigation. However, given the presence of microbial destruction even in bones with apparently exceptional preservation from Matera, perhaps the best approach is to select for bones which have not derived from interred corpses. We note in passing that because animals are rarely interred as intact corpses they display much less evidence of porosity attributed to putrefaction (Jans *et al.*, in press), which may explain why they are proving a more promising substrate for aDNA study than human material.

5.4.2 The persistence of contamination in samples

It is intriguing that porosity in general, and microbial porosity in particular seems to be the main factor influencing the contamination of human bone, yet although all samples that were subjected to handling and washing regimes contained contaminant human DNA prior to our study, they could not be further contaminated with human DNA. Widespread contamination in archaeological specimens that were excavated in the past is not surprising. In many cases freshly excavated samples were cleaned in dirty water that was not changed between samples. In addition, one tool of choice to aid cleansing were old and used human toothbrushes. Thus these would have provided an immediate source of human and bacterial DNA. However, a lack of further contamination suggests that samples are most susceptible to contamination just after excavation, when they are still damp from the burial environment. If the presence of water is a precondition for successful permeation of contaminants,

samples from dry soils may also still be at risk as they are likely to contain some water - due to the surface tension of water, samples containing small pores (those below 4 nm) will contain water even below soil wilting point (Hedges & Millard 1995). Furthermore, it is possible that deposits (potentially deriving from sediment in water used to wash samples) are laid down within the porous regions of the samples as samples dry over time, hindering further contamination. Alternatively, it is possible that some drying may seal pore interconnectivity through collapsing collagen bundles or drying fronts of other dissolved species. Indirect evidence for this comes from one measure of preservation that inversely correlates with contamination – 'cracking' – which Hedges (2002) has hypothesised results from bone swelling, due to mineral deposition within Haversian canals. Such permeation may also explain the low success of soaking samples in liquid (*e.g.* bleach) to remove contaminant DNA. In addition, 'infilling' may also explain why artificial contamination experiments using naked ΦX174 DNA are more successful than those relying on handling if the larger particles represented by DNA within a handler's skin cells cannot permeate as far as the naked ΦX174 DNA. If the above hypothesis is true, then it is possible that the disruption of internal blockages through techniques which powder the samples prior to cleansing warrant further investigation. An alternative approach to consider is partial demineralization of the bone or teeth with EDTA prior to the bleaching. This may increase both accessibility to contaminant DNA through the action of EDTA removing diagenetic calcium and iron salts, and the action of bleach opening up collapsed regions of collagen. A final method that may increase permeation of bleach into bone, is its application under a vacuum.

5.5 Conclusions

The results of this study highlight serious inadequacies in techniques used to decontaminate archaeological samples from contaminant DNA, and have implications for any aDNA study where contamination is a risk. Importantly, this also includes studies of pathogen DNA in ancient samples, which may be at risk from DNA originating from environmental organisms (chapter 7), as well as studies on animal bones that have been treated with preservatives and glues containing animal DNA (*c.f.* Nicholson *et al.* 2002). The simplest remedy for human DNA contamination is the controlled excavation of human samples, preventing direct handling and washing of samples that are to be used for aDNA analyses. Unfortunately as most samples cannot be freshly excavated, and many important specimens were collected in the past, the majority are likely to be already contaminated. Thus further studies are required to investigate the possibility of new decontamination techniques on such specimens. In addition other sources of aDNA that may be more resilient to contamination such as hair need to be investigated (chapter 6).

The findings of this study also suggest that a critical reader must consider three key points before gauging the reliability of results from previous human aDNA studies. Firstly the preservation of the samples. For example, Clisson *et al.* (2002) report human DNA retrieval from a 2000+ year old specimen from Kazakhstan. The excellent preservation reported suggests that contamination is not an issue. Secondly, sequence authenticity. For example, although not well preserved, the distinct nature of Neanderthal sequences helps suggest authenticity of results (*c.f.* Krings *et al.* 1997). Thirdly, and most importantly, a knowledge of the history of the sample handling *prior to the analysis* is critical. For example, in the important early Australian 'Mungo man' study (Adcock *et al.* 2001), although the researchers took

full anti-contamination precautions, it was not mentioned that the sample had been excavated in the 1970's and handled numerous times since (thus likely to be contaminated), contained 'negligible organic preservation' (Gillespie and Roberts 2000), and is considered too fragmentary to sex reliably (Brown 2000; Thorne and Curnoe 2000) – in short, of very poor preservation. Without such information, it is very difficult to comment objectively on the reliability of results.

5.6 Acknowledgements

The authors would like to thank Dott.ssa Maria Luisa Nava (Soprintendente Archeologo della Basilicata) for giving permission to sample the Matera specimens, and Dott.ssa Brunella Bruno (the excavation supervisor at Matera) and Dott.ssa Erminia Lapadula for assistance in obtaining the Matera samples. The authors would also like to thank Arnab Chatterjee, Sam Turvey, Malgosia Nowak-Kemp, Mai Yasue and Rikke Joo Vienberg for their help in handling sample in order to investigate contamination. M.T.P.G. and A.C. would like to thank the Wellcome Trust for funding the research. EW and AJH would like to thank the VILLUM KANN RASMUSSEN FONDEN, Denmark for funding their research. The diagenetic parameters of the Matera bone samples were analysed as part of the 'Degradation of Bone as an Indicator in the Deterioration of the European Archaeological Property' project funded by the EU (ENV4-CT98-0712)

Chapter 6

Hair shafts as a reliable source of ancient DNA

Modified version in review at 'Current Biology'

M Thomas P Gilbert[1], Andrew S Wilson[2,3], Eske Willerslev[4], Anders J Hansen[4], Michael Bunce[1], Beth A Shapiro[1], Tom Higham[5], Mike P Richards[2], Tamsin C O'Connell[5], Desmond J Tobin[3], Rob C Janway[2], Alan Cooper[1].

[1]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, University of Oxford, South Parks Rd, Oxford OX1 3PS, United Kingdom
[2]Department of Archaeological Sciences, University of Bradford, Bradford, West Yorkshire, BD7 1DP, United Kingdom
[3]Department of Biomedical Sciences, University of Bradford, Bradford, West Yorkshire, BD7 1DP, United Kingdom
[4]Department of Evolutionary Biology, Zoological Institute, University of Copenhagen, 15 Universitetsparken, Copenhagen Ø, DK-2100, Denmark
[5]Research Laboratory for Archaeology, University of Oxford, 6 Keble Road, Oxford OX1 3QJ, United Kingdom

Summary

Hair shafts have received little attention as a DNA source in ancient DNA studies due to the paucity of samples in the archaeological record, and relatively low concentrations of DNA present in modern samples. In this chapter, it is demonstrated that small samples of hair shafts over 64,800 years old contain amplifiable mitochondrial DNA. In direct contrast to teeth and bones (chapter 5), it is possible to decontaminate hair samples of any contaminant DNA, even those that are badly degraded or have been handled many times. It is suggested that this relates to the hydrophobic and impermeable nature of the hair cortex. Consequently, hair samples represent an important new source of ancient DNA.

6.1 Introduction

Hair has received little attention from the ancient DNA (aDNA) community, with a few exceptions (Bonnichsen *et al*. 2001), despite the use of hair fibres as a source of DNA in many forensic studies (Allen *et al*. 1998). This may be partly due to the widespread belief that DNA concentrations within hairs are much lower than in other tissue sources. Higuchi *et al*. (1988) report that a single, fresh, human hair will contain no more than 200ng of extractable DNA, while levels in naturally shed hairs often fall below 10ng. After several months, levels in shed hairs have been observed to fall further still, to less than 1ng (Allen *et al*. 1998). There have also been reports of problems using hair as a DNA source, such as allelic dropout in microsatallite studies (Gagneux *et al*. 1997). Further issues include the inhibitory properties of hair melanin in PCR reactions (Uchichi *et al*. 1992; Wilson and Budowle 1993).

Despite these difficulties forensic scientists have successfully amplified DNA from almost all sources of human hair – including eyebrow, pubic and torso hair (Baker *et al*. 2001). Analyses demonstrate that nuclear DNA is found mainly in the root (*c.f.* Pilkington *et al*. 1987; Vigilant *et al*. 1989; Thomson *et al*. 1992; Uchici *et al*. 1992), and only negligible amounts are present in hair shafts (Hukkelhoven *et al*. 1981), although successful amplifications have been reported (*c.f.* Uchichi *et al*. 1992). Mitochondrial DNA is present throughout much of the length of the shed hair fibre, and has been amplified from both root (*c.f.* Savolainen and Lundeberg 1999; Allen *et al*. 1998) and shaft sources (*c.f.* Higuchi *et al*. 1988; Baker *et al*. 2001). Additionally, mtDNA has been successfully extracted from both degraded and old hair samples, including burnt specimens (Baker *et al*. 2001), 100 year old Native American samples (Baker 2001), and Bighorn sheep (*Ovis canadensis*) dating to over 9,400 years old (Bonnichsen *et al*. 2001).

Hair offers a number of benefits as a source of aDNA, including the possibility of using a small and unobtrusive sample. For example, less than 2 cm of morphologically degraded (though relatively recent) hair shaft has proved a suitable source of mitochondrial DNA (Baker *et al.* 2001). More importantly, Wilson *et al.* (1985) and Jehaes *et al.* (1998) have shown that is possible to clean modern hair shafts contaminated with human saliva and blood. Considering the greater difficulty in decontaminating other sources of ancient human DNA from modern DNA present on the sample, such as teeth and bones (chapter 5), this suggests that further analysis of the long-term preservation of DNA in hair is worthwhile.

This paper presents an in-depth study of the suitability of hair as a reliable source of contamination-free ancient DNA. By comparing the properties of degraded and old hair samples in handling and decontamination experiments, it is shown that hair offers a useful, and in many cases, superior source of ancient mitochondrial DNA to teeth and bone.

6.2 Materials and Methods

It is exceedingly difficult to authenticate ancient human DNA sequences, and it may even be impossible in certain cases (Cooper 1997). Several studies have demonstrated how even stringent controls can fail to prevent or detect contamination (Handt *et al.* 1996; Kolman and Tuross 2000). Consequently, the methods and samples were chosen to maximize the chance of preventing or limiting contamination.

6.2.1 Degraded human hair samples

To examine how hair samples might become contaminated, and whether it is possible to decontaminate them if so, a series of experiments were carried out using 12 human hair samples from a previous study by A.S.W. (Table 6.1). The hairs were degraded as part of an experiment at Shelf, West Yorkshire, England by A.S.W., to model how hair and textiles survive and/or alter within the temperate 'buried body' environment (Janaway 2000; Wilson *et al.* 2003). The hairs have all been histologically screened (for details refer to Appendix 3a), and histology appears the best current indicator of sample preservation (Wilson 2002). This allows a comparison between hair preservation and susceptibility to contamination. Each sample was divided, and half of each sample was vigorously handled by M.T.P.G. in order to simulate the contamination which occurs when hair shafts are handled. After handling, the samples were decontaminated by immersion in 50% strength commercial bleach solution for 30 seconds followed by a rinse in sterile ethanol.

6.2.2 Long-term DNA survival in hair

Long-term DNA survival was analysed in twelve hair samples aged between 60 and more than 65,000 years (Table 6.2). The samples were as follows:

Table 6.1: Experimentally degraded hair samples used in artificial contamination experiment.

Sample[1]	Histology[2]	Clones[3]	DNA[4]	Contamination[5]
AW_8	0	12	Yes	No
AW_26	3	12	Yes	No
AW_27	2	12	Yes	No
AW_47	1	12	Yes	No
AW_63	1	12	Yes	No
AW_71	2	12	Yes	No
AW_94	4	12	Yes	No
AW_95	4	12	Yes	No
AW_H081	5	12	Yes	No
AW_H082	5	12	Yes	No

[1]*Sample* Authors' DNA extract identification. [2]*Histology* Histological score following Wilson (2002). Briefly, in concordance with bone histological scoring *(c.f.* Nielsen-Marsh and Hedges, 2000), scores range between 5 (Very good condition) and 0 (Very bad condition). [3]*Clones* Number of cloned amplicons sequenced. [4]*DNA* Presence of amplifiable DNA within DNA extract. [5]*Contamination* Presence of non-endogenous human DNA within DNA extract as evidenced through molecular cloning.

Table 6.2: Details of ancient hair samples analysed.

Sample[1]	Species	Details[2]	Age[3]	DNA[4]	Damage[5]	Source
Tg415	H.sapiens	Onge	50‡‡	Yes	0.0011	Lehrmann
Tg468	H.sapiens	Newton?	361-276††	No		Woolsthorpe Manor
Tg469	H.sapiens	Newton?	361-276††	Yes	0.0007	Cullum Collection
Tg471	H.sapiens	Newton?	361-276††	No		Royal Society
Tg472	H.sapiens	Newton?	361-276††	Yes	0.0019	Lord Portsmouth†
Tg473	H.sapiens	Newton?	361-276††	Yes	0.0015	Lord Portsmouth¶
Tg474	H.sapiens	Newton?	361-276††	Yes	0.0011	American Philosophical Society
Tg491	B.bison	Dawson, Yukon	>64,500	Yes	0.0014‡ 0.0050§	Cooper
Pazyryk 1	E.caballus	Ak-Alakha3	2800-2200	Yes	0.0033	Molodin and Polos'mak
Pazyryk 4	E.caballus	Verkh Kaljin II	2800-2200	Yes	0.0033	O'Connell
Pazyryk 7	E.caballus	Ak-Alakha3	2800-2200	Yes	0.0030	O'Connell
Pazyryk 8	E.caballus	Ak-Alakha3	2800-2200	Yes	0.0010	O'Connell

[1]*Sample* Authors' sample DNA extraction identification. [2]*Details* Sample name. [3]*Age* Sample age in years. [4]*DNA* Presence of amplifiable DNA in extract. [5]*Damage* Damage measured as independent number of miscoding lesions per total bases of sequence amplified (excluding primer) for the region amplified (following Gilbert *et al*. 2003). Where multiple extractions and amplifications have been performed, the average number is given. For full sequence details refer to Appendix 3. ‡‡ Sample collection date. †† Age assuming hairs are authentic. ‡Bison HVR sequence §Bison 16s sequence †Hair attributed to Sir Isaac Newton as a youth ¶Hair attributed to Isaac Newton as an old man.

6.2.2.1 Andaman Islander

Two DNA extractions were made from hair cuttings of an Onge tribal individual. sampled in the Andaman Islands (DNA extraction Tg415a, Tg415b). The hair is believed to have been collected in 1953 by a Harold Lehrmann and has been stored by his family at room temperature in a paper envelope since. The hair fibres were matted together with a red clay-like substance. A portion of the matted hair sample was sent to E.W. and A.J.H. in Copenhagen for independent replication.

6.2.2.2 Sir Isaac Newton

Six human hairs purportedly from one individual (Sir Isaac Newton 1642-1727) were obtained from different museum and private collections. All are believed to have been stored at room temperatures since collection, and are likely to have been handled extensively. Although the samples are attributed to Sir Isaac Newton, isotope analyses of the specimens suggest that the hairs are likely to have originated from multiple individuals (AS Wilson *et al*. unpublished data). Due to the limited number of samples, extractions were limited to one per specimen.

6.2.2.3 Pazyryk putative horse hairs

Four hair samples were obtained from Pazyryk culture burial sites, from the Ukok plateau in the Altai region, Russia. The burial sites are dated between the 6th and 2nd century BCE (Polos'mak and Molodin 2000), and were excavated in the early 1990's by NV Polos'mak and VI Molodin from the Institute of Archaeology, Novosibirsk. Cold temperatures and ice lenses in many graves in this area have contributed to the excellent preservation of human/faunal remains (Polos'mak and Molodin 2000), and DNA has been successful retrieved from naturally mummified humans in the region

(Voevoda *et al.* 2000). Three of the samples were obtained from ceremonial horse burials at 2 sites, Verkh Kaljin II, (Pazyryk 2) and Ak-Alakha 3 (Pazyryk 3 and 4), while the fourth (Pazyryk 1) was sampled from the wig of a female at the site of Ak-Alakha 3. Carbon and nitrogen isotope analyses of Tg515 suggest the hair is very unlikely to be human, but more likely that of a herbivore, probably horse (O'Connell *et al.* 2003). These samples were blind extracted by B.S. in Oxford. As with the Newton specimens, the limited amount of samples permitted only one extraction per specimen.

6.2.2.4 Christie Mine putative bison hair

Four DNA extractions (Tg491,Tg492,Tg494, Tg495) were made from hair samples of a naturally mummified putative bison (*Bos bison*) obtained from permafrost deposits at Christie Mine, near Dawson, Yukon Territory, Canada by A. Cooper. The specimen was recovered by the miners below a volcanic tephra dating to approximately 190,000 years ago (Froese *et al.* 2002). Sample of the hairs were sent to Copenhagen for independent replication of the DNA results by E.W and A.J.H.

Hairs from the sample were AMS ^{14}C dated at the Oxford Radiocarbon Accelerator Unit (ORAU), University of Oxford, following (Hedges *et al.* 1989; Hedges and van Klinken 1992; Bronk Ramsey *et al.* 2000). Briefly, 22.8 mg of hair sample was rinsed at 80°C in 1M HCl for one hour, 0.1 M NaOH at room temperature for one hour, and a repeat 1M HCl rinse. Between each reaction, the hair was rinsed to neutrality with distilled water. Finally, the pre-treated hair was lyophilised yielding a 10.2 mg product. A sub-sample of 2.95 mg was combusted and analysed using mass spectrometry with a Europa ANCA Roboprep interfaced to a Europa 20/20 Mass Spectrophotometer operating under continuous-flow mode.

The histological preservation of the bison hair was assessed using Scanning Electron Microscopy (SEM), High-Resolution Light Microscopy (HRLM) and Transmission Electron Microscopy (TEM) following Wilson (2002). The carbon to nitrogen ratio of the bison samples was assessed after Richards and Hedges (2003).

6.2.3 DNA extraction and analysis

DNA preparation and extraction methods followed strict ancient DNA specific requirements (*c.f.* Cooper and Poinar 2000). Prior to the DNA extraction, samples of hair shafts (<2cm per sample) were decontaminated as detailed above. Samples were digested in a 0.5ml extraction buffer (modified from Barnes *et al.* 2002) containing 0.01M Tris buffer, 0.01M NaCl solution, 1% SDS, 0.5mg/ml proteinase K, 10mg/ml DTT and 0.001M PTB (N-phenacylthiazolium bromide). After incubation at 55°C for 24 hours, the DNA was extracted following a standard phenol:chloroform DNA extraction protocol (after Barnes *et al.* 2002), concentrated by centrifugal dialysis (Millipore, UK) to yield a final volume of approximately 100µl. One blank control extraction was performed for each four samples extracted.

The DNA content within each sample (both human and animal) was investigated using PCR with mitochondrial Hypervariable Region 1 (HVR1) primers L16209-H16356 (Handt *et al.* 1996). In the putative horse and bison samples a mitochondrial 16s DNA fragment of approximately 140bp in size (depending on species) was targeted using generic mammal primers 16sMam1/16sMam2 (Taylor 1996). As no bison 16s mtDNA sequence currently exists in the public domain for comparison, an extract of an ancient bison bone (DNA extract BS200) prepared for a separate study (Shapiro 2003) was also sequenced once the analysis on the hair sample was completed .

A 234 bp fragment of mitochondrial control region DNA was also amplified between *Bos bison* Reference Sequence positions (BbRS) 37-292 (chapter 4) from the putative bison sample using primers Seg1bigF/Seg1bigR (5' ACC CCC AAA GCT GAA GTT CT / 5' CTT GCT TAT ATG CAT GGG GC) (Shapiro 2003). PCR conditions for the reaction were: Enzyme activation 94ºCx1min30, 40 cycles of (94ºCx45 seconds denaturation, 56ºCx45 seconds annealing, 68ºCx90 seconds extension), final extension 68ºCx10 minutes. An attempt was also made to amplify small nuclear DNA fragments of the nuclear, single copy Cyp19 gene, encoding aromatase cytochrome P450, from the putative bison extracts. This gene has previously been used to study the phylogeny of the tribe Bovini (Pitra *et al.* 1997), and bison and cows can be distinguished based on SNPs within the region. Primers used were CypEx10_40 / CypEx10_72 (5'-GAT GAG TGA ATT TTC CAA GTA TTA AT / 5'-GGC ATT CAG CAG ATA CAC TTG T) amplifying a 32bp fragment (without primers) between positions 3040-3072 (GenBank file Z32741, Vanselow and Furbass 1995) and CypEx10_98 / CypEx10_136 (5'-AGT GTA TCT GCT GAA TGC CAC A/ 5'-GGT TGA CTC AAT GGC AAA CA) amplifying a 38bp fragment (without primers) between positions 3098 and 3036. Amplification conditions were identical to those used above.

Amplifications used a high-fidelity polymerase (Platinum Taq Hi-Fidelity, Invitrogen UK) after Cooper *et al.* (2001). However, initial attempts at PCR indicated that reactions were more efficient if Bovine Serum Albumin (BSA) (commonly used in aDNA amplifications to counter inhibitors to the PCR reaction) was excluded from the reaction. Negative control amplifications were performed at a ratio of one control to one sample. Amplified products were extensively cloned using the Topo TA cloning system (Invitrogen, UK), and up to 12 colonies were amplified and

sample following chapter 3. Cloned sequences were aligned manually with the respective reference sequences using the software SEQUENCE NAVIGATOR v 1.0.1 (Applied Biosystems Inc. USA). Consensus sequences and contaminating DNA was determined as in chapter 3.

6.3 Results

Cloned DNA sequences were obtained from 20 out of 22 samples (Table 6.1, 6.2). The remaining two samples did not yield PCR products. Sequences derived from cloned PCR products of ancient DNA extracts often contain miscoding lesions that arise due to DNA damage (*c.f.* Hoftreiter *et al.* 2001a). Levels of post-mortem hydrolytic damage-derived miscoding lesions observed in the samples investigated in this study (Table 6.2) were lower than have been observed in previous studies on bone and teeth (*e.g.* chapter 2,3,4). Authenticity of the data is suggested by the reproducibility of results among multiple extractions and independent PCRs per sample, by the results of the blind extractions, and by the results of independent replication. In addition, no contamination was observed in the PCR or extraction blanks, and the only variation observed among cloned sequences was that which could be attributed to hydrolytic damage-derived miscoding lesions. Finally, all obtained sequences are phylogenetically plausible.

6.3.1 Hair decontamination

Despite extensive contact via the handling experiments, the subsequent bleach treatment of the morphologically degraded hair samples appeared to remove all deposited DNA. In addition, despite the fact the Newton samples are fairly high profile and have been handled on numerous occasions (Malcolm Neame, pers. comm.), again no intra-clone variation was observed that would indicate sample contamination. Lastly no animal hairs yielded human DNA. Thus the data suggest that even well handled, and degraded hairs can be easily decontaminated using a simple bleach treatment protocol.

6.3.2 DNA from Andaman Islander hair

DNA could be extracted from the hair shafts despite being over 60 years old, having been collected from a tropical burial site, and having been stored at warm temperatures. The sequence (16223c-t, 16319g-a, 16344c-t with reference to the Cambridge Reference Sequence (Anderson *et al.* 1981)) is consistent with that previously reported from Andaman Islander mtDNA (Endicott *et al.* 2003). Identical results were independently obtained by E.W. and A.J.H. For full details of the cloned sequence refer to Appendix 3.

6.3.3 DNA from putative Sir Isaac Newton hairs

These samples would be over 270 years old (if authentic), and have been stored at room temperature and handled on many occasions. Nevertheless, four out of seven samples yielded amplifiable DNA. Analysis of the cloned sequences (Appendix 3b) reveals that each sample contains small amounts of damage-derived sequence variation around a consensus sequence. Interestingly, no two hair samples contained the same DNA sequence - despite claims that all hairs were derived from Sir Isaac Newton. This is in agreement with the isotope analyses of the samples (A. Wilson unpublished data). These results can either be interpreted as the hairs containing an excess of contaminant DNA from one source, or that the samples originated from four different individuals. As the hairs have been handled by numerous people, it seems unlikely that only a single contaminant sequence would survive per hair. Thus it is reasonable that the sequences are authentic to the samples, but at least three hairs do not derive from Isaac Newton.

6.3.4 DNA from Pazyryk horse hairs

130 bp of mitochondrial 16s DNA (including primer) matching *Equus caballus* was successfully obtained from all four Pazyryk samples, including that derived from a human wig. This confirms the previous hypothesis (based on the results of isotope analysis) that all four hair samples were horse, (O'Connell *et al.* 2003). Although from the same archaeological location as samples Pazyryk 1 and 8, the sequence from sample Pazyryk 7 contained a unique t-c transition (Appendix 3b)

6.3.5 DNA from bison hairs.

AMS ^{14}C dating revealed this sample to be over 64,800 years old (OxA-12056), and archaeological evidence suggests that the sample may be as old as 190,000 years (Duane Froese, pers. comm.). SEM, HRLM and TEM analysis of the sample demonstrates that apart from some slight cuticular loss and adherent deposits, the hair is remarkably well preserved (Fig 6.1). This is supported by the stable carbon and nitrogen isotope analysis, which had a C:N ratio of 3.47, similar to modern mammal hair (O'Connell and Hedges 1999a; O'Connell and Hedges 1999b) and isotope data ($\delta^{13}C$ = -21.9 ‰, $\delta^{15}N$ = 4.3 ‰) similar to modern and ancient herbivores (Richards and Hedges 2003). A 292 bp HVR1 fragment and a 131bp mtDNA 16s fragment (including primer) was successfully amplified from extracts of this sample. The 16s mtDNA sequence matched that amplified from the bison bone (BS200), and differed from *B.taurus* at 4 sites. The HVR1 sequence differed from *B.taurus*, and phylogenetically grouped among sequences obtained on a large data set of ancient bison (B.Shapiro, in preparation). For sequences refer to Appendix 3. No nuclear DNA could be amplified from the sample.

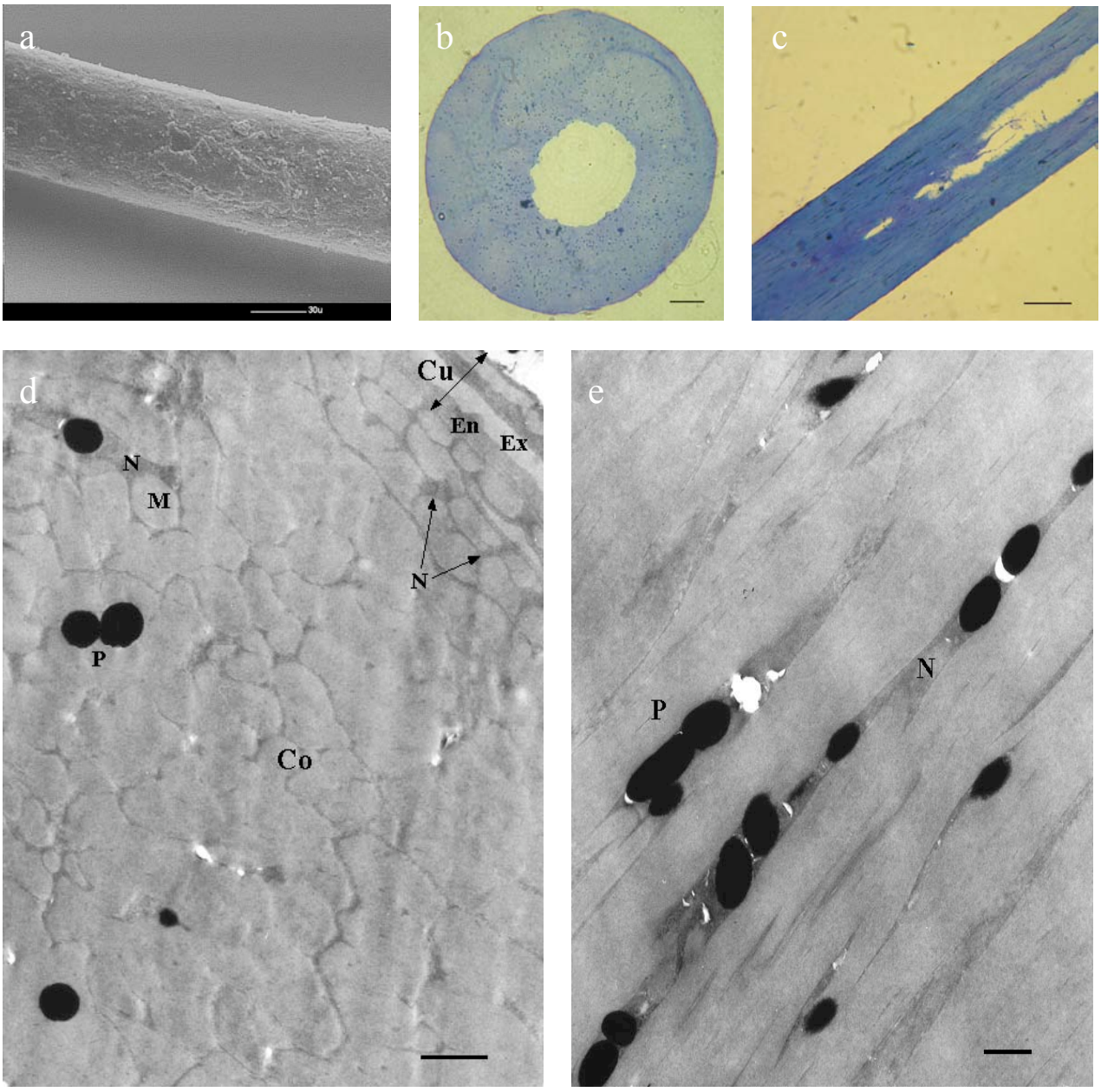


Figure 6.1: Good histological preservation of the 'Bison' hair, a – scanning electron micrograph showing adherent deposits to the otherwise superficially intact Bison hair shaft (bar 30 μm); b – high resolution light micrograph (HRLM) transverse section showing intact cortex with even distribution of well preserved pigment granules and an open central core which presumably held a medulla (bar 20 μm); c – HRLM longitudinal section showing an interrupted medullary space (bar 30 μm); d – transmission electron micrograph (TEM) transverse section showing a portion of cuticle (Cu) comprising intact exocuticle (Ex) and endocuticle (En), and well preserved cortex (Co) with macrofibrils (M), melanin pigment granules (P), and nuclear remnants/ intermacrofibrillar matrix (N) (bar 0.5 μm); e – TEM longitudinal section showing well preserved cortex with melanin pigment granules (P) and intermacrofibrillar matrix (N) (bar 0.5 μm); HRLMs (0.5 μm thick) stained with toluidine blue in borax; TEMs (100 nm thick) stained with osmium tetroxide, uranyl acetate and Reynold's lead citrate.

6.4 Discussion

Bonnichsen *et al*. (2001) have previously reported the amplification of authentic ancient DNA from a hair sample dated at 9,400 years old. This study extends the timeframe by at least seven-fold, placing the DNA from the bison hair among the oldest authentic DNA retrieved from bones and teeth. The findings demonstrate that DNA can be retrieved from samples stored over long periods at room temperature. Finally, the results demonstrate that contaminated, morphologically degraded hairs can yield authentic, uncontaminated DNA. This is in stark contrast to bone and teeth samples where pervasive contamination is extremely difficult to remove (see chapter 5). It is probable that these observations derive in large part from the unique manner in which hair grows during life.

The hair fibre (Fig 6.2, Fig 6.3) is the elaboration of several concentric layers of up to 15 distinct interacting cell sub-populations that emerge from a highly complex set of interactions between ectodermal. mesodermal and neuroectodermal components (Hardy 1982). The follicle's fibre is formed via a process of keratinisation that includes keratinocyte division, protein synthesis, catabolic breakdown of nucleic acids and cell organelles, such that the fibre is completely keratinised about 1 mm from the follicle base or bulb. This reflects a phenomenally rapid cell turnover, with only 2.5 days taken from "birth" of the hair fibre cell to it becoming fully keratinised, and contrasts with the up-to-10 days required for keratinocytes in the skin (Forslind *et al*. 1966).

The hair fibre is constructed of 2-3 distinct components (Fig 6.2), the bulk of which consists of compacted cortical cells that are peripherally encased by cuticular cells. Large (so-called terminal) hairs tend also to have a third central component, the medulla. This forms the fibre's core and may be either continuous or discontinuous

Figure 6.2: Histological structure of hair root.

