LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

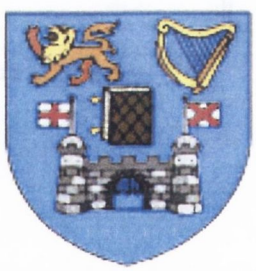
Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

GENETIC VARIATION IN DOMESTICATED CATTLE AND WILD AUROCHSEN USING ANCIENT AND MODERN DNA

Ceiridwen Jane Edwards



A thesis submitted to the University of Dublin
for the degree of Doctor of Philosophy

Department of Genetics,
Trinity College,
University of Dublin.

March 2003

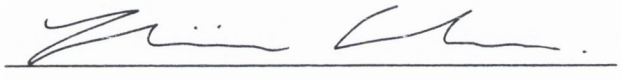
TRINITY COLLEGE
2 4 NOV 2015
LIBRARY DUBLIN

Thesis 7309

Declaration

I hereby certify that this thesis, submitted to the University of Dublin for the degree of Doctor of Philosophy, has not been submitted as an exercise for a degree at any other university. I also certify that the work described here is entirely my own, unless stated as otherwise.

This thesis may be made available for consultation with the University Library and may be photocopied or loaned to other libraries for the purposes of consultation.



Ceiridwen J. Edwards.
March 2003

Acknowledgements

A thousand million thank yous to Dan - the best supervisor anyone could ever wish for, in particular for his support, help and patience throughout my Ph.D., and most of all for offering me the opportunity to do what I enjoy most, while also gaining a qualification and being paid. I am greatly looking forward to the next instalment as the lab's new postdoc!

A huge thanks to the present Boviners, as well as those that have moved on to bigger and better things. Particular thanks to Brian, Dave, Elaine, Emma, John, Laoise, Ruth, Stephen and Teresa for their friendship, selflessness in dropping everything to help, and making it a pleasure to come to work each day. Also thanks to Ashie, Emmy and the Identi-girls for the craic, Ronan Loftus for assistance with the Pustertaler study, and to Dave MacHugh for his assistance with the papers and his wizard ways with all that Mac graphic programs have to offer. I am also extremely grateful to Jill, for her advice, saintly patience and, above all, friendship.

Heartfelt thanks to the Prep Room posse, Dave, Louis, Paul and Brenda – an essential part of the Department who were always on hand to solve problems and lend me their expertise whenever "*I was just wondering if* ... ", not least Louis's ingenuity in developing the Liquid Nitrogen Ladle™. Also thanks to Sue for loaning me the use of the Office printer in the mad dash to meet the thesis deadline.

Heaps of thanks to Declan and Nicola, for making my fourth house in Dublin my first home in Dublin; to Jacinda – always only an e-mail away; and my good friend these past 12 years, Claire – thanks for listening!

Extensive thanks go to all the archaeologists involved with the ancient DNA sections of this thesis, for not only sourcing and allowing destructive sampling of the skeletal material analysed, but for also helping me with archaeological questions that, in certain cases, may have seemed excessive. I am most grateful to Andrew Chamberlain, for supplying both aurochs and wild boar/pig remains, as well as answering my seemingly unending questions about the Neolithic in Britain; to Nigel Monaghan for the two Irish elk specimens; and to the following collaborators for donation of both aurochs and domestic cattle remains: Rick Schulting, László Bartosiewicz, Erich Pucher, Jean-Denis Vigne, Finbar McCormick, Patrick Wallace, Susan McIntosh, Keith Dobney, Louise Martin, Anne Tresset, Daniel Helmer, and Liora Horwitz.

For assistance with modern sample collection, additional acknowledgements go to: A.H. Al-Haboby, M. Diallo, M.A.A. El-Barody, O. Ertrugrul, T. Goodchild, T. Grupp, H. Halila, A.H. Harba, J. Knox, J.A. Lederer, A. Swaid and E. Thomson. Thanks also to co-authors of papers published as a result of this thesis: D.R. Bancroft, J. Connellan, G. Dolf, E. Eythórsdóttir, C. Gaillard, C. Looft, and I. Olsaker. For access to his unpublished data, as well as helpful discussions and scientific advice, many thanks to Juha Kantanen.

My Ph.D. was supported financially by the Enterprise Ireland Basic Research Grants Scheme (project no. SC/1999/409). The ancient DNA work was part-funded by The Wellcome Trust (grant no. 047485/Z/96/Z), and a Eurocores OMLL Programme Grant via CNRS, France. J.R. Knox, of Knox Farms N.S.W. Australia, provided funding for the Pustertaler project.

Lastly, and most importantly, special thanks to my parents' – without their support (both emotional and financial), none of this would have been possible ..... I can get a '*proper*' job now I'm Dr. Edwards, Dad! Also, thanks to the Minton Clan (no longer '*we three*' but now '*the four of us*') – big sister, Rowena, brother-in-law Bernard, plus the two most beautiful nieces ever, Arianwen and Gwenllian.

Summary

(1) Archaeological Genetics

A total of 101 *Bos taurus* teeth and bones aged between 1,000 to 9,000 years old, and 45 *B. primigenius primigenius* bones aged between 3,720 to 12,380 years old, were assessed for the presence of verifiable mitochondrial sequences. It was possible to reproducibly amplify and sequence mitochondrial control region DNA extracted from twelve of the cattle and 24 of the aurochs samples. The sequences obtained from cattle were either identical, or closely related, to the reference sequence for modern cattle, highlighting the proximity of the ancient DNA cattle sequences to modern Near Eastern, European and African *B. taurus*, as well as regional continuity through time. As those samples taken from more arid areas gave considerably lower numbers of reproducible amplification, this analysis leads to the conclusion that remains from this type of excavation are less promising for archaeological genetics. Seven of the 24 aurochs control region sequences generated (29.2%) exhibited taurine-like haplotypes. Although this may be due to greater diversity of aurochsen prior to the last glaciation, which encompassed both the *B. p. primigenius* sampled here and the taurine haplotypes captured by domestication, it is more plausible that cross-breeding was occurring between early domestic cows and wild aurochs bulls. No clear divisions between those individuals with taurine-like or aurochs-like D-loop haplotypes were found when a region of the cytochrome *b* gene was sequenced in a subset of 19 aurochs. Comparisons between the 19 aurochs-like control region sequences and published extant data from Near Eastern *B. taurus* seem to suggest that the star-like patterns seen around the four main taurine haplotypes, previously explained as domestication-induced population expansion, may be due in part to capture of a greater diversity of extinct *B. p. primigenius* haplotypes at the domestication event.

The appendices contain additional ancient analyses of British *Sus scrofa* remains from a 6,000 BP site in Derbyshire, and two *Megaloceros giganteus* (giant Irish deer) bone samples from Irish sites dating between 30,000 and 10,600 years ago. Control region product was amplified from six out of seven of the pig bones amplified. The four domestic ancient pig samples grouped with three Israeli wild boars, and may represent the earliest domestic forms in Britain. The haplotypes shown by the two wild boar samples could describe the original mitochondrial D-loop pattern of wild boar in Britain at the beginning of the Neolithic, however, introgression may have distorted this picture. Both control region and cytochrome *b* products were amplified from the Irish deer samples, and DNA appears to be authentically from a previously unsequenced cervid. In the preliminary data shown here, *Megaloceros* clusters more closely with red deer (*Cervus elaphus*), rather than fallow deer (*C. dama*), contrary to the pattern expected from morphological similarity between these cervids, although caution must be exercised with regard to the reliability of this statement. As far as

can be ascertained, these studies are the first ancient DNA analyses of British *S. scrofa* and giant Irish deer remains.

Amplification of nuclear DNA from ancient remains was tested by using microsatellite markers. Reliable genotypes were successfully obtained from 11 out of 19 cattle bones, from tenth and early eleventh century Viking Dublin, using three autosomal markers. A comparative analysis was performed using data from the same three markers in eleven extant British, Irish and Nordic cattle breeds. The results indicate a probable local origin for the ancient Irish cattle as the remains group more closely with breeds from the British Isles than with those from Scandinavia. The data collected suggest that microsatellites may be useful for the further study of ancient cattle.

(2) Cattle Biogeography

A panel of four Y-specific microsatellite markers was assessed for polymorphism in a range of cattle and related species, three of which displayed putative taurine- and zebu-specific alleles that can be useful indicators of male-mediated gene flow in hybrid populations. Once polymorphism at the four loci had been established, genetic relationships between 73 European, African, Indian and Creole breeds were investigated. The Y-chromosome data confirm the depth of separation between zebu and taurine cattle, thus supporting the independent domestications of these sub-species. European cattle breeds exhibited exclusively taurine haplotypes, with the Creole, African, Near Eastern and South Russian groups containing a mixture of both zebu and taurine haplotypes. When compared to the other groups, analysis of the distribution of taurine haplotypes in the Near East supports the suggestion that this region was a major centre of cattle domestication. The possible local independent domestication of African aurochs, proposed from both archaeological and genetic evidence, is supported by the haplotype diversity exhibited by the African breeds. European cattle groups share the majority of their haplotypes with the Near Eastern group, which relates to the idea that European cattle are a subset of an initially diverse Near Eastern domestic population. Analysis of the Iberian group shows the presence of putative African haplotypes. Local introgression of aurochs could explain the diverse haplotypes contained within the Scandinavian group. The results shown here appear to offer evidence of the two migratory routes of domesticated Near Eastern cattle into Europe. The speed of migration - slow across the Danubian route, quick along the Mediterranean route – is highlighted by the elevated diversity values in the southern and north-western groups when compared to the eastern and central groups.

In addition to the Y-chromosome analysis, 20 nuclear microsatellite markers were amplified in 27-50 unrelated individuals from populations of the endangered German Pustertaler Sprinzen cattle breed and the Pinzgauer, Vosges and Simmental breeds. From the genetic relationships and within-breed diversity estimated for the four breeds, the Pustertaler clustered most closely with the Pinzgauer breed.

Contents

Title Page *i*

Declaration *ii*

Acknowledgements *iii*

Summary *iv*

Contents *vi*

Tables *x*

Figures *xii*

Abbreviations *xiv*

Chapter 1: General Introduction 1

1.1 General Introduction to Domestication 2

1.1.1 *The Fertile Crescent* 2

1.1.2 *Domestication of Animals – Archaeological Evidence* 3

Adaptation of animals to domestication 3

Evidence for animal domestication 4

1.1.3 *Domestication of Animals – Genetic Evidence* 5

1.2 Overview of the history and biology of cattle 8

1.2.1 *Classical archaeological work on cattle diversity and origins* 8

1.2.2 *Biological work on cattle diversity and origins* 11

1.2.3 *Cattle diversity and domestication* 12

1.3 General Introduction to Ancient DNA 14

1.3.1 *A definition of ancient DNA* 14

1.3.2 *The early history of ancient DNA* 14

1.3.3 *Technological aspects of ancient DNA* 15

Preservation of ancient DNA 15

Degradation of ancient DNA 15

The use of bone remains for ancient DNA 16

The use of mitochondrial DNA in ancient studies 16

1.3.4 *Ancient DNA, archaeology and population genetics* 17

Geological (or palaeontological) ancient DNA 17

Archaeological (or Quaternary) ancient DNA 18

Forensic ancient DNA 19

Chapter 2: Ancient DNA Analysis Of 101 Domestic Cattle Remains: Limits And Prospects **21**

2.1 Abstract 22

2.2 Introduction 23

2.2.1 *Patterns of mitochondrial DNA diversity in Bos taurus* *23*

2.2.2 *Ancient DNA – criteria for rating success or failure* *25*

2.2.3 *Archaeological site information* *25*

2.3 Materials & Methods 26

2.3.1 *Archaeological material collection* *26*

2.3.2 *DNA extraction and amplification of archaeological specimens* *27*

2.3.3 *Statistical and phylogenetic analysis* *30*

2.4 Results 31

2.4.1 *Overall amplification success* *31*

2.4.2 *Amplification success - region-by-region* *31*

2.4.3 *Phylogenetic results - region-by-region* *38*

2.4.4 *Genetic diversity analyses* *38*

2.5 Discussion 42

Chapter 3: Ancient DNA Analysis of European Aurochsen **45**

3.1 Abstract 46

3.2 Introduction 47

3.2.1 *The origins of aurochsen* *47*

3.2.2 *Classification of the aurochs* *47*

3.2.3 *The aurochs in Europe* *48*

3.2.4 *Genetic studies of aurochs mitochondrial DNA* *49*

3.3 Materials & Methods 50

3.4 Results 53

Osteological data 63

3.5 Discussion 66

Chapter 4: Y-specific Microsatellite Polymorphism Analyses **70**

4.1 Abstract 71

4.2 Introduction 73

4.2.1 *Simple tandem repeats: microsatellites* *73*

Microsatellite evolution 73

4.2.2 *Review of the Y-chromosome literature* *74*

Brief description of the Y-chromosome 74

Y-chromosome studies in humans 74
Sex chromosomes in cattle 75
Y-chromosome work in cattle 76

4.3 Materials & Methods 78

4.3.1 *Polymorphism screening in a range of bovid species* *78*

4.3.2 *Four-marker analysis of cattle breeds and related species* *80*

4.4 Results 86

4.4.1 *Polymorphism screening in bovids* *86*

4.4.2 *Diversity within cattle geographic groups* *89*

Allele frequencies 89
Mean number of alleles (MNA) and genetic diversity 89
Pairwise mismatch distributions 93

4.4.3 *Network analyses* *99*

4.4.4 *Relationships between geographic groups* *107*

FST values and multidimensional scaling (MDS) 107
AMOVA analysis 109
Geographic analysis using synthetic map drawing 111

4.5 Discussion 113

4.5.1 *Distribution of zebu and taurine haplotypes* *113*

4.5.2 *Trends in diversity among groups* *114*

Africa 115
Europe 116

4.5.3 *Movement of cattle across Europe in the Neolithic* *117*

4.5.4 *Future work* *118*

Chapter 5: Relationships between the endangered Pustertaler-Sprinzen and three related European cattle breeds as analysed with 20 microsatellite loci **119**

5.1 Abstract 120

5.2 Introduction 121

5.3 Materials & Methods 122

5.4 Results & Discussion 123

Chapter 6: Feasibility and utility of microsatellite markers in archaeological cattle remains from a Viking Age settlement in Dublin **126**

6.1 Abstract 127

6.2 Introduction 128

6.3 Materials & Methods 132

6.3.1 *Archaeological material collection from the Fishamble Street excavations* *132*