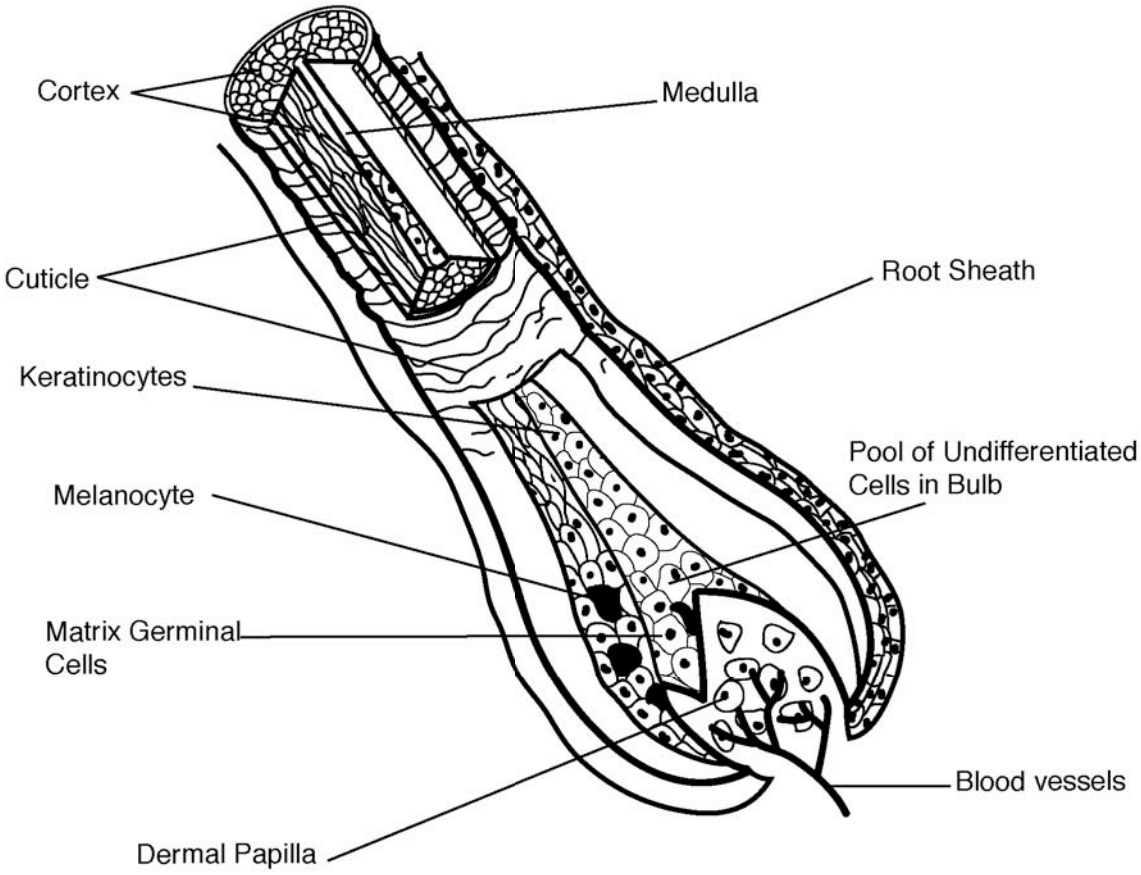
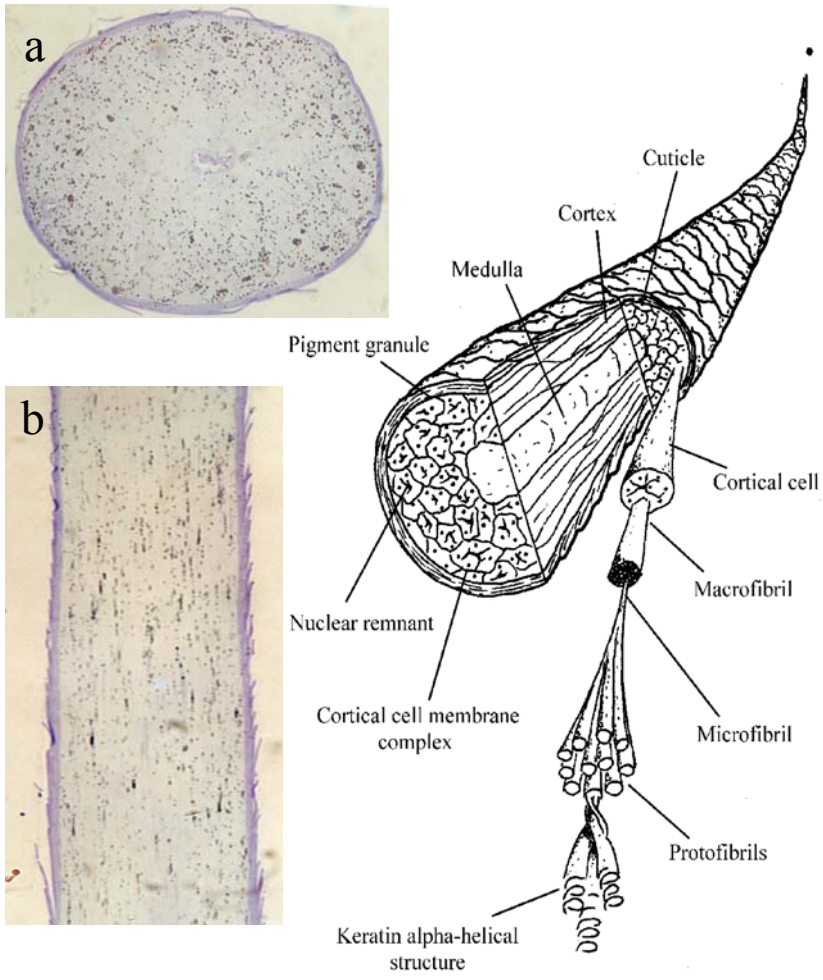


Figure 6.3:Hair shaft exploded view with high resolution light micrographs showing a) transverse and b) longitudinal structural morphology. After Wilson *et al.* (2001a).



throughout the fibre length, depending on body location. Moreover, the medulla can be either intact or collapsed, the latter resulting in the formation of insulating air-spaces (Wilson *et al*. 1994). Pigmented hair fibres have an additional non-keratinaceous component that lends pigmented hair its colour. Hair pigment derives from melanin granules transferred from the hair follicle's neuroectodermal cell population, the melanocytes, to become deposited between keratin bundles in the hair fibre's cortex (Linch *et al*. 2001). Variation in hair colour is a product of the melanin type and level with eumelanin contributing mostly to browns and blacks and pheomelanin to reds and yellows, both of which may act as inhibitors of PCR (Wilson *et al*. 1994).

The bulk of the hair shaft, the cortex, consists of large numbers of 7-8 nm intermediate hard keratin filaments that orient in the hair growth direction. These are distributed within a matrix of non-filamentous proteins and bundled together as macrofibrils within the overlapping and interdigitating spindle-shaped cortical cells of the hair cortex. Although during cortex formation the cells undergo loss of cell cytoplasm, organelle destruction and dehydration, the nuclear and mitochondrial remnants may be retained between keratin fibrils. Linch (2001) has postulated that the undamaged mitochondrial membranes may offer protection for the mtDNA molecules located in the internal inner matrix of these organelles, unlike the nuclear DNA that may be more easily exposed during nuclear fragmentation and lysis. This is likely to be the predominant explanation for why studies on even fresh hair shafts have encountered problems with amplifying nuDNA (Higuchi *et al*.1988). It is interesting to note that cell death due to terminal differentiation of keratinocytes (*i.e*. keratinisation) has recently been likened to apoptosis – where death of the cell is

programmed. A hallmark of apoptosis is the protracted retention of organelle integrity, most specifically mitochondrial integrity (Hacker 2000).

Mitochondria numbers tend to be greatest in highly metabolically active cells (Robin and Wong 1988). The keratinocyte of the hair bulb matrix ranks second only to the gut epithelium for rate of cell turnover (van Scott *et al*. 1963). This high baseline level of mitochondria may be increased still further in pigmented hairs after keratinocytes incorporate melanin granule-containing fragments of melanocytes which may also contain mitochondria (Birbeck *et al*. 1956; Toda *et al*. 1975; Olsen 1994). In combination, the pigmented hair cortex may, in contrast to other tissues, have a bias in DNA preservation due to the unique events involved in hair shaft development.

Outside of the cortex (*i.e*. in the cuticle and medulla), the DNA contribution of other hair shaft components is not likely to be significant, due to their much reduced contribution to the fibre's bulk. However, the hydrophobic nature of the proteins filling the cells (due to chemical cross-links along alpha-helical polypeptide chains, *e.g*. S-S, glutamic acid-lysine isopeptide bonds, and the keratin packing of the cells (Fraser *et al*. 1972;Valkovic 1988)) helps provide a watertight seal around the cortex (Fraser *et al*. 1972), and suggests a plausible explanation as to how samples may be resisting the penetration of contaminant DNA.

With histological degradation some of the protection against water and contaminants conferred by the medulla and cortex will be lost. Nevertheless samples exhibiting low levels of histological preservation do not appear any more susceptible to contamination. This is possibly is due to the highly keratinised (thus impermeable) cortex. In addition the pores that form with the degradation of hair, may provide an

easy route for bleach to access, and thus degrade, any contaminant DNA on/within the hair sample

Mitochondrial DNA from the bison sample has been preserved over extremely long time periods. Although the mitochondrial content of hair is enriched, overall levels are still significantly lower than in tissue (Higuchi 1988). Thus at comparable rates of degradation, mtDNA in hairs would be expected to degrade to sub-amplifiable levels before mtDNA in bones. However, mtDNA survival would be extended should rates of degradation be slower in hairs. The presence of water has been observed to correlate with aDNA degradation (*e.g.* Pääbo 1989; Lindahl 1993a; Waite *et al.* 1997; Rogers *et al.* 2000; Götherström *et al.* 2002; Geigl 2002), likely due to hydrolytic depurination of the DNA. Thus the naturally desiccated nature of hair shafts provides a reasonable explanation as to why DNA survival may be prolonged in hair samples. Indirect evidence for this is provided by the low numbers of miscoding lesions observed among sequenced clones, which derive from hydrolytic deamination of cytosine and adenine bases (another form of damage associated with the interaction of DNA and water).

In this study it was not necessary to use BSA to counter inhibition, despite the facts that melanin is a known inhibitor of PCR (Uchichi *et al.* 1992) and aDNA studies that utilise bones or teeth as DNA sources are often hindered by environmentally-derived PCR inhibitors (such as fulvic or humic acids) (*e.g.* Pääbo *et al.* 1988). This is most likely also due to the lack of porosity in hairs, which in bones and teeth aids uptake of the inhibitors (Tuross 1994).

Without long-term sample preservation the survival of DNA in hair is of little consequence. The amino acid content of hair is reported to be relatively stable over time (Lubec *et al.* 1987; Lubec *et al.* 1994; Macko *et al.* 1999). Additionally, the hard

cuticle surface, strong disulfide–bonded molecular structure, and hydrophobic nature of hair proteins retards microbial degradation (Taylor *et al.* 1994; Taylor *et al.* 1995), makes hair somewhat resistant to enzyme attack (Lubec *et al.* 1987; Yu *et al.* 1993; Lubec *et al.* 1994; Macko *et al.* 1999) and render it insoluble in water, dilute acids and alkali, and various organic solvents at ambient temperatures (Barnett and Sognnaes 1962; Taylor *et al.* 1995). Thus hair should remain durable in environmental conditions. Nevertheless with the exception of mummified remains, hair is rarely reported at archaeological sites (*e.g.* Benfer *et al.* 1978; Massa and Fuhrman 1984; Bonnichsen and Bolen 1985; Lubec *et al.* 1987; Bonnichsen *et al.* 1994; Lubec *et al.* 1994; Chrisman *et al.* 1996; Loy and Dixon 1999; Wilson *et al.* 2001b). However extensive collections exist of hair in museums, both natural and in the form of ancient textiles (in particular those based on mammalian hair, *e.g.* wool). One possible problem with such samples is reports that some artificial hair treatments (*e.g.* peroxide bleaching) are damaging to DNA yields (Yoshii *et al.* 1992; Wilson *et al.* 1995; Baker 2001).

Although hair represents an excellent source of mtDNA, those aiming to use it must be aware of one potentially confounding factor. The enrichment of keratinocyte mitochondrial numbers through donation from melanocytes has been identified as a possible, if not probable source of mitochondrial heteroplasmy in hair shaft samples (Linch *et al.* 2001) - potentially arising as replication errors associated with the increased rate of mitochondrial reproduction, or simply as an effect of the large mitochondrial population. When coupled with the low DNA levels associated with ancient samples, this may result in erroneous results through the sequencing of heteroplasmous DNA extracts. In addition, recombination among different mitotypes through 'jumping PCR' (Pääbo *et al.* 1990) may further complicate this issue.

6.5 Conclusions

This study suggests that the exclusion of water from hair cells during the keratinisation of the hair fibre slows overall rates *post mortem* DNA degradation, despite *in vivo* containing far lower levels of mtDNA than other tissues. Additional benefits are conferred due to the hair cortex's cuticular encasement, both minimising the uptake of contaminant DNA, enabling simple and rapid sample decontamination, and resulting in the lack of absorption to environmental compounds which may act as inhibitors of PCR. This in turn removes the need of BSA or RSA in the PCR reaction (a further source of contaminant DNA). Finally, the relatively small quantities of samples used in the analyses ensure that minimal sample destruction is required in order to generate authentic aDNA results. With the development of formal sample pre-screening tools hair should become a very useful source of ancient DNA.

6.6 Acknowledgements

The authors would like to thank Malcolm Neame and Blakeway productions for researching and securing permission to sample the hairs attributed to Sir Isaac Newton, Prof Paul Lehrman for the Andaman islander sample, Duane Froese and Gimlex Mining, Dawson, for the Christie Mine Bison specimen, and Vyacheslav Molodin & Natalia Polos'mak at the Institute of Archaeology, Novosibirsk for the Pazyryk hair samples. M.T.P.G., A.S.W., T.C.O'C. and A.C. would like to thank the Wellcome Trust for funding their research. EW and AJH would like to thank VILLUM KANN RASMUSSEN FONDEN, Denmark for funding their research.

Chapter 7

Absence of *Y. pestis*-specific DNA in human teeth from five European excavations of putative plague victims

Modified from the 'Microbiology' publication

'Absence of Y. pestis-specific DNA in human teeth from five European excavations of putative plague victims'

Microbiology 150:341-354 (2004)

M Thomas P Gilbert[1], Jon Cuccui[2], William White[3], Niels Lynnerup[4], Richard W Titball[5], Alan Cooper[1], Michael B Prentice[2]

[1]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS
[2]Department of Medical Microbiology, Bart's and the London Medical School, 64 Turner St. London E1 2AD, UK
[3]Museum of London, 46 Eagle Wharf Road, London N1 7ED. UK
[4]Laboratory of Biological Anthropology, Institute of Forensic Pathology, University of Copenhagen, 1017, Copenhagen, Denmark
[5]Defence Science and Technology Laboratory, Porton Down, Salisbury, Wiltshire, SP4 0JQ, UK

Summary

This study reports the results of a collaborative study undertaken by two independent research groups to a) confirm recent PCR-based detection of *Yersinia pestis* DNA in human teeth from medieval plague victims in France, and b) to extend these observations over five different European burial sites believed to contain plague victims dating from the late 13th to 17th centuries. Several different sets of primers were used, including those previously documented to yield positive results on ancient DNA extracts. No *Y. pestis* DNA could be amplified from DNA extracted from 108 teeth belonging to 61 individuals, despite the amplification of numerous other bacterial DNA sequences. Several methods of extracting dentine prior to the DNA extraction were also compared. PCR for bacterial 16S rDNA indicated the presence of multiple bacterial species in 23 out 27 teeth DNA extracts where dentine was extracted using previously described methods. In comparison, positive results were obtained from only 5 out of 44 teeth DNA extracts for which a novel contamination-minimising embedding technique was used. Therefore, high levels of environmental bacterial DNA are present in DNA extracts where previously described methods of tooth manipulation are used. To conclude, the absence of *Y. pestis*-specific DNA in an exhaustive search using specimens from multiple putative European plague burial sites does not allow us to confirm the identification of *Y. pestis* as the aetiological agent of the Black Death and subsequent plagues. In addition, the utility of the published tooth-based ancient DNA technique used to diagnose fatal bacteraemias in historical epidemics still awaits independent corroboration.

7.1 Introduction

The Black Death is the name given to a pandemic which killed up to a third of the European population between 1347 and 1352 (Cohn 2002; Scott and Duncan 2001). Over the next three hundred years this pandemic was followed by further plagues (deadly epidemics or pestilences) of lesser mortality. Whether these historically described diseases correspond to bubonic plague, whose aetiological agent was identified to be the bacterium now known as *Yersinia pestis* by Alexandre Yersin in 1894 (Yersin 1894) has been the subject of controversy (Cohn 2002; Scott and Duncan 2001). Recently, DNA specific for *Y. pestis* was amplified from 16th and 18th century human teeth believed to be from French plague victims (Drancourt *et al.* 1998) and 14th century French Black Death victims (Raoult *et al.* 2000). The lead authors of these reports believe that the consideration of any cause for the Black Death other than *Y. pestis* is now speculative (Raoult and Drancourt, 2002).

The aetiology of the Black Death is of major historical interest, but there are other significant consequences of resolving the issue of attribution. Knowledge of aetiology is important to understand any possible evolutionary impact of the associated mortality (Stephens *et al.* 1998) and to assess our ability to control any similar contemporary disease. Many scientific resources are now available to fight a future re-emerging human pandemic of *Y. pestis* infection. Well-established protocols exist to diagnose, treat and prevent transmission (WHO 1999), the genome has been sequenced (Parkhill *et al.* 2001; Deng *et al.* 2002), and a subunit vaccine is in clinical trials (Titball and Williamson 2001). These resources would be ineffective if a re-emerged Black Death had a different aetiology. Therefore, for historical and public health reasons alone, independent replication of the observations of *Y. pestis* DNA in

ancient remains from a small geographical area seems desirable. The nature of ancient DNA evidence makes independent confirmation a necessity.

The study of ancient DNA (aDNA) involves extraction and analysis of DNA from the remains of organisms preserved as fossils, skeletons, or mummified tissues. Pathogen DNA has been reported in a range of ancient animal and human remains. Many of the reports are from skeletons exhibiting paleopathological evidence of the disease. This includes reports of *Mycobacterium leprae* DNA from skeletal remains (Rafi *et al.* 1994), and DNA from various members of the *Mycobacterium tuberculosis* complex from human skeletal (Salo *et al.* 1994; Taylor *et al.* 1996) bison skeletal (Rothschild *et al.* 2001) and human mummified tissue (Nerlich *et al.* 1997; Fletcher *et al.* 2003a). In most of these cases structurally un-modified bone also yielded mycobacterial DNA. In the case of *Yersinia pestis*, fatal infection would not be expected to leave any specific bony changes, so no osteological confirmation is available and any retrospective diagnosis is completely DNA-based.

Ancient DNA studies are hampered by extremely low levels of DNA preservation, often coupled with the presence of much greater levels of modern contaminants. Characteristically, only short ancient DNA fragments (less than 300 bp) can be amplified (Richards *et al.* 1995; Höss *et al.* 1996; Hofreiter *et al.* 2001b) and easy amplification of longer fragments is an indication that contamination has occurred. Contaminants normally arise from three sources; a) modern equivalents of the source species, b) previously PCR amplified DNA (amplicons) from the source species, or c) similar species in the environment (of especial importance in the study of microorganisms) (Gutierrez and Marin 1998). Unrecognised contamination as a source for positive results is so insidious and difficult to prevent that in many cases it only comes to light when results from one laboratory cannot be confirmed by other

groups and a laboratory-specific contaminant is revealed (Austin *et al.* 1997; Gutierrez and Marin, 1998). A series of strict criteria have been proposed for research standards in this field (Cooper and Poinar, 2000). Possibly the most important of these criteria is the independent replication of results by other groups.

Two studies from the same research group have been published reporting the successful extraction, amplification and direct sequencing from PCR products of *Y. pestis*-specific DNA retrieved from the dental pulp cavities of plague victims (Drancourt *et al.* 1998; Raoult *et al.* 2000). Findings that pathogen-specific DNA can be recovered from this source in systemically infected animals (Aboudharam *et al.* 2000) have led Drancourt *et al.* (1998) to hypothesise that teeth provide a lasting, contamination-free refuge where pathogen aDNA may survive.

This study presents the results of attempts by two independent research groups to amplify and sequence ancient *Y. pestis* DNA from human teeth. Samples were obtained from five archaeological sites in Northern Europe, and were analysed using a range of PCR primers that were designed to be *Y. pestis*-specific. Three methods of aDNA extraction are also both directly and indirectly compared in order to assess the suitability of each method in preventing contamination of the DNA source.

7.2 Materials and methods

7.2.1 Samples

Teeth were chosen as sources of aDNA as in previous studies. Most teeth used were erupted i.e. teeth from adults. In the first publication from the Marseille group, unerupted teeth i.e. mainly teeth from non-adults were used because these were thought to be better protected from environmental contamination (Drancourt *et al*. 1998), but the second Marseille paper reported satisfactory results with adult erupted teeth from 14th century remains (Raoult *et al*. 2000) and we attempted replication on this type of specimen. A range of samples from five archaeological sites were analysed to help ensure that at least some of the individuals died as a result of a plague infection (Table 7.1). The majority of samples (77/108) originated in two well-documented plague burial sites in Copenhagen and London. Four human teeth samples from a Viking burial site in York, that antedate the arrival of the Black Death in Britain (pre-11th century), and two human teeth from Farringdon (London) buried long after the last local plague outbreak (19th century), were also analysed as negative controls. Samples from these archaeological sites are referred to as 'YO' and 'FA' respectively in this study.

7.2.1.1 Copenhagen Site

Building construction in Copenhagen, Denmark, led to the excavation of a cemetery in 1991. Archaeological identification of this cemetery as the historically-recorded Vodroffsgaard plague cemetery was possible due to the geographic location, historical records, and burial finds such as contemporary coffins, (Ringboel Bitsch 1991). This cemetery dates from the last major plague epidemic in Denmark, which struck Copenhagen in 1711-1712 AD. At that time Vodroffsgaard was a farmstead

Table 1: Summary of samples examined in study.

Site*	Country†	Abbreviation‡	Details§	Individuals⇑	Teeth¶	Source#
Copenhagen	Denmark	CP	Plague pit	34	50	N.Lynnerup
Royal Mint, London	S England	RM	Plague pit	8	27	W.White
Verdun	N France	VE	Plague pit	13	14	E.Carniel, F Adam
Angers	N France	AN	Plague pit/Catastrophe grave	1	5	H.Mollaret
Spitalfields, London	S England	SP	Plague pit/Catastrophe grave	5	12	W.White
York	N England	YO	Negative Control	2	4	A. Cooper
Farringdon, London	S England	FA	Negative Control	1	2	W.White

**Site* Archaeological site of origin of specimens. †*Country* Country of origin of specimens. ‡*Abbreviation* Site name abbreviation. §*Details* Specific details of site. ⇑*Individuals* Number of individual skeletons from which teeth were extracted. ¶*Teeth* Total number of teeth investigated for DNA. *#Source* Original source of samples.

outside the city walls, and the farm and property were requisitioned by the authorities to be used as a plague hospital and cemetery (Von Kohl 1911). The excavations in 1991 revealed 54 graves. Subsequent anthropological analyses showed that the extracted human remains represented 57 individuals (13 children and adolescents, 20 adult males and 24 adult females (Lynnerup 1992). Copenhagen samples are referred to as 'CP' in this paper.

7.2.1.2 Royal Mint Site

Purchase and consecration of land for a burial ground at East Smithfield (to the east of the walled city of London) is recorded during the first Black Death epidemic of 1349 (Hawkins 1990). The East Smithfield cemetery on the site of the old Royal Mint was excavated between June 1986 and June 1988 revealing a mass burial pit, two mass burial trenches and 14 rows of stratigraphically contemporary individual graves. A total of 600 individuals were excavated and these were the source for our specimens. Previous studies on these remains have provided much useful archaeological and anthropological data on Black Death plague victims (Hawkins 1990; Waldron 2001). Royal Mint samples are referred to as 'RM' in this paper.

7.2.1.3 Verdun Site

Excavations in Verdun, north-eastern France, of the Hospice Sainte Catherine on a monastery site revealed several multiple graves and two burial pits containing 21, and 26 individuals, respectively. Burial pits of this kind on ancient hospital sites are associated with fatal disease epidemics. The teeth processed were from the larger burial pit, dated from clothing remnants and other artefacts to be from the late 17th/early 18th century. Monastery archives document 16 outbreaks of plague on the

site from the 16th to the 18th century, making this a probable plague pit. Verdun samples are referred to as 'VE' in this paper.

7.2.1.4 Angers Site

Minor earthworks in the Place de la Paix, Angers, France in 2001 revealed a mass grave of up to 1000 individuals that was not recorded in the city archives. Although archaeological analysis has not yet been undertaken, the large scale and unrecorded nature of the burial site represents a catastrophic event in a city noted for multiple severe plague outbreaks (Scott and Duncan 2001), making this a possible plague pit. Angers samples are referred to as 'AN' in this paper.

7.2.1.5 Spitalfields Site

The Spitalfields site is located just outside the walls of Roman and medieval London. The main medieval findings relate to the priory and hospital of St. Mary Spital founded in 1197 and dissolved by Henry VIII in 1538. Excavation by the Museum of London 1999-2001 revealed a large burial ground adjacent to the hospital that appears to have been in use throughout this period. One region of the burial ground is a mass grave with multiple skeletons buried together in a disorderly fashion suggesting mortality in a major epidemic. Carbon-14 dating of remains from this part of the excavation assigns them to the late 13th century (*i.e.* slightly earlier than 1349 when the Black Death was first recognised in London). Spitalfields samples are referred to as 'SP' in this study.

7.2.2 DNA extraction and analysis

Samples were analysed independently by two groups to control for laboratory-specific failures, Oxford University's Ancient Biomolecules Centre (ABC) and St. Bartholomew's Hospital, London (SBH). The ABC is a dedicated aDNA facility where no work on modern DNA is ever undertaken. At SBH, most aDNA extraction and PCR setup was undertaken in a specialised virology laboratory where no bacterial work is undertaken. Strict aDNA protocols were followed by both groups (Cooper and Poinar 2000). This included the independent analysis of pairs of teeth from eight individuals excavated at three archaeological sites by both groups.

7.2.2.1 Assessment of dentine extraction techniques to minimise DNA extract contamination

Teeth samples must be powdered prior to DNA extraction. To remove surface contamination teeth were treated with bleach – those used at the ABC through immersion for 10 minutes in 50% bleach solution, those at SBH were wiped but not immersed in 50% bleach. Post-cleansing all teeth were exposed to UV light at λ=325nm for 20 minutes. Three methods of tooth dentine extraction were then performed at the ABC to compare efficiency at a) preventing contamination entering the DNA extract, and b) retrieving DNA.

1) Eighteen teeth (14VE/2YO/2CP) were directly ground into powder ('ground') using a microdismembrator (Braun, Germany).
2) Nine teeth (5VE/2YO/2CP) were powdered using the method pioneered by Drancourt *et al.* (1998). This consists of pulp cavity removal by scraping with a dental tool following longitudinal fracturing of teeth ('scraped').

3) Fifty-three teeth (5AN/2RM/2SP/44CP) were encased in silicone rubber ('silicone') prior to the removal of dentine with a dental drill as reported in chapter 3 (Fig 3.1). This method of powder extraction was developed due to concerns about the efficiency of contaminant removal from the porous outer surface teeth, and that surviving contaminant DNA may be transferred to extractions as the teeth are handled.

The efficiency of the methods was compared and contrasted by performing extractions using more than one method on some teeth, and testing multiple teeth extracted from single skulls to ensure that the results were not sample or site specific. A total of 69 different teeth from possible plague sites were therefore processed at the ABC.

To directly compare extracts from 'scraped' and 'ground' teeth, seven teeth from seven individuals were first 'scraped', following which the remains were 'ground'. To compare all three methods, two teeth were taken from each of two individuals. The first tooth from each individual was 'scraped' to remove some dentine, after which the remainder of the tooth was 'ground' (as above). Dentine was removed from the second tooth using the 'silicone' method. As a further control, two of the samples that were 'scraped' then 'ground' were from the burial site at York that is not believed to have contained victims of the plague.

7.2.2.2 DNA extractions at the ABC

DNA extractions on the powdered samples followed a standard phenol:chloroform method modified for aDNA samples, as in Cooper *et al*. (2001).

7.2.2.3 Dentine removal at SBH

In total 39 medieval teeth were examined at St Bartholomew's Hospital. Teeth were wiped with bleach and then exposed to UV light as at the ABC. The 'scrape' method used in the initial *Y. pestis* aDNA study (Drancourt *et al*. 1998) was used to obtain dentine for DNA extraction from 17 medieval specimens and two 19th century Farringdon control teeth (DNA extractions BT1-24). Subsequently the ABC 'silicone' encasement technique was adopted for 22 further teeth (DNA extractions BT25-46).

7.2.2.4 DNA extractions at SBH

DNA extractions BT1-10 employed the technique described in the previous study (Drancourt *et al*. 1998). DNA extractions BT12-24 employed a different method, whereby dentine was placed into 1.5 ml DNA extraction buffer (5M Guanidinium isocyanate, 1.3% Triton X –100, 0.1M Tris-HCl pH 6.4, 0.02M EDTA pH 8.2) and agitated at 48°C overnight. Subsequently DNA recovery followed a silica-based extraction protocol (Boom *et al*. 1990) modified for aDNA (Höss *et al*. 1996). An intermediate pre-amplification step after DNA extraction and before PCR using random 15-mers following the method of Zhang *et al*. (1992) was also applied to these specimens only (BT12-24) and not to any other specimens processed**.** The remaining DNA extractions (BT25-46) were undertaken using a commercial kit (GENECLEAN Kit for Ancient DNA, Qbiogene).

7.2.3 PCR amplification

7.2.3.1 Primers

The presence of *Y. pestis* DNA in DNA extracts was monitored using PCR amplification, followed by cloning and sequencing where relevant. Primers, taken from both published aDNA studies, and of new design, were employed against three bacterial targets. One of these was not specific for *Y. pestis* – targeting part of the 16S ribosomal RNA gene. The other two were *Y. pestis* specific: the plasmid-encoded plasminogen activator gene (*pla*), and the chromosomal RNA polymerase β-subunit encoding gene (*rpoB*). All primer sets were optimised on a small sample of extractions. Samples were also assayed for endogenous human DNA using a selection of human mitochondrial Hyper Variable Region 1 (HVR1) and nuclear β-globin DNA primers. For full primer details refer to Table 7.2. Specific details of PCR cycles are provided as Appendix 4a.

7.2.3.2 Positive controls

The use of positive controls was avoided at the ABC in order to reduce the problem of sample cross-contamination that has been reported in aDNA studies (Kolman, 1999). However, two positive controls were initially used at SBH (in a separate building to that used for DNA extraction) for optimising PCRs. One control was designed to be used with published primers (Drancourt *et al*. 1998) (PlaB control), the other was designed for this study (PlaA control). Both consisted of target DNA, though with a central deletion removing a *DraI* site and inserting a *HindIII* site, thus yielding smaller amplicons than the target.

Table 7.2: Primers used in the study.

Group[1]	Primer[2]	Code[3]	°C[4]	Sequence[5]	+ve control[6]	Source
ABC	Yprpob1¶	A	58	aac acc tta tcg tcg tgt acg t	N	Drancourt *et al*. 1998
ABC	Yprpob2¶	A	58	aat ctt cta aaa agc ggc ctt ca	N	Drancourt *et al*. 1998
ABC	Yprpob267¶	B	58	gga aac acc tta tcg tcg tgt acg	N	This study
ABC	Yprpob403¶	B	58	cca aat ctt cta aaa agc ggc c	N	This study
ABC	Yp16s387H#	C	58	ccg cgc tta acg tgg gaa c	N	This study
ABC	Yp16s604L#	C	58	gtc gac atc gtt tac agc	N	This study
ABC	Yp16s586H#	D	58	gct gta aac gat gtc gac	N	This study
ABC	Yp16s786L#	D	58	ctg tgg atg tca aga gta gg	N	This study
ABC/SBH	YPpla728H (YP12D) ◊	E	52	cag cag gat atc agg aaa ca	N	Raoult *et al*. 2000
ABC/SBH	YPpla876L (YP11R) ◊	E	52	gca ag tcc aat ata tgg cat ag	N	Raoult *et al*. 2000
SBH	L16131α	F	56	cac cat gaa ta t gta cgg t	N	Handt *et al*. 1996
SBH	H16218α	F	56	tgt gtg ata gtt gag ggt tg	N	Handt *et al*. 1996
ABC/SBH	H16209α	G	56	cca tgc tta caa gca ag	N	Handt *et al*. 1996
ABC/SBH	H16356α	G	56	gtc atc cat ggg gac gag aa	N	Handt *et al*. 1996
SBH	Pla A (K1) ◊	H	59	aag gag tgc ggg taa tag gt	Y	This study
SBH	Pla A (K2) ◊	H	59	gat gtc ttc tca cgg aaa gt	Y	This study
SBH	Pla B (Pla1) φ	I	53	ctt gga tgt tga gct tcc ta	Y	Hinnebusch *et al*. 1993
SBH	Pla B (Pla2) ◊	I	53	gag atg ctg ccg gta ttt cc	Y	Drancourt *et al*. 1998
SBH	Pla C (7242F)(YP11D) ◊	J	52	cta tgc cat ata ttg gac ttg c	N	Raoult *et al*. 2000
SBH	Pla C (7389R)(YP10R) ◊	J	52	gag ccg gat gtc ttc tca cg	N	Raoult *et al*. 2000
SBH	PCO3⊥	K	56	aca caa ctg tgt tca cta gc	Y	Greer *et al*. 1991
SBH	PCO4⊥	K	56	caa cta cat cca acg ttc acc	Y	Greer *et al*. 1991
SBH	blaTEM 1-20 χ	L	55	atg agt att caa cat ttc cg	N	Essack *et al*. 2001
SBH	blaTEM 858-840 χ	l	55	cca atg ctt aat cag tga c	N	Essack *et al*. 2001
SBH	Yp16S805#	M	55	agc ggt gga gca tgt ggt tt	N	This study
SBH	Yp16S956#	M	55	cca aca ttt cac aac acg ag	N	This study

[1]*Group* Research group where primer pair was used. [2]*Primer* Designated primer name. Further details on primers are as follows: ¶*Y.pestis* RNA polymerase gene. #*Y.pestis* β-subunit of the16s ribosomal RNA gene. ◊*Y.pestis* plasminogen activator encoding gene (pla). α*H.sapiens* mitochondrial HVR-1 region. φ*Y.pestis* transcriptional regulator adjacent to pla gene. ⊥*H.sapiens* nuclear β-globin. χBacterial plasmid RP4. For PCR cycle parameters see appendix 4b. [3]*Code* Primer pair abbreviation code. [4]*°C* Annealing temperature of primers. [5]*Sequence* 5'→3' sequence of primer. [6]*+ve control* Use of positive control at any time in study with this primer set. Further PCR cycle parameters are given in Appendix 4a.

7.2.3.3 PCR amplification, cloning and sequence analysis - ABC

The amplification, reamplification, cloning and PCR methodology used at the ABC are as reported elsewhere (Cooper *et al*. 2001). Where PCR products yielded multiple DNA bands, those of the expected size were excised from the gel for purification and reamplification. All PCR products of the expected size were sequenced. Twenty-three PCR products from the ABC were also cloned prior to sequencing. Attempts were made to identify all sequences using the NCBI BLAST tool.