6.3.2 *DNA extraction and amplification of archaeological specimens* *132*

6.3.3 *Amplification of modern specimens* *133*

6.3.4 *Statistical and phylogenetic analysis* *134*

6.4 Results 136

6.3.1 *Authenticity of archaeological results* *136*

6.3.2 *Genetic diversity and phylogenetic analyses* *138*

6.5 Discussion 145

Literature cited **149**

Appendices **173**

Appendix A Publications arising from this thesis **174**

Appendix B Ancient DNA Analysis of Wild Boar and Domestic Pig **175**

B.1 Abstract 175

B.2 Introduction 175

B.2.1 *Classification of pigs and related species* *175*

B.2.2 *The origins of domesticated pigs* *175*

B.2.3 *History of pigs in Britain* *177*

B.2.4 *Carsington Pasture Cave (CPC) site details* *177*

B.2.5 *Genetic studies of modern pig mitochondrial DNA* *178*

B.3 Materials & Methods 179

B.4 Results 180

B.5 Discussion 183

Appendix C Ancient DNA Analysis of Giant Irish Deer **187**

C.1 Abstract 187

C.2 Introduction 187

C.3 Materials & Methods 189

C.4 Results 191

C.5 Discussion 199

Tables

Chapter 2

Table 2.1 Annealing tube composition for direct sequencing reaction 28

Table 2.2 Labelling reaction for direct sequencing reaction 29

Table 2.3 Table of archaeological *Bos taurus* samples studied, with associated codes, skeletal element used, date excavated and extra information 32

Table 2.4 The variable positions in control region sequences of archaeological *Bos taurus* samples aligned to the European consensus haplotype (T3) 35

Table 2.5 Population linearised *FST* values, with P-values 41

Chapter 3

Table 3.1 Table of archaeological *B. primigenius* samples studied 54

Table 3.2 The variable positions in control region sequences of archaeological *B. primigenius* samples 56

Table 3.3 The variable positions in cytochrome *b* sequences of archaeological *B. primigenius* samples 63

Table 3.4 Measurements of *B. primigenius* specimens from Carsington Pasture Cave 64

Table 3.5 Osteological measurements from CPC-04 65

Chapter 4

Table 4.1 Y-specific microsatellite loci used to assay haplotypic variation 79

Table 4.2 Breed information, including geographical grouping for the 38 cattle breeds sampled as part of this study, plus 35 breeds analysed by J. Kantanen 81

Table 4.3 Allele size ranges and frequencies for *INRA124*, *INRA126*, *INRA189* and *BM861* in a variety of cattle types and related species 88

Table 4.4 Allele size ranges and frequencies for *INRA124*, *INRA126*, *INRA189* and *BM861* in each of the twelve geographic groupings 90

Table 4.5 Direct haplotype counts of all geographic groups 91

Table 4.6 Genetic diversity and mean number of alleles (MNA) calculated for each geographic group 92

Table 4.7 Hierarchical AMOVA analysis, showing the percentage of variation at each of three levels of population hierarchy 99

Table 4.8 Genetic distance estimates between the geographic groupings using: **a)** all the data; **b)** geographic groups containing individuals with zebu haplotypes only; and **c)** geographic groups containing individuals with taurine haplotypes only 108

Chapter 5

Table 5.1 Breeds included in the analysis and their country of origin, including the number of individuals sampled 123

Table 5.2 Genetic distance (D_A) estimates between Pustertaler and each of the other three breeds 123

Chapter 6

Table 6.1 Cattle calcaneus bones analysed from the Fishamble Street excavations 137

Table 6.2 Genetic distance (D_A) matrix obtained from the frequencies of 3 microsatellite markers among 13 breeds 139

Table 6.3 Populations included in the analysis and their country of origin, including the total number of individuals sampled 141

Appendix B

Table B.1 Details for the seven *Sus scrofa* samples from Carsington Pasture Cave (CPC) 179

Table B.2 Variable positions in 377 bp of the 5' end of the mitochondrial DNA control region among wild boars and domestic pigs, including six ancient sequences 182

Appendix C

Table C.1 Primer details for amplification of the mtDNA control region and cytochrome *b* gene of Irish deer 190

Table C.2 Information about extraction from giant Irish deer 192

Table C.3 An 80 base pair fragment of the mitochondrial control region from *Megaloceros giganteus* 193

Table C.4 An 187 base pair fragment of the cytochrome *b* gene from *M. giganteus* 194

Table C.5 Protein information for 186 bp of the cervid cytochrome *b* region 195

Table C.6 Genetic distance estimates between the giant Irish deer ancient sequences using: **a)** 80 bp of the control region; **b)** 187 bp of the cytochrome *b* gene; and **c)** both regions together 196

Figures

Chapter 1

Figure 1.1 Unrooted neighbor-joining mtDNA phylogenies for five Eurasian domestic ungulate species and, inset, for domesticated goat 6

Figure 1.2 Systematic classification of wild and domestic species within the Bovini tribe 9

Chapter 2

Figure 2.1 Reduced median phylogenetic networks constructed for *Bos taurus* mtDNA sequences sampled in Europe, Africa, the Near East and Japan, using 240 bp of the control region 24

Figure 2.2 Reduced median network of ancient and extant regional groups 37

Figure 2.3 A neighbor-joining tree summarising *FST* genetic distances between seven extant cattle groupings and three groups comprising the 12 ancient sequences obtained 40

Chapter 3

Figure 3.1 Approximate geographic location of aurochsen sampled 50

Figure 3.2 Reduced control region network of *B. primigenius* sequences obtained 57

Figure 3.3 Map showing the relative proportions of aurochs and taurine mitotypes in aurochs remains from 14 European archaeological sites 58

Figure 3.4 Comparison of central aurochs haplotype (TP65) to taurine and zebu cattle 59

Figure 3.5 Reduced control region networks of cattle and bison for comparison with aurochs network 60

Figure 3.6 Pairwise differences graphs for 17 aurochsen, 37 Middle Eastern plus 43 Anatolian cattle, 13 wood bison and 40 plains bison sequences 62

Figure 3.7 Hypothetical movements of aurochs and cattle mitotypes 68

Chapter 4

Figure 4.1 Locations of all the breeds sampled as part of this study 83

Figure 4.2 Electronic scans of autoradiograms showing each of the four microsatellite markers amplified in a range of animals 87

Figure 4.3 Graphical representations of the statistical data 94

Figure 4.4 Frequency distributions of the number of mismatches observed in pairwise comparisons within geographic groups, for: **a)** all individuals typed; **b)** zebu haplotypes only; and **c)** taurine haplotypes only 95

Figure 4.5 Frequency distributions of the number of mismatches observed in pairwise comparisons within those geographic groupings that contain both zebu and taurine haplotypes 97

Figure 4.6 Frequency distributions of the number of mismatches observed in pairwise comparisons using taurine haplotypes from ten geographic groupings 98

Figure 4.7 Reduced skeleton network of Y-chromosome haplotypes using four Y-specific microsatellites markers 100
Figure 4.8 Reduced networks of eleven geographical groups 103
Figure 4.9 Two-dimensional MDS plots, drawn using data from four microsatellites, summarising sum of size differences among groups of modern cattle breeds, including: **a)** all individuals typed; **b)** zebu haplotypes only; and **c)** taurine haplotypes only 110
Figure 4.10 **a)** Central locations of the nine geographic groups of cattle breeds; synthetic maps illustrating the taurine geographic variation of: **b)** the first principal coordinate; **c)** genetic diversity; **d)** mean number of alleles (MNA); and **e)** the mean number of mutations calculated from mismatch distributions 112

Chapter 5

Figure 5.1 Neighbor-joining dendrograms constructed from: **(a)** genetic distances among the four cattle breeds; **(b)** allele-sharing distances among the 138 individual animals 124

Chapter 6

Figure 6.1 A two-dimensional MDS plot, drawn using data from three microsatellites, summarising D_A genetic distances among 11 extant European breeds and the medieval Fishamble Street population 140
Figure 6.2 Summary of the clustering results for the cattle data, assuming two populations 143
Figure 6.3 Neighbor-joining dendrogram constructed from allele-sharing distances among 242 individual animals from 11 extant European breeds and the medieval Fishamble Street population 144

Appendix B

Figure B.1 Reduced median network of the two European clades seen in wild boar and domestic pig, putting ancient sequences in a modern context 184
Figure B.2 Neighbor-joining trees of extant and ancient wild boar and domestic pig mitochondrial D-loop haplotypes based on 377 bp of the control region 185

Appendix C

Figure C.1 Neighbor-joining phylogenies constructed using: **a)** 80 bp of cervid control region sequence; **b)** 187 bp of the cytochrome *b* region; and **c)** 267 bp comprising both D-loop and cytochrome *b* data 197
Figure C.2 Manipulated cytochrome *b* tree using MacClade 198

Abbreviations

3'	3 prime end of DNA sequence
5'	5 prime end of DNA sequence
aDNA	ancient DNA
AD	*Anno domini*
AFLP	amplified fragment length polymorphism
AMOVA	analysis of molecular variance
bc	before Christ (uncalibrated date)
BC	Before Christ
bp	base pairs (nucleotides)
bp	before present (uncalibrated date)
BP	years before present (calibrated radiocarbon date)
c.	circa
cal.	Calibrated
Ci	Curie, a measure of radioactivity equal to 3.7×10^{10} Becquerels
D-loop	displacement loop
DNA	deoxyribonucleic acid
dNTP	dinucleotide
EDTA	ethylene-diamine-tetra-acetic acid
HCl	hydrochloric acid
HWE	Hardy-Weinberg equilibrium
IAM	infinite alleles model
IBD	identical by descent
IIS	identical in state
indel	insertion/deletion
kb	kilobase pairs
KCl	potassium chloride
kg	kilograms
LE	linkage disequilibrium
$MgCl_2$	magnesium chloride
MDS	multi-dimensional scaling
µl	microlitre
mM	milli-molar
mtDNA	mitochondrial DNA
Mya	million years ago
n/a	not applicable
nd	not done
ng	nanograms
°C	temperature degrees Celsius
OTU	operational taxonomic unit
PC1	first principal coordinate
PCA	principal component analysis
PCR	polymerase chain reaction
RAPD	random amplified polymorphic DNA
RFLP	random fragment length polymorphism
SSCP	single strand conformation polymorphism
SMM	stepwise mutation model
Tris	Trizma® base
UV	ultraviolet light

CHAPTER 1

General Introduction

1.1 General Introduction to Domestication

Domestication has been described as "*the human creation of a new form of plant or animal – one that is identifiably different from its wild ancestor and extant wild relatives*" (Smith 1995). Wild species were independently domesticated in nine primary centres (Diamond 2002). The earliest evidence occurs in the Fertile Crescent of the Near East around 10,000 years ago, followed by China around 8,500 to 8,000 BP. Between 5,000 and 4,000 BP (Smith 1995), domestication began in Mesoamerica, south-central Andes/Amazonia, eastern United States, Sahel, tropical West Africa, Ethiopia and New Guinea. Recent studies of these areas (reviewed by Pringle 1998a) point to the almost simultaneous implementation of horticulture and pastoralism around 10,000 years ago, with full-scale farming only being developed thousands of years later. Here, only domestication in southwest Asia will be discussed.

1.1.1 *The Fertile Crescent*

The Fertile Crescent in southwest Asia is an area of grasslands and open woodlands arching from the Levant (Israel) in the west, through south-east Anatolia (Turkey and Syria) to the north, on to the slopes of the Zagros mountains (Iran) to the east. The first evidence of farming in the archaeological record originates here around 10,000 years ago, and the general consensus is that this coincided with climatic change (for example, Cole 1967; Byrne 1987; Moore & Hillman 1992; Bar-Yosef & Meadow 1995; Tudge 1998; Diamond 2002). Around 13,000 years BP, during the late glacial of the Pleistocene, the onset of warmer, wetter weather improved growing conditions for wild grasses. The increased seasonality, occurring during the Pleistocene/Holocene transition, was characterised by higher rainfall in winter and a more intense summer drought (Byrne 1987). Rising sea levels from 11,000 to 10,500 BP (Fairbanks 1989), associated with the release of previously ice-trapped water, led to movement of peoples inland (Tudge 1998) towards the upland fertile stands of wild barley and wild emmer and einkorn wheat. With the concentration of population, and the depletion of natural wild stands of grasses, farming would have become a necessity to feed people, and the ease of collection of these crops would have made them amenable to cultivation. Expansion of woodland habitats would have led to an increase in important animal species, including wild populations of sheep, goats and aurochsen, as well as gazelles, asses, pigs and deer. Following the cold, dry period of the Younger Dryas, which lasted from approximately 13,500 to 11,500 years

BP in the Near East (Pringle 1998a), the Fertile Crescent would have been inhabited by a variety of plant and animal resources, and the time was ripe for the transition to an agricultural way of life.

The seven primary Near Eastern domesticates were barley, emmer and einkorn wheat, goats, sheep, pigs and cattle – a list that includes what are still some of the most valuable crops and livestock of the modern world. The transition to agriculture would not have been one specific event in one specific place (Lev-Yadun *et al.* 2000), but rather different domesticates were more likely adapted by different societies over the eastern, central and western regions of the Fertile Crescent over the course of two millennia.

1.1.2 *Domestication of Animals – Archaeological Evidence*

Adaptation of animals to domestication

The agricultural revolution had a series of isolated, independent beginnings involving different areas of the world, different time periods, different peoples and different domesticate species. Only certain habitats hold the potential for domestication, and few wild species can be considered as *preadapted* to domestication (Smith 1995). Indeed worldwide, at the time of domestication, there were 148 species of large terrestrial mammalian herbivores and omnivores, and 200,000 wild species of higher plants, but only 14 animal and around 100 plant species have actually been domesticated (Diamond 1997).

In the case of animals, Francis Galton (1865 [cited in Clutton-Brock 1981]) postulated criteria that need to be fulfilled for a wild species to make the transition to a domesticated life. Domestication was only achieved in those species that exhibited the physiological and behavioural attributes to fulfil these conditions, which include the ability to breed freely under diverse conditions and usefulness to man. As with plants, there are two key elements of the initial domestication of animals: it is most likely that humans first separated a few individuals from wild populations, followed by intervention in the life cycles of the captive populations. Domestication is a process that requires sustained breeding of tame animal populations over a period of time (Uerpmann 1996), and the transformation from wild to domestic would not have occurred within the generation time of the first individuals captured from the wild. This leads to the definition, postulated by Juliet Clutton-Brock (1999), that a domestic animal is "*one that has been bred in captivity for purposes of*

economic profit to a human community that maintains total control over its breeding, organization of territory, and food supply".