7.2.3.4 PCR amplification, cloning and sequence analysis - SBH

At SBH initial PCR (DNA extracts BT1-10) was performed with AmpliTaq (Applied Biosystems), later switching to AmpliTaq Gold (Applied Biosystems) (DNA extracts BT12-24) and latterly Platinum Taq Hifidelity (Invitrogen) (DNA extracts BT 25-46) as at the ABC. In all cases PCR was performed in 25µl volumes with 1µl of DNA extract. For DNA extracts BT1-24 the PCR mix comprised 2.5 units of Taq in Perkin Elmer PCR buffer with 1.5mM $MgCl_2$, 200µM of each dNTP and 2µM of each oligonucleotide primer

7.3 Results

7.3.1 Suitability of dentine extraction techniques

The amplification results from seven teeth where dentine was both 'scraped' and 'ground' prior to DNA extraction highlight that neither method reduces the presence of amplifiable contaminant DNA (Table 7.3). However the direct comparison between DNA extracts from samples where dentine was extracted from 'scraped' and 'silicone' teeth shows a clear reduction in levels of contaminants in 'silicone' treated teeth (Table 7.3). Following this observation all further ABC and SBH work utilised the 'silicone' method.

7.3.2 *Y. pestis* PCR results

In total, DNA extractions were performed on 18 'ground' (14VE/2YO/2CP), 9 'scraped' (5VE/2YO/2CP) and 53 'silicone' teeth (5AN/2RM/2SP/44CP) at the ABC (Table 7.4: this table includes the teeth processed with multiple methods listed in Table 7.3), and 19 'scraped' (17RM/2FA) and 22 'silicone' teeth (8RM/10SP/4CP) from SBH (Table 7.5). This represents a total of 115 extractions on 108 different teeth. All extracts were tested for specific *Y. pestis* pesticin plasmid sequences and 71 were tested for *Y. pestis rpoB* chromosomal DNA. A summary of results for teeth from individuals tested in both laboratories is presented in Table 7.6. In both laboratories no extraction blanks or negative controls yielded PCR products with *Y. pestis*-specific PCR primers. At SBH, a positive result was obtained from a tooth DNA extract using *pla* assay 'H' (see Tables 7.2 and 7.5) when controls remained negative. Cloning and sequencing revealed that this band was identical to the engineered positive control fragment that contained a central indel compared with the native sequence. This sample was one of twelve samples which underwent

Table 7.3: Direct comparison of the amplimers generated using 4 primer sets, from DNA extracts obtained using the 'Silcone', 'Scraped' and 'Ground' extraction methods.

				Primer Sets⇑			
Skeleton*	Site†	Tooth	Method‡	A	B	D	E
V1	VE	Molar	Scraped	-	m	-	M
			Ground	s	m	S	S
V10	VE	Molar	Scraped	-	m	S	-
			Ground	-	m	S	s
V4	VE	Molar	Scraped	-	m	S	S
			Ground	m	m	S	S
V3	VE	Molar	Scraped	s	-	S	M
			Ground	-	m	S	s
V7	VE	Molar	Scraped	-	-	-	-
			Ground	s	m	S	s
4245	YO	Molar	Scraped	-	m	S	-
			Ground	s	m	S	S
2654	YO	Molar	Scraped	M	m	S	S
			Ground	M	m	S	-
AR	CP	Molar	Scraped	-	m	m	-
			Ground	s	m	m	-
			Silicone	-	-	-	-
CG	CP	Molar	Ground	s	m	m	-
			Scraped	-	m	-	-
			Silicone	-	-	-	-

**Skeleton* Specimen from which teeth were extracted. †*Site* Archaeological site of origin of sample. For abbreviation details refer to Table 1. ‡*Method* Extraction method used to extract dentine prior to DNA extraction. Briefly 'Silicone' refers to samples encased in silicone rubber prior to dentine removal, following chapter 3. 'Ground' refers to dentine recovery following complete grinding of specimens. 'Scraped' refers longitudinal fracturing of teeth after which dentine was recovered from the pulp cavity by scraping with a dental pick (following Drancourt *et al*. 1999). ⇑*Primer Sets* Primer pairs used for PCR. For abbreviation details refer to Table 7.2.

Further notes on amplifications: s=single PCR product amplified, S=single PCR product of expected size amplified, m=multiple PCR products amplified, M=multiple PCR products amplified, containing band of expected size, -=no PCR product amplified.

Table 7.4: Results of PCR amplification undertaken at the ABC.

				Primer Sets						
Skeleton	Site	Tooth	Method	A	B	C	D	E	F	G
V2	VE	Molar	Ground	m	m	-	S¶	s	S	N/A
V8	VE	Molar	Ground	m	m	-	S¶	S§	S	N/A
V14	VE	Molar	Ground	m§	m	-	S¶	-	S	N/A
V9	VE	Molar	Ground	-	m	S¶	-	s	S	N/A
V1	VE	Molar	Scraped	-	m	-	-	M§	S	N/A
V1	VE	Molar	Ground	s	m	-	S¶	S~	-	N/A
V6	VE	Molar	Ground	-	m	-	S¶	S§	S	N/A
V10	VE	Molar	Scraped	-	m	-	S¶	-	S	N/A
V10	VE	Molar	Ground	-	m	s	S¶	s¶	S	N/A
V12	VE	Molar	Ground	m	S	S¶	S¶	s¶	-	N/A
V4	VE	Molar	Scraped	-	m	S¶	S¶	S§	S	N/A
V4	VE	Molar	Ground	m§	m	-	S¶	S§	S	N/A
V5	VE	Molar	Ground	M§	m	S¶	S¶	-	S	N/A
V3	VE	Molar	Scraped	s	-	-	S¶	M§	S	N/A
V3	VE	Molar	Ground	-	m	-	S¶	s	S	N/A
V7	VE	Molar	Ground	s	m	S¶	S¶	s¶	S	N/A
V11	VE	Molar	Ground	M§	m	S¶	S¶	s¶	S	N/A
V13	VE	Molar	Ground	m	m	S¶	S¶	s¶	S	N/A
4245	YO	Molar	Scraped	-	m	S¶	S¶	-	S	N/A
4245	YO	Molar	Ground	s	m	-	S¶	S~	-	N/A
2654	YO	Molar	Scraped	M§	m	-	S¶	S~	S	N/A
2654	YO	Molar	Ground	M§	m	-	S¶	-	S	N/A
V7	VE	Molar	Scraped	-	-	-	-	-	-	N/A
S32_BK	CP	Molar	Silicone	M§	N/A	N/A	S§	-	N/A	S
AE	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	-
XA	CP	Molar#	Silicone	-	N/A	N/A	-	-	N/A	S
BN	CP	Molar#	Silicone	-	N/A	N/A	-	-	N/A	-
CR	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	-
CP	CP	Molar#	Silicone	-	N/A	N/A	-	-	N/A	-
CV	CP	Molar	Silicone	-	N/A	N/A	S§	-	N/A	-
AD	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	-
BN	CP	Molar	Silicone	-	N/A	N/A	S§	-	N/A	-
DF	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	-
DB	CP	Molar	Silicone	M§	N/A	N/A	S§	-	N/A	S
XA	CP	Molar#	Silicone	-	N/A	N/A	S§	-	N/A	S§
BT	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
BR	CP	Premolar	Silicone	-	N/A	N/A	-	-	N/A	S
AO	CP	Molar#	Silicone	-	N/A	N/A	-	-	N/A	S
AP	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
CY	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
S34_L	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
BN	CP	Molar#	Silicone	-	N/A	N/A	-	-	N/A	-
CH	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
AI	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S

				Primer Sets						
Skeleton	Site	Tooth	Method	A	B	C	D	E	F	G
CA	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
DE	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
AS	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	-
CM	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
XA	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
CM	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
XA	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
AH	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S§
S60b	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
BU	CP	Molar#	Silicone	-	N/A	N/A	-	-	N/A	S
AO	CP	Molar#	Silicone	-	N/A	N/A	-	-	N/A	-
XD	CP	Premolar	Silicone	-	N/A	N/A	-	-	N/A	S§
BU	CP	Molar	Silicone	-	N/A	N/A	-	-	N/A	S
BN	CP	Molar#	Silicone	-	N/A	N/A	-	-	N/A	s¶
BU	CP	Molar#	Silicone	-	N/A	N/A	-	-	N/A	s¶
6181	RM	Molar	Silicone	-	N/A	N/A	-	-	N/A	s¶
3780	SP	Molar	Silicone	-	N/A	N/A	-	-	N/A	s¶
3817	SP	Molar	Silicone	-	N/A	N/A	-	-	N/A	s¶
9830	RM	Molar	Silicone	-	N/A	N/A	-	-	N/A	-
C1	AN	Premolar	Silicone	N/A	N/A	N/A	N/A	-	N/A	-
C1	AN	Premolar	Silicone	N/A	N/A	N/A	N/A	-	N/A	-
C1	AN	Premolar	Silicone	N/A	N/A	N/A	N/A	-	N/A	-
C1	AN	Canine	Silicone	N/A	N/A	N/A	N/A	-	N/A	-
C1	AN	Premolar	Silicone	N/A	N/A	N/A	N/A	-	N/A	-
AR	CP	Molar	Silicone	-	N/A	N/A	N/A	-	N/A	S¶
CG	CP	Molar	Silicone	-	N/A	N/A	N/A	-	N/A	-
CB	CP	Molar	Silicone	-	N/A	N/A	N/A	-	N/A	-
CN	CP	Molar	Silicone	-	N/A	N/A	N/A	-	N/A	-
AR	CP	Molar	Scraped	-	m	N/A	m	-	N/A	N/A
AR	CP	Molar	Ground	s	m	N/A	m	-	N/A	N/A
CG	CP	Molar	Ground	s	m	N/A	m	-	N/A	N/A
CG	CP	Molar	Scraped	-	m	N/A	-	-	N/A	N/A
AR	CP	Molar	Silicone	-	-	N/A	-	-	N/A	N/A
CG	CP	Molar	Silicone	-	-	N/A	-	-	N/A	N/A

Table subheadings and abbreviations are as in Table 7.3 with the addition of the following: N/A=not tested. § Denotes amplicons that were cloned and sequenced. ¶ Denotes amplicons that were directly sequenced. ~ Denotes amplicons that could not be cloned. # Denotes unerupted molar. For PCR primer set abbreviation details refer to Table 7.2.

Table 7.5: Results of PCR amplification, SBH.

Sample*	Skeleton	Site	Tooth	Method	PCR primer pair					
					E	G	H	I	J	K
BT1/A1	11314	RM	Molar	Scraped	N/A	N/A	N/A	-	N/A	-
BT2/A2	11314	RM	Molar	Scraped	N/A	N/A	N/A	-	N/A	S
BT4/M1	1533	FA	Molar	Scraped	N/A	N/A	N/A	-	N/A	-
BT6/A3	11143	RM	Canine	Scraped	N/A	N/A	N/A	-	N/A	-
BT8/A4	11314	RM	Incisor	Scraped	N/A	N/A	N/A	-	N/A	-
BT9/M2	1533	FA	Canine	Scraped	N/A	N/A	N/A	-	N/A	-
BT10/A5	11314	RM	Incisor	Scraped	N/A	N/A	N/A	-	N/A	-
BT12/A6	11034	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT13/A7	11034	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT14/A8	11034	RM	Incisor	Scraped†	N/A	N/A	-	N/A	N/A	S
BT15/A9	9830	RM	Molar	Scraped†	N/A	N/A	M‡§	N/A	N/A	S
BT17/A10	9830	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT18/A11	5383	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT19/A12	11034	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT20/A13	9830	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT21/A14	6181	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT22/A15	9830	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT23/A16	6181	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT24/A17	8440	RM	Molar	Scraped†	N/A	N/A	-	N/A	N/A	-
BT25/A18	9859	RM	Molar	Silicone	N/A	s	N/A	N/A	-	N/A
BT26/A19	7440	RM	Molar	Silicone	N/A	s	N/A	N/A	-	N/A
BT27/A20	19009	SP	Molar	Silicone	N/A	-	N/A	N/A	-	N/A
BT28/A21	3780	SP	Molar	Silicone	N/A	-	N/A	N/A	-	N/A
BT29/A22	11039	RM	Incisor	Silicone	N/A	-	N/A	N/A	-	N/A
BT39/A23	9830	RM	Incisor	Silicone	N/A	s	N/A	N/A	-	N/A
BT31/A24	3816	SP	Molar	Silicone	N/A	-	N/A	N/A	-	N/A
BT32/A25	3815	SP	Incisor	Silicone	N/A	-	N/A	N/A	-	N/A
BT33/A26	5383	RM	Molar	Silicone	N/A	-	N/A	N/A	-	N/A
BT34/A27	3817	SP	Molar	Silicone	N/A	-	N/A	N/A	-	N/A
BT35/A28	5383	RM	Molar	Silicone	-	-	N/A	N/A	-	-
BT36/A29	9830	RM	Incisor	Silicone	-	S	N/A	N/A	-	-
BT37/A30	19009	SP	Molar	Silicone	-	-	N/A	N/A	-	-
BT38/A31	3816	SP	Molar	Silicone	-	M	N/A	N/A	-	-
BT39/A32	3817	SP	Molar	Silicone	-	-	N/A	N/A	-	-
BT40/A33	6181	RM	Incisor	Silicone	-	M‡	N/A	N/A	-	-
BT41/A34	3780	SP	Incisor	Silicone	-	-	N/A	N/A	-	-
BT42/A35	3815	SP	Premolar	Silicone	-	-	N/A	N/A	-	-
BT43/A36	BR	CP	Incisor	Silicone	-	S	N/A	N/A	-	-
BT44/A37	CB1	CP	Premolar	Silicone	-	M	N/A	N/A	-	-
BT45/A38	AR	CP	Premolar	Silicone	-	M‡	N/A	N/A	-	-
BT46/A39	AE	CP	Premolar	Silicone	-	-	N/A	N/A	-	-

Headings and notation as in Tables 2, 3 and 4, with the exception of: **Sample* DNA extraction name, †Samples extracted with Guanidinium isocyanate-silica method. ‡Cloned and sequenced. §Identified as contaminated with PCR-positive control DNA.

Table 7.6: Summary of presence (+) and absence (-) of amplifiable *Y. pestis* DNA within teeth DNA extracts taken from skeletons investigated by both research groups.

	Primer Sets							
Skeleton	M†	H†	I†	J†	E†	E‡	D‡	A‡
6181	n/a	-	n/a	n/a	n/a	-	n/a	N/a
3780	n/a	-	n/a	-	n/a	-	n/a	N/a
3817	n/a	-	n/a	-	n/a	-	n/a	N/a
9830	n/a	+§	n/a	n/a	n/a	-	n/a	N/a
BR	-	n/a	-	-	-	-	-	-
CB1	-	n/a	-	-	-	-	-	-
AR	-	n/a	-	-	-	-	-	-
AE	-	n/a	-	-	-	-	-	-

Headings and abbreviations as in Tables 7.2 and 7.3 with the addition of the following: †primers sets used at SBH, ‡primer sets used at ABC, +§Positive control contaminant amplified from one extract (Three other teeth from the same individual tested negative with the same primers).

preamplification at SBH as specified above. Preamplification was abandoned at SBH following this episode. At ABC 5 samples (4 ground, 1 scraped) yielded a single band of the expected size with *Y. pestis* pPst (pesticin plasmid)-specific primers. Four of these bands could be cloned and sequenced and none contained *Y. pestis*-specific sequence. No further bands were obtained with these primers at ABC or SBH following adoption of the silicone embedding technique. One tooth extracted by the silicone method produced multiple bands in the *rpoB* PCR assay. The correct size band was cloned and the sequence of one clone contained a short match to *Y. pestis rpoB,* but this was found to represent a primer dimer of one of the PCR primers (Yprpob1).

Of four control samples at the ABC from the burial site not thought to contain plague victims (York), two (one ground, one scraped) yielded PCR products with *Y. pestis*-specific primers but these products could not be cloned. No directly sequenced PCR products yielded readable DNA, indicating the presence of multiple DNA sequences of similar size in each amplified extract. Such products yielded up to 10 different sequences each when cloned at the ABC. No sequences were *Y. pestis*-specific DNA. All sequences obtained were searched using BLAST. The majority were reported as unknown, though some were loosely similar to other bacteria genera. Within the clones of two samples amplified using 16S rDNA primers, single sequences matching, but not discriminating between, numerous proteobacteria including *Y. pestis, Y. pseudotuberculosis* and *Y. enterocolitica* were found. A data file containing all cloned bacterial DNA sequences and database matches is included as Appendix 4b.

7.3.3 Human PCR results

Human amplification products of the expected size were obtained in 53 out of 89 mitochondrial amplifications performed at the ABC, and 3 out of 39 nuclear amplifications performed at SBH. No extraction blanks at the ABC yielded human DNA. Nevertheless, no attempt was made to sequence human products from samples extracted without silicone at the ABC, due to the obvious bacterial, and therefore likely human sample contamination. All human sequences are commonly found among Europeans, and do not match any member of the Oxford laboratory (data not shown). At SBH, 9 out of 22 mitochondrial DNA amplifications yielded bands of the expected size. However, extraction blanks and the PCR control also yielded bands with this PCR. Cloning and sequencing of a contaminating band from one of the extraction controls and bands obtained from two teeth from different test subjects showed the contaminating band to be human mtDNA distinct both from that found in the two test subjects and also the individual who performed the DNA extraction and PCR. As this study has only a marginal interest in human DNA, the authenticity of the mtDNA amplified and cloned at the ABC has not been as exhaustively tested as would be appropriate in a study of human evolution.

At SBH no positive amplification results were obtained in the beta-globin control PCR following adoption of the silicone-encasement technique and the use of a commercial kit for DNA extraction (specimens BT25-46). One ml of DNA extract was added to a control PCR for TEM beta lactamase (Essack *et al*. 2001) containing 4.3 pg of target DNA (PCR carried out in 25ml volumes with Platinum Taq DNA polymerase as for the ABC standard method) to ensure that negative results in this assay and assays for *Y. pestis*-specific DNA were not due to PCR inhibitors in the DNA extracts. PCR inhibition was detected in two out of 22 tested extracts.

7.4 Discussion

7.4.1 Previous extraction techniques are unsuited to preventing bacterial contamination of the DNA extract

Almost all previous work that has employed teeth as a source of aDNA has used either the 'ground' (*c.f.* Oota *et al.* 1995) or 'scraped' (*c.f.* Drancourt *et al.* 1998; Raoult *et al.*, 2000) methods described to recover dentine. Dental enamel, as a highly mineralised material, is extremely resistant to diagenesis, but may be permeable to contaminating environmental DNA. The extractions performed here on 'ground' samples, where DNA from environmental sources was amplified, despite UV light and bleach pretreatment, suggest that contamination persists within external layers of teeth despite external decontamination. Our extractions of 'scraped' teeth meanwhile point to the relative ease with which these contaminants may subsequently enter the DNA extraction, even when high levels of preventative care are taken. A plausible route is via contact with gloves, when the teeth are manipulated. Encasing teeth in silicone appears to act as a barrier to such movement, and may explain the observed reduction in contamination levels with this method.

7.4.2 The absence of authentic *Y. pestis* DNA in the samples

No evidence for surviving *Y. pestis* DNA was found in this study, despite the examination of a large number of samples from five mass graves, including two well documented plague pits and several other probable plague victim burial sites. This result strongly contrasts with previous studies (Drancourt *et al.* 1998; Raoult *et al.* 2000). PCR amplification using supposedly *Y. pestis*-specific primers did produce amplicons, though not of *Y. pestis* DNA. Similarly, cloned amplicons from the non-

specific bacterial 16S rDNA revealed sequences that matched a variety of bacteria. While two cloned 16S rDNA sequence fragments obtained with one set of primers resembled *Y. pestis*, and although this short amplicon incorporates a variable loop that can be used to distinguish bacterial species, the sequence was shared by at least seven different prokaryotes belonging to three genera found on the Ribosomal Database project website (http://rdp.cme.msu.edu/html/) and numerous other bacterial sequences in the NCBI BLAST database (http://www.ncbi.nlm.nih.gov/BLAST/). The database matches of most 16S rRNA sequences obtained, suggest the contaminating DNA probably originates in contemporary soil bacteria, rather than authentic aDNA sources (see supplementary data). Evidence to support this hypothesis includes the lack of *post mortem* damage-driven sequence variation observed among multiple clones of the same species (Pääbo 1989), the large size of fragments obtained (Pääbo 1989), and most importantly the presence of microorganism DNA where human DNA cannot be amplified.

The human DNA bears characteristics that suggest authenticity, including the observed consistency of amplified sequences and the observed spectrum of damage within the human cloned sequences (Pääbo *et al.* 1989; chapter 2), the pattern of nuclear to mitochondrial DNA survival (Hofreiter *et al.* 2001), and the appropriate behaviour of negative controls and extraction blanks (Cooper and Poinar 2000). However it is very difficult to guarantee the authenticity of aDNA from human samples (*c.f.* Handt *et al.* 1996; Cooper 1997; Kolman and Tuross 2000; Hofreiter *et al.* 2001) thus it is possible that the human results are derived from contaminant DNA. Such a case would add support to the conclusion that the amplified bacterial DNA is from modern environmental contaminants.

In contrast to our findings, previous studies reported successful direct sequencing of *Y. pestis*-specific PCR products (Drancourt *et al.* 1998; Raoult *et al.* 2000) from ancient teeth. This implies a low level of contaminating non-*Y. pestis* bacterial DNA, despite using a dentine extraction method that this study has demonstrated to be contamination-prone. Two questions therefore need to be answered. Firstly why such levels of contaminating DNA from bacteria other than *Y. pestis* are found in our study, even when *Y. pestis* specific primers, high-fidelity enzymes, dedicated aDNA facilities, rigorous cleaning and established extraction techniques are used. Secondly, why it was not possible to amplify *Y. pestis*-specific DNA from samples of plague victims that yield what appears to be authentic human DNA.

A tempting explanation for the discrepancy between the results is heterogeneity between archaeological sites. Samples studied in this paper were obtained from north-western European locations that differ environmentally from the relatively warmer and drier southern French locations of the previous studies. It is possible therefore, that the diagenetic conditions at the southern locations were conducive to ancient *Y. pestis* survival, and that local environmental bacteria do not share amplifiable DNA similarities with *Y. pestis*. Several possible flaws can be identified with this hypothesis. Ancient DNA studies have repeatedly demonstrated that an inverse correlation exists between average temperature of archaeological site, humidity and aDNA retrieval (Höss *et al.* 1996). It is surprising that samples used in the two successful studies were from warmer locations than those used here. An alternative environmental variable to consider is that the samples analysed in this study can be expected to have experienced more groundwater. Although this has not been directly implicated in aDNA survival, Nielsen-Marsh and Hedges (2000) note an inverse correlation of sample exposure to water and aDNA survival. However the ability to

amplify what appears to be endogenous human aDNA from extracts suggests host DNA survival is not an issue. The reduced rate of amplification of DNA associated with environmental bacteria from the pulp of silicone-encased and drilled teeth compared with glove-handled split teeth suggests that the pulp is indeed relatively protected from environmental contamination, whether arising due to water penetration while in the soil or as a result of handling during processing. The wetter northern environments may harbour a specific range of micro-organisms that share amplifiable similarities with *Y. pestis*. However with increasing numbers of scientists suggesting global dispersion of microorganisms, (Finlay *et al.* 1999a; Finlay *et al.* 1999b), it appears strange the two (globally) close environments should differ so much in microorganism content.

A further explanation is that a) the individuals from whom the samples derive were either infected by a *Y. pestis* strain lacking the plasmid-located sites for amplification, or b) were not infected with *Y. pestis* (because they were not victims of the Black Death, or c) because the infection did not seed the pulp cavity, or d) because the Black Death and subsequent plagues were not caused by *Y. pestis*). Hypothesis (a) is unlikely, as although some *Y. pestis* plasmids may vary between strains, the plasmid containing the *pla* gene is a consistent feature of contemporary *Y. pestis* isolates (Filippov *et al.* 1992). In addition a *Y. pestis* chromosomal target (*rpoB*) was employed for the ABC-processed teeth. If hypothesis (b) were true, and the 66 selected individuals were not infected by *Y. pestis* but were surrounded by infected individuals, then this would be a most unusual sampling error. The third hypothesis, (c), is plausible. There is no guarantee that bacteria causing a systemic infection entered the teeth of infected individuals, even if individuals were infected with *Y. pestis*. Contemporary microbiological studies with dental pulp (primarily but not

exclusively in the context of dental caries) are able to both directly culture and obtain DNA by PCR from various anaerobes and Streptococci (Hoshino *et al.* 1992; Conrads *et al.* 1997; Bate *et al.* 2000), i.e. typical oral bacteria. It has been demonstrated that DNA specific to *Coxiella burnetii* could be detected in dental pulp of five out of ten guinea pigs 15-20 days after intraperitoneal injection with the bacteria (Aboudharam *et al.* 2000). However, in no animal were blood cultures and dental pulp PCR positive at the same time. Consequently, the relationship between bacteraemia and pulp colonisation is not straightforward, and it is possible that *Y. pestis* may not have been present in the teeth specimens but *Y. pestis* infection still caused death.

The fourth hypothesis, (d), that the Black Death was not caused by *Y. pestis,* is controversial (Scott and Duncan 2001; Cohn 2002), but cannot be immediately discounted. Recent re-examinations of the epidemiology of the Black Death from contemporary descriptions and mortality records have suggested that it does not correspond to the illness referred to as bubonic plague, and that the organism responsible was therefore not *Y. pestis* (Scott and Duncan 2001; Cohn 2002). These authors argue that instead of the cumbersome rat-flea vector system of *Y. pestis* infections, a more direct person-to-person spread of the disease is required to cause the predominantly city-based outbreaks with subsequent rapid spread over long distances. They also postulate that the illness itself, presenting multiple haemorrhagic skin lesions (tokens) and killing 40-50% of the population, does not convincingly resemble microbiologically-confirmed bubonic plague in the pre-antibiotic era. The proponents of these hypotheses regard the molecular evidence of the Marseille group as either relevant only to distinct local outbreaks of bubonic plague in Marseille and other parts of Provence (Scott and Duncan 2001) or in need of independent

corroboration (Cohn 2002). Scott and Duncan (2001) instead suggest that the Black Death was a form of viral haemorrhagic fever, a group of emerging infections is caused predominantly by RNA Filoviruses (Khan *et al.* 1998). If so, they would be impossible to identify by current techniques on available remains from Black Death victims (viral RNA extraction and successful reverse transcriptase PCR from historical specimens has only been recorded from frozen or formalin-preserved organs (Taubenberger *et al.* 1997; Reid *et al.* 1999; Basler *et al.* 2001)).

In contrast to our findings, all but one individual tested from three different locations in Southern France in the previous studies yielded positive PCR amplifications of fragments of the *Y. pestis pla* or *rpoB* genes (Drancourt *et al.* 1998; Raoult *et al.* 2000) The authors also report in their second publication a higher percentage of the older erupted teeth to be positive for *Y. pestis*-specific DNA (20 out of 23 teeth positive) than the more recent unerupted teeth used in the first paper (8 out of 13 teeth positive). This pattern of results seems surprising for several reasons. Firstly, during an epidemic, not all deaths can be expected to arise due to the specific pathogen (Kiple 1993). Thus as increasing numbers of specimens are examined, the chance of analysing authentic uninfected remains increases. Therefore it seems unlikely that among the 8 individuals sampled in these two publications, only one negative for *Y. pestis* DNA was observed. Secondly, preservation of *Y. pestis* DNA in nearly all teeth from infected individuals is unlikely due to the nature of DNA degradation. Pfeiffer *et al.* (1999) observe that teeth storage in soil for only six weeks leads to a decrease in extractable endogenous DNA by 90%. Therefore aDNA tests for bacterial infection can be expected to demonstrate a proportion of false negative results. For example, independently replicated *M. tuberculosis* aDNA assays on comparably old specimens that display characteristic diagnostic bony lesions, did not yield 100% positive results

(Haas *et al*. 2000) despite the robust successfulness of assays for mycobacterial aDNA in various laboratories (Zink *et al*. 2000; Haas *et al*. 2000; Rothschild *et al*. 2001; Spigelman *et al*. 2002; Fletcher *et al*. 2003a; Fletcher *et al*. 2003b). We have found no published data on long-term preservation of aDNA from Enterobacteriaceae in human specimens, other than the Marseille group's papers on *Y. pestis*. Evidence of rapid *Y. pestis* DNA degradation over short periods of time is provided in the paper from which the original *pla* PCR assay was developed - an assay for *Y. pestis* in infected fleas (Hinnebusch and Schwan 1993). After five months storage, sensitivity of the original assay dropped from 100% to 90% in fleas stored at -20°C, and to 55% in fleas stored at room temperature in ethanol (Hinnebusch and Schwan 1993). Although the target in this original assay was larger (478 bp) than the 300 and 150 bp targets used by ourselves and the Marseille group, *pla* target aDNA preservation over 5 centuries in soil at Mediterranean ambient temperatures would have to be remarkably better than this to yield the positivity rate of 20 out of 23 14th century teeth that they describe (Raoult *et al*. 2000). Another factor suggesting excellent DNA preservation in southern France is the detection of *rpoB* in the first publication (Drancourt *et al*. 1998), albeit by two rounds of PCR, One of the reasons *pla* is used as a target for current *Y. pestis* detection assays is the high copy number of the pPst (pesticin) plasmid on which it is located – over a hundred per bacterium (Parkhill *et al*. 2001). In contrast *rpoB* is a single copy chromosomal gene, (Parkhill *et al*. 2001) a relatively poorly represented target which is not used as the basis for any current assays for *Y. pestis* preculture diagnosis.

Poor laboratory technique may explain our results, as could differences in the methodologies used here and previously. However similar results are found by two independent laboratories, one of which is a facility dedicated to aDNA research and

has published numerous replicated aDNA studies (*e.g.* Cooper *et al.* 2001; Shapiro *et al.* 2002; Barnes *et al.* 2002; Endicott *et al.* 2003; chapter 2; chapter 3). Secondly the lack of contaminant DNA sequence variation observed between samples from one archaeological site, and the much larger sequence variation between different sites points to the presence of contaminants in samples prior to extraction. Lastly, the techniques used here include the use of the modified PCR enzyme Platinum Taq Hi-fidelity (as opposed to 'standard' Taq polymerases). The increased yields and successes of aDNA amplifications performed using this enzyme has been noted previously (Willerslev *et al.* 1999; Hansen *et al.* 2001; chapter 2) especially in amplifying low copy number DNA.

It is possible that the DNA sequences presented previously (Drancourt *et al.* 1998; Raoult *et al.* 2000) derive from the contamination of DNA extracts with formerly amplified or extracted *Y. pestis* DNA. Other examples exist of pioneering bacterial aDNA studies that could not be replicated and may have resulted from unsuspected contamination. Christner *et al.* (2000) reports unsuccessful attempts at replicating aDNA detection from bacteria within Greenland ice cores (Catranis and Starmer, 1991; Ma, 1999). Graur and Pupko (2001) and Nickle *et al.* (2002) have been unable to replicate detection of Vreeland *et al.*'s (2002) 250 million-year-old halotolerant bacterium.

The previous *Y. pestis* aDNA studies were undertaken in a busy facility not dedicated to aDNA work, where numerous other bacteriological studies are undertaken, and initially using a positive control of modern *Y. pestis* DNA, all of which are risk factors for contamination. However occurrence of contamination in ancient DNA work even under the most stringent conditions is well documented (Handt *et al.* 1996; Kolman, 1999; Kolman and Tuross, 2000; Hofreiter *et al.* 2001). In our own study the

use of a modified positive control in the SBH hospital laboratory (a non-dedicated ancient DNA set up comparable to that used by the Marseille group) has resulted in contamination problems with bacterial DNA. This clear occurrence of selective contamination, where importantly, negative controls remained blank is especially revealing. This phenomena, termed the 'carrier effect', and the dangers of relying on negative controls are discussed elsewhere (Cooper 1994; Handt *et al*. 1994; Kolman, 1999). Raoult *et al*. (2000) developed 'suicide PCR' to avoid the possibility of amplicon contamination, but this technique is not resilient to contamination by fragments of extracted modern DNA, from any organism sharing the DNA sequence of interest.

7.5 Conclusion

This study has failed to replicate previous reports of specific *Y. pestis* DNA amplification from dental pulp residues extracted from historical plague victims. Analysis of 16S rDNA PCR products reveal a wide variety of bacterial DNA in the extracts from teeth prepared using the previous method. An improved method has been developed for pulp cavity sampling, which resulted in a reduced number of amplicons from apparent environmental contaminants, and it is recommended that future studies consider adopting the approach, which is cheap, simple and effective.

It is difficult to interpret the previous reports of almost uniform positive results, when similar techniques reveal the non-specificity of supposedly *Y. pestis*-specific primers in this application, the ease of positive control cross-contamination, and the apparent lack of *Y. pestis* DNA in samples that yield human DNA, from well documented archaeological plague sites. Only a minority of published ancient pathogen DNA studies clearly indicate that they have involved independent replication by separate laboratories. Most of the replicated reports concern successful amplification of *M. tuberculosis* from skeletal remains, or mummified tissue (Zink *et al*. 2000; Haas *et al*. 2001; Rothschild *et al*. 2001; Spigelman *et al*. 2002; Fletcher *et al*. 2003a; Fletcher *et al*. 2003b). The special properties of *M. tuberculosis* that enable lengthy persistence in the body as a latent infection have been suggested to preserve its DNA after the death of its host (Fletcher *et al*. 2003a). There is no evidence that *Y. pestis* can cause persistent latent disease. Independent replication seems particularly necessary when attempting novel retrospective diagnosis of diseases that cause rapid death, and leave no specific bony changes. Authoritative aDNA evidence offers the only conclusive method to match current pathogens with ancient epidemics of disease, but misleading DNA data is very easily generated from contemporary bacteria. For these reasons we

believe that until an independently replicated, successful study on *Y. pestis* aDNA is undertaken in a suitable, controlled, environment, meeting best practice guidelines in ancient DNA research (Cooper and Poinar 2000), it is premature to claim (Drancourt and Raoult 2002; Raoult and Drancourt, 2002) that aDNA studies have unequivocally proved *Y. pestis* to be the cause of the Black Death and subsequent historical plagues. The aetiology of one of the major pandemics of the last millennium remains unproven by molecular techniques.

7.6 Acknowledgements

The authors would like to thank Professor H Mollaret (AN) and Frédéric Adam (VE) for samples used in this project, and Mark Achtman, Elizabeth Carniel and Edward Holmes for facilitating this project. M.T.P.G. and A.C. are grateful for the Wellcome trust for funding their research. M.B.P. and J.C. acknowledge funding from the Joint Research Board of St Bartholomew's Hospital and technical assistance from Ken Voong and Richard Prior for part of this work.

Chapter 8

The long-term survival of ancient DNA in Egypt

Modified from the article in press in the 'American Journal of Physical Anthropology'

'News and Comments: The long-term survival of ancient DNA in Egypt: Response to Zink and Nerlich 2003'

Reply to A. Zink and A.G. Nerlich 'Molecular analyses of the Pharaos: Feasibility of molecular studies in ancient Egyptian material' Am J Phys Anth 121:109-111 (2003)

M Thomas P Gilbert[1], Ian Barnes[2], Matthew J Collins[3], Colin Smith[4], Julie Eklund[5], Jaap Goudsmit[6,7], Hendrik Poinar[8], Alan Cooper[1]

[1]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, Oxford University, South Parks Rd, Oxford OX1 3PS, UK
[2]The Center for Genetic Anthropology, Department of Biology, Darwin Building, University College London , Gower Street, London, WC1E 6BT
[3]Fossil Fuels and Environmental Geochemistry, (Postgraduate Institute); NRG, Drummond Building, University of Newcastle, Newcastle upon Tyne NE1 7RU, United Kingdom
[4] Department of Palaeobiology, Museo Nacional de Ciencias Naturales (CSIC),C/José Gutiérrez Abascal. 228006, Madrid, Spain
[5] Institute of Archaeology, University College London, 31-34 Gordon Square, London WC1H 0PY, UK
[6] Crucell, Archimedesweg 4, 2333 CN LEIDEN, The Netherlands
[7] Center for Poverty-related Communicable Diseases, Meibergdreef 15, 1105 AZ Amsterdam, the Netherlands'
[8] Max Planck Institute for Evolutionary Anthropology, Inselstarsse 22, D-04103 Leipzig

8.1 Introduction

Marota *et al*. (2002) have proposed that PCR amplifiable DNA is unlikely to survive in ancient Egyptian specimens, using the degradation of papyrus chloroplast DNA as a general model of molecular degradation in Egypt. If correct, this suggests that the numerous studies claiming to have successfully amplified host and/or pathogen DNA from ancient Egyptians must be re-evaluated.

Zink and Nerlich (2003) criticise the argument of Marota *et al*. (2002) on two fronts in defense of ancient Egyptian DNA. They suggest both that DNA degradation may be slower than that calculated by Marota *et al*. (2002), and that the results of several successful molecular studies prove that ancient DNA (aDNA) can survive in Egypt.

We believe that the arguments used by Zink and Nerlich (2003) are either flawed or inconclusive, and wish to re-assert the premise that in most, if not all, ancient Egyptian remains, aDNA does not survive to a level that is currently retrievable.

8.2 Arguments for a reduced rate of DNA degradation

DNA degradation in biological samples is principally by hydrolysis of purine bases, and is known to depend on a number of variables, including temperature, pH and the availability of oxygen and chemically free water (Lindahl and Nyberg 1972; Pääbo *et al*. 1989; Lindahl 1993a). Zink and Nerlich (2003) argue that several factors result in DNA degradation rates in Egypt being lower than those calculated by Marota *et al*. (2002):

1) Temperatures within tombs are considerably lower than the 35ºC used as an estimate by Marota *et al*. (2002).

2) Most ancient Egyptian samples that have previously failed to yield DNA were associated with environments that fluctuated between wet and dry, due to annual Nile

flooding, and these are not representative of many other, constantly dry, burial sites.

3) Mummified Egyptian specimens were extremely desiccated as a result of preparation using natron, and as a result, less subject to decomposition.

Each of these issues is addressed in turn:

8.3 Temperature

Zink and Nerlich (2003) argue that although the external environment in Egypt is too warm for long term DNA survival, the micro-environment within tombs is considerably cooler, and more suited to DNA survival. Two problems can be identified with this argument; first the implication that the tomb environment has been constant since the deposition of the mummy, and second that the tomb temperature is significantly lower than the average outdoor temperature.

Zink and Nerlich (2003) assume that environmental conditions, including temperature and humidity, have remained in equilibrium for the period that any tomb has been sealed (Stoppelaere 1942). However, the likelihood of a tomb remaining sealed from the time its intended (initial) inhabitant was placed within the chamber is slight, as evidenced by the rarity of intact tombs discovered in the recent past. A number of events may interrupt the equilibrium of the burial environment, including episodes of looting, the re-use or multiple use of a tomb for other burials which could have occurred several dynasties later, the accidental cross-cutting of tombs, and geological activity resulting in the creation of fissures and cracks into tomb chambers, thereby exposing the tomb to external weather factors. For example, in several of the successful aDNA studies cited by Zink and Nerlich (2003: Zink *et al*. 2000a; Zink 2001) the authors acknowledge that the tombs at Abydos and Thebes-West from which their samples were obtained had experienced several intrusive burials.

Once tombs have been opened it might be argued that protection from direct sunlight will maintain cooler daily temperatures than the external environment, although the effect and extent is debatable. Zink and Nerlich (2003) argue that temperatures in the tomb of Nefertari in the Valley of the Queens have been reported as significantly cooler than the 35ºC used in Marota *et al.* (2002), and vary between 15ºC -25ºC annually. We are unable to find these data in the reference quoted (Maekawa and Preusser 1993), which along with an earlier study (Hideo 1988) report considerably warmer temperatures, ranging between 26ºC and 29ºC. Winter temperature data derived for six tombs in the Valley of the Kings show a significant range of values, with temperatures markedly higher deep within tombs than in the entrance way (12-24ºC for entrance, 21-29ºC for the far end of the tombs - Lucas 1924).

It might be suggested that recently measured tomb temperatures could have been elevated by the impact of tourist traffic, and that sealed tombs are at much lower temperatures. Quantitative temperature data from freshly opened tombs is not available, however anecdotal evidence supports high temperatures even prior to re-opening. In perhaps the most famous account of the opening of a tomb, that of Tutankhamun (KV62), Howard Carter wrote; '*At first I could see nothing, the hot air escaping from the chamber causing the candle flames to flicker…*' (Carter and Mace 1923). In another famous account, similar conditions are noted at the opening of the temple of Ramesses II at Abu Simbel on the 1st August 1817, by Giovanni Belzoni. '*The heat was so great in the interior of the temple, that it scarcely permitted us to take any drawings, as the perspiration from our hands soon rendered the paper quite wet.'* (Belzoni 1821).

The most logical estimate for tomb temperature is the local annual mean air temperature, and this should be used in order to calculate a limit on DNA survival time. Tombs can be categorised into three basic types; those that are dug in to the ground, and those that are cut into a rock formation, form two. A third type, consisting of a burial chamber either within or below a constructed superstructure, such as pyramid or mastaba tombs, are not considered here, as few remains from such locations have been used in DNA studies. Nevertheless similar concerns would apply. It is reasonable to assume that over time tombs sealed at a reasonable depth will reach equilibrium with the mean soil or rock temperature, which in turn will be in equilibrium with the mean air temperature. Fluctuation around this mean will decrease with the depth of burial or the tomb, such that in most cases it is likely to be constant. Periodic opening of the tombs (as mentioned above) would upset equilibrium, and complicate models of tomb temperature. The thermal regime is likely to approximate to that of a cave for unsealed tombs cut into rock faces, where movement of air into the cave controls the climate. The mean air temperature is generally considered a good approximation of local deep cave air temperatures (Bogli 1980), and deep cave temperatures are known to be static with little seasonal variation. There is more temperature fluctuation at the cave entrance.

For tombs cut into the soil that slope downwards it may be anticipated that the temperature may be slightly cooler than the mean temperature, as cooler denser air is always drawn into the cave. Nevertheless this is not likely to account for an arbitrary reduction of 10-20ºC below the mean air temperature (Zink and Nerlich 2003).

Data available from on the NOAA (National Oceanic and Atmospheric Administration) web site (Baker *et al.* 19940), from local weather stations in Egypt (*c.f.* Aswan approx 26 ± 12, Dakhla approx 23 ± 10, Kharga approx 24 ± 10) indicate

that average temperatures are likely to be in the region of 25ºC, not 35ºC as quoted by Marota *et al*. (2002). Nevertheless, it is still doubtful whether DNA could survive until the present even if an Eyptian burial site could be found with temperatures as low as those argued by Zink and Nerlich (2003). Marota *et al*. (2002) estimated the survival limit of chloroplast DNA in papyri at 35ºC to be ~ 800 years. This study used a depurination reaction rate at 15ºC calculated by Pääbo & Wilson (1993), which supposed a simplistic trebling of the reaction rate per 10ºC temperature increase, thereby effectively using an activation energy of 76 kJ m^{-1}not 127kJ m^{-1} (Lindahl & Nyberg 1972). Smith *et al*. (*submitted*) use the original Lindahl & Nyberg (1972) kinetics and a modified approach to calculate the maximum time for the survival of a 105bp fragment at 35ºC as only 88 years (84 and 96 years at 5^{th} and 95^{th} percentiles respectively). For a temperature of 25ºC (as suggested above), the Smith *et al*. (*submitted*) calculations suggest the maximum DNA survival based upon depurination is theoretically 462 years (437 to 505 years at 5^{th} and 95^{th} percentiles). These calculations follow the depurination assumptions of Pääbo and Wilson (1993) – naked DNA in solution at pH 7. The higher pH of bone and the binding of hydroxyapatite to DNA may result in slightly reduced depurination rates (Lindahl and Nyberg 1972; Lindahl 1993), thus longer survival of the DNA, but it is doubtful if survival will be long enough to ensure retrieval of ancient Egyptian DNA.

It should also be mentioned that these theoretical calculations assume the rate determining step in DNA degradation is hydrolytic depurination; processes such as dehydration or encapusulation may retard this process, whist others such as oxidation, wetting / drying cycles would accelerate DNA damage. As an example, DNA has been successfully recovered in coprolites from extremely arid environments in SW United States (Poinar *et al*. 2003), which have thermal ages five times greater than

Neanderthal 1, the supposed limit of DNA recovery from bone (Smith *et al*. 2001).

8.4 Water and humidity

Zink and Nerlich note that periodic flooding of burials by the inundation of the Nile is a likely cause of the degradation of organic materials in tombs, and that those sites distant from rivers should be targeted for DNA recovery. However, the complexities of climatic processes must be considered in making such a statement. At least one location suggested by the authors to be dry, the Valley of the Kings, has been subject to occasional dramatic flooding, which is considered to be a severe source of damage to tombs in the area. Recent flash-flooding episodes were reported in 1992 and 1994 (Wüst and Schlüchter 2000), resulting in standing water in many chambers and elevated relative humidity for long periods of time. Whilst such extreme conditions are unlikely to be common, tombs are often very humid even when not flooded, and DNA decay is accelerated in the presence of water and by fluctuations in moisture level (Waite *et al*. 1997; Nielsen-Marsh and Hedges 2000). For example, Hideo (1988) reports relative humidity levels within the Nefarti tomb as between 30-40% Relative Humidity (RH), while Stoppelaere (1942) reports at least 15% variation in RH over winter in the Valley of the Kings tombs of Amenhotep II (d. approx 1401 BCE) and Seti II (d. approx 1220 BCE). Lucas (1924) also reports a significant range of RH values, from 18-72% at entranceways, and 15-47% at the far ends, of six tombs.

8.5 Natron and embalming

Although Zink and Nerlich (2003) discuss the merits of desiccation due to the sodium carbonate content of the naturally occurring salt mixture, natron ($Na_2CO_3.10H_2O$ with

traces of sodium bicarbonate, sodium chloride, and sodium sulphate), they fail to consider that the use of natron varied through time. Pre-dynastic burials did not use natron as a preservative, but relied on less sophisticated techniques such as the hot sandy matrix of the burial location for preservation (Aufderheide 2003). This time period includes the purportedly successful aDNA studies cited by Zink and Nerlich (2003) from specimens excavated at Adaïma (Crubezy *et al.* 1998) and possibly that from Abydos (Zink *et al.* 2000b). Whilst this raises further questions about the temperatures of burial sites, the general inadequacy of such methods is demonstrated by the survival of only a few remains from this period, rarely consisting of more than bones (Fleming *et al.* 1980). This includes the samples from Adaïma and Abydos mentioned above.

Research into mummification techniques is discouraging with reference to the likelihood of DNA survival. In addition to the paucity of predynastic remains, few specimens survive prior to the introduction of evisceration in the mummification procedure around the 4th dynasty (2575-2465 BCE) (Aufderheide 2003). Therefore it appears that evisceration is a requisite to prevent sample putrefaction, regardless of the external application of natron (Brier 1998). Furthermore, during much of the Old Kingdom, in particular between the 3rd and 6th dynasties (2650-2150 BCE), it was common practice to encase mummies in plaster (slaked lime) casts (Fleming *et al.* 1980). Goudsmidt and colleagues (*in preparation*) remark that specimens from both the Baboon gallery at Saqqara and Memphis show sample carbonization resulting from such treatment. This is most likely a result of a reaction taking place between the slaked lime and the sodium carbonate in natron in the presence of water, – $Ca(OH)_2 + Na_2CO_3 \Rightarrow 2\ NaOH + CaCO_3$. Both the strongly exothermic nature of this reaction, and the resulting potential for alkylation damage and strand cleavage, reduce

the potential for DNA survival.

One development in mummification that would seem to confer a degree of protection to the specimen was the application of resin to the exterior of the sample during the 12th dynasty of the Middle Kingdom (1991-1783 BCE). Even so, evidence suggests that it was not until the 18th-20th dynasties of the New Kingdom (1549-1069 BCE) that this became a popular treatment (Fleming *et al*. 1980; Buckley and Evershed 2001). It is interesting to note that while Zink and Nerlich (2003) comment on the benefits for DNA survival that natron treatment may convey through an elevation of the pH of the body (presumably through the action of sodium bicarbonate in the natron) acidic resins will counteract this pH change.

While resin treatment may confer some external protection, the effects on internal degradation are still questionable, and rely on complete desiccation by natron. Prior to the 18th-20th dynasties, bags of natron are believed to have only been applied to the insides of a specimen for a 40 day period, before removal from the body cavity (Fleming *et al*. 1980). Over time specimens treated in this way would rehydrate in the humid conditions of many tombs, restarting DNA degradation through hydrolysis. By the 18th-20th dynasties matters had improved, with natron left inside the cavity at the end of the mummification period, although it is questionable about how effective this was in keeping the specimen dry over the next 3000 years. Experimental research into human mummification conducted by Brier and Wade (Brier 1998) demonstrates that frequent changes of natron are required to fully desiccate specimens, so the relatively humid conditions of many tombs are likely to undermine the effectiveness of a limited quantity of natron.

The collapse of the New Kingdom at the end of the 20th dynasty (1196-1070 BCE) saw a significant decline in the quality of mummification techniques, and natron was

replaced with common salt during the turmoil of the 21st- 24th dynasties that constitute the Third Intermediate Period (1070-712 BCE).

A further problem associated with DNA survival and the use of natron also needs to be considered, assuming the current interpretation of the techniques is correct. It is believed that the fully developed mummification procedures of the 18th-20th dynasties were a 70 day procedure, with 40 days of desiccation and a subsequent 30 days for the mummy to be prepared for burial (Brier 1998). This procedure is estimated to have been undertaken at environmental conditions of around 40ºC and 30% RH (Brier 1998).

In summary, any attempt to use a single estimate of burial environment conditions for a complex and long-lived culture such as that of Ancient Egypt will lead to inaccurate predictions of DNA preservation. The variable burial location and post-mortem practices, determined by geographical location, period of internment and social class, will have a great effect on the potential for DNA survival. Any attempt to mathematically model DNA decay should be conducted on the basis of the individual specimen history, rather than by generalisation.

8.6 The survival of biomolecules in Egyptian remains.

Zink and Nerlich imply that a 'good preservation state of the organic matrix' within some mummified remains supports the survival of DNA. However a variety of specimens, apparently well preserved at the microscopic level, have failed to yield authentic, verifiable DNA, *e.g.* Miocene leaf specimens (Logan *et al.* 1993), amber entombed insects (Austin *et al.* 1997), ethanol preserved museum specimens (Barnes *et al.* 2000), Cretaceous bone (Woodward *et al.* 1994) and Lindow Man, a preserved bog body (Hughes and Jones 1986). Quantified analysis of bone histological

preservation suggests some relationship with DNA recovery, but not a direct correspondence (Haynes *et al*. 2002; Colson *et al*. 1997).

A more suitable alternative is the analysis of quantifiable molecular markers that have been calibrated as proxies for DNA survival. In fact both bone protein and histological preservation might be used as markers for degradation, but only if employed in a quantitative manner. Similarly, multi-copy mitochondrial DNA has been suggested as an appropriate marker for the survival of nuclear or pathogen DNA, as well as a test for contamination (Barnes *et al*. 2000; Cooper and Poinar 2000).

Zink and Nerlich (2003) draw upon several published reports of DNA from ancient Egyptian material (*e.g*. Crubezy *et al*. 1997; Nerlich *et al*. 1997; Zink *et al*. 2000a; Zink *et al*. 2000b; Zink *et al*. 2001), which include the recovery of single and multi-copy sequences of both human and pathogen origin. In light of the biochemical evidence presented above, it seems implausible that they originate from the specimen. As Zink and Nerlich (2003) are aware, the analysis of DNA from archaeological material remains both extremely difficult and prone to misinterpretation. The potential for exogenous contamination of DNA analyses, especially of human material, is extremely high. These problems are compounded by a variety of phenomena peculiar to ancient DNA, such as the 'carrier' effect (Cooper 1994; Handt *et al*. 1994), nucleotide base damage (*e.g*. Pääbo *et al*. 1989), 'hotspot' localisation of that damage (chapter 3) and 'jumping' PCR (Pääbo *et al*. 1990).

8.7 Conclusion

To conclude, we find that the thermal history of most, if not all, ancient Egyptian material argues against the recovery of DNA. Consequently, such claims should continue to be considered skeptically until the concerns outlined here are addressed.

8.8 Acknowledgments

The authors would like to thank Dr Salima Ikram, Dr Shin Maekawa and Dr Stephen Quirke for helpful discussion, and Jaques Kinnear and his excellent web site www.ancient-egypt.org for provision of up to date dynastic datings. I.B., M.T.P.G. and A.C. are funded by the Wellcome Trust.

Chapter 9

Conclusions and Future Directions

‘Science is simply common sense at its best, that is, rigidly accurate in observation and merciless to fallacy in logic.’

Thomas Henry Huxley (1825-95)

9.1 Thesis summary

The field of ancient DNA (aDNA) is rich in unfulfilled promise and potential. It represents the chance for molecular biology to investigate an exciting range of previously un-approachable questions: observations of DNA evolution through snapshots of the past; the phylogeny and genetics of extinct species; the history of pathogens and disease epidemics to name just a few. However, it is first necessary to develop a clear understanding of the nature of aDNA in order to achieve these goals.

DNA degradation and damage is a complex phenomenon. Chapter 2 describes new insights into the mechanisms behind one small section of this damage, the hydrolytic deamination events that lead to DNA sequence modifications. The insights gained from the results will help research into remedies for the damage, for example the removal or reversal of adenine→hypoxanthine formation.

In addition the findings of this thesis point to a mitochondrial DNA strand bias during hydrolytic deamination (chapter 2), as well as the curious spatial distribution of such damage (chapters 3,4). These results indicate the weakness of results from human (and other) aDNA analyses that rely on un-cloned DNA sequences. They also paradoxically provide further insights into the mutational processes that affect mtDNA sequences *in vivo*. Examples include the effects of sequence protection (either through co-binding with other molecules or through local DNA conformational secondary structure), and causes for the heterogenous mutation rate observed across the mitochondrial Hypervariable Region. This in turn helps validate the presence of mutational hotspots, and thus provides an argument against theories

that promote the existence of recombination in mitochondrial DNA (*c.f.* Hagelberg 2003).

Numerous human aDNA studies rely on the physical structure of bones and teeth to both preserve DNA, and to provide a contamination-free niche. The results of chapter 5 indicate that such hypotheses are irrational, and have arisen through ignorance of the innate porosity of bone and tooth samples. The data also point to the necessity of persuading field archaeologists of the need to abstain from sample washing and handling if DNA testing is to be attempted at a later date. Finally, correlations of contamination with biomolecular and histological preservation (in particular microbial attack-derived porosity) suggest that it should be possible for archaeologists to screen excavated samples in order to determine which should be kept safe for future aDNA testing.

The results of chapter 6 indicate that hair is more resilient to contamination than teeth and bones. They also indicate that endogenous mtDNA survives within hair shafts for very long time periods. This surprising result further illuminates how little is known about the nature of DNA damage, and the lack of ability to accurately predict DNA survival. Naturally, the use of hair as an aDNA source is associated with a major problem - how much 'ancient' hair is there to sample? While retrieval from the archaeological record may be scarce, there are sizeable museum collections that are worth exploring.

Many aDNA studies have investigated paleopathology. To date, two studies have reported on the successful retrieval of ancient *Yersinia pestis* DNA from human remains (Drancourt *et al.* 1998; Raoult *et al.* 2000). These findings have been used to categorically argue that *Y. pestis* was the causative organism of the 1st and 2nd great plague epidemics (Drancourt *et al.* 2002; Raoult *et al.* 2002). In chapter 7, as part of a

joint study we fail to replicate the results. Nevertheless two important findings are identified. Firstly, the 'carrier effect' (Handt *et al.* 1994a) is a ready source of positive control-to-extract contamination during ancient pathogen DNA PCR setup. Secondly, the field is wide open to problems associated with a general lack of knowledge about the genetic variation in, and geographic spread of, micro-organisms. In combination these results suggest that there is more to that which meets the eye in previous 'successful' ancient pathogen studies.

The survival of DNA in ancient Egypt is addressed in chapter 8, with particular reference to the validity of arguments that support its existence. This chapter highlights that the environment (and thus environmentally-induced DNA damage and contamination) is more complex than described elsewhere (*c.f.* Zink and Nerlich 2003).

The suitability of current aDNA criteria aimed at the generation of reliable results (*c.f.* Cooper and Poinar 2000) has not yet been questioned. While these criteria are admirable, and have helped stem the tide of 'dubious' aDNA research, they also run the risk of alienating some decent studies that do not follow the criteria. For example, studies that a) do not extract and sequence DNA from humans and pathogens (and thus are relatively difficult to contaminate), b) have been produced in carefully controlled environments, c) on organisms where the relevant DNA sequences are already clearly defined, and d) that produce reasonable data, may not need to adhere to all the criteria. In particular are the requirements for independent replication, the expensive and technically demanding processes of DNA quantification (some samples that are well preserved may be rejected as the DNA is too high quality), and the need for amino acid racemisation analyses (the reliability of which as an indicator of DNA degradation is still under debate - Collins *et al.* 1999). On the other hand,

due to the nature of sample contamination *prior* to the aDNA analysis, some studies that follow all criteria may be published, despite claiming results that may be dubious. Of particular reference here are pathogen studies from ancient Egypt (*c.f.* Zink *et al*. 2000a; Zink *et al*. 2000b; Zink *et al*. 2001; Zink *et al*. 2003a; Zink *et al*. 2003b), and various human studies that report the retrieval of DNA sequences which, although suggest reasonable results, are similar to the majority of haplogroups also found in the area in modern populations (thus may be better explained through sample contamination (*c.f.* Caramelli *et al*. 2003)).

There is no easy solution to this problem. However, one potential requirement that will help elucidate matters is the routine publication of full details of sample handling since excavation, and of environmental conditions at the burial and subsequent storage sites. With such information, reviewers will have more information on which to make suitable decisions as to how likely findings are, through an assessment of sample risk of contamination and DNA preservation.

9.2 Future directions in ancient DNA research

The future of aDNA requires close collaboration with other research fields – and in particular with those that address tissue morphology, histology and degradation, and the extraction, damage and repair of DNA. In the light of the results of this thesis, several potential areas of research are of particular interest.

The study of damage characteristic to aDNA is a nascent field. As more accurate predictions are made of where and how DNA is surviving (*c.f.* Smith *et al.* 2001) new possibilities will arise for extracting and repairing such DNA. It may thus be useful to use various techniques that are currently used in cancer research in order to screen for aDNA damage. Ideal candidates with which to start this process include DNA-end labelling with Polynucleotide kinase (PNK) (which enables the quantification of DNA fragmentation), measurements of the refractivity of samples to heat denaturation using the double-stranded DNA binding marker Picogreen© (Molecular Probes) (providing quantitative measures of DNA cross-linking), and the use of DNA cleavage enzymes which specifically cleave DNA at sites of specific damage, such as Oxoguanine glycosylase (Ogg), Endonuclease III (Nth), and FaPy glycosylase (Fpg) (D.Mitchell, *pers. comm.*).

Several new studies have also suggested methods to retrieve such damage. It has recently been demonstrated that the binding of DNA to hydroxyapatite is positively correlated with the long-term survival of DNA (*c.f.* Geigl 2002; Götherström *et al.* 2002). Investigation is therefore required into methods of releasing the DNA from the DNA-hydroxyapatite complex. In addition, there is also a need to further investigate the nature of the DNA extraction agent 'PTB' and its action on increasing DNA yield in ancient DNA extracts. Although a model of its action has been suggested (Vasan *et al.* 1996), subsequent studies have queried the plausibility of the model (Thornalley

and Minhas 1999). Furthermore, PTB is not currently commercially available, thus research into the efficacy of its more-readily available chemical analogues (*e.g.* 3-phenacyl-4,5-dimethylthiazolium chloride, more commonly known as PTC or ALT-711 (Alteon Pharmaceuticals)) will be worthwhile.

A further area of research that warrants investigation is the development of techniques that are effective in decontaminating bone and tooth specimens. The relationship of sample porosity to contamination suggests a simple solution – the reduction of samples to powders that are fine enough to allow bleach (or other liquids that may destroy DNA) to fully perforate the sample. Nevertheless, if the assumption that DNA is predominantly surviving in osteocytes is correct, a balance will need to be struck to prevent bleach also permeating these and destroying endogenous DNA. An alternative approach is to investigate other potential decontaminating agents with lower surface tensions than bleach that may better penetrate the porous systems of bones and teeth.

The preservation and degradation of DNA in hair remains to be explored. Not only will this have implications for forensic science (currently the field taking greatest advantage of hair as a source of DNA), but it is possible that with further study a method may be developed to recover nuclear DNA. Nuclei in human cortical cells can be seen to lyse during keratinisation, but the extent of nuclear DNA degradation is not understood, and it is possible that traces of nuclear DNA may remain. In addition, as the structure and development of hair is very variable between different organisms, nuclear DNA may survive better in some species. Of related interest is the survival of DNA in other keratinous tissues – for example feathers. Recently we have succeeded in the extraction of mitochondrial DNA from small fragments of extinct

Moa (*Dinornis* spp.) feather tips, dating to over 500 years (MTP Gilbert, M Bunce, unpublished data).

9.3 Conclusions

Despite its glamorous sheen, the field of ancient DNA is no different to other scientific disciplines. Thus its studies should be undertaken with the same scientific rigour that is expected elsewhere. As aDNA research has progressed, many unintentionally-erroneous results have be published. This can be argued to be partly due to an eagerness of high profile journals to publish aDNA studies, whatever the quality of the research. Nevertheless, with a positive and open attitude, critical assessment of current protocols, and inter-discipline collaboration future aDNA research can help provide remedies to these past problems. However, if these issues are not addressed the field runs the risk of being relegated to where many already view aDNA currently belongs, in the company of discarded and dubious scientific disciplines such as phrenology and eugenics. Due to the considerable financial cost of ancient DNA research, misleading results should rightly be viewed as the result of studies that detract desperately needed funding from other more valuable areas of science. However, with the advent of an understanding of the nature of the field, and the routine use of suitable checks, the field should produce results that are both ground breaking and useful.

Literature cited

Aboudharam G, La Scola B, Raoult D, Drancourt M. 2000. Detection of *Coxiella burnetii* in the dental pulp during experimental bacteremia. Microb Pathog 28: 249-254

Adcock G, Dennis E, Easteal S, Huttley G, Jermelin L, Peacock W, Thorne A. 2001. Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. Proc Natl Acad Sci USA 98:537-542

Allard MW, Young D, Huysen Y. 1995. Detecting dinosaur DNA. Science 268:1192

Allen M, Engström A-S, Meyer S, Handt O, Saldeen T, von Haeseler A, Pääbo S, Gyllensten U. 1998. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: Sensitivity and matching probabilities. J Forensic Sci 43:453-464

An C-C, Li Y, Zhu Y-X. 1995. Molecular cloning and sequencing of the 18S rDNA from specialized dinosaur egg fossil found in Xixia Henan, China. Acta Sci Nat Univ Pekinensis 31:140-147

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

Anderson S, De Bruijn MHL, Coulson AR, Eperon EC, Sanger F, Young IG. 1982. Complete sequence of bovine mitochondrial DNA. J Mol Biol 156:683-717

Ånenson 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

Ansorge W, Sproat BS, Stegemann J, Schwager C. 1986. A non-radioactive automated method for DNA sequence determination. J Biochem Biophys Methods 13:315-323

Anzai T, Naruse TK, Tokunaga K, Honma T, Baba H, Akazawa T, Inoko H. 1999. HLA genotyping of 5,000-and 6,000-year-old ancient bones in Japan. Tissue Antigens 54:53-58

Aris-Brisou S, Excoffier L. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol 13:494-504

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, Janke A. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc Natl Acad Sci USA 99:8151-8156

Arnason U, Gullberg A. 1993. Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. J Mol Evol 37:312-322

Arnason U, Gullberg A, Janke A. 1997. Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between *Xenarthra* (*Edentata*) and *Ferungulates*. Mol Biol Evol 14:762-768

Arnason U, Gullberg A, Janke A. 1999. The mitochondrial DNA molecule of the aardvark, *Orycteropus afer*, and the position of the Tubulidentata in the eutherian tree. Proc R Soc Lond B 266:339-345

Arnason U, Gullberg A, Johnsson E, Ledje C. 1993. The nucleotide sequence of the mitochondrial DNA molecule of the grey seal, *Halichoerus grypus*, and a comparison with mitochondrial sequences of other true seals. J Mol Evol 37:323-330

Arnason U, Gullberg A, Xu X. 1996a. A complete mitochondrial DNA molecule of the white-handed gibbon, *Hylobates lar*, and comparison among individual mitochondrial genes of all hominoid genera. Hereditas 124:185-189

Arnason U, Johnsson E. 1992. The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. J Mol Evol 34:493-505

Arnason U, Xu X, Gullberg A. 1996b. Comparison between the complete mitochondrial DNA sequences of *Homo* and the common chimpanzee based on nonchimeric sequences. J Mol Evol 42:145-152

Arneborg J, Heinemeier J, Lynnerup N, Nielsen HL, Rud N, Sveinbjörnsdóttir ÁE. 1999. Change of diet of the Greenland Vikings determined from stable carbon isotope analysis and ^{14}C dating of their Bones. Radiocarbon 41:157-168

Arriaza BT, Salo W, Aufderheide AC, Holcomb TA. 1995. Pre-Columbian tuberculosis in northern Chile - molecular and skeletal evidence. Am J Phys Anthropol 98: 37-45

Aufderheide A. 2002. The Scientific Study of Mummies. CUP:Cambridge

Austin JJ, Arnold NE. 2001. Ancient mitochondrial DNA and morphology elucidate an extinct island radiation of Indian Ocean giant tortoises (*Cylindraspis*). Proc R Soc Lond B 268:2515-2523

Austin JJ, Ross AJ, Smith AB, Fortey RA, Thomas RH. 1997a. Problems of reproducibility – does geologically ancient DNA survive in amber-preserved insects? Proc Roy Soc Lond B 264:467-474

Austin JJ, Smith AB, Thomas RH. 1997b. Paleontology in a molecular world: The search for authentic ancient DNA. Trends Ecol Evol 12:303-306

Bachmann L, Scholz M, Broghammer M, Giddings IK, Pusch C. 2000. Voltage-induced release of nucleic acids from paleontological samples. Electrophoresis 21:1488-1492

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 354:77-87

Bailey JF, Richards MB, Macaulay VA, Colson IB, James IT, Bradley DG, Hedges RE, Sykes BC. 1996. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. Proc R Soc Lond B 263:1467-73

Baker CB, Eischeid JK, Karl TR & Diaz HF. 1994. The quality control of long-term climatological data using objective data analysis. Preprints of AMS Ninth Conference on Applied Climatology, Dallas, TX. January 15-20, 1995.

Baker LE. 2001. Mitochondrial DNA haplotype and sequence analysis of historic Choctaw and Menominee hair shaft samples. PhD Thesis. University of Tennessee, Knoxville.