Evidence for animal domestication

To determine whether an animal species at a particular site is domestic, a distinction must be made between wild and domestic forms. The criteria used to determine the status of animal remains have been summarised by Meadow (1989 [from Bökönyi 1969]):

Evidence for animal husbandry at a prehistoric settlement –

1. The proportion of age groups of a domestic species is different to that found in wild population.
2. The ratio of male to female remains is different to that occurring naturally in the wild.
3. Domesticated species appear in a region where no wild ancestors exist.
4. Morphological changes appear in domesticated animals.

After a distinction has been made between wild and domestic animals, further direct indications are required to determine whether an animal species has been domesticated locally at the site. This involves evidence of a progressive change from wild to domesticated animals over time.

Evidence for local domestication in faunal remains at a site –

1. Both wild and domestic remains are present at the site.
2. Transitional forms between the wild ancestor and domesticate exist.
3. Temporal changes are evident in age and sex ratios of the wild form.

There are problems with applying each of these criteria. Firstly, demographic evidence of animal domestication is fraught with difficulties, not least the impossibility of precluding selective hunting as an explanation of age and sex ratios in faunal assemblages (Meadow 1989). Even the presence of animal remains outside the natural range of their modern wild relatives does not offer solid evidence of domestication, especially as it is difficult to determine the distribution of wild populations in the past. If wild and domestic animals are found together at a site, this may be due to the continued hunting of wild stock at the same time as domestic animals were being imported from elsewhere, rather than any indication of local domestication (Meadow 1989).

Many changes that occur as part of the adaptive process of domestication are morphological, and it is these observable changes that enable archaeologists to determine whether any species has been domesticated. One example of this is the distinctive change in the cross sections of the horns of domestic and wild goats and sheep (Zeuner 1955; Cole 1967; Smith 1995; Wenke 1999). In many cases, morphological markers are harder to distinguish in early domesticates. Bökönyi (1976) considers that well-defined morphological changes do not occur for about 30 generations. However, at least two kinds of changes can occur as early as the first or second domestic generation (Meadow 1989). The first is the appearance of bone pathologies, reflecting the physical trauma, poor diet and higher infection rates caused by enforced confinement. The second is a decrease in the size of animals, most likely caused by stress, both general and nutritional, and hormonal changes. However, size change in mammals need not be associated with domestication, but may rather reflect temperature-related size diminution (Jarman 1969; Meadow 1989), or that animals had either been transported beyond their original range or were no longer being bred with the wild type (M. Zeder [quoted by Pringle 1998b]).

1.1.3 *Domestication of Animals – Genetic Evidence*

Mitochondrial DNA (mtDNA) has been extensively used to test hypotheses about the domestication of animals proposed by archaeologists. Mammalian mitochondrial chromosomes show clonal transmission, a relatively rapid rate of mutation, and geographic inertia linked to maternal inheritance and low migration of females under managed breeding. The clonal transmission of mtDNA haplotypes from mother to daughter allows determination of discrete maternal lineages within domestic populations that may have complex genetic histories. Sequences that originate from different captures of a diverse wild species can maintain a phylogenetic distinction even after millennia of domestic inbreeding, which allows the resolution of pre- and post-domestic patterns of sequence diversity.

Figure 1.1 shows phylogenetic trees summarising mtDNA sequence diversity in six domestic ungulate species (taken from MacHugh & Bradley 2001 and Luikart *et al.* 2001). With the exception of the horse and goat, the sequence divergence of domestic cattle, water buffalo, sheep and pig cluster into two groups, separated by a dominant internal branch; a topology that resembles a *double-headed broomstick*. Also, the distinct clades identified in these species are spatially-distributed along an east-west division. The phylogenetic divisions within cattle and water buffalo follow the morphological

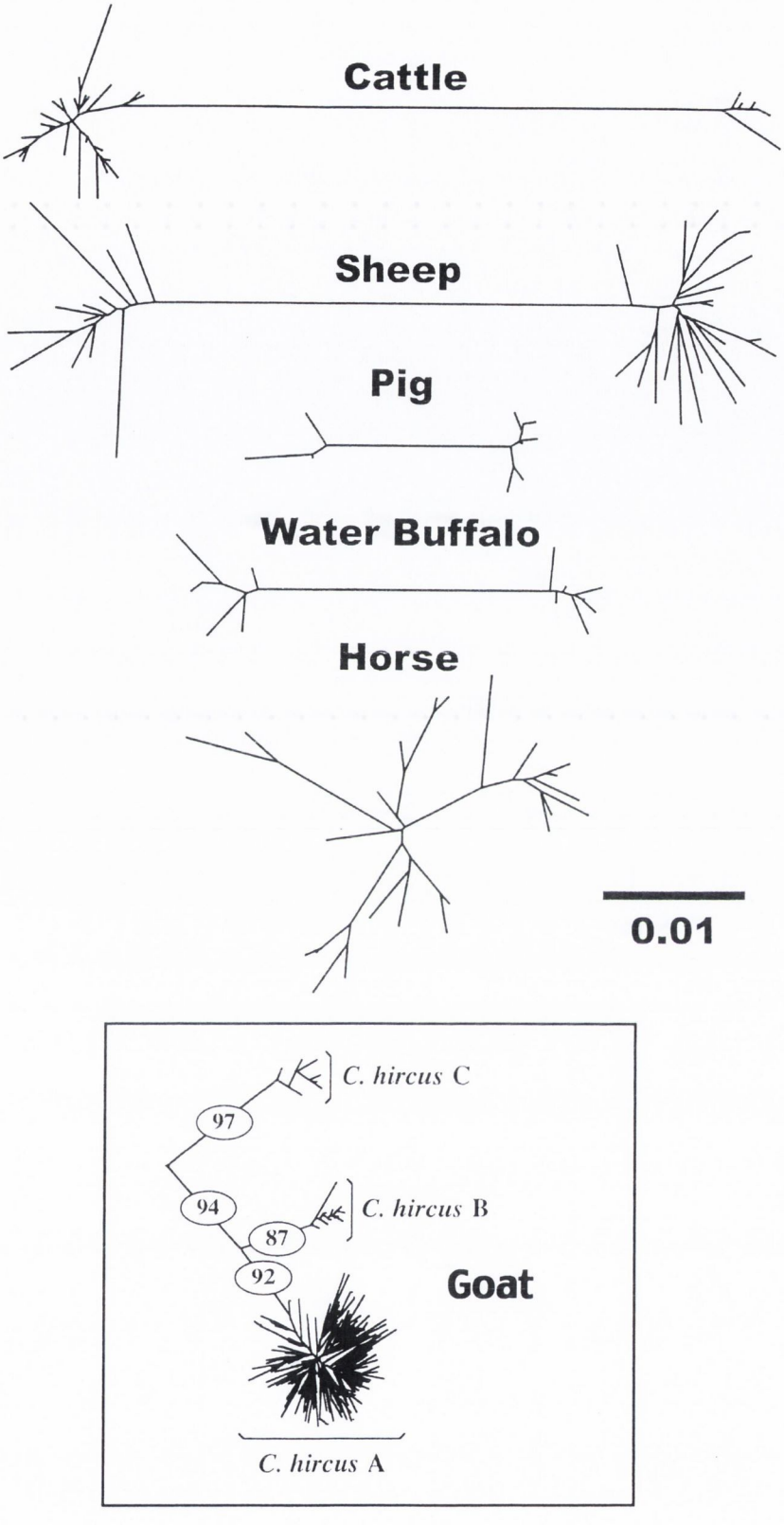


Figure 1.1 Unrooted neighbor-joining mtDNA phylogenies for five Eurasian domestic ungulate species (reproduced from MacHugh & Bradley 2001) and, inset, for domesticaed goat (reproduced from Luikart *et al.* 2001). A duality in diversity is evident in cattle, sheep, pig and water buffalo. In contrast, there are three domestic lineages in goat, and multi-divergent lineages in horse. The scale bar is a measure of genetic relatedness.

distinctions noted between taurine and indicine cattle (Loftus *et al.* 1994a; Mannen *et al.* 1998), and river and swamp buffalo (Tanaka *et al.* 1996; Lau *et al.* 1998) respectively. For pig, the divide occurs between Asian and European animals (Watanobe *et al.* 1999; Giuffra *et al.* 2000), with the sheep divide distinguishing between one clade of European individuals alone, and the other containing samples of both European and Asian provenance (Wood & Phua 1996; Hiendleder *et al.* 1998, 2001). Using mtDNA sequence data, it is possible to estimate the coalescence time between the two clades in each species. Time-depth calibrations for these four domestic species have been calculated to be in the order of several hundreds of thousand of years, thus pre-dating domestication and supporting the hypothesis that cattle, water buffalo, sheep and pig were domesticated at least twice.

The phylogenetic distributions seen in domestic goats suggest at least three geographically disparate captures of wild female progenitors during the formation of early domestic herds (Luikart *et al.* 2001). The more complex pattern of diversity demonstrated by the domestic horse populations are indicative of numerous widespread introductions of wild maternal lineages, associated with extensive domestication events on the Eurasian steppes (Vilà *et al.* 2001).

1.2 Overview of the History and Biology of Cattle

1.2.1 *Classical archaeological work on cattle diversity and origins*

Cattle are the most economically important of domesticated animal species and number over 1.2 billion on a world-wide scale. Cattle are classified in the genus *Bos*, family Bovidae, suborder Pecora (from the Latin *pecus/pecoris*, meaning cattle), of the large order Artiodactyla (even-toed ungulates, or *hoofed* mammals). The family Bovidae consists of 47 extant genera (including antelope, bison, buffalo, cattle, goats and sheep) and 138 species (Nowak 1991) with a natural distribution covering most of Eurasia and North America, and the whole of Africa. They represent the most diverse group of large mammals originating since the Miocene. *Bos* (cattle) have been grouped into the subfamily Bovinae with the following related genera: *Boselaphus* (Nilgai), *Cephalophus* (small forest antelopes), *Tragelaphus* (kudus, bushbucks, etc.), *Bison* (both European and American bison), *Bubalus* (including the Asian water buffalo), *Syncerus* (African buffalo), *Taurotragus* (elands), and *Tetracerus* (for example, the four-horned antelope) (Gentry 1992). Cattle are classified as belonging to the tribe Bovini, which includes bison (both the American *Bison bison*, and European *Bison bonasus* forms), kouprey (*Bos sauveli*), and both the wild and domestic forms of banteng (*Bos* [*Bibos*] *banteng* and *B.* [*Bibos*] *javanicus*), gaur (*Bos* [*Bibos*] *gaurus* and *B.* [*Bibos*] *frontalis*) and yak (*Bos* [*Pöephagus*] *mutus* and *B.* [*Pöephagus*] *grunniens*). A systematic classification of the Bovini tribe is depicted in **Figure 1.2**.

The zhelestids, rat-sized ungulate forebears, have been identified as far back as 85 million years ago in the fossil record (Archibald 1996). The earliest bovid is the *Eotragus* found in the Lower and Middle Miocene, with the first identifiable remains of the Bovini evident towards the end of the Upper Miocene (5 – 11 million years BP) (Pilgrim 1947; Gentry 1992). Although the fossil record is relatively incomplete for the period covering the Pliocene, it is believed that the genus *Bos* developed along similar lines to the buffalo, from *Perimia* through *Proleptobos* to *Leptobos*. During the Middle and Upper Pleistocene, the aurochsen (*B. primigenius*), thought to be descended from *B. acutifrons* in Asia (Pilgrim 1947), were widespread throughout the Old World and three continental races have been identified: *B. primigenius primigenius* (Europe); *B. p. namadicus* (Asia) and *B. p. opisthonomus* (North Africa).

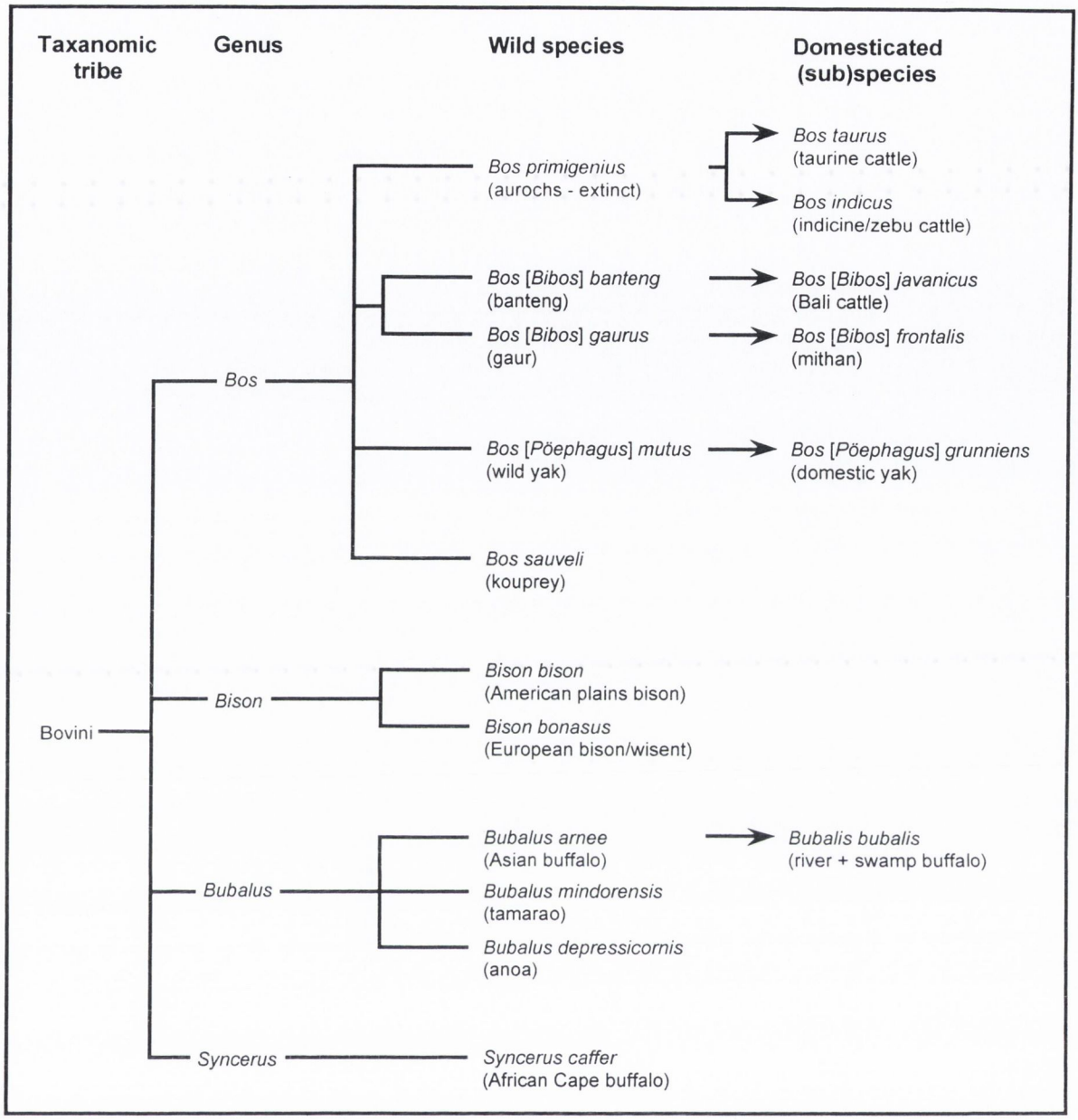


Figure 1.2 Systematic classification of wild and domestic species within the Bovini tribe. Diagram adapted from Payne (1991) considering the phylogenetic inferences of Ritz *et al.* (2000) and Buntjer *et al.* (2002). It has been proposed that the Bovini be divided into four genera: *Bos*, *Bison*, *Bubalus* and *Syncerus*. The lineage leading to domesticated cattle is shown in red. The clustering of the buffalo taxa (*Bubalus* and *Syncerus*), as well as the groupings of the two domestic cattle taxa (*Bos taurus* and *Bos indicus*), the banteng with the gaur, and both *Bison* taxa, are supported by both microsatellite and genetic fingerprint analyses. However, the placement of the yak in the *Bos* genus by Ritz *et al.* (2000) is not well supported, and Buntjer *et al.* (2002) found yak to group more closely with the *Bison* species.