Baker LE, McCormick WF, Matteson KJ. 2001. A silica-based mitochondrial DNA extraction method applied to forensic hair shafts and teeth. J Forensic Sci 46:126-130

Bandelt H-J, Lahermo P, Richards M, Macaulay V. 2001. Detecting errors in mtDNA data by phylogenetic analysis. Int J Legal Med 115:64-69

Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150-1160

Bar W, Kratzer A, Machler M, Schmid W. 1988. Postmortem stability of DNA. Forensic Sci Int 39:59-70

Barnes I, Holton K, Vaira D, Spigelman M, Thomas MG. 2000a. An assessment of the long-term preservation of the DNA of a bacterial pathogen in ethanol-preserved archival material. J Pathol 192:554-559

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A. 2002. Dynamics of mammal population extinctions in Eastern Beringia during the last glaciation. Science 295, 2267-2270

Barnes I, Young JPW, Dobney KM. 2000b. DNA-based identification of goose species from two archaeological sites in Lincolnshire. J Archeol Sci 27:91-100

Barnett RJ, Sognnaes RF. 1962. Histochemical distribution of protein-bound sulfhydryl and disulfide groups in vertebrate keratins. In: Butcher EO and Sognnaes RF (Eds) Fundamentals of Keratinization. Association for the Advancement of Science:Washington DC. Pp 27-43

Baron H, Hummel S, Herrman B. 1996. *Mycobacterium tuberculosis* complex DNA in ancient human bones. J Archeol Sci. 23:667-671

Basler CF, Reid AH, Dybing JK, Janczewski T A, Fanning TG, Zheng H, Salvatore M, Perdue ML, Swayne DE, Garcia-Sastre A, Palese P, Taubenberger JK. 2001. Sequence of the 1918 pandemic influenza virus non-structural gene (NS)

segment and characterization of recombinant viruses bearing the 1918 NS genes. Proc Natl Acad Sci USA 98:2746-2751

Bate AL, Ma JK, Pitt Ford TR. 2000. Detection of bacterial virulence genes associated with infective endocarditis in infected root canals. Int Endod J 33:194-203

Belzoni GB. 1921. Narratives of the Operations and Recent Discoveries within the Pyramids, Temples, Tombs and Excavations in Egypt and Nubia. London:John Murray

Bendall KE, Macaulay VA, Baker JR, Sykes BC. 1996. Heteroplasmic point mutations in the human mtDNA Control Region. Am J Hum Genet 59:1276-1287

Benfer RA, Typpo JT, Graf VB, Picket EE. 1978. Mineral analysis of ancient Peruvian hair. Am J Phys Anthropol 48:277-282

Beresford JN. 1989. Osteogenic stem cells and the stromal system of bone and marrow. Clin Orthop 240 :270-280

Birbeck MSC, Mercer EH, Barnicott NA. 1956. The structure and formation of pigment granules in human hair. Exp Cell Res 10:505-514

Blin N, Stafford DW. 1976. A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res 3:2303-2308

Bogli A. 1980. Karst Hydrology and Physical Speleology. (Translated by JC Schmid) New York: Springer Verlag

Bonnichsen R, Beatty MT, Turner MD, Stoneking M. 1994. What can be learned from hair? A hair record from the Mammoth Meadow locus, southwestern Montana. In: Akazawa T, Szathmary E (Eds) Prehistoric Mongoloid Dispersions. OUP:Cambridge. Pp 201-213

Bonnichsen R, Bolen CW. 1985. On the importance of hair for Pleistocene studies. Current Research in the Pleistocene 3:63-64

Bonnichsen R, Hodges L, Ream W, Field K, Kirner DL, Selsor K, Taylor RE. 2001. Methods for the study of ancient hair:Radiocarbon dates and gene sequences from individual hairs, J Archeol Sc 28:775-785

Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM, van der Noordaa J. 1990. Rapid and simple method for purification of nucleic acids. J Clin Microbiol 28: 495-503

Braun M, Cook DC, Pfeiffer S. 1998. DNA from *Mycobacterium tuberculosis* complex identified in North American, pre-Columbian human skeletal remains. J Archeol Sci 25:271-277

Brier B. 1998. The Encyclopedia of Mummies. New York:Checkmark Books

Bronk Ramsey C. 1995. Radiocarbon calibration and analysis of stratigraphy: The OxCal Program. Radiocarbon 37:425-430

Bronk Ramsey C, Pettitt PB, Hedges REM, Hodgins GWL, Owen DC. 2000. Radiocarbon dates from the Oxford AMS system: archaeometry datelist 29. Archaeometry 42:243-254

Brown P. 2000. Australian Pleistocene variation and the sex of Lake Mungo 3. J Hum Evol 38:743-749

Bruno B. 2001. Archeologia nei Sassi di Matera. In: Patitucci Uggeri S (Ed) Scavi Medievali in Italia, 1996-1999. Herder:Rome. Pp 137-148

Buckley SA, Evershed RP. 2001. Organic chemistry of embalming agents in Pharonic and Graeco-Roman mummies. Nature 413:837-841

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

Cano RJ, Borucki MK, Higby-Schweitzer M, Poinar HN, Poinar GO Jr, Pollard KJ. 1994. Bacillus DNA in fossil bees: an ancient symbiosis? Appl Environ Microbiol 60:2164-167

Cano RJ, Poinar H, Poinar Jr GO. 1992a. Isolation and partial characterisation of DNA from the bee *Problebeia dominicana* (*Apidae*:*Hymenoptera*) in 25-40 million year old amber. Med Sci Res 20:249-251

Cano RJ, Poinar HN, Pieniazek NJ, Acra A, Poinar Jr GO. 1993. Amplification and sequencing of DNA from a 120-135-million-year-old weevil. Nature 363:536-538

Cano RJ, Poinar HN, Roubik DW, Poinar Jr GO. 1992b. Enzymatic amplification and nucleotide sequencing of portions of the 18S rRNA gene of the bee *Problebeia dominicana* (*Apidae*:*Hymenoptera*) isolated from 25-40 million year old Dominican amber. Med Sci Res 20:619-622

Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, Chiarelli B, Dupanloup I, Bertranpetit J, Barbujani G, Bertorelle G. 2003. Evidence for a genetic discontinuity between Neanderthal and 24,000-year-old anatomically modern Europeans. Proc Natl Acad Sci USA 100:6593-6597

Carracedo A, Bar W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M. 2000. DNA commission of the international society for forensic genetics:guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79-85

Carter H, Mace AC. 1927–1933. The Tomb of Tutankhamun. 3 vols. London:Cassell and Co

Catranis C, Starmer WT. 1991. Microorganisms entrapped in glacial ice. Antarc J US 26:324-326

Cattaneo C, Craig OE, James NT, Sokol RG. 1997. Comparison of three DNA extraction methods on bone and blood stains up to 43 years old and amplification of three different gene sequences. J Forensic Sci 42:1126-1135

Chang D, Clayton D. 1985. Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc Natl Acad Sci USA 82:351-355

Chrisman DR, MacNeish S, Mavalwala J, Savage H. 1996. Late Pleistocene human friction skin prints from Pendejo Cave, New Mexico. American Antiquity 61: 357-376

Christidis L, Leeton PR, Westerman M. 1996. Were bowerbirds part of the New Zealand fauna? Proc Natl Acad Sci USA 93:3898-3901

Christner BC, Mosley-Thompson E, Thompson LG, Zagorodnov V, Sandman K, Reeve JN. 2000. Recovery and identification of viable bacteria immured in glacial ice. Icarus 144:479-485

Cipollaro M, Di Bernardo G, Forte A, Galano G, De Masi L, Galderisi U, Guarino FM, Angelini F, Cascino A. 1999. Histological analysis and ancient DNA amplification of human bone remains found in Caius Iulius Polybius house in Pompeii. Cro Med J 40:392-397

Cipollaro M, Di Bernardo G, Galano G, Galderisi U, Guarino F, Angelini F, Cascino A. 1998. Ancient DNA in human bone remains from Pompeii archaeological site. Biochem Biophys Res Commun 247:901-904

Clisson I, Keyser C, Francfort HP, Crubézy E, Samashev Z, Ludes B. 2002. Genetic analysis of human remains from a double inhumation in a frozen kurgan in Kazakhstan (Berel site, Early 3rd Century BC). Int J Legal Med 116:304-308

Cohn S. 2002. The Black Death transformed: Disease and culture in early Renaissance Europe. London:Arnold

Collins MJ, Waite ER, van Duin ACT. 1999. Predicting protein decomposition: the case of aspartic-acid racemization kinetics. Philos Trans Roy Soc Lond B 354:51-64

Collura RV, Stewart CB. 1995. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. Nature 378:485-489

Colson IB, Richards MB, Bailey JF, Sykes BC, Hedges REM. 1997. DNA analysis of seven human skeletons excavated from the Terp of Wijnaldum. J Archeol Sci 24:911-917

Conrads G, Gharbia SE, Gulabivala K, Lampert F, Shah HN. 1997. The use of a 16S rDNA directed PCR for the detection of endodontopathogenic bacteria. J Endod 23:433-438

Cooper A. 1994. DNA from museum specimens. In: Ancient DNA. Hermann B, Hummel S (Eds) New York: Springer-Verlag. Pp 149-165

Cooper A. 1997. Reply to Stoneking: Ancient DNA - How do you really known when you have it? Am J Hum Genet 60:1001-1002

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

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

Cooper A, Poinar H. 1994. Ancient DNA Newsletter 2(1):5-7

Cooper A, Poinar H. 2000. Ancient DNA: Do it right or not at all. Science 289:1139

Cooper A, Rambaut A, Macaulay V, Willerslev E, Hansen AJ, Stringer C. 2001b. Human origins and ancient human DNA. Science 292:1655-1656

Cooper RR, Milgram JW, Robinson RA. 1966. Morphology of the osteon: an electron microscopic study. J Bone Joint Surg 48A:1239-1271

Crichton M. 1991. Jurassic Park. London:Century

Crubézy É, Ludes B, Poveda JD, Clayton J, Crouau-Roy B, Montagnon D. 1998. Identification of *Mycobacterium* DNA in an Egyptian Pott's disease of 5,400 years old. C R Acad Sci III 321:941-951

Currey J. 2002. Bones. Structure and Mechanics. Princeton:Princeton University Press

Curtis TA, Ashgrafi SH, Weber DF. 1985. Canalicular communication in the cortices of human long bones. Anat Rec 212:336-344

Cymbron T, Loftus RT, Malheiro MI, Bradley DG. 1999. Mitochondrial sequence variation suggest an African influence in Portugese cattle. Proc R Soc Lond B 266:597-603

Demarchi DA, Panzetta-Dutari GM, Colanonio SE, Marcellino AJ. 2001. Absence of the 9-bp deletion of mitochondrial DNA in pre-Hispanic inhabitants of Argentina. Hum Biol 73:757-582

Deng W, Burland V, Plunkett III G, Boutin A, Mayhew GF, Liss P, Perna NT, Rose DJ, Mau B, Zhou S, Schwartz DC, Fetherston JD, Lindler LE, Brubaker RR, Plano GV, Straley SC, McDonough KA, Nilles NL, Matson JS, Blattner FR, Perry RD. 2002. Genome sequence of *Yersinia pestis* KIM. J Bacteriol 184:4601-4611

de Pancorbo MM, Castro A, Alonso S, Fernandez-Fernandez I, Barbero C, Garcia-Orad A, Izaguirre N, Iriondo M, de la Rua C. 1995. Genetic typing with HUMTH01, HUMVWA31A and HUMFES/FPS short tandem repeat loci, D1S80 variable number tandem repeat locus and HLA-DQ alpha of recent and from XII-XIII centuries spongy bone. Electrophoresis:1612-1616

DeSalle R, Barcia M, Wray C. 1993. PCR jumping in clones of 30-million-year-old DNA fragments from amber preserved termites (*Mastotermes electrodominicus*). Experientia 49:906-909

DeSalle R, Gatesy J, Wheeler W, Grimaldi D. 1992. DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. Science 257:1933-1936

DeSalle R, Grimaldi D. 1994. Very old DNA. Curr Opin Genet Dev 4:810-815

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

Di Bernardo G, Del Gaudio S, Cammarota M, Galderisi U, Cascino A, Cipollaro M. 2002. Enzymatic repair of selected cross-linked homoduplex molecules enhances nuclear gene rescue from Pompeii and Herculaneum remains. Nucleic Acids Res. 30:e16 1-6

Dinner A, Blackburn G, Karplus M. 2001. Uracil-DNA glycosylase acts by substrate autocatalysis. Nature 413:752-755

Doda ND, Wright CT, Clayton DA. 1981. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc Natl Acad Sci USA 10: 6116-6120

Donoghue HD, Spigelman M, Zias J, Gernaey-Child AM, Minnikin DE. 1998. *Mycobacterium tuberculosis* complex DNA in calcified pleura from remains 1400 years old. Lett Appl Microbiol 27:265-269

Douglas MP, Rogers SO. 1998. DNA damage caused by common cytological fixatives. Mutat Res 401:77-88

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

Drancourt M, Raoult D. 2002. Molecular insights into the history of plague. Microbes Infect 4:105-109

Drummond AJ, Rambaut A (2003). BEAST v1.0, Available from http://evolve.zoo.ox.ac.uk/beast/

Dubeau L, Chandler LA, Gralow Jr, Nichols PW, Jones PA. 1986. Southern blot analysis of DNA extracted from formalin-fixed pathology specimens. Cancer Res 46:2964-2969

Dumolin-Lapegue S, Pemonge H-M, Gielly L, Taberlet P, Petit RJ. 1999. Amplification of oak DNA from ancient and modern wood. Mol Ecol 8:2137-2140

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:IRL Press/Oxford University Press. Pp 225-244

Endicott P, Gilbert MTP, Stringer C, Lalueza-Fox C, Willerslev E, Hansen AJ, Cooper A. 2003. The genetic origins of the Andaman islanders. Am J Hum Genet 72:178-184

Eshleman J, Smith DG. 2001. Use of DNAase to eliminate contamination in ancient DNA analysis. Electrophoresis 22:4316-4319

Essack SY, Hall LM, Pillay DG, McFayden ML, Livermore DM. 2001. Complexity and diversity *of Klebsiella pneumoniae* strains with extended-spectrum beta-lactamases isolated in 1994 and 1996 at a teaching hospital in Durban, South Africa. Antimicrob Agents Chemother 45:88-95

Excoffier L, Yang Z. 1999. Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol 16:1357-1368

Faerman M, Filon D, Kahila G, Greenblatt C, Smith P, Oppenheim A. 1995. Sex identification of archaeological human remains based on amplification of the X and Y amelogenin alleles. Gene 167:327-332

Faerman M, Kahila G, Smith P, Greenblatt C, Stager L, Filon D, Oppenheim A. 1997a. DNA analysis reveals the sex of infanticide victims. Nature 385:212-213

Faerman M, Jankauskas R, Gorski A, Bercovier H,
Greenblatt CL. 1997b. Prevalence of human tuberculosis in a Medieval population of Lithuania studied by ancient DNA analysis. Ancient Biomol 1:205-214

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

Filippov AA, Solodovnikov NS, Kookleva LM, Protsenko OA. 1990. Plasmid content in *Yersinia pestis* strains of different origin. FEMS Microbiol Lett 55:45-48

Fily M-L, Crubézy É, Courtaud P, Keyser C, Ébrard D, Ludes B. 1998. Analyse paléogénétique des sujets de la grotte sépulcrale d'Elzarreko Karbia (Bronze ancien, Pays Basque). C R Acad Sci Paris série III 321:79-85

Fincham AG, Moradian-Oldak J, Simmer JP. 1999. The structural biology of the developing dental enamel matrix. J Struct Biol 126:270-299

Finlay BJ, Clarke KJ. 1999. Ubiquitous dispersal of microbial species. Nature 400:828

Finlay BJ, Esteban GF, Olmo JL, Tyler PA. 1999. Global distribution of free-living microbial species. Ecography 22:138-144

Finnilä S, Lehtonen M, Majamaa K. 2001. Phylogenetic network for European mtDNA. Am J Hum Genet 68:1475-1484

Fish SA, Shepherd TJ, McGenity TJ, Grant WD. 2002. Recovery of 16S ribosomal RNA gene fragments from ancient halite. Nature 417:432-436

Flagstad O, Walker CW, Vila C, Sundqvist AK, Fernholm B, Hufthammer AK, Wiig O, Koyola I, Ellegren H. 2003. Two centuries of the Scandinavian wolf population: patterns of genetic variability and migration during an era of dramatic decline. Mol Ecol 12:869-880

Fleming S, O'Connor D, Fishman B, Silverman D. 1980. The Egyptian Mummy. Secrets and Science. Philadelphia:The University Museum, University of Pennsylvania

Fletcher HA, Donoghue HD, Holton J, Pap I, Spigelman M. 2003a. Widespread occurrence of *Mycobacterium tuberculosis* DNA from 18th-19th century Hungarians. Am J Phys Anthropol 120:144-152

Fletcher HA, Donoghue HD, Taylor GM, van der Zanden AG, Spigelman M. 2003b. Molecular analysis of Mycobacterium tuberculosis DNA from a family of 18th century Hungarians. Microbiology 149:143-151

Forsland B, Swanbeck G. 1966. Keratin formation in the hair follicle. Exp Cell Res 43:191-209

Forster L, Forster P, Lutz-Bonengel S, Willkomm H, Brinkmann B. 2002. Natural radioactivity and human mitochondrial DNA mutations. Proc Natl Acad Sci USA 99:13950-13954

Frank RM, Nalbandian J. 1989. Structure and ultrastructure of dentine. In: Okshe A, Vollrath L (Eds) Teeth. Berlin:Springer Verlag. Pp. 173-247

Frasca P. 1981. Scanning-electron microscopy studies of "ground substance" in the cement lines, resting lines, hypercalcified rings and reversal lines of human cortical bone. Acta Anatomica 109:114-121

Fraser RB, MacRae TP, Rogers GE. 1972. Keratins: Their composition, structure and biosynthesis. Springfield:Charles E. Thomas

Fricker EJ, Spigelman M, Fricker CR. 1997. The detection of *Escherichia coli* DNA in the ancient remains of Lindow Man using the polymerase chain reaction. Lett Appl Microbiol 24:351-354

Froese DG, Westgate J, Preece S, Storer J. 2002. Age and significance of Dawson tephra in eastern Beringia. Quat Sci Rev 21:2133-2138

Frost HM. 1960. Presence of microscopic cracks in vivo in bone. Henry Ford Hospital Bulletin 8:25-35

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

Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sbisa E, Saccone C.1989. The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J Mol Evol 28:497-516

Gaensslen RE, Berka KM, Herrin G Jr, Mishler RE, Sveum LK, Lee HC. 1993. Amplification of a genomic sequence in 19th century human bone DNA. Naturwissenschaften 80:80-81

Gagneux P, Boesch C, Woodruff DS. 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. Mol Ecol 6:861-868

Geigl E-M. 2001. Inadequated use of moleular hybridization to analyze DNA in Neanderthal fossils. Am J Hum Genet 68:287-291

Geigl E-M. 2002. On the circumstances surrounding the preservation and analysis of very old DNA. Archaeometry 44:337-342

Gernaey AM, Minnikin DE, Copley MS, Dixon RA, Middleton JC, Roberts CA. 2001. Mycolic acids and ancient DNA confirm and osteological diagnosis of tuberculosis. Tuberculosis 81:259-265

Gerstenberger J, Hummel S, Schultes T, Hack B, Herrman B. 1999. Reconstruction of a historical genealogy by means of STR analysis and Y-haplotyping of ancient DNA. Eur J Hum Genet 7:469-477

Gilbert MTP, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, Cooper A. 2003a. Characterisation of genetic miscoding lesions caused by post mortem damage. Am J Hum Genet 72:48-61

Gilbert MTP, Willerslev E, Hansen AJ, Rudbeck L, Barnes I, Lynnerup N, Cooper A. 2003b. Distribution patterns of post mortem damage in human mitochondrial DNA. Am J Hum Genet 72:32-47

Gilbert MTP, Cuccui J, White W, Lynnerup N, Titball RW, Cooper A, Prentice MB 2004a. Absence of *Y. pestis*-specific DNA in human teeth from five European excavations of putative plague victims. Microbiology 150:341-354

Gilbert MTP, Barnes I, Collins MJ, Smith C, Eklund J, Goudsmit J, Poinar H, Cooper A. 2004b. News and Comments: The long-term survival of ancient DNA in Egypt: Response to Zink and Nerlich 2003. Am J Phys Anth (*in press*)

Gillespie R, Roberts RG. 2000. On the reliability of age estimates for human remains at LakeMungo. J Hum Evol 38:727-732

Gissi C, Gullberg A, Arnason U. 1998. The complete mitochondrial DNA sequence of the rabbit, *Oryctolagus cuniculus*. Genomics 50:161-169.

Glenn Smith D, Malhi RS, Eshlemann J, Lorenz JG, Kaestle FA. 1999. Distribution of mt DNA haplogroup X among native north americans. Am J Phys Anthropol 110:271-284

Gocke C, Benko F, Rogan P. 1998. Transmission of mitochondrial DNA heteroplasmy in normal pedigrees. Hum Genet 102:182-186

Goebel AM, Simmons JE. 1993. Preserving DNA in liquid-preserved musuem specimens. In: Snyder AM (Ed) The 1992 American Society of Ichthylologists and Herpetologists Workshop on Collections Care and Management Issues. Kansas :Publication of the division of Herpetology. Pp 34-52

Golenberg EM. 1991. Amplification and analysis of Miocene plant fossil DNA. Philos Trans R Soc Lond B 333:419-26; discussion 426-7

Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, Zurawski G. 1990. Chloroplast DNA sequence from a miocene Magnolia species.Nature 344:656-658

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

González-Olivier A, Márquez-Morfin L, Jiminéz JC, Torre-Blanco A. 2001. Founding amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo. 2001. Am J Phys Anthropol 116:230-235

Götherström A, Collins MJ, Angerbjörn, Lidén K. 2002. Bone preservation and DNA amplification. Archaeometry 44:395-404

Grauer D, Pupko T. 2001. The Permian bacterium that isn't. Mol Biol Evol 18:1143-1146

Greenwood AD, Capelli C, Possner G, Pääbo S. 1999. Nuclear DNA sequences from late Pleistocene megafauna. Mol Biol Evol 16:1466-1473

Greenwood AD, Lee F, Capelli C, DeSalle R, Tikhonov A, Marx PA, MacPhee RD. 2001. Evolution of endogenous retrovirus-like elements of the woolly mammoth (*Mammuthus primigenius*) and its relatives. Mol Biol Evol 18:840-847

Greer S, Zamenhof S. 1962. Studies on depurination of DNA by heat. J Mol Biol 4:123

Guhl F, Jaramillo C, Vallejo GA, Yockteng R, Cardenas-Arroyo F, Fornaciari G, Arriaza B, Aufderheide AC. 1999. Isolation of *Trypanosoma cruzi* DNA in 4,000-year-old mummified human tissue from Northern Chile. Am J Phys Anthropol 108:401-407

Guhl F, Jaramillo C, Yockteng R, Vallejo GA, Cardenas-Arroyo F. 1997. *Trypanosoma cruzi* DNA in human mummies. Lancet 349:1370

Gutierrez G, Marin A. 1998. The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. Mol Biol Evol 15:926-929.

Haas CJ, Zink A, Molnar E, Szeimies U, Reischl U, Marcsik A, ArdagnaY, Dutour O, Palfi G, Nerlich A. 2000. Molecular evidence for different stages of tuberculosis in ancient bone samples from Hungary. Am J Phys Anthropol 113:293-304

Hacker G. 2000. The morphology of apoptosis. Cell Tissue Res 301:5-17

Hackett C. 1976. Microscopical focal destruction (tunnels) in exhumed human bone. Medicine Science and the Law. Vol 12(4). Pp243-264

Haddrath O, Baker AJ. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proc R Soc Lond B Biol Sci 268:939-945

Hagelberg E. 2003. Recombination of mutation rate heterogeneity? Implications for Mitochondrial Eve. Trends Genet 19:84-90

Hagelberg E, Bell L, Allen T, Boyde A, Jones S, Clegg J. 1991. Analysis of ancient bone DNA: techniques and applications. Philos Trans Roy Soc Lond B 333:399-407

Hagelberg E, Clegg JB. 1991. Isolation and characterization of DNA from archaeological bone. Proc Roy Soc Lond B 244:45-50

Hagelberg E, Sykes B. 1989. Ancient bone DNA amplified. Nature 342:485

Hagelberg E, Thomas MG, Cook Jr CE, Sher AV, Baryshnikov GF, Lister AM. 1994. DNA from ancient mammoth bones. Nature 370:333-334

Halliwell B. 1991. DNA damage by oxygen-derived species. Its mechanisms and measurement in mammalian systems. FEBS Lett 281:9-19

Hamazaki S, Koshiba M, Habuchi T, Takahashi R, Sugiyama T. 1993. The effect of formalin fixation on restriction endonuclease digestion of DNA and PCR amplification. Path Res Pract 189:553-557

Hancox N. 1956. The osteoclast. In: The Biochemistry and Physiology of Bone. Bourne GH (Ed) London:Academic Press. Pp 213-250

Handt O, Höss M, Krings M, Pääbo S. 1994a. Ancient DNA:methodological challenges. Experientia 50:524-529

Handt O, Richards M, Trommsdorf M, Kilger C, Simanainen J, Georgiev O, Bauer K, Stone A, Hedges R, Schaffner W, Utermann G, Sykes B, Pääbo S. 1994b. Molecular genetic analyses of the Tyrolean Ice Man. *Science* 264:1775-1778

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

Hänni C, Laudet V, Coll J, Stehelin D. 1994a. An unusual mitochondrial DNA sequence variant from an Egyptian mummy. Genomics 22:487-489

Hänni C, Laudet V, Sakka M, Bègue A, Stéhelin D. 1990. Amplification de fragments d'ADN mitochondrial à oartur de debts et d'os humains anciens. CR Acad Sci Paris 310: 365-370

Hänni C, Laudet V, Stehelin D, Taberlet P. 1994b. Tracking the origins of the cave bear (*Ursus spelaeus*) by mitochondrial DNA sequenceing. Proc Natl Acad Sci USA 91:12336-12340

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

Hardy C, Callou C, Vigme J-D, Casane D, Dennebouy N, Mounolou JC, Monnerot M. 1995. Rabbit mitochondrial DNA diversity from prehistory to modern times. J Mol Evol 40:227-237

Hardy C, Casane D, Vigme J-D, Callou C, Dennebouy N, Mounolou JC, Monnerot M. 1994. Ancient DNA from bronze age bones of European rabbit (*Oryctolagus cuniculus*). Experientia 22:487-489

Hardy MH. 1982. The secret life of the hair follicle. Trends Genet 8:55-61

Hasegawa M, Di Rienzo A, Kocher T, Wilson A. 1993. Towards a more accurate time scale for the human mitochondrial gene tree. J Mol Evol 37:347-354

Hasegawa M, Horai S. 1991. Time of the deepest root for polymorphism in human mitochondrial DNA. J Mol Evol 32:37-42

Hassanin A. 2002. Ancient specimens and DNA contamination: a case study from the 12S rRNA gene sequence of the "Linh Duong" bovid (*Pseudonovibos spiralis*). Naturwissenschaften 89:107-110

Hauf J, Waddell PJ, Chalwatzis N, Joger U, Zimmermann FK. 2000. The complete mitochondrial genome sequence of the African elephant (*Loxodonta africana*), phylogenetic relationships of *Proboscidea* to other mammals and D-loop heteroplasmy. Zoology 102:184-195

Hauswirth WW, Dickel CD, Lawlor DA. 1994. DNA analysis of the Windover population. In: Hermann B, Hummel S (eds.) Ancient DNA. New York:Springer-Verlag. pp104-121.

Hawkins D. 1990. The Black Death and the new London cemeteries of 1348. Antiquity 60, 637-642

Haynes S, Searle JB, Bretman A, Dobney KM. 2002. Bone preservation and ancient DNA: The application of screening methods for predicting DNA survival. J Archeol Sci 29:585-592

Hedges REM. 2002. Bone diagenesis: an overview of processes. Archaeometry 44:319-328

Hedges REM, Law LA, Bronk CR, Housley RA. 1989. The Oxford accelerator mass spectrometry facility technical developments in routine dating. Archaeometry 31:99-113

Hedges REM, Millard AR. 1995 Bones and groundwater - towards the modeling of diagenetic processes. J Archaeol Sci 22:155-164

Hedges REM, Millard AR, Pike AWG. 1995. Measurements and relationships of diagenetic alteration of bone from three archaeological sites, J Archeol Sci 22:201-209.

Hedges REM, van Klinken GJ. 1992. A review of current approaches in the pretreatment of bone for radiocarbon dating by AMS. Radiocarbon 34:306-311

Hedges SB, Schweitzer MH. 1995. Detecting dinosaur DNA. Science 268:1191-1192

Heller A. 2000. On the hypothesis of cathodic protection of genes. Faraday Discuss 116:1-13

Henikoff S. 1995. Detecting Dinosaur DNA. Science 268:1192

Hermann B, Hummel S (Eds). 1994. Ancient DNA. New York:Springer Verlag

Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D. 2001. Phylogenetic and familial estimates of mitochondrial substitution rates: Study of control region mutations in deep-rooting pedigrees. Am J Hum Genet 69:1113-1126

Hideo A. 1988. Nefertari ohi bo no biseibutsu ni tsuite (On microorganisms in the tomb of Nefertari, Egypt). Hozon kagaku 27: 13-20

Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, and extinct member of the horse family. Nature 312:282-284

Higuchi R, von Beroldingen CH, Sensabaugh GF, Erlich HA. 1988. DNA typing from single hairs. Nature 332:543-545.

Higuchi R, Wilson AC. 1984. Fedn Proc 43:1557.

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. Sunderland (Massachusetts):Sinauer Associates, Inc. Pp 321-381

Hinnebusch J, Schwan TG. 1993. New method for plague surveillance using polymerase chain reaction to detect Yersinia pestis in fleas. J Clin Microbiol 31:1511-1514

Hofreiter M, Capelli C, Krings M, Waits L, Conard N, Munzel S, Rabeder G, Nagel D, Paunovic M, Jambresic G, Meyer S, Weiss G, Pääbo S. 2002. Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late pleistocene cave bears. Mol Biol Evol 19:1244-1250

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

Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, Possnert G, Pääbo S. 2000. A molecular analysis of ground sloth diet through the last glaciation. Mol Ecol 9:1975-1984

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. 2001b. Ancient DNA. Nat Rev Genet 2:353-358

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92:532-536

Horai S, Hayasaka K, Murayama K, Wate N, Koike H, Nakai N. 1989. DNA amplification from ancient human skeletal remains and their sequence analysis. Proc Jpn Acad B 65:229-233

Hoshino E, Naomi A, Michiko S, Kohichi K. 1992. Bacterial invasion of non-exposed dental pulp. Int Endod J 25, 2-5

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

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

Höss M, Pääbo S, Vereshchagin NK. 1994. Mammoth DNA sequences. Nature

370:333

Howell N, Smejkal CB, Mackay DA, Chinnery PF, Turnbull DM, Herrnstadt C. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: There is a difference between phylogenetic and pedigree rates. Am J Hum Genet 72:659-670

Howland DE, Hewitt GM. 1994. DNA analysis of extant and fossil beetles. In: Biomolecular paleontology:Lyell meeting volume. NERC Earth Sciences Directorate:Special Publication N. 94/1. Pp 34-51

Hughes MA, Jones DS. 1985. Body in the bog but no DNA. Nature 323:208

Hukkelhoven MWAC, Vromans E, Markslag AMG, Vermoken AJM. 1981. Anticancer Res 1:341-344

Hummel S, Herrmann B. 1991. Y-chromosome-specific DNA amplified in ancient human bone. Naturwissenschaften 78:266-267

Hummel S, Schmidt D, Kahle M, Herrmann B. 2002. ABO blood group genotyping of ancient DNA by PCR-RFLP. Int J Leg Med 116:327-333

Hunter JAA. 1979. Transmission electron microscopy. In: Marks R (Ed) Investigative Techniques in Dermatology. Oxford:Blackwell Scientific Publications. Pp 151-173

Jackson PJ, Hugh-Jones ME, Adair DM, Green G, Hill KK, Kuske CR, Grinberg LM, Abramova FA, Keim P. 1998. PCR analysis of tissue samples from the 1979 Sverdlovsk anthrax victims: the presence of multiple *Bacillus anthracis* strains in different victims. Proc Natl Acad Sci USA 95:1224-1229

Janaway RC. 2002. Degradation of clothing and other dress materials associated with buried bodies of both archaeological and forensic interest. In: Haglund WD, Sorg MH (Eds) Advances in Forensic Taphonomy. Boca Raton:CRC Press. Pp 279-402

Janczewski D, Yukhi N, Gilbert D, Jefferson G, O'Brien S. 1992. Molecular phylogenetic inference from saber-toothed cat fossils of Rancho La Brea. Proc Natl Acad Sci USA 89:9769-9773

Janke A, Feldmaier-Fuchs G, Thomas WK, von Haeseler A, Pääbo S.1994. The marsupial mitochondrial genome and the evolution of placental mammals. Genetics 137:243-256

Janke A, Gemmell NJ, Feldmaier-Fuchs G, von Haeseler A, Pääbo S. 1996.The mitochondrial genome of a monotreme - the platypus (*Ornithorhynchus anatinus*).J Mol Evol 42:153-159

Janke A, Xu X, Arnason U. 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among *Monotremata*, *Marsupialia*, and *Eutheria*. Proc Natl Acad Sci USA. 94:1276-1281

Jans MME, Kars H, Nielsen-Marsh CM, Smith CI, Nord AG, Arthur P, Earl N. 2002. *In situ* preservation of archaeological bone: a histological study within a multidisciplinary approach. Archaeometry 44:343-352

Jehaes E, Gilissen A, Cassiman JJ, Decorte R. 1998. Evaluation of a decontamination protocol for hair shafts before mtDNA sequencing. Forensic Sci Int 94:65-71

Jehaes E, Toprak K, Vanderheyden N, Pfeiffer H, Cassiman JJ, Brinkman B, Decorte R. 2001. Pitfalls in the analysis of mitochondrial DNA from ancient specimens and the consequences for forensic DNA analysis: the historical case of the putative heart of Louis XVII. Int J Leg Med 115:135-141

Jones AS, Mian AM, Walker RT. 1966. The action of alkali on some purines and their derivatives. J Chem Soc C:692-695

Junqueira ACM, Lessinger AC, Azeredo-Espin AML. 2002. Methods for the recovery of mitochondrial DNA sequences from musuem specimens of myiasis-causing flies. Med Vet Entomol 16:39-45

Kaestle F, Glenn Smith D. 2001. Ancient mitochondrial DNA evidence for prehistoric population movement: The numic expansion. Am J Phys Anthropol 115:1-12

Kalmar T, Bachrati CZ, Marcsik A, Rasko I. 2000. A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. Nucleic Acids Res 28:E67

Karlsen F, Kalantari M, Chitemerere M, Johansson B, Hagmar B. 1994. Modifications of human and viral deoxyribonucleic acid by formaldehyde fixation. Lab Invest 71:604-611

Karran P, Lindahl T. 1980. Hypoxanthine in deoxyribonucleic acid: Generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. Biochemistry 19:6005-6011

Kars EAK, Kars H (Eds). 2002 The Degradation of Bone as an Indicator for the Deterioration of the European Archaeological Property. Final Report. Amersfoort:Rijksdienst voor het Oudheidkundig Bodemonderzoek

Kaufman DS, Manley WF. 1998. A new procedure for determining DL amino acid ratios in fossils using reverse phase liquid chromatography. Quaternary Science Reviews, 17:987-1000

Khan AS, Sanchez A, Pflieger AK. 1998. Filoviral haemorrhagic fevers. Br Med Bull 54:675-692

Kim KS. Lee SE, Jeong HW, Ha JH. 1998. The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. Mol Phylogenet Evol 10:210-220

Kiple KF. 1993. The Cambridge World History of Human Disease. Cambridge: Cambridge University Press

Kodaka T, Debari K, Yamada M. 1991. Physiochemical and morphological studies of horse dentin. J Electron Microsc 40:385-391

Kolman CJ. 1999. Molecular Anthropology – Progress and Perspectives on Ancient DNA Technology. In: Papiha, Deka, Chakraborty (Eds) Genomic Diversity: Applications in Human Population Genetics. New York: Academic/Plenum Publishers

Kolman CJ, Centurion-Lara A. Lukehart SA, Owsley DW, Tuross N. 1999. Identification of *Treponema pallidum* subspecies *pallidum* in a 200-year-old skeletal specimen. J Infect Dis 180:2060-2063

Kolman CJ, Tuross N. 2000. Ancient DNA analysis of human populations. Am J Phys Anthropol 111:5-23

Koshiba M, Ogawa K, Hamazaki S, Sugiyama T, Ogawa O, Kitajima T. 1993. The effect of formalin fixation on DNA and the extraction of high-molecular- weight DNA from fixed and embedded tissue biopsies. PCR Methods Appl 2:175-176

Krajewski C, Buckley L, Westerman M. 1997. DNA phylogeny of the marsupial wolf resolved. Proc R Soc Lond B Biol Sci 264:911-917

Krajewski C, Driskell A, Baverstock P, Braun M. 1992. Phylogenetic relationships of the thylacine (Mammalia: Thylacinidae) among dasyuroid marsupials: evidence from cytochrome b DNA sequences. Proc R Soc Lond B 250:19-27

Krettek A, Gullberg A, Arnason U. 1995. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the *Lipotyphla*. J Mol Evol 41:952-957

Krings M, Geisert H, Schmitz R, Krainitzki H, Pääbo S. 1999. DNA sequence of the mitochondrial hypervariable region II from the Neanderthal type specimen. Proc Natl Acad Sci USA 96:5581-5585

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

Küch M, Rohland N, Betancourt JL, Latorre C, Steppan S, Poinar HN. 2002. Molecular analysis of an 11,700-year-old rodent midden from the Atacama Desert, Chile. Mol Ecol 11:913-924

Kühn R, Schröder W, Rottman O. 2001. Sequencing mtDNA of the cave bear *Ursus spelaeus* from the Bavarian Alps is feasible by nested and touchdown PCR. Acta Theriologica 46:61-68

Kurosaki G, Matsushita T, Ueda S. 1993. Individual DNA identification from ancient human remains. Am J Hum Genet 53:638-643

Kuznetsov GV, Kulikov EE, Petrov NB, Ivanova NV, Lomov AA, Kholodova MV, Poltaraus AB. 2001. The "Linh Duong" *Pseudonovibos spiralis* (*Mammalia, Artiodactyla*) is a new buffalo. Naturwissenschaften 88:123-125

Kwok S, Higuchi R. 1989. Avoiding false positives with PCR. Nature 339:237-238

Lalueza C, Péréz- Péréz A, Prats E, Cornudella L, Turbón D. 1997. Lack of founding Amerindian mitochondrial DNA lineages in extinct aborigines from Tierra del Fuego-Patagonia. Hum Mol Genet 6:41-46

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

Lalueza-Fox C, Gilbert MT, Marinez-Fuentez AJ, Calafell F, Bertranpetit J. 2003. Mitochondrial DNA from pre-Columbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean. Am J Phys Anthropol 121:97-108

Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, Baroni C. 2002. Rates of evolution in ancient DNA from Adelie penguins. Science 295:2270-2273

Lau A, Wyatt M, Glassner B, Samson L, Ellenberger T. 2000. Molecular basis for discrimination between normal and damaged bases by the human alkyladenine glycosylase, AAG. Proc Natl Acad Sci USA 97:13575-13578

Lawlor DA, Dickel CD, Hauswirth WW, Parham P. 1991. Ancient HLA genes from 7,500-year-old archaeological remains. Nature. 349:785-787.