Although they do in fact inter-breed to produce fertile offspring, domestic cattle are classified with binomial Linnean names: *B. taurus* denoting the humpless taurine cattle and *B. indicus* denoting the humped zebu or indicine cattle (following convention and the work of Baker & Manwell 1980 and Grigson 1980). Normally, however, taurine and zebu cattle are considered to be sub-species.

One classical interpretation of domestication states that the idea originated in the Fertile Crescent area of southwest Asia and spread from there by diffusion of people and/or ideas. In the past it had not been considered fruitful to search for alternative centres of domestication outside this region, even though apparent evidence for domestication of cattle from wild populations did exist elsewhere. For example, Neolithic cattle remains predating any found in the Fertile Crescent have been reported from Cyprus (Vigne 1998). It was due to this general reluctance to look beyond the classical interpretation that left the interactions between man and cattle prior to the Neolithic, and the subsequent domestication process, unexplored in great detail until recently.

Several criteria have been postulated to distinguish domesticates (Neolithic *B. taurus*) from their wild ancestors (Mesolithic *B. primigenius*) and thus offer evidence of domestication at dated archaeological sites; for example, changes in population structure, morphological changes and a decrease in size with the onset of domestication. Cattle are remarkably adaptable to changing environments and, while their range has indeed been expanded by recent human intervention, as a species they have ranged over a vast area of the globe encompassing many varying habitats over the last 100,000 years. Jarman (1969) challenges the view that a change in size is necessarily linked to domestication. He references Degerbøl (1962) who found that size change was associated with a long-term trend towards decreased size in cattle earlier than the critical *wild-to-domestic* transition. Cattle from the Pleistocene are smaller that those in the post-glacial and a general decrease in size can be traced through the Bronze and Iron Ages to the Neolithic. This size change is reflected in carnivore species and is most likely due to a climatic change following the glacial. The reason the phenomenon of decreased size came to be associated with the domestication process is more an accident of prehistory rather than any actual domestication events, and Jarman (1969) believes the argument to be circular: "*Mesolithic and Palaeolithic people are known to be hunter-gatherers, and therefore have no domestic stock; Neolithic people are known to have been farmers, and therefore have domestic stock; Neolithic cattle are smaller than Mesolithic and Palaeolithic cattle, therefore small*

cattle are domestic, large cattle wild". He accuses archaeologists of dismissing any remains of large Neolithic cattle as being residual wild populations that were hunted and continues in his paper to offer alternative explanations for other domestication criteria, concluding that the search for the earliest domestication of a species is "*as hopeless as it is futile*" (Jarman 1969).

B. indicus, the humped zebu of India, differ from *B. taurus* in several respects; for example, the presence of a muscular fatty hump of variable size, positioned either on the shoulders (cervico-thoracic) or the neck (cervical); a larger, more pendulous dewlap, drooping ears; and tolerance to arid conditions. Zebu cattle are native to the Indian subcontinent, the Near East and most of Africa, whereas taurine cattle predominate in the North and West of Africa and almost the whole of the Eurasian landmass, from North-west Europe to Japan. Archaeologists have questioned the assumption that *B. indicus* were derived from domesticated forms of *B. p. primigenius*. The idea that *B. p. namadicus* is the likely ancestor of *B. indicus* is supported by craniological work undertaken by Grigson in 1980. Bones believed to be those of *B. indicus* have been found at Neolithic sites in Southern India at Piklihal, Utnur and Maski, and at Mehrgarh in Pakistan dating to the sixth millennium bc (Meadow 1984; Grigson 1985).

B. indicus cattle breeds were thought to have entered sub-Saharan Africa between 700 and 1,500 AD (Payne 1964), prior to and during Arab invasions. However evidence from Ngamuriak, a Neolithic site in Kenya, suggests that these cattle breeds were present in East Africa at least 1,500 years earlier, between the 2nd century BC and the 1st century AD (Marshall 1989). The earliest domestic cattle in Africa are thought to be the humpless longhorn cattle, a breed of *B. taurus*, which were present in North Africa by about 7,000 BC (Gautier 1987).

1.2.2 *Biological work on cattle diversity and origins*

In 1980, Baker & Manwell assessed the classification of 216 cattle breed groups by comparing data collected from approximately 1000 protein polymorphism papers. Biochemical polymorphisms act as a relatively precise measure of genetic relationship and seemed to separate most of the conventional major breed groups to a significant degree, agreeing in the main with the existing geographic and morphological criteria of breed definition (see Baker & Manwell 1991 for a review).

Work has also been undertaken into looking at the systematics of the Bovinae with orthologous mtDNA sequences (for example, Allard *et al.* 1992; Janecek *et al.* 1996). Using mtDNA, the *Boselaphus/Bovini/Tragelphus* cluster has been found to be highly robust, in concordance with immunological, morphological and ecological work, but intertribal bovid relationships cannot be resolved by this method due to rapid cladogenesis, which allows little time for mutation. Mitochondrial work has confirmed the theory that the major bovid lineages originated between the Early Miocene (16 – 17 million years BP) and the Late Oligocene (23 – 28 million years ago), and has narrowed down the bovid radiation to between 5 and 7 million years ago. Recently, Buntjer *et al.* (2002) have attempted to determine robust phylogenies of the bovine species by utilising amplified fragment length polymorphisms (AFLPs). With this method, three consistent clusters were identified: *Syncerus/Bubalus*; *Bos taurus/Bos indicus*; and *Bison bison/Bison bonasus*; with a further indeterminate cluster of *Bos gaurus/Bos banteng*.

A correspondence between geographic distance and genetic distance has been found using classic biological techniques. Gene frequencies from 23 classical polymorphisms (blood group systems and blood and milk proteins) studied in European cattle breeds indicate wide-spread cattle population movements in the Neolithic (Medjugorac *et al.* 1994). This correlates with archaeological evidence supporting the hypothesis that farming spread via demic diffusion into Europe from Anatolia in the Neolithic, about 9000 BP (Zohary 1988).

1.2.3 *Cattle diversity and domestication*

Wide ranging surveys of genetic variation in modern cattle populations have been instrumental in showing that they all derive from either taurine or zebu progenitors, or are hybrids of the two types (Bradley *et al.* 1998). The divergence seen between *B. taurus* and *B. indicus* is indicative of a pre-domestication separation and differences seen between European and African taurine cattle hint at the possibility of separate domestication events from native aurochsen populations on each continent. The pattern of genetic variation seen suggests extensive hybridisation between the two sub-species in Northern India, South-east Europe and most of Africa. Heavy admixture has been uncovered in modern Middle Eastern populations (Loftus *et al.* 1999), and this is most likely due to zebu introgression that has occurred since the Neolithic. These migrations of cattle from India were most probably brought about as a consequence of climatic change in the Middle East leading to increased aridity. Similar hybridisation of zebu and taurine cattle has been studied in African breeds using autosomal AFLPs (Nijman *et al.* 1999), Y-chromosome

microsatellites (Hanotte *et al.* 2000) and nuclear microsatellites (Hanotte *et al.* 2002), showing an east-west and north-south clines of zebu alleles across the continent. Studies of molecular genetic variation provide a more sensitive tool with which to assess breed and population relationships than measurements of morphological characters. When an ancestral population splits, the frequencies of alleles represented within each separate group will slowly change over time due to the processes of mutation, random genetic drift and selection. It is these differences that can be assessed to determine the time depth of the divergence event.

The most striking aspect of bovine genetic diversity studies has been the repeated support for a deep divergence between of *B. taurus* and *B. indicus*. Y-chromosomal polymorphisms (Bradley *et al.* 1994), mtDNA restriction fragment length polymorphism (RFLP) analysis (Loftus *et al.* 1994b), mitochondrial sequence data (Loftus *et al.* 1994a; Bradley *et al.* 1996), and autosomal microsatellite allele frequency variation (MacHugh *et al.* 1996, 1997) all provide clear evidence for a theory that the two major types of cattle (taurine and zebu/indicine) were domesticated independently by two separate Neolithic cultures from distinct races of the wild aurochs, *B. p. primigenius* and *B. p. namadicus*.

Based on palaeontological evidence the estimated time since the divergence of *Bos leptobos* from *Bison* is at least one million years ago; this increases to 1.4 million years when allozyme data are used (Hartl *et al.* 1988). These dates have been used to estimate the split between the *B. taurus* and *B. indicus* lineages, using autosomal microsatellite data, at 610,000 – 850,000 years (MacHugh *et al.* 1997). The time since divergence of the European and African taurine clades has also been dated as either 22,000 – 26,000 years or 180,000 – 250,000 years (Bradley *et al.* 1996; MacHugh *et al.* 1997), which provides tentative evidence for two separate domestication events for the cattle of these two continents as the time since divergence pre-dates the Neolithic. Troy *et al.* (2001) found different ancestral haplotypes present in African and European domesticates, lending weight to this theory, with the recent expansions of African and European cattle (possible ranges of 9,000 and 5,000 BP) linked to domestication events.

1.3 General Introduction to Ancient DNA

1.3.1 *A definition of ancient DNA*

The study of ancient DNA (aDNA) is a sub-discipline of molecular palaeontology, which encompasses extraction of any biomolecule (such as DNA, blood proteins and collagen fibres, resins, fats and oils) from ancient tissues. The discovery that DNA can survive in ancient remains and be amplified by the polymerase chain reaction has added a temporal dimension to evolutionary and population genetics studies. Initial reports suggested that the time period open to investigation extended back to the Cretaceous period. However, attempts to replicate these results, involving DNA purported to be over a million years old, have not been successful. Rather, theoretical studies suggest that DNA is unlikely to survive intact for more than about 100,000 years. However, even over this time period, many evolutionary questions can be addressed, including systematics, palaeoecology, palaeopathology, and evolutionary processes at the population level.

1.3.2 *The early history of ancient DNA*

George Poinar first realised the potential of fossil DNA in 1982 while studying insects entombed in amber (Palmer 1992), but the recognised genesis of the field of ancient DNA study came with the ground breaking work published by Higuchi *et al.* in 1984. This group proved that DNA could survive after death of the organism by cloning mtDNA sequences from the quagga, a member of the horse genus *Equus*, which resided in South Africa until its extinction in 1883. The DNA extracted from dried muscle tissue was of low molecular weight (229 bp fragment) and was in amounts approaching 1% of that expected from fresh muscle. However, this was the first demonstration that clonable DNA sequences could not only be recovered from extinct species, but actually be of phylogenetic importance. aDNA research started to progress rapidly with the advent of the polymerase chain reaction (PCR). Allan Wilson's group at Berkeley was one of the first to utilise PCR in 1985, with Svante Pääbo applying the technique to the quagga skins looked at by Higuchi *et al.* (1984), thus confirming the sequence data obtained (Pääbo 1993).

1.3.3 *Technological aspects of ancient DNA*

Preservation of ancient DNA

Reliable large-scale sampling of genetic information from skeletal material is highly dependent upon a better understanding of the depositional conditions that are conducive to DNA preservation. The degree of DNA preservation is not always closely linked to the gross morphological/macroscopic condition of the sample. This is especially true of the sites that are, or once were, water-logged. Wet sites, such as peat bogs, often yield well-preserved material, but the likelihood of retrieval of DNA is dependent on the pH (neutral pH is desirable; Thomas & Pääbo 1993). However, the general consensus is that cold and dry depositional conditions are best (for example, Hauswirth *et al.* 1994). Thorough dehydration will provide greater protection for DNA against water-requiring autolytic cellular enzymes. Temperature is also a crucial factor; although samples may have been excavated in an extremely arid environment, unless the ambient temperature is relatively low, DNA is unlikely to survive. A recent comparison of the effect of environmental factors upon DNA preservation identified that cool cave conditions, even with the presence of water, gave the highest yield of DNA (Burger *et al.* 1999).

Degradation of ancient DNA

Lindahl (1993) has persuasively argued that the damage that occurs to DNA molecules over time can severely limit the study of aDNA from very old material. However, it is feasible that DNA can survive for tens of thousands of years, especially if the remains have been consistently at low temperatures. Partial dehydration of the sample may also facilitate DNA survival over time, although *dry* DNA is more prone to base damage since the helical structure of the DNA is compromised.

Degradation of target DNA is a problem for PCR. DNA from archaeological specimens is usually of low average molecular size and low copy number due to various forms of damage, including oxidative processes (which cause chemical modification of sugar residues and pyrimidines, particularly thiamine, leading to greatly reduced C and T content; Pääbo *et al.* 1989), breakage of strands, and formation of cross-links (Sykes 1991; Pääbo 1993). This damage may prevent useful sequence data being obtained or can give a false result. Damage may also cause the DNA polymerase to stall, slowing down initial amplification cycles, therefore allowing modern non-specific contaminants or primer-dimer to be preferentially amplified. For these reasons, primers that amplify short

sequences are recommended. It is also desirable, where possible, to use primers that are designed to amplify species- or genus-specific sequences.

The use of bone remains for ancient DNA

aDNA was first extracted from bone in 1989, with Hagelberg *et al.* (1989) and Horai *et al.* (1989) both publishing results of successful extractions of DNA from human bones, aged between 300 and 6,000 years old. Hagelberg *et al.* (1989) were able to amplify mtDNA up to 600 bp in length. This was a breakthrough in aDNA research, for although skeletal material is more likely to be devoid of amplifiable DNA (Hagelberg *et al.* 1991b), those bones that do contain original DNA allow for longer amplifications than soft tissues (Thomas & Pääbo 1993; Richards *et al.* 1995). Also, bones are more abundant than soft tissue, and are generally better preserved (Hagelberg & Clegg 1991).

Pääbo (1993) hypothesises that DNA binds to bone minerals, and must therefore presumably be found in the osteoblasts and the osteoclasts, cells that reside within compact bone and remodel it throughout the organism's life. One reason for the increased fragment size in bone could be that the inorganic surface of the bone provides mechanical protection and a sterile barrier against bacterial and enzymatic decay. DNA binds to the inorganic component of bone called hydroxyapatite, which is a mineral consisting mainly of calcium phosphate, with fluorine and/or chlorine, that crystallises in a hexagonal form (Davis 1987). Lindahl (1993) reports that this adsorption of DNA to hydroxyapatite results in a two-fold decrease in depurination, making DNA preserved in bones more stable.

The use of mitochondrial DNA in ancient studies

Several loci have been utilised in the search for ancient DNA; for example, the X- and Y-chromosomes (Hummel & Herrmann 1991; Mannucci *et al.* 1994; Stone *et al.* 1996) and DNA fingerprinting (Hagelberg *et al.* 1991b). However, most ancient DNA studies concentrate on short mtDNA sequences. The mtDNA molecule is ideally suited for archaeological genetic studies. The mitochondrion is present at a very high copy number in all nucleated cells (in the order of 500-1,500 copies in an average somatic cell; Robin & Wong 1988). This implies that the stochastic survival of intact mtDNA fragments during archaeological diagenesis is much more likely than single-copy nuclear DNA. Other important factors, are that mtDNA is thought not to recombine (Hayashi *et al.* 1985), has a fast rate of evolution (roughly ten times faster than that in the nuclear genome; Brown *et al.* 1979), and a nearly exclusively maternal mode of inheritance, all of which can help in

phylogenetic analysis and relationship studies. It has become the target of choice for studies of evolutionary relationships and genetic diversity. GENBANK contains a huge repository of mtDNA sequences characterised from a wide range of species, particularly those from the most rapidly evolving segment, the control region.

1.3.4 *Ancient DNA, archaeology and population genetics*

Bailey (1998) divided aDNA analysis into three rough timeframes:

1) geological DNA (over 60,000 years old);
2) archaeological DNA (between 100 to approximately 60,000 years old); and
3) forensic DNA (up to 99 years old).

Geological (or palaeontological) ancient DNA

Contentious claims for extraction and amplification of very ancient DNA, that is material in the order of millions of years old, seem to have plagued the field of ancient DNA almost since its inception. In 1990, Golenberg *et al.* claimed to have extracted authentic DNA from a still green 17 million year old magnolia leaf compression fossil. Since then, there have been several other declarations of million year old DNA, including two sequences from 80 million year old dinosaur bone fragments (Woodward *et al.* 1994), and various results from amber-preserved remains: for example, a 25-30 million year old fossil termite (DeSalle *et al.* 1992), a 25-40 million year old leaf from an extinct plant (Poinar *et al.* 1993), and the purported oldest amplifiable DNA yet recovered, that from an 120-135 million year old weevil (Cano *et al.* 1993).

All of these spectacular assertions for the survival of DNA have, unsurprisingly, come under heavy criticism and thorough investigation, suggesting that these publications are erroneous. It was the claim of reproducible DNA from dinosaurs that attracted most comment. Several 134 bp segments of the mitochondrial cytochrome *b* gene were reportedly amplified, and a homology of 60% in amino-acid sequence to mammals and birds was calculated, placing the sequences equidistant from birds, mammals, and reptiles (Woodward *et al.* 1994). Subsequent phylogenetic analysis (Allard *et al.* 1995; Hedges & Schweitzer 1995; Henikoff 1995) showed the sequences clustered more closely with mammals than birds, with whale being the closest fit (Aldhous 1995). Human nuclear inserts of the mitochondrial genome were also detected (Zischler *et al.* 1995) which showed quite comprehensively that the sequences obtained were more likely to be modern human, rather than dinosaur, in origin.

Additional analysis was also undertaken for the Miocene magnolia remains looked at by Golenberg *et al.* (1990). Sidow *et al.* (1991) tried to replicate the work and, although there was no amplifiable plant DNA present, a large amount of contemporary bacterial DNA was found. In fact, since 1995, no reliable reports of DNA past the million mark have been published, and the prevailing view now is that specimens over 100,000 years in age are unlikely to contain endogenous DNA that can be amplified (Lindahl 1993; Poinar *et al.* 1996; Lindahl 1997).

Archaeological (or Quaternary) ancient DNA

aDNA work opens the possibility of molecular evolution studies including ancestral taxa. Such studies may show how a gene pool changes over time, and whether a reduction in genetic diversity is linked to extinctions. For example, a study of 3,500 year old rabbit populations around the Mediterranean sea showed the abrupt introduction of a new mitochondrial haplotype approximately 2,000-1,000 years ago, most probably due to human intervention (Hardy *et al.* 1994).

aDNA can also help in the construction of molecular phylogenies of extinct species. For example, Cooper *et al.* (1992, 2001) showed, by the examination of the flightless moa family, extinct by 1500 AD, that ancestors of the ratite bird group colonised New Zealand twice. In similar studies, the Smilodon sabre-toothed cat was placed with the Felidae (Janczweski *et al.* 1992), and the 13,000 year old giant ground sloth was linked to two species of living two-toed sloths (Höss *et al.* 1996).

aDNA also has the power to answer archaeological questions about the relationships between ancient human populations (for example, the use of sex identification for examining the burial customs and relationships of gender in archaic populations; Brown 1998), and the effects of disease in prehistory (for example, *Mycobacterium tuberculosis*, isolated from a lung lesion of 1,000 year old mummified remains, showed that tuberculosis was present in pre-Columbian America; Salo *et al.* 1994). Also of interest are ancient human population movements, which can be most easily followed using extant mitochondrial sequences with aDNA providing a temporal aspect to the conclusions; for example, peopling of the Easter Islands by Polynesians (Hagelberg *et al.* 1994), and colonisation of the Americas (Merriwether *et al.* 1994).

The development of animal husbandry can also be traced using ancient DNA. Modern analysis of domesticated dogs and their wild ancestors (Vilà *et al.* 1997) seemed to place domestication at 100,000 years ago, although this has not been widely accepted. Further ancient analysis of Old and New World dogs shows that the dogs that entered America were already domesticated (Leonard *et al.* 2002), and therefore domestication must have occurred before 14,000 years ago. MtDNA sequences from Pleistocene horses, dating from 28,000 to 12,000 BP, were compared to contemporary analyses, thus establishing that horses had been domesticated several times over a wide geographic area (Vilà *et al.* 2001). The species identification of skeletal remains of sheep and goats are often indistinguishable, therefore Loreille *et al.* (1997) developed a mtDNA-based test to differentiate bone fragments based on their cytochrome *b* sequence. This tool may prove useful in understanding the movement of domesticated sheep and goats into northern Europe. Bar-Gal *et al.* (2002) also claim to be able to differentiate between wild and domestic forms of goat, and their analysis of remains from Israel indicate that this transition may have taken less than 2000 years for completion. In certain cases, it is also difficult to distinguish the archaeological remains of domestic cattle from wild aurochsen. Successful attempts have been made to extract aDNA from British aurochsen remains dating to the Pleistocene (Bailey *et al.* 1996), as well as early Neolithic aurochs (Troy *et al.* 2001; Chapter **3**), and cattle (MacHugh *et al.* 1999; Chapters **2** and **6**) from across Europe. This topic is explored more thoroughly in Section **3.1**.

Similar studies with plant remains have investigated the advance of agriculture. These include maize (Rollo *et al.* 1991; Goloubinoff *et al.* 1993), wheat (Brown *et al.* 1993, 1994; Allaby *et al.* 1994, 1997) and sorghum (Deakin *et al.* 1998). By modern analysis of extant einkorn samples, Heun *et al.* (1997) have proposed the Karacadag mountains of southeast Turkey as the putative site of wheat domestication, an assertion that could be tested with aDNA (Brown 1999).

Forensic ancient DNA

The extraction of DNA from small amounts of tissue, skin, hair, and bone has proved to be of great value to forensic science. The positive identification of the remains of a murder victim Karen Price, who had been killed eight years previously, was the first successful typing of skeletal remains by DNA analysis (Hagelberg *et al.* 1991b), and the first case of PCR-derived forensic evidence being accepted by a British court of Law (Davies 1991). During their investigation, Hagelberg *et al.* (1991b) noted that DNA becomes fragmented,

due to decay, within the first 24 hours after death, consistent with general diagenetic effects on the corpse thought to be more dependent on environment than time (Hedges *et al.* 1995). Hagelberg has also been involved in verification of the Auschwitz physician, Josef Mengele, whose alleged skeletal remains were exhumed in 1985 (Jeffreys *et al.* 1992), and of the identities of the putative remains of the Romanov family, unearthed eleven years ago in Eastern Russia (Gill *et al.* 1994). In work of this sort, contamination can be ruled out if the sequence obtained is identical to the putative relative. This type of aDNA can therefore provide a very useful rigorous setting for which to test methods and techniques (Hagelberg 1994).

CHAPTER 2

Ancient DNA Analysis Of 101 Domestic Cattle Remains: Limits And Prospects

2.1 Abstract

A total of 101 *Bos taurus* teeth and bones, from 13 Near Eastern, European and African archaeological sites aged between 1,000 to 9,000 years old, were assessed for the presence of verifiable mitochondrial sequences. It was possible to reproducibly amplify and sequence mitochondrial control region DNA extracted from twelve of the samples. The results were compared with published extant data by constructing phylogenetic networks. The sequences obtained were either identical to the reference sequence for modern cattle or closely related to it, and they included three sequences not previously documented. The network analysis of the ancient data highlights the proximity of the ancient DNA cattle sequences to modern Near Eastern, European and African *B. taurus*, as well as regional continuity.

2.2 Introduction

2.2.1 *Patterns of mitochondrial DNA diversity in Bos taurus*

Ancient and extant mtDNA diversity in cattle shows geographical structure that can be related to the process of domestication (Troy *et al.* 2001). Three divergent families of sequences have been described: 1) a South Asian group, which occurred by domestication of the *Bos indicus* progenitor; 2) those found in modern *Bos taurus* and believed to originate from Western Asia; and 3) a separate clade that has to date only been encountered in six wild aurochs (*B. primigenius*) bones sampled in Britain (Bailey *et al.* 1996; Troy *et al.* 2001).

Analyses of previously described 536 *B. taurus* mtDNA control region sequences have revealed a strong phylogenetic structuring of mtDNA haplotypes throughout Eurasia and Africa, with all individuals rooting to the taurine phylogeny through one of five common 240 base pair (bp) haplotypes (Loftus *et al.* 1994a; Bradley *et al.* 1996; Mannen *et al.* 1998; Cymbron *et al.* 1999; Troy *et al.* 2001; Magee *et al.* 2002). These haplotypes are designated as T, T1, T2, T3 and T4, and represent five geographically distributed clusters (**Figure 2.1**). Haplogroup T1 almost exclusively describes African diversity and is found only at low frequencies elsewhere (Troy *et al.* 2001). Haplogroup T4 has thus far only been detected in Japanese cattle (Mannen *et al.* 1998). The other three variants occur at appreciable frequencies in the Near East, with one of these, T3, predominating across Europe. The predominance of this T3 haplogroup in Europe, a subset of the variation encountered in the Near East, is suggestive of a derived Near Eastern origin for European cattle. *B. taurus* mtDNA sequence diversity is highest in Near Eastern cattle populations and this, along with the phylogenetic distinction of British aurochs sequences, has argued against a local origin for Western European cattle. However, the near absence of the T1 haplogroup in the Near East, and the positioning of haplogroup T4 in Japanese animals alone, provides some support that these African and East Asian-specific *B. taurus* variants may reflect regional incorporation of *B. primigenius*. Interestingly, haplotypes in all regions show star-burst network phylogenies, which are characteristic of past population expansions, here likely to be associated with the domestication process.

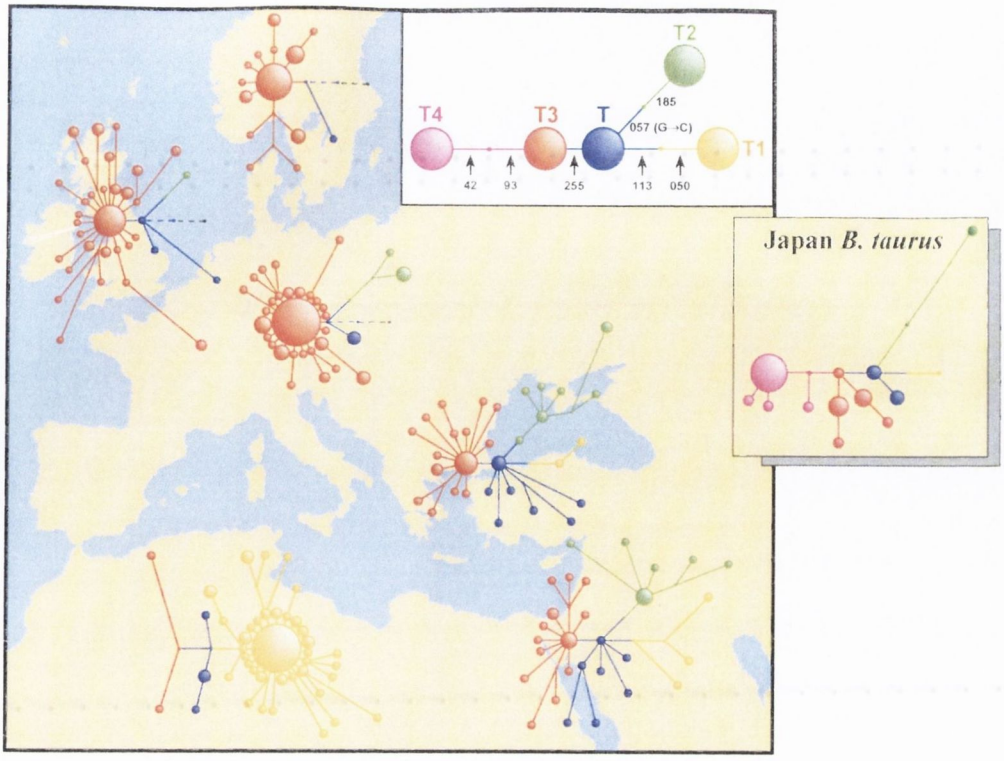


Figure 2.1 Reduced median phylogenetic networks constructed for *Bos taurus* mtDNA sequences sampled in Europe, Africa, the Near East and Japan, using 240 bp of the control region (Mannen *et al.* 1998; Troy *et al.* 2001). Inset: the relationships between the five central, primary *B. taurus* haplotypes, T, T1, T2, T3 and T4. The ancestral haplotype, T, is defined by a transition at position 16,255 from the reference sequence of Anderson *et al.* (1982); T1 by transitions at 16,050, 16,113 and 16,255; T2 by transitions at 16,185 and 16,255 plus a G to C transversion at 16,057; T3 is identical to the reference sequence; and T4 is distinguished by transitions at 16,042 and 16,082. The spatial arrangement of the skeleton network and the colour codes are preserved in the full networks (all networks are placed in the region of origin on the background map, apart from the Japanese population). Seven regional haplotype groups are represented: the Middle East, Anatolia, mainland Europe, Britain, western-fringe Europe, Africa and Japan. Haplotypes encountered in each region (coloured circles) and unsampled intermediate nodes (small points) are shown. Circle areas are proportional to haplotype frequencies and colour coding indicates which of the five skeleton network haplotypes they root to. Geographical distribution of the four haplogroups is evident. T3 predominates in Europe and, along with T and T2, comprises almost all Near Eastern variation. Haplogroup T1 dominates African diversity but is scarcely represented elsewhere. T4 is only seen in Japan. Diagram modified from original drawn by D.A. Magee, and adapted from Troy *et al.* (2001).