Lee W-J, Conroy J, Huntting Howell W, Kocher TD. 1995. Structure and evolution of teleost mitochondrial Control Regions. J Mol Evol 41:54-66

Leonard JA, Wayne RK, Cooper A. 2000. Population genetics of Ice Age brown bears. Proc Natl Acad Sci USA 97:1651-1654

Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillén S, Vilà C. 2002. Ancient DNA Evidence for Old World Origin of New World Dogs. Science 298:1613-1616.

Li H-C, Fujiyoshi T, Lou H, Yashiki S, Sonoda S, Cartier L, Nunez L, Munoz U, Horai S, Tajima K. 1999. The presence of ancient human T-cell lymphotropic virus type I provirus DNA in an Andean mummy. Nature Medicine 5:1428-1432

Li J, Liao X, Yang H. 2000. Molecular characterization of a parasitic tapeworm (*Ligula*) based on DNA sequences from formalin-fixed specimens. Biochem Genet 38:309-321

Li Y, An C-C, Zhu Y-X. 1995. DNA isolation and sequence analysis of dinosaur DNA from Cretaceous dinosaur egg in Xixia Henan, China. Acta Sci Nat Univ Pekinensis 31:148-152

Linch CA, Whiting DA, Holland MM. 2001. Human hair histogenesis for the mitochondrial DNA forensic scientist. J Forensic Sci 46:844-853.

Linch CA, Smith SL, Prahlow JA. 1998. Evaluation of the human hair root for DNA typing subsequent to microscopic comparison. J Forensic Science 43:305-314

Lindahl T. 1979. DNA glycosylases, endonucleases for apurinic/apyrimidinic sites and base excision-repair. Prog Nucleic Acid Res Molec Biol 22:135-192

Lindahl T. 1993a. Instability and decay of the primary structure of DNA. Nature 362:709-715

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

Lindahl T, Andersson A. 1972. Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. Biochemistry 11:3610-3618

Lindahl T, Karlström O. 1973. Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution. Biochemistry 12:5151-5154

Lindahl T, Nyberg B.1972. Rate of depurination of native deoxyribonucleic acid. Biochemistry 12:5151-5154

Lindahl T, Nyberg B. 1974. Heat induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry 12:3405-3410

Logan GA, Boon JJ, Eglington G. 1993. Structural biopolymer preservation in Miocene leaf fossils from the Clarkia site, northern Idaho. Proc Natl Acad Sci USA 90:2246-2250

Lopez JV, Cevario S, O'Brien SJ. 1996. Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. Genomics 33:229-246

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

Loreille O, Roumat E, Verneau O, Bouchet F, Hänni C. 2001. Ancient DNA from *Ascaris* : extraction amplification and sequences from eggs collected in coprolites. Int J Parasitol 31:1101-1106

Loy TH, Dixon EJ. 1998. Blood residues on fluted points from Eastern Beringia. American Antiquity 63:21-46

Lubec G, Nauer G, Seifert K, Strouhal E, Porteder H, Szilvassy J, Teschler M. 1987. Structural stability of hair over three thousand years. J Archeol Sci 14:113-120

Lubec G, Weniger M, Anderson SR. 1994. Racemization and oxidation studies of hair protein in the Homo Tirolensis. The FASEB Journal 8:1166-1169

Lucas A. 1924. Note on the temperature and humidity of several tombs in the valley of the tombs of the kings at Thebes. Annales du Service des antiquités de l'Égypte 24:12-14

Lundström R, Tavaré S, Ward R. 1992. Estimating substitution rates from molecular data using the coalescent. Proc Natl Acad Sci USA 89:5961-5965

Lynnerup N. 1992. Anthropological Report on human remains fromVodroffsgaard, Copenhagen, AS 3/92. On file at: The Laboratory ofBiological Anthropology, University of Copenhagen, Copenhagen.

Lynnerup N. 1998. The Greenland Norse. A Biological-anthropological Study. Meddelelser Om Grønland, Man & Society 24. Copenhagen

Ma L-J, Rogers SO, Catranis CM, Starmer WT. 1999. Detection and characterization of ancient fungi entrapped in glacial ice. Mycologia 92, 286-295.

Macaulay V, Richards M, Forster P, Bendall K, Watson E, Sykes B, Bandelt H-J. 1997. mtDNA mutation rates - No need to panic. Am J Hum Genet 61:983-986

MacHugh DE, Troy CS, McCormick F, Olsaker I, Eythorsdottir E, Bradley DG. 1999. Early medieval cattle remains from a Scandinavian settlement in Dublin:genetic analysis and comparison with extant breeds. Philos Trans R Soc Lond B 354:99-108

Macko SA, Engel MH, Andrusevich V, Lubec G, O'Connell TC, Hedges REM. 1999. Documenting the diet of ancient human populations through stable isotope analysis of hair. Philos Trans R Soc Lond B 354:65-76

Maekawa S, Preusser FD. 1993. Environmental monitoring at the Tomb of Nefertari. In: ICOM Committee for Conservation, Tenth Triennial Meeting, Washington, DC, 22-27 August 1993. Pp 616-623

Malhi RS, Glenn Smith D. 2002. Brief Communication: Haplogroup X confirmed in prehistoric North America. Am J Phys Anthropol 119:84-86

Malyarchuk BA, Rogozin IB, Berikov VB, Derenko MV. 2002. Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. Hum Genet 111:46-53

Marota I, Basile C, Ubaldi M, Rollo F. 2002. DNA decay rate in papyri and human remains from Egyptian archaeological sites. Am J Phys Anthropol 117:310-318

Marotti G, Zallone AZ. 1980. Changes in the vascular network during the formation of the Haversian systems. Acta Anatomica 106:84-100

Marotti G, Ferretti M, Remaggi F, Palumbo C. 1995. Quantitative evaluation on osteocyte canalicular density in human secondary osteons. Bone 16:125-128

Massa ER and Fuhrman AMC. 1978. Early Egyptian mummy hairs: tensile strength tests, optical and scanning electron microscope observation. J Hum Evol 9:133-137

Mays S, Fysh E, Taylor GM. 2002. Investigation of the link between visceral surface rib lesions and tuberculosis in a Medieval skeletal series from England using ancient DNA. Am J Phys Anthropol 119:27-36

Mays S, Taylor GM, Legge AJ, Young DB, Turner-Walker G. 2001. Paleopathological and biomolecular study of tuberculosis in a medieval skeletal collection from England. Am J Phys Anthropol 114:298-311

Meijer H, Perizonius WRK, Geraedts JPM. 1992. Recovery and identification of DNA sequences harboured in preserved ancient human bones. Biochem Biophy Res Commun. 183:367-374

Merriweather D, Rothhammer F, Ferrel R. 1994. Genetic variation in the New World: ancient teeth, bone, and tissue as sources of DNA. Experientia 50:592-601

Meyer E, Wiese M, Bruchhaus H, Claussen M, Klein A. 2000. Extraction and amplification of authentic DNA from ancient human remains. Forensic Sci Int 113:87-90

Meyer S, Weiss G, von Haeseler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103-1110

Meyer S, von Haeseler A. 2003. Identifying site-specific substitution rates. Mol Biol Evol 20:182-189

Minakata K, Nozawa H, Suzuki KW, Suzuki O. 2002. Paternity diagnosis by using umbilical cords preserved for periods ranging from 9 months to 44 years. Int J Legal Med 116:314-315

Monsalve MV, Stone AC, Lewis CM, Rempel A, Richards M, Straathof D, Devinde DV. 2002. Brief communicatinon: molecular analysis of the Kwadaya Dan Ts'finchi ancient remains found in a glacier in Canada. Am J Phys Anthropol 199:288-291

Montagnon D, Ravaorimananana B, Rakotosamimanana B, Rumpler Y. 2001. Ancient DNA from *Megaladapis edwardsi* (Malagasy Subfossil): preliminary results using partial cytochrome b sequence. Folia Primatol 72:30-32

Montiel R, Malgosa A, Francalacci P. 2001. Authenticating ancient human mitochondrial DNA. Hum Biol 73:689-713

Moraga M, Aspillaga E, Santoro C, Standen V, Carvallo P, Rothhammer F. 2001. Análisis de ADN mitocondrial en momias del norte de Chile avala hipótesis de origen amazónico de poblaciones andinas. Revista Chilena de Historia Natural 74:719-726

Moritz C, Dowling TE, Brown W. 1987. Evolution of animal mitochondrial DNA:Relevance for population biology and systematics. Ann Rev Ecol Syst 18:269-292

Mullender MG, Huiskes R, Verslayen H, Buma P. 1996. Osteocyte density and micromorphometic parameters in cancellous bone of the proximal femur in five mammalian species. J Orthop Res 14:972-979

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

Mumm S, Whyte MP, Thakker RV, Buetow KH, Schlessinger D. 1997. MtDNA analysis shows common ancestry in two kindreds with X-linked recessive hypoparathyroidism and reveals a heteroplasmic silent mutation. Am J Hum Genet 60:153-159

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien S. 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409:514-618

Nerlich AG, Haas CJ, Zink A, Szeimies U, Hagedorn HG. 1997. Molecular evidence for tuberculosis in an ancient Egyptian mummy. Lancet 350:1404

Nicholson GJ, Tomiuk J, Czarnetzki A, Bachmann L, Pusch L. 2002. Detection of bone glue treatment as a major source of contamination in ancient DNA analyses. AJPA 118:117-120

Nickle DC, Learn GH, Rain MW, Mullins JI, Mittler JE. 2002. Curiously modern DNA fro a "250 million-year-old" bacterium. J Mol Evol 54:134-137

Nielsen-Marsh C, Hedges R. 2000. Patterns of diagenesis in bone I: Effects of site environments. J Archeol Sci 27:1139-1150

Noro M, Masuda R, Dubrovo IA, Yoshida MC, Kato M. 1998. Molecular phylogenetic inference of the wooly mammoth *Mammuthus primigenius*, based on complete sequences of mitochondrial cytochrome b and 12S ribosomal RNA genes. J Mol Evol 46:314-326

Notari RE. 1967. A mechanism for the hydrolytic deamination of cytosine arabinoside in aqueous buffer. J Pharm Sci 56:804

Nuorala E. 1999. Tuberculosis on the 17th century Man-of-War Kronan. Int J Osteoarchaeol 9:344-348

O'Connell TC, Hedges REM. 1999a. Isotopic composition of hair and bone: Archaeological analyses. J Archaeol Sci 26: 661-665

O'Connell TC, Hedges REM. 1999b. Investigations into the effect of diet on modern human hair isotopic values. Am J Phys Anthropol 108:409-425

O'Connell TC, Levine MA, Hedges REM. 2003. The importance of fish in the diet of Central Eurasian peoples from the Mesolithic to the Early Iron Age. In: Levine MA, Renfrew C, Boyle K (eds.) Prehistoric Steppe Adaptation and the Horse. Cambridge:McDonald Institute Monograph. Pp 253-268

O'Rourke DH, Carlysle SW, Parr RL. 1996. Ancient DNA: methods, progress, and perspectives. Am J Hum Biol. 8:557-571

O'Rourke DH, Hayes MG, Carlyle SW. 2000. Ancient DNA studies in physical anthropology. Ann Rev Anthropol 29:217-242

O'Rourke DH, Parr RL, Carlyle SW. 1999. Molecular genetic variation in prehistoric inhabitants of the Eastern Great Basin. In: Hemphil BE, Larsen CS (Eds) Understanding prehistoric lifeways in the great basin wetlands: Bioarchaeological reconstruction and interpretation. Salt Lake City:Univ Utah Press. Pp 84-102.

Ohno K, Tanaka M, Suzuki H, Ohbayashi T, Ikebe S, Ino H, Kumar S, Takahashi A, Ozawa T. 1991. Identification of a possible control element, mt5, in the major noncoding region of mitochondrial DNA by intraspecific nucleotide conservation. Biochem Int 24:263-272

Olsen EA. 1994. Disorders of the hair shaft. New York:McGraw-Hill Inc

Oota H, Kurosaki K, Pookajorn S, Ishida T, Ueda S. 2001. Genetic study of the paleolithic and neolithic Southeast Asians. Hum Biol 73: 225-231

Oota H, Saitou N, Matsushita T, Ueda S. 1995. A genetic study of 2,000 year old human remains from Japan using mitochondrial DNA sequences. Am J Phys Anthropol 98:133-145

Oota H, Saitou N, Matsushita T, Ueda S. 1999. Molecular genetic analysis of remains of a 2,000-year-old human population in China - and its relevance for the origin of the modern Japanese population. Am J Hum Genet 64:250-258

Orlando L, Bonjean D, Bocherens H, Thenot A, Argant A, Otte M, Hänni C. 2002. Ancient DNA and the population genetics of cave bears (*Ursus spelaeus*) through space and time. Mol Biol Evol 19:1920-1933

Ovchinnikov I, Goodwin W. 2003. Ancient human DNA from Sungir? J Hum Evol. 44:389-392

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

Owen M, Ashton B. 1986. Osteogenic differentiation of skeletal cell populations. In: Ali SY (Ed) Cell Mediated Calcification and Matrix Vesicles. Amsterdam:Elsevier. Pp 279-284.

Okazaki M, Yoshida Y, Yamaguchi S, Kaneno M, Elliott JC. 2001. Affinity binding phenomena of DNA onto apatite crystals. Biomaterials 22:2459-2464

Ozawa T, Hayashi S, Mikhelson VM. 1997. Phylogenetic position of mammoth and Steller's sea cow within *Tethytheria* demonstrated by mitochondrial DNA sequences. J Mol Evol 44:406-413

Pääbo S. 1985a. Preservation of DNA in ancient Egyptian mummies. J Archeol Sci 12:411-417

Pääbo S. 1985b. Molecular cloning of ancient Egyptian mummy DNA. Nature. 314:644-645

Pääbo S. 1986. Molecular genetic investigations of ancient human remains. Cold Spring Harbour Symp Quant Biol. 51:441-446

Pääbo S. 1989. Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci USA 86:1939-1943

Pääbo S, Gifford JA, Wilson AC. 1988. Mitochondrial DNA sequences from a 7000-year-old brain. Nucleic Acids Res 16:9775-9787

Pääbo S, Higuchi R, Wilson A. 1989. Ancient DNA and the Polymerase Chain Reaction. J Biol Chem 264:9709-9712

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

Page R, Holmes E. 1998. Molecular Evolution - A Phylogenetic Approach. Oxford:Blackwell Science

Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebaihia M, James KD, Churcher C, Mungall KL, Baker S, Basham D, Bentley SD, Brooks K, Cerdeno-Tarraga AM, Chillingworth T, Cronin A, Davies RM, Davis P, Dougan G, Feltwell T, Hamlin N, Holroyd S, Jagels K, Karlyshev AV, Leather S, Moule S, Oyston PC, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature 413:523-527

Parr RL, Carlyle SW, O'Rourke DH. 1996. Ancient Dna analysis of Fremont Amerindians of the Great Salt Lake Wetlands. Am J Phys Anthropol 99:605-609

Parsons TJ, Muciec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM. 1997. A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363-368

Parsons TJ, Weedn VW. 1996. Preservation and recovery of DNA in postmortem specimens and trace samples. In: Haglund WD, Sorg MH (Eds) Forensic Taphonomy: The Postmortem Fate of Human Remains. Boca Raton FL:CRC. Pp109-138

Pedersen AGU, Thomsen TR, Lomstein BA, Jorgensen NOG. 2001. Bacterial influence on amino acid enantiomerization in a coastal marine sediment. Limnol Oceanogr 46: 1358-1369

Pesole G, Saccone C. 2001. A novel method for estimating substitution rate variation among sites in a large dataset of homologous DNA sequences. Genetics 157:859-865

Pfeiffer H, Huhne J, Seitz B, Brinkmann B. 1999. Influence of soil storage and exposure period on DNA recovery from teeth. Int J Legal Med 112, 142-144

Phillips MJ, Lin Y-H, Harrison GL, Penny D. 2001. Mitochondrial genomes of a bandicoot and a brushtail possum confirm the monophyly of Australidelphian Marsupials. Proc R Soc Lond B 268:1533-1538

Pilkington S, Summers C, Thein SL, O'Connor NT, Wainscoat JS. 1987. Hair root DNA: a source of constitutional DNA in leukaemia. Lancet 8524:112

Pitra C, Fürbass R, Seyfert HM. 1997. Molecular phylogeny of the tribe *Bovini* (*Mammalia*:*Artiodactyla*): alternative placement of the *Anoa*. J Evol Biol 10:589-600

Poinar H. 2002. The genetic secrets some fossils hold. Acc Chem Res 35:676-684

Poinar H, Cano R, Poinar G. 1993. DNA from an extinct plant. Nature 363:677

Poinar H, Hofreiter M, Spaulding G, Martin P, Stankiewicz A, Bland H, Evershed R, Possnert G, Pääbo S. 1998. Molecular coproscopy: Dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. Science 281:402-406

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

Poinar HN, Küch 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

Polos'mak NV, Molodin VI. 2000. The Pazyryk culture:Burial sites on the Ukok plateau. Archaeol Ethnol Anthropol Eurasia 4:66-87

Polotoraus AB, Kulikov EE, Lebedeva IA. 2000. The molecular analysis of DNA from the remains of three individuals from Sungir site. In: Alexeeva TI, Bader NO (Eds) *Homo sungirensis*. Upper Paleolithic Man: Ecological and Evolutionary Aspects of the Investigation. Moscow:Scientific World. Pp 351-358

Prince AM, Andrus L. 1992. PCR: How to kill unwanted DNA. BioTechniques 12:358-360

Pritchard JJ. 1956. The osteoblast. In: Bourne GH (Ed) The Biochemistry and Physiology of Bone. New York and London:Academic Press. Pp 179-211

Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA, Baumeister K. 1987. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science 238:336-341

Pumo DE, Finamore PS, Franek WR, Phillips CJ, Tarzami S, Balzarano D. 1998 Complete mitochondrial genome of a neotropical fruit bat, *Artibeus jamaicensis*, and a new hypothesis of the relationships of bats to other eutherian mammals. J Mol Evol 47:709-717

Pusch C, Giddings I Scholz M. 1998. Repair of degraded duplex DNA from prehistoric samples using *Escherichia coli* DNA polymerase I and T4 DNA ligase. Nucleic Acids Res. 26:857-859

Pusch CM, Kayademir T, Prangenberg K, Conard NJ, Czarnetzki A, Blin N. 2002. Documenting ancient DNA quality via alpha satellite amplification and assessment of clone sequence diversity. J Appl Genet 43:351-364

Pusch C, Nicholson GJ, Bachmann L, Scholz M. 2000. Degenerate oligonucleotide-primed preamplification of ancient DNA allows the retreival of authentic DNA sequences. Anal Biochem 279:118-122

Rafi A, Spigelman M, Stanford J, Lemma E, Donoghue H, Zias J. 1994. *Mycobacterium leprae* DNA from ancient bone detected by PCR. Lancet 343:1360-1361

Ramos MD, Lalueza C, Girbeau E, Péréz-Péréz A, Quevedo S, Turbon D, Estivell X. 1995. Amplifying dinucleotide microsatellite loci from bone and tooth samples of up to 5000 years of age:more inconsistency than usefulness. Hum Genet 96:205-212

Randi E, Lucchini V. 1998. Organization and evolution of the mitochondrial DNA control region in the Avian Genus *Alectoris*. J Mol Evol 47:449-462

Raoult D, Aboudharam G, Crubèzy E, Larrouy G, Ludes B. Drancourt M. 2000. Molecular identification by "suicide PCR" of *Yersinia pestis* as the agent of medieval Black Death. Proc Natl Acad Sci USA 97:12800-12803

Raoult D, Drancourt M. 2002. Cause of Black Death. Lancet Infect Dis 2:456

Reed FA, Kontanis EJ, Kennedy KAR, Aquadro CF. 2003. Ancient DNA prospects from Sri Lankan highland dry caves support an emerging global pattern. AJPA 121:112-116

Reid AH, Fanning TG, Hultin JV, Taubenberger JK. 1999. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. Proc Natl Acad Sci USA 96:1651-1656

Ribeiro-dos-Santos A, Santos S, Lucia Machado A, Guapindaia V, Zago M. 1996. Heterogeneity of mitochondrial DNA haplotypes in pre-Columbian natives of the amazon region. Am J Phys Anthropol 101:29-37

Richards M, Macaulay V. 2000. Genetic data and the colonization of Europe: genealogies and founders. In: Renfrew C, Boyle K (Eds). Archaeogenetics: DNA and the Population Prehistory of Europe. Cambridge:McDonald Institure for Archaeological Research. Pp139-151

Richards M, Sykes B, Hedges R. 1995. Authenticating DNA extracted from ancient skeletal remains. J Archeol Sci 22:291-299

Richards MP, Hedges REM. 2003. Bone collagen $d^{13}C$ and $d^{15}N$ values of fauna from Northwest Europe reflect palaeoclimatic variation over the last 40,000 years. Palaeogeogr Palaeoclimatol Palaeoecol 193:261-267

Rickards O, Martínez-Labarga C, Favaro M, Frezza D, Mellegni F. 2001. DNA analyses of the remains of the Prince Branciforte Barresi family. Int J Legal Med 114:141-146

Ringboel Bitsch B. 1991. Archaeological Excavation Report"Vodroffsgaard", KBM 836. On file at: Copenhagen City Museum, Copenhagen

Ristaino JB, Groves CT, Parra G. 2001. PCR amplification of the Irish potato famine pathogen from historic specimens. Nature 411, 695-697

Robin ED, Wong R. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136:507-513

Robinson TJ, Bastos AD, Halanych KM, Herzig B. 1996. Mitochondrial DNA sequence relationships of the extinct blue antelope *Hippotragus leucophaeus*. Naturwissenschaften 83:178-182

Rogan P, Salvo JJ. 1992. Study of nucleic acids isolated from ancient remains. Yearb Phys Anthropol 33:195-214

Rogers SO, Langenegger K, Holdenrieder O. 2000. DNA changes in tissues entrapped in plant resins (the precursors of Amber). Naturwissenschaften 87:70-75

Roon DA, Waits LP, Kendall KC. 2003. A quantitative evaluation of two methods for preserving hair samples. Mol Ecol Notes 3:163-166

Rothschild BM, Martin LD, Lev G, Bercovier H, Bar-Gal GK, Greenblatt C, Donoghue H, Spigelman M, Brittain D. 2001. *Mycobacterium tuberculosis* complex DNA from an extinct bison dated 17,000 years before the present. Clin Infect Dis 33:305-311

Saccone C, Pesole G, Sbisá E. 1991. The main regulatory function of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. J Mol Evol 33:83-91

Sagan, C. 1995. The Demon-Haunted World: Science as a Candle in the Dark. Random House

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

Sallares R, Gomzi S. 2001. Biomolecular archaeology of malaria. Anc Biomol 3:195-213

Salo WL, Aufderheide AC, Buikstra J, Holcomb TA. 1994. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. Proc Natl Acad Sci USA 91:2091-2094

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular Cloning: A Laboratory Manual. 2nd Edition.Cold Spring Harbour NY: Cold Spring Harbour Lab

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

Savolainen P, Lundeberg J. 1999. Forensic evidence based on mtDNA from dog and wolf hairs. J Forensic Sci 44:77-81

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C. 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125-140

Schaffler MB, Burr DB, Frederickson RG. 1987. Morphology of the osteonal cement line in human bone. Anat Rec 217:223-228

Schmitz RW, Serre D, Bonani G, Feine S, Hillgruber F, Krainitzki H, Paabo S, Smith FH. 2002. The Neanderthal sype site revisited: Interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. Proc Natl Acad Sci USA 99:13342-13347

Schultes T, Hummel S, Hermann B. 1999. Amplification of Y-chromosomal STRs from ancient skeletal material. Hum Genet 104:164-166

Schuster H. 1960. Die Reaktionsweise der Desoxyribonucleinsäure mit salpetriger Säure. Z Naturforsch B 15:298-304

Scott S, Duncan C. 2001. Biology of Plagues: Evidence from Historical Populations. Cambridge:Cambridge University Press

Shaaper RM, Kunkel TA, Loeb LA. 1983. Infidelity of DNA synthesis associated with bypass of apurinic sites. Proc Natl Acad Sci USA 80:847-891

Shapiro B. 2003. Inferring Evolutionary History and Processes Using Ancient DNA. PhD Thesis, University of Oxford

Shapiro B, Sibthorpe D, Rambaut A, Austin J, Wragg GM, Bininda-Emonds OR, Lee PL, Cooper A. 2002. Flight of the dodo. Science 295:1683

Shapiro HS, Klein RS. 1966. The deamination of cytidine and cytosine by acidic buffer solutions. Mutagenic Implications. Biochemistry 5:2358-2362

Shapiro R. 1981. Damage to DNA caused by hydrolysis. In: Seeberg E, Kleppe K (Eds) Chromosome Damage and Repair. New York:Plenum Publishing Corp. New York. Pp 3-12

Sidow A, Wilson AC, Pääbo S. 1991. Bacterial DNA in Clarkia fossils. Philos Trans Roy Soc Lond B 333:429-433

Sigurðardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P. 2000. The mutation rate in the human mtDNA control region. Am J Hum Genet 66:1599-1609

Slott EF Jr, Shade RO, Lansman RA. 1983. Sequence analysis of mitochondrial DNA in a mouse cell line resistant to chloramphenicol and oligomycin. Mol Cell Biol 3:1694-1702

Smith BC, Fisher DL, Weedn VW, Warnock GR, Holland MM. 1993. A systematic approach to the sampling of dental DNA. J Forensic Sci 38:1194-1209

Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, Collins MJ. 2001. Neanderthal DNA: not just old but old and cold? Nature 410:772-773

Smith CI, Nielsen-Marsh CM, Jans MME, Arthur P, Nord AG, Collins MJ. 2002. The strange case of Apligliano:early 'fossilization' of medieval bone in southern Italy. Archaeometry 44:405-416

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C,Connell CR, Heiner C, Kent SBH, Hood LE. 1986. Fluorescence detection in automated DNA sequence analysis. Nature 321:674-679

Sofer WH. 1991. Intoduction to Genetic Engineering. Massachutes:Butterworth-Heinemann

Soltis PS. 1995. Fossil DNA:Its potential for biosystematics. In: Hoch PC, Stephenson AG (Eds) Experimental and Molecular Approaches to Plant Biosystematics. Missouri:Missouri Botanical Garden. pp 1-13.

Sorenson MD, Cooper A, Paxinos EE, Quinn TW, James HF, Olson SL, Fleischer RC. 1999. Relationships of the extinct moa-nalos, flightless Hawaiian waterfowl, based on ancient DNA. Proc R Soc Lond B 266:2187-2193

Spigelman M, Lemma E. 1993. The use of the polymerase chain reaction to detect *Mycobacterium tuberculosis* in ancient skeletons. Int J Osteoarch. 3:137-143

Spigelman M, Matheson C, Lev G, Greenblatt C, Donoghue H. 2003. Confirmation of the presenc of *Mycobacterium tuberculosis* complex-specific DNA in three archaeological specimens. Ing J Osteoarchaeol 12:393-401

Stankiewicz B, Poinar H, Briggs D, Evershed R, Poinar G. 1998. Chemical preservation of plants and insects in natural resins. Proc R Soc Lond B 265:641-647

Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, Huttley GA, Allikmets R, Schriml L, Gerrard B, Malasky M, Ramos MD, Morlot S, Tzetis M, Oddoux C, di Giovine FS, Nasioulas G, Chandler D, Aseev M, Hanson M, Kalaydjieva L, Glavac D, Gasparini P, Kanavakis E, Claustres M, Kambouris M, Ostrer H, Duff G, Baranov V, Sibul H, Metspalu A, Goldman D, Martin N, Duffy D, Schmidtke J, Estivill X, O'Brien SJ, Dean M. 1998. Dating the origin of the CCR5 Delta32 AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet 62: 1507-1515

Stone AC, Milner GR, Pääbo S, Stoneking M. 1996. Sex determination of ancient skeletons using DNA. Am J Phys Anthropol 99:231-238

Stone AC, Stoneking M. 1998. mtDNA analysis of a prehistoric Oneota population: Implications for the peopling of the New World. Am J Hum Genet 62:1153-1170

Stoneking M. 1995. Ancient DNA: how do you know when you have it and what can you do with it? Am J Hum Genet 57:1259-1262

Stoneking M. 2000. Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029-1032

Stoppelaëre A. 1942. Dégradations et restaurations des peintures murales Egyptiennes. Ann Serv Antiq. l'Egypte 40: 941-950

Stuiver M, Reimer PJ, Bard E, Beck WE, Burr GS, Hughen KA, Kromer B, McCormac FG, v.d. Plicht J, Spurk M. 1998. INTCAL98 radiocarbon age calibration 0-24,000 BP. Radiocarbon 40:1041-1083

Suyama Y, Kawamoto K, Kinoshita I, Yoshimura K, Tsumura Y, Takahara H. 1996. DNA sequence from a fossil pollen of *Abies spp* from Pleistocene peat. Genes Genet Systems 71:145-149

Swofford, DL. 2002. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA), Version 4.0b10.