Presence of modern extensive reference data facilitates the study of ancient mtDNA in cattle. Also, the mtDNA molecule is present at a very high copy number in most somatic cells (approximately 1,000 copies), implying that survival of intact mtDNA fragments during archaeological diagenesis is more likely than that of single-copy nuclear DNA. In addition, the segment of the D-loop analysed here is the most variable region of the mitochondrial genome. The 157 bp section amplified in these ancient specimens contains

66 sites that vary in aurochs and modern *B. taurus* and *B. indicus* cattle, with the larger 176 bp section defining a further 31 sites (Bailey *et al.* 1996; Bradley *et al.* 1996; Troy *et al.* 2001).

2.2.2 *Ancient DNA – criteria for rating success or failure*

Several researchers in the field have set out various criteria to measure the authenticity of ancient (aDNA) results (for example, Cooper & Poinar 2000; MacHugh *et al.* 2000). We believe the most important criterion to be reproducibility. If the sequence data obtained are to be accepted as authentic then it must be reproducible. In our opinion, this condition is fulfilled when consistent sequences are obtained in three or more amplifications, with at least one from each of two separate extractions. A physically isolated pre-amplification laboratory is used for extractions and set-up of PCR, thus avoiding carry-over contamination. As shorter length fragments are easier to amplify from ancient samples than longer ones (Bailey *et al.* 1996; MacHugh *et al.* 1999), we target small fragments using species-specific primers, as suggested by Richards *et al.* (1995). Extractions and amplification reactions are run in tandem with ancient specimens from distantly related species for which the primers should not work (Thomas & Pääbo 1993). All sequence data obtained have to be phylogenetically consistent before the results are accepted as genuine.

2.2.3 *Archaeological site information*

The material tested here vary in date, from a Bronze Age site in Syria, Neolithic sites in Anatolia, Israel and France, and Iron Age sites from Western Africa. The climatic conditions of the sites also vary in macroclimatic condition, from semi-arid Near Eastern and African areas, to arid Mediterranean (Anatolia), wet Mediterranean (Southern France) and temperate (Paris Basin) regions. Thirteen sites were sampled: seven from the Near East (Çatalhöyük in Anatolia; five sites around Israel – Tel Hreiz, Atlit Yam, Newe Yam, Kfar Hahoresh and Abu Gosh; and Tell Brak in Syria); three from France (two in the Ardèche region – Baume d'Oullins and Combe Obscure; and one in Paris, Bercy); and three sites from Western Africa (Jenné-jeno and Kaniana in Mali; plus Sincu Bara in Senegal). These sites also display different conditions of deposit, from open field (African and most Near-Eastern sites) to karstic caves (Ardèche), to waterlogged areas (Bercy, Paris; Tel Hreiz, Atlit Yam and Newe Yam, Israel). Details of these sites and associated information are shown in **Table 2.3**.

2.3 Materials & Methods

2.3.1 *Archaeological material collection*

101 cattle teeth and bones, sourced by Keith Dobney, Liora Horwitz and Louise Martin (Middle East), Daniel Helmer, Anne Tresset and Jean-Denis Vigne (France), and Susan McIntosh (Africa), were retrieved from various archaeological sites. Based on archaeological context, 23 Bronze Age *B. taurus* samples were extracted, alongside 62 Neolithic and 16 Iron Age samples. In addition, three *Equus* (horse) teeth from Çatalhöyük, three *Sus scrofa* (pig) teeth from Israel, and one *Sus scrofa* (wild boar) vertebra from the 5,000 year-old site of Carsington Pasture Cave in Derbyshire, were extracted and used as controls.

Previous studies of aDNA survival have found a good correspondence between relatively subjective assessments of bone preservation and the presence of amplifiable endogenous DNA (Richards *et al.* 1995; Colson *et al.* 1997). Most of the Near Eastern bovine bone material was light in weight, sandy coloured, with damaged external surfaces. The majority of these remains came from open field sites in semi-arid areas, apart from those bones from Tel Hreiz, Atlit Yam and Newe Yam, which were all from submerged sites in the Mediterranean Sea. None of the Near Eastern remains could be considered well preserved using the gross morphological criteria outlined by Hagelberg *et al.* (1991a). On the other hand, teeth and bones from the two French caves, Baume d'Oullins and Combe Obscure, were very well preserved and, with reference to radiocarbon dates and previous aDNA data from caves in Southern France (Hardy *et al.* 1995), could be expected to have both well preserved collagen and aDNA. Bercy is located on the edge of a palaeochannel of the Seine (Bocherens *et al.* 1997), and all bones used in this study were taken from areas within fluctuating water levels. Bones from immersed areas were brown-black in colour, and this staining was assumed to mainly be due to humic acids. Stable isotopes analyses carried out previously on these bones have revealed good collagen preservation in the majority. Extraction yields were relatively high, with carbon and nitrogen amounts, as well as the C/N ratios, being similar to those obtained on modern materials (Balasse *et al.* 1997). The African bones were, on the whole, dense and dark brown in colour, with much trapped loose soil within the bone cavities.

2.3.2 *DNA extraction and amplification of archaeological specimens*

A range of aDNA extraction methods have been developed specifically for bone and tooth material. These procedures are required to separate and purify the DNA from proteins and other organic components of bone or teeth, which may inhibit DNA amplification via PCR. The original silica-based method (Höss & Pääbo 1993), adapted from Boom *et al.* (1990), is prone to contamination and can shear DNA leading to low recovery. However, this chaotropic salt bridge method does remove potential inhibitors that co-extract, and therefore Richards *et al.* (1995) adapted this to include a phenol-chloroform step prior to use of a commercially available silica column. Recently, Yang and colleagues (1998) compared four different methods of extracting DNA from proteinase K-digested bone solutions. They suggest buffer extraction followed only by QIAquick purification as, in a single step, the DNA is concentrated and separated from contaminating substances that can inhibit the PCR. However, when the starting volume of the bone digest is large, it is suggested that the volume be decreased using an additional Centricon concentration step prior to purification. The bone remains sampled here were prepared using a modified procedure based upon Yang *et al.*'s (1998) experiments.

Before extraction, each bone or tooth was sandblasted to remove external contamination that can out-compete endogenous DNA (Richards *et al.* 1995). This protocol involves the following: 0.5 to 5 gram samples of bone or tooth powder was drilled from each sample, and dissolved in 8 ml extraction buffer (0.5 M EDTA pH 8.0, 0.5% sodium dodecyl sulfate, 100 µg/ml proteinase K). This solution was incubated in a shaking waterbath at 55°C overnight, followed by 24 hours at 37°C. The extraction solution was centrifuged at 2,000*g* for 5 minutes, and the supernatant transferred to a 30 kDa molecular weight cut-off 4 ml centrifugal filter tube to reduce the digest volume and remove excess EDTA. After centrifugation for an hour, the flow-through was discarded and the remaining supernatant was washed with 10 mM Tris-HCl at pH 7.5 and spun once again. The supernatant was then transferred to a 2.0 ml tube and mixed with five volumes of QIAquick PB buffer. Using a sterile disposable pipette, 750 µl was loaded directly onto a QIAquick column and centrifuged at 12,800*g* for 1 minute. The flow-through was discarded and the process repeated until all of the extract had passed through the column. The DNA was washed by adding 500 µl of QIAquick PE buffer, followed by centrifugation for 1 minute. The flowthrough was again discarded and the DNA was eluted from the column by addition of 100 µl warmed TE buffer, and centrifugation for 1 minute. The DNA was then ready to be amplified by PCR. Varying numbers of DNA extractions were performed for each

specimen. In the cases where only one extract was attempted, this was because no amplification products were obtained from several attempts using the first extract.

PCR set-up was conducted in a laboratory dedicated solely to pre-amplification ancient work. PCR conditions were as described in MacHugh *et al.* (1999). The region analysed was a defined, highly variable region of the mtDNA control region between bases 16,022 and 16,262 (Loftus *et al.* 1994a). Amplification was attempted for both a 157 base pair fragment, $AN2_{FOR}$ (16,022–16,041) – $AN1_{REV}$ (16,178–16,159); and a 176 bp region, $AN1_{FOR}$ (16,159–16,178) – $AN3_{REV}$ (16,334–16,314). Primer details are as follows: $AN2_{FOR}$ 5'-GCCCCATGCATATAAGCAAG-3'; $AN1_{REV}$ 5'-CACGCGGCATGGTAATT AAG-3'; $AN1_{FOR}$ 5'-CTTAATTACCATGCCGCGTG-3'; $AN3_{REV}$ 5'-CGAGATGTCTTA TTTAAGAGG-3'. These primers are species-specific so do not amplify human DNA.

Direct sequencing conditions were carried out as follows. PCR products were purified from excess primers, unincorporated nucleotides and other buffer components using Concert™ PCR purification columns. After elution and re-suspension in TE buffer to a final concentration of approximately 50-200 ng/µl, the following annealing reaction was assembled in standard 0.5 ml PCR tubes (**Table 2.1**).

Component	Quantity (x1)	Final amount
Purified PCR amplicon	2.0 µl	100-400 ng
10 µM PCR primer	2.0 µl	20 moles
Sequenase™ buffer	2.0 µl	---
50% DMSO	1.0 µl	5% DMSO (v/v)
Purified water	3.0 µl	---

Table 2.1 Annealing tube composition for direct sequencing reaction.

This mixture was denatured by heating to 95°C for 3-4 minutes in a pre-heated PCR machine. The annealed samples were then snap-cooled by immediately plunging into liquid nitrogen. While this annealed primer-template mix was left sitting in liquid nitrogen, a labelling master-mix for the number of templates required was assembled (**Table 2.2**).

Component	Quantity (x1)	Final amount
100 mM DTT	1.50 µl	~ 30 mM
50% DMSO	0.50 µl	~ 5% v/v
Sequenase™ labelling mix	0.40 µl	---
Purified water	2.10 µl	---
a-^{35}S–dATP	0.50 µl	~ 5 µCi
T7 Sequenase™ enzyme	0.25 µl	~ 3 units

Table 2.2 Labelling reaction for direct sequencing reaction.

Each of the four Sequenase™ termination mixes (G, A, T and C) was added separately to 50% DMSO solution in a 9:1 ratio to give a working concentration of 5% DMSO. The amount of termination mix required depends on the number of samples. 2.5 µl of each DMSO-containing termination mix was aliquoted into a labelled 60-well Nunclon Microwell™ (Terasaki) plate. The annealed template-primer tubes were then removed quickly from the liquid nitrogen and 5 µl of labelling reaction master-mix added to the side of each tube and flick-spun in a microcentrifuge to simultaneously thaw and mix. The labelling reaction was allowed to continue proceed at room temperature for at least 2 minutes before 3.5 µl from each tube was added to the appropriate termination mixes on the Terasaki tray. These termination reactions were carried out at room temperature for a further 5 minutes and then stopped by the addition of 4 µl of Sequenase™ stop solution. The reactions on the trays were then heated to 90°C before gel fractionation, using standard polyacrylamide gel electrophoresis.

Second-round PCR was not undertaken on any samples that did not amplify in the first-round. All non-amplifiable samples were tested for presence of inhibitors that may have been impeding PCR amplification of endogenous DNA. This involved spiking each negative sample extract with aurochs DNA (from 6,000 BP Carsington Pasture Cave) in a ratio of 1:4. PCR was continued as described above, and those samples that yielded a product from the spiked sample were then designated as containing little or no endogenous DNA.

The criteria for authenticating mitochondrial haplotypes were as follows. For working samples, at least two independent extractions were undertaken from those samples that produced amplifiable DNA. Subsequently, a range of PCR amplifications and subsequent

direct sequencing analyses were performed from each purified DNA extract. Reproducible samples were designated as those that gave consistent sequences in three or more amplifications, with at least one from each of two separate extractions. In order to be considered authentic, any mutations observed had to be replicated in sequences from two separate extracts. The mtDNA sequences thus derived from each sample were therefore verified through independent extractions, amplifications and sequence determinations.

2.3.3 *Statistical and phylogenetic analysis*

mtDNA sequences were aligned by eye and reduced median networks were constructed using the median algorithm of Bandelt *et al.* (1995). The networks include published data from the following regions: Anatolia (43), Middle East (37), Mainland Europe (91), and Africa (95). These data comprise that analysed by Troy *et al.* (2001), truncated to the 117 bp segment presented here. In the network, all sequences were shaded corresponding to which of the five *B. taurus* haplogroups they group with (T, T1, T2, T3 or T4). The skeleton network estimates the phylogenetic relationship between these five haplogroups and the British *B. primigenius* reference sequence (GENBANK accession number AF336746; Troy *et al.* 2001). The segment amplified in the ancient sequences does not contain the 16,225 base pair change that differentiates T from T3 (as designated by Troy *et al.* 2001), and therefore these two haplogroups have been amalgamated into the T/T3 group discussed here.