Taanman J-W. 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochem Biophys Acta 1410:103-123

Tani N, Tsumura Y, Sato H. 2003. Nuclear gene sequences and DNA variation of *Cryptomeria japonica* samples from the postgalcial period. Molecular Evology 12:859-868

Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG. 1997. Initial genetic characterization of the 1918 "Spanish" Influenza virus. Science 275:1793-1796

Taylor GM, Rutland P. Mollesion T. 1997. A sensitive polymerase chain reaction method for the detection of *Plasmodium* species DNA in ancient human remains. Anc Biomol 1:193-203

Taylor GM, Widdison S, Brown IN, Young D, Molleson T. 2000. A medieval case of lepromatous leprosy from 13-14th century Orkney, Scotland. J Archeol Sci 27:1133-1138

Taylor PG. 1996. Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. Mol Biol Evol 13:283-285

Taylor RE. 1994. Radiocarbon dating of bone using Accelerator Mass Spectrometry: current discussion and future directions. In: Bonnichsen R, Steele DG (Eds) Method and Theory for Investigating the Peopling of the Americas. Corvallis:Center for the Study of the first Americans. Pp 26-44

Taylor GM, Crossey M, Saldanha J, Waldron T. 1996. DNA from *Mycobacterium tuberculosis* identified in medieval human skeletal remains using PCR. J Archaeol Sci 23:789-799

Taylor GM, Goyal M, Legge AJ, Shaw RJ, Young D. 1999. Genotypic analysis of *Mycobacterium tuberculosis* from medieval human remains. Microbiology 145:899-904

Taylor RE, Hare PE, Prior CA, Kirner DL, Wan L, Burky RB. 1995. Radiocarbon dating of biochemically characterized hair. Radiocarbon 37:1-11

Thomas RH, Schaffner W, Wilson AC, Pääbo. 1989. DNA phylogeny of the extinct marsupial wolf. Nature 340:465-7

Thomson DM, Brown NN, Clague AE. 1992. Routine use of hair root or buccal swab specimens for PCR analysis: advantages over using blood. Clin Chim Acta 207:169-174

Thorne A, Curnoe D. 2000. Sex and significance of Lake Mungo 3: reply to Brown "Australian Pleistocene variation and the sex of Lake Mungo 3". J Hum Evol 39:587-600

Thornalley PJ, Minhas HS.1999. Rapid hydrolysis and slow α,β-dicarbonyl cleavage of an agent proposed to cleave glucose-derived protein crosslinks. Biochem Pharm. 57:303-307

Thuesen I, Engberg J. 1990. Recovery and analysis of human genetic material from mummified tissue and bone. J Archaeol Sci 17:679-689

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

Titball RW, Williamson ED. 2001. Vaccination against bubonic and pneumonic plague. Vaccine 19:4175-4184

Toda K, Ishibashi U, Hori Y, Morikawa F (Eds). 1975. Biology and disease of the hair. Baltimore:University Park Press

Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus M-L, Bonné-Tamir B, Scozzari R. 1998. MtDNA analysis reveals a major late paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137-1152

Trewick SA. 1996. Flightlessness and phylogeny among endemic rails (*Aves:Rallidae*) of the New Zealand region. Philos Trans R Soc Lond B 352:429-446

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain AT, Sykes BC, Bradley DG. 2001. Genetic evidence for Near-Eastern origins of European cattle. Nature 410:1088-1091

Turner-Walker G, Nielsen-Marsh CM, Syversen U, Kars H, Collins MJ. 2002. Sub-micron spongiform porosity is the major ultra-structural alteration occurring in archaeological bone. Int J Osteoarch12:407-414

Tuross N. 1993. The other molecules in ancient bone: noncollagenous proteins and DNA. In: Lambert JB, Grupe G (Eds) Prehistoric Human Bone: Archaeology at the Molecular level. New York:Springer-Verlag. Pp. 275-292

Tuross N. 1994. The biochemistry of ancient DNA in bone. Experientia 50:530-535

Uchichi R, Tanaki K, Kojima T, Yamamoto T, Katsumata Y. 1992. Deoxyribonucleic acid typing of human leukocyte antigen (HLA)-DQA1 from single hairs in Japanese. J Forensic Sci 37:853-859

Ursing BM, Arnason U. 1998. The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). J Mol Evol 47:302-306

Vachot A-M, Monnerot M. 1996. Extraction, amplification and sequencing of DNA from formalin-fixed specimens. Ancient Biomol 1:3-16

Valkovic V. 1988. Human hair: Volume I, Fundamentals and methods for measurement of elemental composition. Raton:CRC Press

van de Peer Y, Baldauf SL, Doolittle WF, Meyer A. 2000. An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. J Mol Evol 51:565-576

van de Peer Y, Neefs JM, Rijk PD, Wachter RD. 1993. Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. J Mol Evol 37:613-623

van der Kuyl AC, 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 phylogenetics. J Mol Evol 40:652-657

van Scott E, Ekel TM, Auerbach R. 1963. Determinants of rate and kinetics of cell division in human scalp hair. J Invest Dermatol 41:269-273

Vanselow J, Furbass R.1995. Novel aromatase transcripts from bovine placenta contain repeated sequence motifs. Gene 154:281-286

Vasan S, Zhang X, Zhang X, Kapurniotu A, Bernhagen J, Teichberg S, Basgen J, Wagle D, Shih D, Terlecky I, Bucala R, Cerami A, Egan J, Ulrich P. 1996. An agent cleaving glucose-derived protein crosslinks in vitro and in vivo. Nature 382:275-278

Vass AA. 2001. Beyond the grave – understanding human decomposition. Microbiol Today 28:109-192

Vaughan J. 1970. The Physiology of Bone. Oxford:Clarendon Press

Vernesi C, Caramelli D, Carbonell I Sala S, Ubaldi M, Rollo F, Chiarelli B. 1999. Application of DNA sex tests to bone specimens from three Etruscan (VII-III century B.C.) archaeological sites. Ancient Biomol 2:295-305

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science 253:1503-1507

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. 1989. Mitochondrial DNA sequences in single hairs from a Southern African population. Proc Natl Acad Sci USA 86:9350-9354

Vila C, Leonard JA, Götherström A, Marklund S, Sandberg K, Liden K, Wayne RK, Ellegren H. 2001. Widespread origins of domestic horse lineages. Science 291:474-477

Voevoda MI, Romashchenko AG, Sitnikova VV, Shulgina EO, Kobsev VF. 2000. A comparison of mitochondrial DNA polymorphism in Pazyryk and modern Eurasian populations. Archaeol Ethnol Anthropol Eurasia 4:88-94

von Kohl C. 1911. Historiske Meddelelser om København. Copenhagen, 1911-12, GEC Gad, 3, 545-614. (in Danish)

Vreeland RH, Rozenwieg WD, Powers DW. 2000. Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. Nature 407, 897-900

Waite ER, Child AM, Craig OE, Collins MJ, Gelsthorpe K, Brown TA. 1997. A preliminary investigation of DNA stability in bone during artificial diagenesis. Bull Soc Geol France 168:547-554

Wakely J. 1993. Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. J Mol Evol 37:613-623

Walden KKO, Robertson HM. 1997. Ancient DNA from amber fossil bees? Mol Biol Evol 14:589-591

Waldron H. 2001. Are plague pits of particular use to palaeoepidemiologists? Int J Epidemiol 30:104-108

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. Baltimore:Johns Hopkins University Press. Pp 910-954

Wandeler P, Smith S, Morin PA, Pettifor RA, Funk SM. 2003. Patterns of muclear DNA degeneration over time - a case study in historic teeth samples. Mol Ecol 12:1087-1093

Wang GH, Lu CC. 1981. Isolation and identification of nucleic acids of the liver from a corpse from the Chang-sha Han tomb. Sheng Wu Hua Hsueh Yu Sheng Wu Li Chin Chan (in Chinese) 39:70

Wang HL, Yan ZY, Jin DY. 1997. Reanalysis of published DNA sequence amplified from cretaceous dinosaur egg fossil. Mol Biol Evol 14:589-591

Wang L, Oota H, Saitou N, Jin F, Matsushita T, Ueda S. 2000. Genetic structure of a 2,500-year-old human population in China and is spatiotemporal changes. Mol Biol Evol 17:1396-1400

Watanobe T, Ishiguro N, Nakano M, Takamiya H, Matusui A, Hongo H. 2002. Prehistoric introduction of domestic pigs onto the Okinawa Islands: ancient mitochondrial DNA evidence. J Mol Evol 55:222-231

Watanobe T, Ishiguro N, Okumura N, Nakano M, Matsui A, Hongo H, Ushiro H. 2001. Ancient mitochondrial DNA reveals the origin of *Sus scrofa* from Rebun Island, Japan. J Mol Evol 32:139-147

Westerman M, Springer MS, Dixon J, Krajewski C. 1999. Molecular relationships of the extrinct pig-footed bandicoot *Chaeropus ecaudatus* (*Marsupialia:Perameloidea*) using 12S rRNA sequences. J Mamm Evol 6:271-288

WHO. 1999. Plague manual: epidemiology, distribution, surveillance and control Geneva: World Health Organisation

http://www.who.int/csr/resources/publications/plague/WHO_CDS_CSR_EDC_99_2_EN/en

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

Willerslev E, Hansen AJ, Binladen J, Brandt TB, Gilbert MTP, Shapiro B, Bunce M, Wiuf C, Gilichinsky DA, Cooper A. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science 300:791-795

Wilson AS. 2002. Biodegradation of Hair and its Effects on the Quality of Biogenic Information in Hair from Inhumation Burials. PhD Thesis, University of Bradford.

Wilson AS. 2003. Survival and alteration – experiments in hair degradation. In: Lynnerup N, Andreason C. Berglund J (Eds.) Mummies in a New Millenium. Greenland:Greenland National Museum and Archives and Danish Polar Centre. Pp. 63-66

Wilson AS, Janaway RC, Pollard AM, Dixon RA, Tobin DJ. 2001b. Survival of human hair- the impact of the burial environment. In: Williams E (Ed) Human Remains Conservation, Retrieval and Analysis. BAR Int Ser 934:119-127

Wilson AS, Janaway RC, Tobin DJ. 1999. Effect of the burial environment on shaft morphology – relevance for archaeology and medico-legal investigations. J Invest Dermatol Symp Proc 4:353

Wilson MR, Budowle B. 1993. Inhibition of Taq polymerase by Melanin:Possible solutions for forensic applications. 45th Annual Meeting of the American Academy of Forensic Sciences, Boston, MA. Colorado Springs:Williams Printing

Wilson MR, D. Polanskey, J. Butler, J.A. DiZinno, J. Replogle, B. Budowle. 1995. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. Biotechniques 18:662-669

Wilson MR, Polanskey D, Butler J, DiZinno JA, Budowle B. 1994. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. BioTechniques 18:662-669

Woodward SR, Weyand NJ, Bunnell M. 1994. DNA sequence from Cretaceous period bone fragments. Science 266:1229-1232

Wu J, Smith RK, Freeman AE, Beitz DC, McDaniel BT, Lindberg GL. 2000. Sequence heteroplasmy of D-Loop and rRNA coding regions in mitochondrial DNA from Holstein cows of independent maternal lineages. Biochem Genet 38:323-335

Wüst RAJ, Schlüchter C. 2000. The origin of soluble salts in rocks of the Thebes Mountains, Egypt: The damage potential to ancient Egyptian wall art. J Arch Sci 27: 1161-1172

Xu X, Arnason U. 1994. The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. Gene 148:357-362.

Xu X, Arnason U. 1996. The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. J Mol Evol 43:431-437

Xu X, Arnason U. 1997. The complete mitochondrial DNA sequence of the white rhinoceros, *Ceratotherium simum*, and comparison with the mtDNA sequence of the Indian rhinoceros, *Rhinoceros unicornis*. Mol Phylogenet Evol 7:189-194

Xu X, Gullberg A, Arnason U. 1996. The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. J Mol Evol 43:438-446

Xu X, Janke A, ArnasonU. 1996. The complete mitochondrial DNA sequence of the greater Indian rhinoceros, *Rhinoceros unicornis*, and the Phylogenetic relationship among *Carnivora*, *Perissodactyla*, and *Artiodactyla* (+ *Cetacea*). Mol Biol Evol 13:1167-1173

Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR. 1998. Technical note:improved DNA extraction from ancient bones using silica-based spin columns. Am J Phys Anthropol 105:539-543

Yang H, Golenberg EM, Shoshani J. 1996. Phylogenetic resolution within the Elephantidae using fossil DNA sequence from the American mastodon (*Mammut americanum*) as an outgroup. Proc Natl Acad Sci USA 93:1190-1194

Yersin A. 1894. La peste bubonique à Hong Kong. Ann Inst Pasteur 8:662-667

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. Nucl Acids Res 25:1570-1577

Yoshii T, Tamura K, Ishiyama I. 1992. Presence of PCR-inhibitors in hairs. Nippon Hoigaku Zasshi 46:313-316

Young DL, Huyen Y, Allard MW. 1995. Testing the validity of the cytochrome b sequence from Cretaceous period bone fragments as dinosaur DNA. Cladistics 11:199-209

Yu J, Yu D, Checkla DM, Freedberg IM, Berolino AP. 1993. Human hair keratins. J Invest Dermatol 101:56-59

Zhang L, Cui A, Schmitt K, Hubert R, Navidi W, Amheim N. 1992. Whole genome amplification from a single cell: Implications for genetic analysis. Proc Natl Acad Sci USA 89, 5847-5851

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-hydroxy-methyluracil- 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. Nucl Acids Res 25:3969-3973

Zierdt H, Hummel S, Herrmann B. 1996. Amplification of human short tandem repeats from medieval teeth and bone samples. Hum Biol 68:185-199

Zink AR, Grabnre W, Reischl U, Wolf H, Nerlich AG. 2003a. Molecular study of human tuberculosis in three geographically distinct and time delineated populations from ancient Egypt. Epidemiol Infect 130:239-249

Zink A, Haas C, Reischl U, Szeimies U, Nerlich A. 2000a. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. J Med Microbiol 50:355-366

Zink A, Nerlich AG. 2003. Molecular analyses of the Pharaos: Feasibility of molecular studies in ancient Egyptian material. Am J Phys Anthropol 121:109-111

Zink A, Reischl U, Wolf H, Nerlich AG. 2000b. Molecular evidence of bacteremia by gastrointestinal pathogenic bacteria in an infant mummy from ancient Egypt. Arch Pathol Lab Med 124:1614-1618

Zink A, Reischl U, Wolf H, Nerlich AG, Miller R. 2001. Corynebacterium in ancient Egypt. Med Hist 45:267-272

Zink AR, Sola C, Reischl U, Grabner W, Rastogai N, Wolf H, Nerlich AG. 2003b. Characterization of *Mycobacterium tuberculosis* complex DNA from Egyptian mummies by spoligotyping. J Clin Microbiol 41:359-367

Zischler H, Geisert H, von Haeseler A, Pääbo S. 1995a. A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature 378:489-492

Zischler H, Höss M, Handt O, von Haeseler A, van der Kuyl AC, Goudsmit J. 1995b. Detecting dinosaur DNA. Science 268:1192-1193

Appendix 1: In vivo mutation and post mortem DNA damage rates within the Bison control region.

Appendix 1a: Damage rates over Bison 'MR' region.

Site	Mutation	Damage	Bos taurus
273	2	3	JW00
274	1	1	
275	0	1	
276	0	0	
277	0	1	
278	0	0	
279	0	0	
280	0	1	
281	1	1	
282	7	0	
283	0	0	
284	0	0	
285	0	0	
286	0	0	
287	1	0	
288	1	0	
289	0	0	
290	0	0	
291	0	1	
292	0	0	
293	0	0	TC99
294	0	0	
295	0	0	
296	16	0	
297	2	0	
298	1	0	
299	24	2	
300	34	0	TC99, CT01
301	32	1	TC99, JW00, CT01
302	29	2	
303	0	3	
304	0	1	
305	0	0	
306	0	1	
307	0	1	
308	39	0	TC99, JW00, CT01
309	4	0	TC99, CT01
310	26	0	
311	0	0	
312	0	3	
313	0	3	
314	15	1	
315	0	1	
316	0	0	
317	0	0	
318	9	0	TC99
319	0	2	
320	1	0	TC99
321	0	2	TC99
322	0	2	
323	0	0	
324	0	0	

Table headings and annotation as in Table 4.3

Appendix 1: In vivo mutation and post mortem DNA damage rates within the Bison control region.

Site	Mutation	Damage	Bos taurus
325	5	1	
326	0	1	TC99, CT01
327	3	2	
328	19	0	
329	0	0	
330	1	0	
331	0	0	TC99
332	0	1	
333	0	0	
334	3	1	
335	0	0	
336	0	1	
337	44	0	TC99, CT01
338	0	0	
339	0	0	
340	2	1	TC99
341	0	2	
342	0	0	
343	0	0	
344	13	0	
345	1	1	
346	0	1	
347	0	0	TC99
348	1	1	
349	0	0	
350	0	0	
351	0	1	
352	0	0	
353	0	0	
354	0	2	
355	0	0	
356	0	0	
357	0	0	
358	0	0	
359	0	1	
360	1	0	
361	8	0	TC99
362	13	1	
363	1	0	
364	9	0	
365	29	2	TC99, CT01
366	0	1	
367	1	0	
368	0	1	
369	2	1	
370	23	1	
371	13	0	
372	0	1	
373	0	0	CT01
374	0	0	
375	0	1	
376	0	0	
377	0	1	
378	1	2	

Table headings and annotation as in Table 4.3

Appendix 1: In vivo mutation and post mortem DNA damage rates within the Bison control region.

Site	Mutation	Damage	Bos taurus
379	2	0	
380	0	0	JW00
381	0	0	
382	16	1	
383	1	3	TC99, CT01
384	1	2	TC99
385	1	2	
386	3	0	TC99, JW00
387	0	0	
388	0	1	
389	0	2	
390	0	0	TC99
391	41	2	
392	0	2	
393	0	2	
394	1	2	
395	0	0	
396	0	1	
397	0	1	
398	1	1	
399	1	1	
400	0	3	
401	0	0	
402	0	2	
403	0	1	
404	0	2	
405	0	0	
406	0	0	
407	0	0	
408	39	6	TC99
409	0	1	
410	0	0	
411	0	3	
412	1	6	
413	0	0	
414	0	0	
415	0	1	
416	0	2	

Table headings and annotation as in Table 4.3

Appendix 1b: Damage rates over Bison 'OR' region.

Site	Mutation	Damage
24	0	1
25	0	0
26	0	0
27	1	0
28	2	0
29	0	0
30	0	0
31	0	1
32	0	0
33	0	0
34	0	0
35	0	0
36	3	0
37	0	0
38	2	0
39	0	0
40	0	0
41	0	1
42	0	0
43	0	0
44	0	0
45	0	2
46	0	0
47	0	1
48	0	1
49	0	0
50	0	1
51	0	1
52	1	0
53	0	0
54	0	0
55	0	0
56	0	0
57	2	1
58	0	1
59	0	1
60	0	1
61	0	0
62	0	0
63	0	1
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0

Site	Mutation	Damage
69	0	0
70	0	1
71	0	1
72	0	1
73	0	0
74	0	0
75	0	1
76	1	0
77	0	0
78	0	0
79	0	0
80	0	1
81	0	1
82	1	1
83	0	0
84	0	1
85	0	2
86	0	0
87	0	3
88	1	0
89	0	1
90	0	1
91	0	5
92	0	0
93	0	2
94	1	1
95	0	3
96	0	0
97	0	0
98	0	0
99	0	0
100	1	1
101	0	2
102	0	0
103	0	0
104	0	0
105	1	0
106	0	0
107	0	0
108	0	1
109	0	0
110	0	0
111	0	0
112	0	0
113	0	0

Table headings and annotation as in Table 4.3

Appendix 1: In vivo mutation and post mortem DNA damage rates within the Bison control region.

Site	Mutation	Damage
114	0	0
115	0	1
116	0	1
117	4	0
118	1	0
119	0	0
120	0	1
121	0	0
122	0	0
123	0	0
124	1	1
125	0	0
126	4.75	2
127	0	0
128	0	1
129	0	0
130	0	4
131	0	1
132	2	2
133	1	0
134	0	0
135	0	1
136	41	2
137	0	1
138	0	1
139	0	0
140	0	0
141	1	2
142	0	0
143	0	2
144	0	1
145	0	0
146	4.75	2
147	0	1
148	0	1
149	1	1
150	0	0
151	0	1
152	0	0
153	0	1
154	0	0
155	0	0
156	6	0
157	0	1
158	33	0
159	0	1
160	0	1

Site	Mutation	Damage
161	0	1
162	0	0
163	1	1
164	0	1
165	0	2
166	0	2
167	0	2
168	0	0
169	0	1
170	0	1
171	2.75	1
172	1.75	0
173	1.75	1
174	1.75	1
175	0	1
176	0	1
177	1	1
178	1	0
179	2	1
180	1	0
181	0	1
182	1	1
183	0	2
184	0	0
185	1	0
186	1	0
187	0	1
188	0	0
189	0	1
190	0	0
191	0	1
192	0	0
193	0	0
194	0	1
195	0	2
196	1	1
197	0	0
198	0	0
199	0	0
200	2	1
201	2	1
202	0	1
203	0	0
204	0	0
205	5	2
206	4	0
207	2	1

Table headings and annotation as in Table 4.3

Appendix 1: In vivo mutation and post mortem DNA damage rates within the Bison control region.

Site	Mutation	Damage
208	19	1
209	0	3
210	1	3
211	0	1
212	0	1
213	0	1
214	13	1
215	0	2
216	23.5	2
217	24.5	1
218	0	0
219	0	0
220	0	2
221	0	0
222	0	0
223	0	1
224	0	1
225	47.25	3
226	0	2
227	0	1
228	0	0
229	0	0
230	0	0
231	0	2
232	0	3
233	0	1
234	0	0
235	0	0
236	11	0
237	0	0
238	0	0
239	8	1
240	3	1
241	0	0
242	0	1
243	0	0
244	0	1
245	0	0
246	3	0
247	1	0
248	0	0
249	0	0
250	0	0
251	0	1
252	1	1
253	0	0
254	0	0

Site	Mutation	Damage
255	0	0
256	0	0
257	0	2
258	0	1
259	0	0
260	0	0
261	0	0
262	0	0
263	0	0
264	0	0
265	5	0
266	1	0
267	0	0
268	1	0
269	0	0
270	7	0
271	0	0
272	0	0
417	0	2
418	1	0
419	0	2
420	0	0
421	0	2
422	0	1
423	0	0
424	0	0
425	0	2
426	0	4
427	0	1
428	0	1
429	1	1
430	0	3
431	0	0
432	0	2
433	1	3
434	0	4
435	0	2
436	0	0
437	1	3
438	0	0
439	33.25	3
440	0	0
441	0	1
442	1	1
443	1	0
444	16	1
445	0	1

Table headings and annotation as in Table 4.3

Appendix 1: In vivo mutation and post mortem DNA damage rates within the Bison control region.

Site	Mutation	Damage
446	0	1
447	0	2
448	36.75	3
449	8	2
450	7	1
451	1	1
452	1	2
453	2	1
454	0	3
455	0	0
456	0	1
457	0	0
458	0	0
459	1	0
460	0	1
461	0	0
462	0	0
463	1	1
464	0	1
465	0	0
466	0	1
467	0	1
468	0	1
469	1	4
470	0	1
471	0	0
472	1	2
473	33	1
474	2	0
475	41.5	3
476	22.75	1
477	0	0
478	0	0
479	0	2
480	0	2
481	0	2
482	0	0
483	0	2
484	0	1
485	0	0
486	1	0
487	1	1
488	0	0
489	0	0
490	1	0
491	0	1
492	1	1

Site	Mutation	Damage
493	1	0
494	0	1
495	0	1
496	0	0
497	0	0
498	2	0
499	27.5	2
500	1	1
501	0	0
502	0	0
503	0	0
504	0	0
505	0	0
506	0	0
507	0	0
508	1	2
509	1	1
510	0	1
511	0	0
512	0	0
513	0	0
514	0	1
515	0	1
516	0	0
517	0	1
518	0	0
519	0	1
520	0	0
521	0	0
522	1	1
523	0	0
524	0	0
525	0	1
526	0	0
527	0	0
528	0	2
529	2	1
530	0	1
531	1	1
532	1	1
533	0	1
534	1	1
535	0	0
536	1	0
537	1	0
538	33.75	3
539	1	0

Table headings and annotation as in Table 4.3

Appendix 1: In vivo mutation and post mortem DNA damage rates within the Bison control region.

Site	Mutation	Damage
540	0	0
541	0	0
542	29	1
543	0	0
544	0	0
545	34.5	1
546	38.75	1
547	0	0
548	21.75	1
549	0	2
550	0	1
551	0	0
552	2	0
553	0	1
554	0	0
555	0	1
556	0	1
557	0	0
558	0	0
559	1	2
560	0	1
561	0	0
562	0	0
563	0	0
564	0	0
565	0	1
566	0	0
567	0	0
568	0	0
569	0	1
570	0	0
571	0	1
572	0	2
573	0	1
574	0	0
575	0	0
576	0	0
577	0	0
578	1	0
579	0	0
580	0	0
581	0	0
582	0	0
583	0	0
584	1	0
585	0	0
586	0	0

Site	Mutation	Damage
587	0	0
588	0	0
589	0	1
590	15	0
591	1	1
592	0	0
593	0	2
594	1	3
595	0	0
596	0	0
597	0	0
598	1	0
599	1	1
600	0	0
601	2	1
602	0	2
603	0	0
604	1	0
605	1	0
606	0	2
607	0	1
608	0	2
609	2	0
610	0	2
611	0	1
612	0	3
613	0	0
614	0	0
615	0	3
616	0	0
617	0	0
618	0	0
619	1	0
620	0	0
621	0	1
622	0	0
623	0	0
624	1	1
625	1	1
626	0	0
627	0	0
628	0	0
629	2	0

Table headings and annotation as in Table 4.3

Appendix 2: Full biochemical preservation data for Matera bones.

Sample	Skeleton	Tissue	Contam	□X174	Histology	Wedl	Linear Long[1]	Budded[2]	Lamellar[3]	Cracking[4]	Collagen	C:N[7]	N[8]	IR SF[9]	C:P	Bulk Density[10]	Skel Density[11]	Total Hg	Hg < 100 nm	Hg < 100 nm - 6 um	Hg > 6 um	Asx
TG458	SLM-3	Tooth	N	N/A	5	0	0	0	0	46	20.9	3.2	3.7	3	0.68	2	2.2	0.0853	0.0244	0.021	0.0399	0.1274
TG460	SLM-3	Tooth	N	N/A	5	0	0	0	0	46	20.9	3.2	3.7	3	0.68	2	2.2	0.0853	0.0244	0.021	0.0399	0.1164
TG425	SLM-4	Tooth	N	N/A	5	0	0	0	0	7	20.8	3.3	3.9	3.3	1.03	1.6	2	0.1535	0.0497	0.0339	0.0699	0.0747
TG453	SLM-4	Tooth	N	N/A	5	0	0	0	0	7	20.8	3.3	3.9	3.3	1.03	1.6	2	0.1535	0.0497	0.0339	0.0699	0.0814
TG447	SLM-5	Tooth	N	N/A	4	0	1	1	0	17	20.4	3.3	3.4	3.5	0.85	1.8	2.5	0.237	0.059	0.097	0.0810	0.0856
TG448	SLM-5	Tooth	N	N/A	4	0	1	1	0	17	20.4	3.3	3.4	3.5	0.85	1.8	2.5	0.237	0.059	0.097	0.0810	0.0810
TG449	SLM-6	Tooth	N	N/A	2	0	1	1	0	11	15.2	3.3	2.3	3.8	0.28	1.5	2.2	0.3117	0.0698	0.1941	0.0478	0.0921
TG450	SLM-6	Tooth	N	N/A	2	0	1	1	0	11	15.2	3.3	2.3	3.8	0.28	1.5	2.2	0.3117	0.0698	0.1941	0.0478	0.0888
TG438	SLM-11	Tooth	N	N/A	4	1	0	0	0	0	22.2	3.2	3.8	2.9	0.66	1.8	2.1	0.123	0.0306	0.0652	0.0272	0.0822
TG455	SLM-11	Tooth	?	N/A	4	1	0	0	0	0	22.2	3.2	3.8	2.9	0.66	1.8	2.1	0.123	0.0306	0.0652	0.0272	0.0873
TG446	SLM-7	Tooth	Y	N/A	3	1	0	1	0	9	13.1	3.3	2.6	3.2	0.29	1.6	2.2	0.2679	0.0767	0.163	0.0282	0.1032
TG461	SLM-7	Tooth	Y	N/A	3	1	0	1	0	9	13.1	3.3	2.6	3.2	0.29	1.6	2.2	0.2679	0.0767	0.163	0.0282	0.1033
TG423	SLM-8	Tooth	Y	N/A	2	1	0	1	0	0	9.5	3.4	1.9	3.8	0.21	1.4	2.3	0.3749	0.089	0.266	0.0198	0.1083
TG435	SLM-8	Tooth	?	N/A	2	1	0	1	0	0	9.5	3.4	1.9	3.8	0.21	1.4	2.3	0.3749	0.089	0.266	0.0198	0.0957
TG439	SLM-10	Tooth	Y	N/A	3	1	0	1	0	0	13.1	3.3	2.8	3	0.34	1.6	2.2	0.2211	0.0449	0.1592	0.0171	0.0898
TG426	SLM-10	Tooth	?	N/A	3	1	0	1	0	0	13.1	3.3	2.8	3	0.34	1.6	2.2	0.2211	0.0449	0.1592	0.0171	0.1023
TG454	SLM-12	Tooth	Y	N/A	2	0	1	1	0	53	13.9	3.3	2	3.7	0.23	1.5	2.3	0.3337	0.0799	0.223	0.0307	0.0944
TG452	SLM-14	Tooth	Y	N/A	2	1	0	1	0	0	7.8	3.3	2	3.5	0.31	1.4	2.3	0.3868	0.0697	0.2401	0.0770	0.1002
TG443	SLM-14	Tooth	?	N/A	2	1	0	1	0	0	7.8	3.3	2	3.5	0.31	1.4	2.3	0.3868	0.0697	0.2401	0.0770	0.0918
TG445	SLM-13	Tooth	Y	N/A	1	1	1	0	0	0	9	3.3	1.8	3.2	0.31	0.8	1.1	0.2285	0.0387	0.1666	0.0232	0.0974
TG457	SLM-1	Tooth	?	N/A	4	1	1	0	0	18	20.2	3.3	3.3	3.3	1.14	2.1	2.6	0.1874	0.1333	0.055	-0.0009	0.0932
TG456	SLM-1	Tooth	?	N/A	4	1	1	0	0	18	20.2	3.3	3.3	3.3	1.14	2.1	2.6	0.1874	0.1333	0.055	-0.0009	0.0991
TG451	SLM-9	Tooth	?	N/A	5	0	0	0	0	0	21.2	3.2	3.9	2.9	0.53	1.9	2.1	0.0826	0.0228	0.032	0.0278	0.0946
TG436	SLM-9	Tooth	?	N/A	5	0	0	0	0	0	21.2	3.2	3.9	2.9	0.53	1.9	2.1	0.0826	0.0228	0.032	0.0278	0.0944
TG497	SLM-5	Femur	N	Y	4	0	1	1	0	17	20.4	3.3	3.4	3.5	0.85	1.8	2.5	0.237	0.059	0.097	0.0810	n/a
TG498	SLM-6	Femur	Y	N	2	0	1	1	0	11	15.2	3.3	2.3	3.8	0.28	1.5	2.2	0.3117	0.0698	0.1941	0.0478	n/a
TG499	SLM-7	Femur	Y	Y	3	1	0	1	0	9	13.1	3.3	2.6	3.2	0.29	1.6	2.2	0.2679	0.0767	0.163	0.0282	n/a
TG500	SLM-8	Femur	Y	Y	2	1	0	1	0	0	9.5	3.4	1.9	3.8	0.21	1.4	2.3	0.3749	0.089	0.266	0.0198	n/a
TG501	SLM-9	Femur	Y	Y	5	0	0	0	0	0	21.2	3.2	3.9	2.9	0.53	1.9	2.1	0.0826	0.0228	0.032	0.0278	n/a

Sample	Skeleton	Tissue	Contam	X174	Histology	Wedl	Linear Long[1]	Budded[2]	Lamellar[3]	Cracking[4]	Collagen	C:N[7]	N[8]	IR SF[9]	C:P	Bulk Density[10]	Skel Density[11]	Total Hg	Hg < 100 nm	Hg < 100 nm - 6 um	Hg > 6 um	Asx
TG503	SLM-11	Femur	N	Y	4	1	0	0	0	0	22.2	3.2	3.8	2.9	0.66	1.8	2.1	0.123	0.0306	0.0652	0.0272	n/a
TG504	SLM-12	Femur	Y	Y	2	0	1	1	0	53	13.9	3.3	2	3.7	0.23	1.5	2.3	0.3337	0.0799	0.223	0.0307	n/a
TG505	SLM-14	Femur	Y	Y	2	1	0	1	0	0	7.8	3.3	2	3.5	0.31	1.4	2.3	0.3868	0.0697	0.2401	0.0770	n/a

Headings as in Table 1 with the exception of:
[1]presence (1) or absence (0) of Linear longitudinal fungal attack,[2]presence (1) or absence (0) of Budded fungal attack,[3]presence (1) or absence (0) of Lamellar fungal attack,[4]evidence of bone microscopic cracking, measured as the percentage of osteons with at least one radial crack. [5]bone collagen carbon to nitrogen ratio,[6]total bone nitrogen content (%),[7]bone infra-red splitting factor, measuring phosphate oxygen bond stretch thus how perfect the crystal lattice is,[8]bulk density (gcm^{-3})of the total bone, including all pores (density in a vacuum on 0), [9]skeletal density (gcm^{-3})(density after all pores have been filled). It is important to note that all measurements represent the values measured in bones from the skeleton in which teeth samples originate. Thus the measurements are not direct measurements taken from teeth samples.

Appendix 3a: Histological screening of hair.

Sections of each hair fibre were histologically screened after processing via fixation, resin infiltration, and embedment in resin blocks, according to standard protocols (Hunter 1979; Tobin 1997). Hair fibers were sectioned both transversely and longitudinally using an ultramicrotome to obtain semi-thin (~0.5μm) and ultra-thin (~100nm) sections. The former were stained with toluidine blue and borax and examined using High Resolution Light Microscopy, while the latter were stained with lead citrate and uranyl acetate and examined by transmission electron microscopy. Changes to histological morphology from normal were first scored according to their presence or absence of specific morphologic criteria, and secondly by severity of each of these criteria using a scale of 0 to 3 with 0 indicating absence, 1 limited change and 3 severe alteration. These scores were summated to give a ranked histological score for each sample (Wilson 2002). In concordance with that adopted for bone histology (Hedges *et al.*1995), 5 represents no change, and 0 is extremely modified.

Appendix 3b: Amplified DNA sequences

DNA extracts Tg491, Tg492, Tg494 and Tg495 (Christie Mine *B.bison* hair) and BS200 (*B.bison* bone) cloned 16SmtDNA sequences. Sequences were amplified between *Bos taurus* reference positions 2605-2736 (primer inclusive) (Anderson et al. 1982). Positions that differ from the *Bos taurus* reference sequence are indicated.