Along with other regional data (32 Japanese, 65 British and 51 Western European Fringe *B. taurus*; Mannen *et al.* 1998; Troy *et al.* 2001), the truncated modern data were included in analyses of inter-population genetic distances between extant and ancient populations. The ancient sequences were regionally grouped. Linearised F_{ST} values, with associated P-values after 1023 permutations, were generated according to Slatkin (1995) using the ARLEQUIN computer program (version 2.000; Schneider *et al.* 1997). The P-value of the test is the proportion of permutations leading to a F_{ST} value larger or equal to the observed one, so low P-values equate to significant genetic distances. The neighbor-joining method (Saitou & Nei 1987) was used to construct a dendrogram of breed relationships from the genetic distances using the PHYLIP package (Felsenstein 1989).

2.4 Results

2.4.1 *Overall amplification success*

Table 2.3 gives full data for amplification success in all 101 samples tested. Reproducible sequences were only obtained from 12 samples: one Çatalhöyük, two Tell Brak, one Baume d'Oullins, one Combe Obscure, three Bercy and four African (**Tables 2.3** and **2.4**). These comprise a total of six haplotypes, differentiated by eight polymorphic sites, three of which have not been previously described. They were placed on reduced modern networks (**Figure 2.2**). Sites prone to repeated mutations have previously been identified (Troy *et al.* 2001), and most of the reductions (11 sites out of 13) were at these sites.

Only one bone, from Çatalhöyük, gave amplification products with the larger 176 bp fragment. Apart from the 12 remains that gave reproducible sequences, varying numbers of amplification products for the smaller 157 bp region were obtained in a further 18 of the 101 samples amplified: eight Çatalhöyük, one Israel, and nine Tell Brak. Some of these 18 samples gave two or more amplicons that could be sequenced, in most cases from separate extractions. However, these were discounted because ostensibly identical sequences contained different substitutions (**Table 2.3**). The inhibition test showed that three of the 13 Bercy mandibles contained unknown substances that were inhibiting DNA amplification. This inhibition was overcome by using five-times the usual concentration of *Taq* polymerase.

2.4.2 *Amplification success - region by region*

Only one reliable sequence was obtained from the 23 Çatalhöyük remains, although nine samples generated products. This is a success rate of 4.3%, which decreases to 2.6% when a further 15 negative samples (D.E. MacHugh unpublished data) are taken into consideration. Only one of the 23 samples from Israel gave an amplifiable product. However, this was excluded as a mosaic sequence. Analysis of a further five bones from Abu Gosh also proved unsuccessful (D.E. MacHugh unpublished data). Of the 23 Tell Brak remains sampled, 11 gave products, of which two were reliable (8.7% success). Two Ardèche samples gave reproducible results from the three bones sampled (66.7%). The Bercy mandibles gave three reliable products out of a total of 13 samples (23.1%), but three others were found to contain PCR inhibition and therefore may harbour intact aDNA. Of the 16 samples from West Africa, four samples gave reliable sequences (25.0%). Two

lab code	archaeological code	skeletal part	date excavated	extra information	PCR amplification: no. of extractions	PCR amplification: 157 bp product	PCR amplification: 176 bp product	PCR amplification: maximum sequence length (bp)	sequence cluster(s)	reasons for success/failure
Catalhöyük - c. 9-8000 BP (early Neolithic)										
CH01	CH1996 #5 Area South/Unit 1668	metatarsal	1996	stored at RT	4	no	no	–	–	–
CH02	CH1996 #4 Area South/Unit 1520	metacarpal	1996	stored at RT	4	1 of 8	no	157	T/T3	non-reproducible
CH03	CH1996 #3 Area South/Unit 1579	metacarpal	1996	stored at RT	3	1 of 6	no	107	T/T3	non-reproducible
CH04	CH1996 #1 Area South/Unit 1832	metacarpal	1996	stored at RT	2	1 of 4	no	124	T/T3	non-reproducible
CH05	CH1996 #X1 Area South/Unit 1816	metatarsal	1996	stored at RT	2	no	no	–	–	–
CH07	CH1996 #4 Area South/Unit 1579	metapodial	1996	stored in freezer	4	3 of 8	no	157	T/T3; aurochs	contradiction
CH08	CH1996 #8 Area South/Unit 1668	metacarpal	1996	stored in freezer	2	1 of 4	no	132	T/T3	non-reproducible
CH09	CH1996 #3 Area South/Unit 1520	metacarpal	1996	stored in freezer	2	3 of 4	2 of 4	157	T/T3; aurochs	contradiction
CH10	CH1996 #3 Area South/Unit 1822	metacarpal	1996	stored in freezer	2	no	no	–	–	–
CH11	CH1996 #X1 Area North/Unit 1430	metacarpal	1996	stored in freezer	3	4 of 6	no	133	T/T3	reproducible
CH99-01	90 2331.F3 South	tooth	1997	CH97-B10	2	no	no	–	–	–
CH99-02	74 2706.F1 South	tooth	1997	CH97-B01	2	no	no	–	–	–
CH99-03	83 2331.F2 South	tooth	1997	CH97-B07	2	no	no	–	–	–
CH99-04	82 2303.F1 South	tooth	1997	CH97-B06	2	no	no	–	–	–
CH99-05	80 1873.F611 South	tooth	1997	CH97-B04	2	no	no	–	–	–
CH99-06	85 2340.F2 South	tooth	1997	CH97-B09	2	no	no	–	–	–
CH99-07	2845.X8 S.11 (S3) South	tooth	1998		2	2 of 4	no	157	T/T3	contradiction
CH99-08	3375.F1 (S6) South	tooth	1998	from Plot S2	1	no	no	–	–	–
CH99-09	2845.X5 (S2) South	tooth	1998		2	1 of 4	no	157	T/T3	non-reproducible
CH99-10	3314.F159 South	tooth	1998		1	no	no	–	–	–
CH99-11	3129.F54 S.3 (S5) South	tooth	1998		1	no	no	–	–	–
CH99-12	3702.X4 (S7) South	tooth	1998		1	no	no	–	–	–
CH99-13	2845.X7 S.11 (S4) South	tooth	1998		1	no	no	–	–	–
Israel - c. 9-6000 BP (Neolithic)										
IS01	Tel Hreiz - 30/93-24/61	radius	1993	submerged in Mediterranean Sea	1	no	no	–	–	–
IS02	Tel Hreiz - 30/93-33	humerus	1993	submerged in Mediterranean Sea	1	no	no	–	–	–
IS03	Tel Hreiz - 30/93-27	humerus	1993	submerged in Mediterranean Sea	1	no	no	–	–	–
IS04	Tel Hreiz - 30/93-24/60	scapula	1993	submerged in Mediterranean Sea	1	no	no	–	–	–
IS05	Newe Yam - 30/93-9/61	humerus	1993	submerged in Mediterranean Sea	1	no	no	–	–	–
IS06	Newe Yam - 28/8/90-5/9/90	pelvis	1990	submerged in Mediterranean Sea	1	no	no	–	–	–
IS07	Newe Yam - 30/93-9/159	tibia	1993	submerged in Mediterranean Sea	1	no	no	–	–	–
IS08	Newe Yam - 30/90-9	metatarsal	1990	submerged in Mediterranean Sea	1	no	no	–	–	–
IS09	Newe Yam - 21/94-47	scapula	1994	submerged in Mediterranean Sea	1	no	no	–	–	–
IS10	Newe Yam - 30/93-9/78	femur	1993	submerged in Mediterranean Sea	1	no	no	–	–	–
IS11	Newe Yam - 30/90-10/125	radius	1990	submerged in Mediterranean Sea	1	no	no	–	–	–
IS12	Atlit Yam - Bld. 11 Ht. 0-120	tibia	1988	submerged in Mediterranean Sea	1	no	no	–	–	–
IS13	Atlit Yam - Bld. 11 Ht. 60-80	femur	1988	submerged in Mediterranean Sea	1	no	no	–	–	–
IS14	Atlit Yam - Bld. 7 No. 14	scapula	1990	submerged in Mediterranean Sea	1	no	no	–	–	–
IS15	Atlit Yam - Bld. 9 No.9	tibia	1990	submerged in Mediterranean Sea	1	no	no	–	–	–
IS16	Atlit Yam - Bld. 9 No.47	tibia	1990	submerged in Mediterranean Sea	1	1 of 2	no	157	aurochs	mosaic
IS17	Atlit Yam - Bld. 20	femur	1989	submerged in Mediterranean Sea	1	no	no	–	–	–
IS18	Kfar Hahoresh - L1005 No.230	tibia	1992		1	no	no	–	–	–
IS19	Kfar Hahoresh - L1005	metatarsal	1992		1	no	no	–	–	–
IS20	Kfar Hahoresh - ZW58d 45/3.88-3.98	radius	1997		1	no	no	–	–	–
IS21	Kfar Hahoresh - Q57a 30/4,012/30 L1155	metatarsal	1997		1	no	no	–	–	–
IS22	Kfar Hahoresh - L1005 No.153	tibia	1992		1	no	no	–	–	–
IS23	Abu Gosh - L146 AA5 Level 3 Ht. 696.02-695.96	scapula	1995		1	no	no	–	–	–

lab code	archaeological code	skeletal part	date excavated	extra information	no. of extractions	PCR amplification: 157 bp product	PCR amplification: 176 bp product	maximum sequence length (bp)	sequence cluster(s)	reasons for success/failure
Tell Brak, Syria - c. 4-2000 BC (Chalcolithic-Bronze Age)										
TB01	TB94 A1186:2/HS	limb bone	1994	4th - 3rd millennium	2	1 of 4	no	157	T/T3	non-reproducible
TB02	TB94 A2047:2/HS	bone	1994	4th - 3rd millennium	2	no	no	–	–	–
TB03	TB94 A1077:2/HS	vertebra	1994	4th - 3rd millennium	2	4 of 4	no	157	T	reproducible
TB04	TB96 A4130:2/HS1	bone	1996	phase II, mid 4th millennium (Middle Uruk)	2	no	no	–	–	–
TB05	TB94 A1023:2/HS	bone	1994	4th - 3rd millennium	2	2 of 4	no	157	T/T3	PCR product from one extract only
TB06	TB95 A6035:2/HS4	vertebra	1995	phases III/IV, late 4th - late 3rd millennium	2	3 of 4	no	157	T/T3	sequence mosaic in one extract
TB07	TB95 A1136:2/HS3	vertebra	1995	phase IV, late 3rd millennium	2	4 of 4	no	157	T/T3	reproducible
TB08	TB95 A6517:3/HS4	bone	1995	phases III/IV, late 4th - late 3rd millennium	2	2 of 4	no	157	T/T3	contradiction
TB10	TB96 A4083:2/HS1	bone	1996	phase II, mid 4th millennium (Middle Uruk)	2	1 of 4	no	115	T2	non-reproducible
TB11	TB95 A756:2/HS6	vertebra	1995	phase I, early-mid 4th millennium (early-mid Uruk)	2	no	no	–	–	–
TB12	TB95 A155:2/HF2	bone	1995	phase III, late 4th - early 3rd millennium (Ninevite V)	1	no	no	–	–	–
TB13	TB95 A9004:2/HF2	bone	1995	phase III, late 4th - early 3rd millennium (Ninevite V)	1	no	no	–	–	–
TB14	TB?? Trench HL	bone	–	phase III, late 4th - early 3rd millennium (Ninevite V)	1	no	no	–	–	–
TB15	TB94 A1023:12/HS3A	bone	1994	phase IV, late 3rd millennium	1	no	no	–	–	–
TB16	TB94 A1023:12/HS3B	bone	1994	phase IV, late 3rd millennium	2	3 of 4	no	157	T/T3	contradiction
TB17	TB94 A62:3/HN	bone	1994	phase V, 2nd millennium	1	no	no	–	–	–
TB18	TB96 A244:2/HN	bone	1996	phase V, 2nd millennium	1	no	no	–	–	–
TB19	TB96 A264:10/HN	bone	1996	phase V, 2nd millennium	1	no	no	–	–	–
TB20	TB94 A1023:12/HS3C	vertebra	1994	phase IV, late 3rd millennium	1	no	no	–	–	–
TB21	TB96 A252:2/HN	limb bone	1996	phase V, 2nd millennium	2	1 of 4	no	157	T/T3	non-reproducible
TB22	TB96 A215:2/HN	vertebra	1996	phase V, 2nd millennium; possibly *Bos indicus*	2	1 of 4	no	157	T/T3	non-reproducible
TB23	TB96 A192:2/HN	bone	1996	phase V, 2nd millennium; possibly *Bos indicus*	2	1 of 4	no	157	T/T3	non-reproducible
TB24	TB96 A4165:2/HS1	limb bone	1996	phase II, mid 4th millennium (Middle Uruk)	1	no	no	–	–	–
Ardèche - c. 7000-7500 cal BP (early Neolithic)										
Ardèche 01	Baume d'Oullins - BA66 ADN; BO83-N12.C46 Z.250	tooth	1983	early Cardial Neolithic (6630 ± 110 bp); karstic cave site	2	2 of 2	no	129	T/T3	reproducible
Ardèche 02	Combe Obscure - CO74 G4 chc 6 37	jaw bone	1974	Cardial Neolithic (earlier than 6400 ± 160 bp); karstic cave site	2	2 of 2	no	129	T/T3	reproducible
Ardèche 03	Combe Obscure - CO74 H5 chc 6A 264	jaw bone	1974	Cardial Neolithic (earlier than 6400 ± 160 bp); karstic cave site	2	no	no	–	–	–
Bercy, Paris - c. 5000 cal BP (middle Neolithic)										
Bercy 01	QS. MXVII 16-C21; MB 10000	right mandible	–	upper part of river bank; good collagen content	2	no	no	–	–	inhibition present
Bercy 02	QS. MXVI 24; MB 20000	right mandible	–	upper part of river bank; good collagen content	2	no	no	–	–	–
Bercy 03	QS. MXV 8; MB 21000	right mandible	–	upper part of river bank; good collagen content	2	no	no	–	–	–
Bercy 04	QS. MXV; MB 23000	right mandible	–	upper part of river bank; good collagen content	2	no	no	–	–	–
Bercy 05	QS. LXII 9-C21; MB 24000	right mandible	–	lower part of river bank; low collagen content	2	no	no	–	–	inhibition present
Bercy 06	QS. MXVIII 13-C21; MB 28000	right mandible	–	upper part of river bank; good collagen content	2	no	no	–	–	–
Bercy 07	QS. MXX 22-C21; MB 29000	right mandible	–	upper part of river bank; good collagen content	2	no	no	–	–	–
Bercy 08	QS. MXX 11-C21; MB 33000	right mandible	–	upper part of river bank; good collagen content	2	no	no	–	–	–
Bercy 09	QS. MXV 21-C21; MB 35000	right mandible	–	upper part of river bank; good collagen content	2	2 of 2	no	157	T/T3	reproducible
Bercy 10	QS. MXV; MB 71000	right mandible	–	upper part of river bank; good collagen content	2	2 of 2	no	117	T2	reproducible
Bercy 11	QS. LXV 9-C21; MB 78000	right mandible	–	lower part of river bank; good collagen content	2	no	no	–	–	–
Bercy 12	QS. LXX 11-C21a; MB 86000	right mandible	–	lower part of river bank; good collagen content	2	no	no	–	–	inhibition present
Bercy 13	QS. MXI 3-C21; MB 89000	right mandible	–	upper part of river bank; good collagen content	2	2 of 2	no	129	T3	reproducible