DNA extraction	Clone	2222222222222222 6666666666666777 3333346677789001 2568921245686 4032
	B.taurus	ACCCCGCCCACCCCCGA
Tg491	E2	G....A...G....AA.
	E3	.....A...G....A.G
	E5	.....A...G....AA.
	E6	.....A...G....A..
	E11	.....A...G...TA..
	E12	.....A...G....AA.
Tg492	F1	.....A...G....A..
	F2	.....A...GT...A..
	F3	.....A...G....A..
	F4	.....A...G....A..
	F6	.....A...G....A..
	F7	.....A...G....A..
	F8	.....A...G....A..
	F10	.....A...G....A..
	F12	.....A...G....A..
Tg494	G4	.....A...G....A..
	G5	.....AT..G....A..
	G6	...TTAT..G....A..
	G7	.....AT..G....A..
	G8	.....A...G....A..
	G10	.....A...G....A..
Tg495	H3	.....A.T.G....A..
	H4	.....A.T.G....A..
	H5	.....A...G.....
	H10	.....A.T.G..T.A..
	H12	.....A.T.G....A..
BS200	a1	.....A..TG....A..
	a3	..T..A...G....A..
	a4	.T...A...G....A..
	a5	.....A..TG....A..
	a6	.....A...G....A..
	a7	.....A...G....A..
	a8	.....A...G....A..
	a9	.....A...G....A..
	a11	...T.A...G.T..A..

Tg491 (*B.bison* hair) Control Region sequence.
Amplified between *Bos bison* Reference Sequence (Table 4.1) positions 37-292.

```
1                                          37
---------- ---------- ---------- -------ATT TAAACTATTC
51
CCTGAACGCT ATTAATATAG TTCCATAAAT GCAAAGAGCC TCACCAGTAT
101
TAAATTTACT AAAAATTCCA ATAACTCAAC ACAAATTTTG TACTCTAACC
151
AAATATTACA AACACCACTA GCTAACGTCA CTCACCCCCA AAATGCATTA
201
CCCAAACGGG GGGGACGTAC ATAATATTAA TGTAATAAAA ACATATTATG
251                     272
TATATAGTAC ATTAAATTAT AT-------- ---------- ----------
```

DNA extracts Tg491, Tg492 and Tg495 (*B.bison* hair) cloned Control Region sequence between *Bos bison* Reference Sequence positions 37-292. Table indicates sites that differ between clones and the consensus sequence.

DNA extraction	Clone	12222 16880127 93487159
BbRS		CTACTGTA
Tg491	c1	....C...
491	g1	....C...
491	b1	....C.C.
491	f1	....C...
491	a1	....C...
491	e1	....C...
491	d1	....C...
Tg492	f2	....C...
492	g2	....C...
492	h2	....C...
492	d2	TC..C...
492	b2	....CA..
492	c2	...TC...
492	e2	....C...
Tg495	d3	....C..G
495	f3	....C...
495	e3	....C...
495	b3	....C...
495	c3	....C...
495	g3	....C...
495	a3	....C...
495	h3	..G.C...

16SmtDNA cloned sequences from samples Pazyryk 1, 4, 7, and 8 (*E.caballus* hair). Positions shown are those that differ from the Equus reference sequence NC_001640 (Xu and Arnason, 1994). Positions are numbered with reference to Equus reference sequence.

	222222222 333333344 002346800 898701389
E.caballusACAACTACA	
Pazyryk 1	
c1	....
c2	.......A.
c3	....
c5	....
c7	....
c8	.....CG..
c9	....
c10	....
c11	....
c12	....
Pazyryk 4	
d2	.......A.
d3	....
d4	....
d5	........G
d8	....
d9	....
d10	....T....
d11	....
d12	....
Pazyryk 7	
e1	..G......
e2	..G......
e3	C.GG.....
e4	..G......
e5	..G......
e6	..G....A.
e8	..G......
e9	..G......
e10	..G......
e11	..G......
e12	..G......

	222222222 333333344 002346800 898701389
E.caballusACAACTACA	
Pazyryk 8	
f1	....
f2	....
f3	....
f4	....
f5	....
f6	....
f7	....
f8	....
f9	....
f10	....
f12	.T.......

Cloned mtDNA control region sequences from putative hairs of Sir Isaac Newton. Table indicates differences in clones with reference to the Cambridge Reference Sequence (CRS) (Anderson et al. 1981) between positions 16209-16356.

		1111111111111111
		6666666666666666
		2222222222223333
		2344556677991123
	Clone ↓	2701341513481823
CRS		CAAAAACATGCTTAAA
Young Newton 1.1		.........T.....
Young Newton 1.2		....G.....T.....
Young Newton 1.3		...G......T.....
Young Newton 1.4		.........T.....
Young Newton 1.5		.........T.....
Young Newton 1.6		.........T..G..
Young Newton 1.7		.........T.....
Young Newton 1.8		.........T.....
Young Newton 1.9		.........T.....
Young Newton 1.10		.........T.....
Young Newton 1.11		.........T.....
Young Newton 2.1		.........T.....
Young Newton 2.2		........A.T.....
Young Newton 2.3		.........T.....
Young Newton 2.4		.........AT.....
Young Newton 2.5		.........T.....
Young Newton 2.6		.........T.....
Young Newton 2.7		.........T.....

Clone	1111111111111111 6666666666666666 2222222222223333 2344556677991123 2701341513481823
CRS	CAAAAACATGCTTAAA
Old Newton 1.1	T.....T.........
Old Newton 1.2	T.....T.........
Old Newton 1.3	T.....T.........
Old Newton 1.4	T.....T.........
Old Newton 1.5	T.....T.........
Old Newton 1.6	T.....T.........
Old Newton 1.7	T.....T.........
Old Newton 1.8	T.....T.........
Old Newton 1.9	T.....T.........
Old Newton 1.10	T.....T....C....
Old Newton 1.11	T.....T.........
Old Newton 1.12	T.....T.........
Old Newton 2.1	T.....T.........
Old Newton 2.2	T.....T.........
Old Newton 2.3	T.....T.........
Old Newton 2.4	T.....T.........
Old Newton 2.5	T.....T.........
Old Newton 2.6	T.....T.C.......
Old Newton 2.7	T.....T.........
Old Newton 2.8	T.....T.......G.
Old Newton 2.9	T.....T..A......
Old Newton 2.10	T.....T.........
Old Newton 2.11	T.....T.........
Old Newton 2.12	T.....T.........
Old Newton 3.1	T.....T.........
Old Newton 3.2	T.....T.........
Old Newton 3.3	T.....T........G
Old Newton 3.4	TG....T.........

		1111111111111111
		6666666666666666
		2222222222223333
	Clone	2344556677991123
	↓	2701341513481823
CRS		CAAAAACATGCTTAAA
Am Phil Soc	1.1	.............C...
Am Phil Soc	1.2	.............C...
Am Phil Soc	1.3	.............C...
Am Phil Soc	1.4	.............C...
Am Phil Soc	1.5	.......G....C...
Am Phil Soc	1.6	.............C...
Am Phil Soc	1.7	.............C...
Am Phil SoC	1.8	.............C...
Am Phil Soc	1.9	.............C...
Am Phil Soc	1.10	..G.........C...
Am Phil Soc	1.11	.............C...
Am Phil Soc	1.12	.............C...

		1111111111111111
		6666666666666666
		2222222222223333
	Clone	2344556677991123
	↓	2701341513481823
CRS		CAAAAACATGCTTAAA
Man House Mus	1.1	...........C....
Man House Mus	1.2	...........C....
Man House Mus	1.3	...........C....
Man House Mus	1.4	.....G.....C....
Man House Mus	1.5	...........C....
Man House Mus	1.6	...........C....
Man House Mus	1.7	...........C....
Man House Mus	1.8	...........C....
Man House Mus	1.9	...........C....
Man House Mus	1.10	....

Cloned mtDNA control region sequences extracted from Andaman Islander, Onge individual, hairs. Table indicates sequence differences from CRS between positions 16209-16356.

	11111111
	66666666
	22222333
	24489014
	33505294
CRS16201	CTCACAGC
1.1	T....GAT
1.2	T.....AT
1.3	T.....AT
1.4	T.....AT
1.5	T.....AT
1.6	T.....AT
1.7	T.....AT
1.8	T..G..AT
1.9	T.....AT
1.10	T.....AT
1.11	TC....AT
1.12	T.....AT
2.1	T.....AT
2.2	T.....AT
2.3	T.....AT
2.4	T.....AT
2.5	T.....AT
2.6	T.....AT
2.7	T.....AT
2.8	T.T...AT
2.9	T.....AT
2.10	T...T.AT
2.11	T.....AT
2.12	T.....AT

Appendix 4a: Specific PCR conditions used in study.

Pla A: 94°C 5 minutes, 35 cycles of (94°C 1 minute, 59°C 1 minute, 72°C 1 minute), 72°C 1 minute.

Pla B: 95°C 1.5 minutes, 40 cycles of (95°C 20 seconds, 53°C 20 seconds, 72°C 30 seconds).

Pla C: 94°C 5 minutes, 40 cycles of (94°C 30 seconds, 52°C 30 seconds, 72°C 90 seconds), 72°C 1 minute.

Beta globin PCO3/4: 94°C 5 minutes, 35 cycles of (94°C 1 minute, 56°C 1 minute, 72°C 1 minute), 72°C 5 minutes.

BlaTEM: 95°C 3 minutes, 35 cycles of (95°C 1 minute, 55°C 1 minute, 72°C 1 minute) 72°C 4 minutes.

Original Bart's Mitochondrial PCR: 94°C 5 minutes, 40 cycles of (94°C 1 minute, 55°C 2 minutes, 72°C 1 minute), 72°C 1 minute.

All primer sets used at the ABC incorporated specific annealing temperatures indicated in Table 7.5 into the following conditions:

94°C 1.5 minutes, 40 cycles of (94°C 45 seconds , annealing 45 seconds, 68°C 1.5 minutes) 68°C 10 minutes.

Appendix 4b: Cloned bacterial sequences obtained at the ABC using primers designed to target *Y.pestis.*

All sequences are presented *without* primers. Sample name interpretation is as follows:

For example - tg9_V2_pla_8

where

tg9 = Authors' sample extraction name
V2 = Original sample name
pla = PCR primer pair abbreviation (see below)
8 = Authors' sequencing reaction reference

PCR primer pair abbreviations

pla = YPpla728H (YP12D)/YPpla876L (YP11R)
rpb = Yprpob1/Yprpob2
16s = Yp16s586H/Yp16s786L

>tg9_V8_pla_8
ACTTGGAATTACTCCTNTAGAGGGCGCAAAAACCCAGAAAACTTTTGCAAAC
TTACTAAAAACGTTCAGAGTCCAAAGTAAGAGCCATGGGGCACTNTCATCCT
GATGANTGGGTTAGTGTACAAAGCAGTTACCTATGTGGGCAGGTGTGATATG
AAAAGGCTGCGGAAAGCAGCATATAACACTCAAAGAATTCCTAGGTGNGTTT
TCAGGTGATAGAGCCAGCGTaGGGGGCATGCAGGTGTTTTCTA

>tg9_V8_pla_8
ATCCGGTCTCAACTGCGGTATG

>tg9_V8_pla_8
ACTTGGAATTACTCCTCTAGAGGGCGCAAAAACCCAGAAAACTTTTGCAAAC
TTACTAAAAACGTTCAGAGTCCAAAGTAAGAGCCATGGGGCACTCTCATCCT
GATGACTGGGTTAGTGTACAAAGCAGTTACCTAGGTGGGCAGGTGTGATATG
AAAAGGCTGCGGAAAGCAGCATATAACACTCAGAGAATTCCTAGGTGAGTTT
TCAGGTGATAGAGCCAGCGTAGGGGGCATGCAGGTGTTTT

>tg9_V8_pla_8
ACTTGGAATTACTCCTCTAGAGGGCGCAAAAACCCAGAAAACTTTTGCAAAC
TTACTAAAAACGTTCAGAGTCCAAAGTAAGAGCCATGGGGCACTCTCATCCT
GATGACTGGGTTAGTGTACAAAGCAGTTACCTATGTGGGCAGGTGTGATATG
AAAAGGCTGCGGAAAGCAGCATATAACACTCAAAGAATTCCTAGGTGAGTTT
TCAGGTGATAGAGCCAGCGTAGGGGGCATGCAGGTGTTTT

>tg13_V1_pla_8
AGAGCCTTCAATCCTGCAGCTCAAGAGATACTGCTCCGGCGTGCTGCCGAGC
TACATGGTGCCGGACCGGTTCAGATTCTTGCCGTCCCTTCCCAAGACGTCGAC
CGACAAGACCGACTATCAGAAGCTGAAAGAGCTCGGATGATGGATCTGAGCC
TTTCAACCGAGGACGCGGAGCTCCGCGAGAAGATCGTGCGGTTTGCACAGAA
GGAGTTGTCGTCCGGTGTGATCGCGCGCGATCGGTCTCACGAATTCTCCCGCG
ACCTTTGGGACAAGTGCGGCCGCATGGGCCTGACCGGGCTTcCTGTCCCCGAG
GC

>tg13_V1_pla_8
TCCGGACGTCACGGAAGCGGCCGTTATCAGCCGCTCCGATTCCGACGGGCAG
GTGAGCGTGGACGCATTTCTTGCCTGGAGCGCGAAAGAAGAGCCTTCAATCC
TGCAGCTCAAGAGATACTGCTCCGGCGTGCTGCCGAGCTACATGGTGCCGGA
CCGGTTCAGATTCTTGCCGTCCCTTCCCAAGACGTCGACCGACAAGACCGACT
ATCAGAAGCTGAAAGAGCTCGGATGATGGATCTGAGCCTTTCAACCGAGGAC
GCGGAGCTCCGCGAGAAGATCGTGCGGTTTGCACAGAAGGAGCTGTCGTCCG
GTGTGATCGCGCGCGATCGGTCTCACGAATTCTCCCGCGACCTTTGGGACAA
GTGCGGCCGCATGGGCCTGACCGGGCTTCCTGTCCCCGAGGC

>tg15_V6_pla_8
CAACGCATTGTCTTTCATCCCCAAACTGCTCCCACGAATGAACCCGACGGGA
AAGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGT
GGCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGC
AACGGA

>tg15_V6_pla_8
CAACGCATTGTCTTTCATCCCCAAACTGCTCCCACGAATGAACCCGACGGGA
AAGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGT
GGCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGC
AACGGA

>tg15_V6_pla_8
CAACGCATTGTCTTTCATCCCCAAACTGCTCCCACGAATGAACCCGACGGGA
AAGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGT
GGCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGC
AACGGA

>tg15_V6_pla_8
CAACGCATTGTCTTTCATCCCCAAACTGCTCCCACGAATGAACCCGACGGGA
AAGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGT
GGCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGC
AACGGA

>tg15_V6_pla_8
CAACGCATTGTCTTTCcTCCCCAAACTGCTCCCACGAATGAACCCGACGGGAA
AGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGTG
GCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGCA
ACGGA

>tg15_V6_pla_8
CAACGCATTGTCTTTCATCCCCAAACTGCTCCCACGAATGAACCCGACGGGA
AAGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGT
GGCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGC
AACGGA

>tg19_V4_pla_8
NAANGCATTGTCTTTCATCCCCAAACTGCTCCCACGAATGAACCCNACGGGA
AAGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGT
GGCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGC
AACGGA

>tg19_V4_pla_8
CAACGCATTGTCTTTCATCCCCAAACTGCTCCCACGAATGAACCCGACGGGA
AAGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGT
GGCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGC
AACGGA

>tg19_V4_pla_8
CAACGCATTGTCTTTCATCCCCAAACTGCTCCCACGAATGAACCCGACGGGA
AAGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGT
GGCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGC
AACGGA

>tg19_V4_pla_8
CAACGCATTGTCTTTCATCCCCAAACTGCTCCCACGAACGAACCCGACGGGA
AAGACCCTGAGCCACTGGCTGGTCACATCATTCATCAACGTCGAAGCGCAGT
GGCCTTCGACGGACGCACAAGCCTTTCCGCCGATCGTTTCTTCACCATGCTGC
AACGGA

>tg20_V4_pla_8
ACTTTTACCCGTTTATGATAAGCCGATGATTTATTATCCTTTGTCTACCCTAAT
GCTGGCTGGGCTAAATGATATTCTCATCATCACGACTCCTCGCGAGTTACCGA
TGTTTAATTATTTATTAGGAGATGGCTCTCATTGGGGGATCTCTCTGCGTTAC
GCGGCACAACAAAATCCGAATGGTATTGCTGAGGCTTTTATTATTGGTGAAG
AATTCATCGGTAATGATAGTGTCTGTCTCATCCTTGGAGATAATATTTTATAT
GGCGATAATTTATCCGTTAAATTACAACAGGCAGCGCAATTAAAAAATGGCG
CCACTATTTTTGGATATTATGTCTCCGATCCAGAACGCTATGGTGTAATTCTTT

TTGATAAAAAAGGTGATGCACTCGATGTAATCGAAAAACCTCCTACGCCGAT
TTCTCA

>tg20_V4_pla_8
ACTTTTACCCGTTTATGATAAGCCGATGATTTATTATCCTTTGTCTACCCTAAT
GCTGGCTGGGCTAAATGATATTCTCATCATCACGACTCCTCGCGAGTTACCGA
TGTTTAATTATTTATTAGGAGATGGCTCTCATTGGGGGATCTCTCTGCGTTAC
GCGGCACAACAAAATCCGAATGGTATTGCTGAGGCTTTTATTATTGGTGAAG
AATTCATCGGTAATGATAGTGTCTGTCTCATCCTTGGAGATAATATTTTATAT
GGCGATAATTTATCCGTTAAATTACAACAAGCAGCGCAATTAAAAAATAGCG
CCACTATTTTTGGATATTATGTCTCCGATCCAGAACGCTATGGTGTAATTCTTT
TTGATAAAAAAGGTGATGCACTCAATGTAATCGAAAAACCTCCTACGCCGAT
TTCTCA

>tg22_V3_pla_8
ACTTTTACCCGTTTATGATAAACCAATGATTTATTATCCTTTGTCTACCCTAAT
GCTGGCTGGATTAAATGATATTCTCATCATCACGACTCCTCGCGAGTTACCGA
TGTTTAATTATTTATTAGGAGATGGCTCTCATTGGGGGATCTCTCTTCGATAT
GCTGCCCAACAAAATCCGAATGGTATTGCCGAGGCTTTTATTATTGGCGAAG
AATTTATCGGTAATGAAAGTGTCTGCCTCATCCTTGGAGATAATATTTTATAT
GGCGATAATTTATCCGTTAAATTACAACAAGCAGCGCAATTAAAAAATGGCG
CCACTATTTTTGGATATTATGTCTCCGATCCAGAACGCTATGGTGTAATTCTTT
TTGATAAAAAAGGTGATGCACTGGATGTAATCGAAAAACCCCCTACGCCGAT
TTCTCA

>tg38_BK_rpb_17
CACCACGGAGCCGCAGGAAGCGCTGCTGCTGGGCGGCCACACCGTCGTCATG
CA

>tg38_BK_rpb_17
GGCGCTGTGGACGGACGACCTGTTCCAGCGCTGCACCCTGTTCCGCCCCGA

>tg38_BK_rpb_17
GGCGCTGTGGACGGACGACCTGTTCCAGCGCTGCACCCTGTTCCGCCCCGA

>tg75_DB_rpb_17
CGTGGAGTGCAGAGAACAACGTCCTCGGCCGCGTCGGGAACGAAGACGAGA
TCGCACAAGCAGTGTTCTATCTCTTCACCAACACGTACACGACGATAAGGTGT
TA

>tg75_DB_rpb_17
CACCACGGAGCCGCAGGAAGCGCTGCTGCTGGGCGGCCACACCGTCGTCATG
CA

>tg38_BK_16s_52
TTGGAGGTTGTTCCCTTGAGGAGTGGCTTCCGGAGCTAACGCGTTAAGTCGAC
CGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCC
CGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACTTA

>tg62_CV_16s_52
TAGCCGTTGGAGACCCTCGTGGTTTTTAGTGGCGAAGCTAACGCGATAAGTC
GACCGCCTGGgAAGTACGGCCGCAAGGTTGAAACTCAAATGAATTGACGGGG
GCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACgcGAAGAACCT
TA

>tg62_CV_16s_52
TAGTTGTCGGAGGGTAACTCCTTCGGTAACGCAGCTAACGCGTGAAGTCGAC
CGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGAC
CCGCACAAGCGGTGGATGATGTGGATTAATTCGATGCAACGCGAAAAACCTT
A

>tg62_CV_16s_52
TAgCCGGTGGAGAcCCTCGTGGTTTTTAGTGGCGAAGCTANCGCGATAAGTCG
ACCGCCTGGGAAGTACGGCCGCAAGGTTGAAACTCAAATGAATTGACGGGG
GCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACC
TTA

>tg62_CV_16s_52
TAGCTGTTGGAGACCCTCGTGGTTTTTAGTGGCGAAGCTAACGCGATAAGTC
GACCGCCTGGGAAGTACGGCCGCAAGGTTGAAACTCAAATGAATTGACGGG
GGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAAC
CTTA

>tg62_CV_16s_52
TTGGAGGGTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGA
CCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGC
CCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTT
A

>tg66_BN_16s_52
TAGCTGTTCGTGGTCTTGTACTGTGAGTAGCGCAGCTAACGCACTAAGTAGAC
CGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAATGAATTGACGGGGGCC
CGCACAAGCGGTGGAACATGTGGTTTAATTCGACGCAACGCGAAGAACCTTA

>tg66_BN_16s_52
TAGCTGTTCGTGGTCTTGTACTGTGAGTAGCGCAGCTAACGCACTAAGTAGAC
CGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAATGAATTGACGGGGGCC
CGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTA

>tg73_DB_16s_52
TAGCCGTTGGTCCTATTTACAGGATTAGTGGCGCAGCTAACGCATTAAGTCGA
CCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGC
CCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTT
A

>tg73_DB_16s_52
TAGCCGTTGGTCCTATTTACAGGATTAGTGGCGCAGCTAACGCATTAAGTCGA
CCGCCCGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGG
CCCACACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAATGCGAAGAACCT
TA

>tg75_XA_16s_52
TAgCCGGTGGAGACCTCGTGGTTTTTAGTGGCGAAGCTANCGCGATAAGTCG
ACCGCCTGGGAAGTACGGCCGCAAGGTTGAAACTCAAATGAATTGACGGGG
GCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACC
TTA

>tg75_XA_16s_52
AGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGGCTGTAAACGATGTCGA

>tg75_XA_16s_52
TTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGA
CCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGC
CCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTT
A

>tg10_V14_rpb_53
CTTCGCCGTGGCCATCTACCGCCGCAAGATCTCCTGACAGTTTTTCCTAGGCG
GTGTCGGATCGGGTGAGGGCGGTTCGTAGCAGGGG

>tg10_V14_rpb_53
GCgCGATGTCGTCTGGGTCGTCCCCGGTGTTCGTCTCGTACGTGCCGTTGCCG
TCGAGGTCCCACTCGTAGCGGATGATCGTCCCGTCGGGATCGGCCGACGCCG
TGGCGTTGAAGGCCACCGACGAGCCGCTGGTCACCGGGTTCGGCGTGGCGG

>tg10_V14_rpb_53
CGCCACCATCGATGTCGCCGCCTGATCGATATGTCTACGATCCTGAACATCCG
ACGCCCGACGACAGGAGGCAGCATCGTATCGTACGTGTACACGCCAGGCAGT
GTGGATGTGAGCGCGATTCAGCGACGATCCGATGTGTTGA

>tg10_V14_rpb_53
TCGTCTCCAGCAGGCGTACATCGCGCGTGGCTGCgACGGTCGACGCNACGCT
CaCCGGGCCGCAGTTCGCCGTCCTCACCGCGGTGGAGACCTATCCCGGCACCG
ACCAGGTACNCTGGCCATGTCGGTCGACCTGGACCGCTCGACCATGGCCGAT
ATCGTCCGCAGGCTGGAGGAACGCGGCCTGATCACGCGCTGTCCCGACGTAC

ACGA

>tg10_V14_rpb_53
CTTCGCCCGTGGGCCATCTACCGCCGCAAGAATCTCCTGACAGTTTTTCCTAG
CGGTGTCGGATCGGGTGAGGGCGGTTCGTAgCAGGGG

>tg10_V14_rpb_53
CGCGTGCCCGCCCTGAATGTCGACCATGG

>tg20_V4_rpb_53
CACCGACAACGGCTGGCTACAGGATCCGAAAAGTTACCACAATGCGCCGAA
ATCGAAGCTCTCGCAGTACGACGCAGGGCATCGGACGCCGATCATGCTGAGC
TGGCCTGGGAGGATCGAGCCGAAGAAGGTCGACGCGCCCATCTCCAGCATCG
ACCTGGCGCCGACGCTCCTGCGTGTCGTCGGCATCGATCCGCCTGTCGAAAT
GACTGGCGTCGACTTGCTCGATCCCGCCGCAGTCGCCGCGCGGAAGCGGGTT
TTCGGCGAAACGTTTTCGCATGACGCCGTCGACGTCGATCGGCCTGCGGCCT
GCCTGGAGTACCGCTGGTGCGTCGAGTGGCCGTGGCGATTGATCGTGCCGAA
CACGGACGTTGTTCCCCAGGGGAAACTCGAGCTGTTCAACTTGGCAGACGAT
CCCTGGGAGGAAAAGAATCTGGCGGAAGTGCATCCTGAAATGGTCGACCGCT
TGCTCGGCGCCTTGGATGGTTGGTGGAATCCGGCGGCGGATCAGTAGCTGCT
AAATGCGCTCAAAATCGCAACAGCTCATTAATTAATTGATGCGGGCGAGTTG
CGGTTGCCTTGCGAGCCAAGGGTAACAAGGTAATGACGAGTAGTAGTCTATA
GGGCCCCCATATTGGCCTTGGCTCGTACGGGAATTACCGTCATAAACAAGCC
GGCGTACTGGCATTTGCGATGACAGTGTGATCTCGTCAGGTGAGTGTAATGA
TTGTAGAACCCCGAGGTATCGTTGCGGCCTTCTTTCTGTGGACGCTTGGATTC
AATAGCTTGCATGCGGCGAATGCGGCCACCATTCGCTATTCATGGGCAGGTC
AAATCACGCCAAACAGCGCAGATCCATGGAAAATTGGTAGTGAATGGTTTGA
TTTTGAGCTATCTCTTCGCGTTGACAGCAATGCACTTGATCTGAATTCGAACG

>tg20_V4_rpb_53
GAGGCGAGCGTCAGGATGATGAGCGAGAAGCGCATCGAGCCGAGGTAGA

>tg20_V4_rpb_53
TGAGGTCGATGGTGTAGATGAAGCACAGTTCGGAGGACTTTTTGCGCGCGTC
GTAGCTGCGCTTGAACAAGGTGAAATCGAGCAGGTCATCGCTGGCGATGCCC
AGGCGCTGCACGATGGCAACGCGCAGGTCTTCTTCGGGATGGTCGATTGGCA
GCTTGAGTTCAGTGATTCGTAACATGGCGAGGATCCGGGTTCGCGGGGCGCA
CAACTGCGCCAGGGCGT

>tg20_V4_rpb_53
TGAACTGATTCCAACAAACGATGATTCGTGCGTCGACTTTTCAGCTGGTCATC
CCAAATTGGCAAAGATCGCGCCAAAGCCCATCTCGGACAGCATCATTGCAAT
ACGAGAATTCAGCCGGATGCTTTTTTGAGCAGATTCGGCATCGCAATTCAGTC
TTCAAGCGTGACCGAATCATATCGCAAGACATAGGGATCGACCACGG

>tg21_V5_rpb_53
ATTGACCTGGAATTCCGACCCGACGGGTTGCTGAGCCGGCGAGCTCCGGCAG
GACCNNTACCAGGAGTCCCAAGGGTCGGACTGCTGCCAGTGTGATGGATATC
TGCAGAATTCGCCCTTAACACCTTATCGTAGTGTACG

>tg21_V5_rpb_53
TAACGCCGCCATGTGCTGGAATTCGCCCTTAACACCTTATCGTCGTGTACG

>tg21_V5_rpb_53
ACAGCATCGTGCAGGACCACCCCGACTGCGCGGACGGCGAGCCGTTCGCGGT
GG

>tg21_V5_rpb_53
TCCCCGGGTCCCGAGACCTGGAGCGACAAACCGGCTGTCCGCAGACATCCAG
AGCGTCTTCGTCATGCCGGAGCAGCGAGGCCGGGGAATCGGCTCAGCACTGG

>tg21_V5_rpb_53
GCTGACGTGTCGGCGCTGGTTCCTGTTATTTCTCAAGTAACCTAAATGGCGGA
AGCGGGTGGGAATCGAACCCACCGCTGCGGGGGTTANCCCGCAGCCGGACG
GCTT

>tg21_V5_rpb_53
GCTGCTGATCTTCGGCGTGTTCGGCTATGTGCTGG

>tg21_V5_rpb_53
TGGCGATGCTGTCTTCGATCTTGCGCGTCCCTCCGTTTCCAACTGAACGGCGG
TCATGCCACTCAGCTGCGCTGTAACGGTGACTACCGGCAGCTCGATATCGGG
CATGTCTTGGATCGGCAGCCGATAGAACCCGTACGCCCCCGCGAGGACCGCG
AGCACAAAGATAACGACAACTGGAATCGGTCGGCGGATCGACCAGGTTGCA
AGATTCATGGCGACGCCTGGAACGCCACGCTGGCCTCGCCGGCGAGACGCGC
ACGAGATCGCCGTCGTGCAGAAAGCCGCCGCCGCTGACGACGATGGCCTGCT
CGACCGACAAGCCACTCAGCACTTCGACATCGGCGCCATTGCGTCGGNCGCT
CTTCACGGAAGTCTGCACTACCTTGTTGTCGGCACCGATAGCGAAGACATACT
GACGCCCGTCACGCATCACGACGCTGCTCATGGGTACGGTGAGCGCCTGCNA
AGCACCGCTGACGATCGTACCGTTGAGATACATGCCGATACGGGCGAGTGCC
TGCGCGTCTNTGCGCAACTTCACGTACACGAACGATA+run out of seq

>tg25_V11_rpb_53
GCCGGTAAAGCTCACNCANGCATTCCGCAATTTCGATTTCCTCTCGCCTTGGG
AAGGCATGACGCTGCCGGGCGATGAAAAAGCGGGCGATCCGAAgCCATGAG
CGCGGGTCAAATCCGTGACGCGACGCCTGTGCTCCTCGTCAAGGACGTGG

>tg25_V11_rpb_53
CCTCATCCAGTTTGAACGGCTTGATGATGGCTTCAATGAGTTTCATCGTGATG
TCTCCTATGAATAGGCCCGCTCATTATGTTGGGTCAGATCCAACCCCGTGGCT
TCTTCGTCTTGTGACACGCGCAAGCCGATCGAACCGTCTACAAGTTTGAGGAT

CACATAGGTCCCGACGAATGAGAACACGATCGTGGAAAAGACCGCGACAGC
T

>tg25_V11_rpb_53
GCTGCCGCGCCTTGCTCGCCGGAACCGCCGTCGGCGCCGTCGGCCTGAGCCT
GGCCGCCTGCGGCAAGGGCAAGAAGACCGACTCCAAGACCGTCACTGTCGTC
ATCCACGACTCCTTCTACGTCTCGGACGACCTCA

>tg25_V11_rpb_53
CGCTCGGCGACATTCCCGTGGGCCTTGGCGTCAACCGCTGGTTCTCCCTGTCG
TTTGACAAGCCGGAATT

>tg25_V11_rpb_53
CGAGGACCGCGTCGTCGATCTTAAGCGTTCCGGCCACGTCGTGAACGCCGCG
CGCCTCGCCTCGGGTGAGACGTTGCGTGGGCAGACATTCGTCAACGCCGCGG
GAGCGTGGTCCGCGCAGCTTTGCGCGATGGCCGGGATGCCGCTGCCCGTCGT
CCCGCTACGCAGGTTCGAGCATTACTTCACCGCAGGCAACCCGATCGAGCGC
CTGCCTTATGTCAAGGATCTCGACCGGCTCGCGTTCCGCTC

>tg25_V11_rpb_53
CATGCTCCAACCCTGTTGCGACATAAGGCGCTGCCCCGGCTCCTGGAGTGCT
GATAGGCGAGACACCACTTGCAGTATAGGCATAGCGAGCATAGCTATGAGCG
GCTGCCCCCGACGGGTGTGTGCCCATTGCTTCCATTGCTCTGATGCTGCATGAT
GCGCCCCTGGAGTTCTGCTGGTTGGAATAGATCGGCATTGACGCTCTCAACGC
GTGCAGTCAACGACTGATCGGTGAGGAAGAGGACCGGGATCTGGTAGCGCTC
CGCCATAT

>tg25_V11_rpb_53
GGTCAACAACAACGGCTCTCTCAACCAGGAGATCCCGCTGG

>tg25_V11_rpb_53
CGAGCTGCCGAAGGTCAACACCGCGG

>tg26_2654_Yprp1/2_53
GGTCAACAACAACGGCTCTCTCAACCAGGAGATCCCGCTGG

>tg26_2654_Yprp1/2_53
GGAGTATACGGGCACCGCCCTTACCGCCATCC

>tg26_2654_Yprp1/2_53
GCTGCTGATCTTCGGCGTGCTCGGCTATGCGCTGG

>tg26_2654_Yprp1/2_53
TACCAACAAATCTACCTACCCTTAACAGTACATAGTACATAAAGCCATTTACC
GTATATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGAC

>tg27_2654_Yprp1/2_53
GGTCAACAACAACGGCTCTCTCAACCAGGAGATCCCGCTGG

>tg27_2654_Yprp1/2_53
GGAGTATACGGGCACCGCCCTTACCGCCATCC

>tg27_2654_Yprp1/2_53
GCTGCTGATCTTCGGCGTGCTCGGCTATGCGCTGG

>tg27_2654_Yprp1/2_53
GTCATCCATGGGGACGAGAAGGGATTTGACTGTAATGTGCTATATACGGTAA
ATGGCTTTATGTACTATGTACTGTTAAGGGTAGGTAGATTTGTTGGTA