lab code	archaeological code	skeletal part	date excavated	extra information	no. of extractions	PCR amplification: 157 bp product	PCR amplification: 176 bp product	maximum sequence length (bp)	sequence cluster(s)	reasons for success/failure
West Africa - c. 2200-600 BP (Iron Age)										
Africa 01	Jenné-jeno, Mali - Jj85 LXN L52	bone	–	located under layer radiocarbon dated to 2090 ± 110 bp	2	no	no	–	–	–
Africa 02	Kaniana, Mali - Jj81 KAN L2	bone	–	dating to c. AD 1000-1200	2	no	no	–	–	–
Africa 03	Sincu Bara, Senegal - SBA1 L12	bone	–	dating to c. AD 700-1000	2	no	no	–	–	–
Africa 04a	Jenné-jeno, Mali - Jj77 M1-20	bone	–	dating to c. AD 400-600	2	no	no	–	–	–
Africa 04b	Jenné-jeno, Mali - Jj77 M1-20	bone	–	dating to c. AD 400-600; burnt bone	2	no	no	–	–	–
Africa 05a	Jenné-jeno, Mali - Jj81 LXN L52	bone	–	located under layer radiocarbon dated to 2090 ± 110 bp	2	no	no	–	–	–
Africa 05b	Jenné-jeno, Mali - Jj81 LXN L52	bone	–	located under layer radiocarbon dated to 2090 ± 110 bp	2	no	no	–	–	–
Africa 05c	Jenné-jeno, Mali - Jj81 LXN L52	bone	–	located under layer radiocarbon dated to 2090 ± 110 bp	2	2 of 2	no	117	T1	reproducible
Africa 05d	Jenné-jeno, Mali - Jj81 LXN L52	bone	–	located under layer radiocarbon dated to 2090 ± 110 bp	2	2 of 2	no	157	T1	reproducible
Africa 06	Jenné-jeno, Mali - Jj81 LXN L31	bone	–	dating to c. AD 850-1100	2	2 of 2	no	117	T/T3	reproducible
Africa 07	Jenné-jeno, Mali - Jj77 M1-15	bone	–	dating to c. AD 600-850	2	2 of 2	no	157	T1	reproducible
Africa 08	Jenné-jeno, Mali - Jj77 M1-5	bone	–	dating to c. AD 1100-1400	2	no	no	–	–	–
Africa 09a	Jenné-jeno, Mali - Jj77 M1-8	bone	–	dating to c. AD 1100-1400	2	no	no	–	–	–
Africa 09b	Jenné-jeno, Mali - Jj77 M1-8	bone	–	dating to c. AD 1100-1400	2	no	no	–	–	–
Africa 10	Jenné-jeno, Mali - Jj8 LXN L47	bone	–	situated above a C14 sample dated 1910 ± 110 bp	2	no	no	–	–	–
Africa 11	Jenné-jeno, Mali - Jj81 LXS L48	bone	–	dating to c. AD 1100-1400; burnt bone	2	no	no	–	–	–

Table 2.3 (on previous three pages) Table of archaeological *B. taurus* samples studied, with associated codes, skeletal element used, date excavated and extra information supplied. Site details and dates are given above each set of specimens. Archaeological dates are based on the context of, and evidence from, each excavation. PCR amplification results and maximum mitochondrial region sequence lengths are included. The last two columns indicate the *B. taurus* cluster(s) to which each sequence belongs (see text), and reasons for success or failure. *n/a*, not known or not applicable; *non-reproducible*, only one sequence obtained from several attempts; *contradiction*, two or more sequences obtained, but results different; *reproducible*, at least three reproducible sequences from two different extractions; *nd*, not done.

	16042	16049	16050	16051	16057	16058	16074	16082	16084	16085	16093	16102	16109	16110	16113	16116	16117	16119	16121	16122	16130	16137	16138	16143	16147	assignment to haplogroup *
T/T3 consensus	T	C	C	T	G	C	T	G	C	T	G	G	T	C	T	T	G	T	G	T	T	T	T	A	T	–
T1 consensus	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	–
T2 consensus	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	–
T4 consensus	C	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	–
British aurochs	.	T	.	C	.	T	C	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	–
CH11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T/T3**
TB03	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T**
TB07	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T/T3**
Ardèche 01	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T/T3**
Ardèche 02	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T/T3**
Bercy 09	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T/T3**
Bercy 10	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T2**
Bercy 13	.	.	.	.	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	**T3**
Africa 05c	.	.	T	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	**T1**
Africa 05d	.	.	T	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	**T1**
Africa 06	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T/T3**
Africa 07	.	.	T	.	.	.	C	.	.	C	.	.	C	.	C	.	.	.	.	.	.	.	.	.	.	**T1**
Bos indicus	.	T	.	.	A	T	C	A	T	C	.	A	C	.	C	C	A	C	A	C	C	C	C	*	C	–

Table 2.4 The variable positions in control region sequences of archaeological *B. taurus* samples aligned to the European consensus haplotype (T3). Differences are indicated with a period (.) denoting identity. Sequence codes are given in the first column and only variable sites are shown. The sequence positions from the BOVMT GENBANK sequence are given above each column (accession number V00654; Anderson *et al.* 1982). The common and putatively ancestral Middle Eastern and European consensus sequence is denoted T/T3 (see text), with T1, T2 and T4 denoting the consensus sequences from Africa, the Middle East/Europe, and Japan respectively. The putative ancestral British aurochs sequence is also included (GENBANK accession number AF336746; Troy *et al.* 2001), as well as a representative *B. indicus* sequence (HA2; Bradley *et al.* 1996). All ancient sequences generated as part of this study are shown in bold type. * each sample has been assigned to one of the six main haplogroups by means of its relative position in the median networks.

of these represent different skeletal elements of the same individual, in concordance with the sequence analysis.

In terms of rough categories of taphonomic contexts, when all sites are considered together, open field sites gave only nine reliable products out of 101 samples (9%), whereas the proportion is more than 23% for the waterlogged context of Bercy, and two out of three for the bones from cave contexts.

Figure 2.2 (on following page) Reduced median network of ancient and extant regional groups. *B. taurus* mtDNA reduced median networks constructed from four regional haplotypes groups, including both extant lineages (taken from Troy *et al.* 2001) and the 12 ancient sequences obtained (here shown in white). Inset: the relationships of the five main *B. taurus* haplotypes, T, T1, T2, T3 and T4, to the British aurochs, *B. primigenius*, haplotype P (GENBANK accession number AF336746). As T is differentiated from the reference sequence, T3 (Anderson *et al.* 1982), at position 16,255, and the region sequenced comprises only 117 bp of the D-loop, from 16,042 to 16,158, the T and T3 consensus haplotypes have been reduced together in the network. T1 is defined by transitions at positions 16,050 and 16,113; T2 by a G to C transversion at 16,057; and T4 (a haplogroup so far only found in Japan; Mannen *et al.* 1998) by transitions at 16,042 and 16,093. The spatial arrangement of the skeleton network and the shading codes are preserved in the full data networks (Figure 2.1). The shading indicates which of the four T haplotypes each ancient sample roots to: blue = T ancestral; yellow = T1 African; green = T2 Middle Eastern/European; and red = T3 Middle Eastern/European. The modern data are grouped as originating in Anatolia, the Middle East, mainland Europe or Africa. Circles represent sequence haplotypes, the area being proportional to the frequency of the haplotypes. Points are theoretical intermediate nodes introduced by the median algorithm, and branches between haplotypes represent mutations. The networks were reduced at the following positions; Anatolian sequences: 16,050, 16,057, 16,074, 16,110, 16,113 and 16,138; Middle Eastern sequences: 16,049, 16,050, 16,058, 16,074, 16,085, 16,113, 16,121 and 16,122; mainland European sequences: 16,110 only; and African sequences: 16,049, 16,056, 16,057, 16,084 and 16,122. Fast mutating sites are underlined above and shaded on the skeleton network. Despite the decreased fragment length represented, the spatial arrangements of the haplotypes have been conserved from Troy *et al.* (2001). The archaeological samples (shown in white) were placed in the different networks based on their geographic location; CH11 (Anatolian); TB03 and TB07 (Middle East); Bercy 09, Bercy 10, Bercy 13, Ardèche 01 and Ardèche 02 (Mainland Europe); and Africa 05c, Africa 05d, Africa 06 and Africa 07 (Africa).

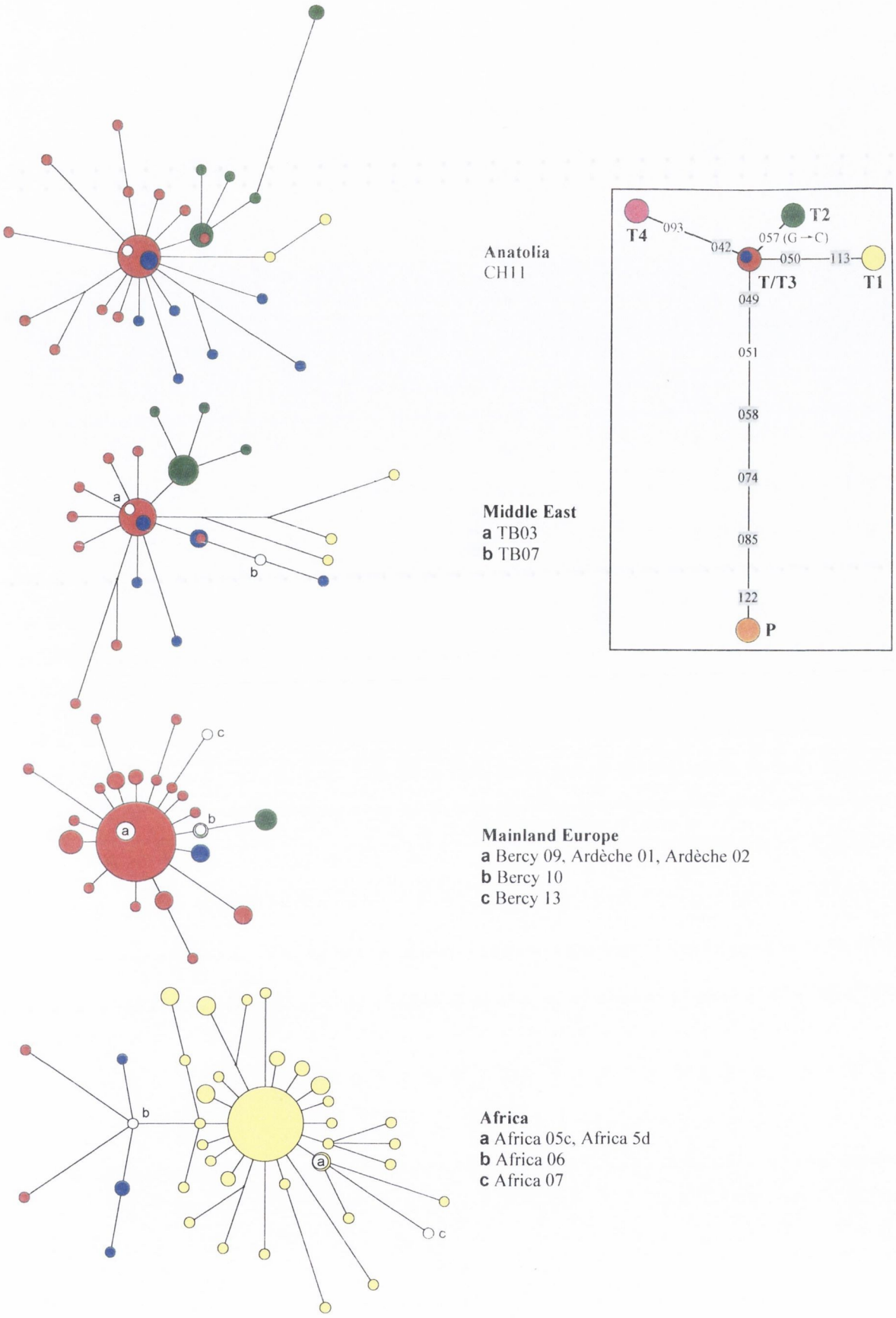


Figure 2.2 Reduced median network of ancient and extant regional groups.

2.4.3 *Phylogenetic results – region by region*

The haplogroups T/T3, T1 and T2 are geographically distributed in extant cattle (**Figure 2.2**). The 12 reproducible sequences obtained from the ancient remains may be diagnosed as belonging to one of these three haplogroups (**Table 2.3**). Of the three Near Eastern sequences, two animals have the ancestral, predominant European and Near Eastern haplotype, T/T3. This group contains 15 Anatolian, 11 Middle Eastern and 54 Mainland European extant animals, as well as animals from additional regions: Britain (25), Western European Fringe (28), and Japan (4). The remaining sample from the Near East coalesces to T, but has not previously been described. Its nearest neighbours all have T or T3 haplotypes.

Of the five reproducible French samples, both of those coming from the two Mediterranean Cardial early Neolithic sites have the same central T/T3 haplotype. Though from the same site (Bercy; middle Neolithic and northern situation), the three other samples gave three different haplotypes: a central T/T3, a previously undescribed one that coalesces to T3, and a T2, which shares a haplotype with four Anatolian and seven Middle Eastern animals, as well as one British and one Western European Fringe animal. As can be seen from the networks (**Figure 2.3**), the T2 haplotype is rare in mainland Europe but, as it does occur at a low frequency, this result is not inconsistent with European variation.

One African sample has the ancestral T/T3 haplotype, which is rare in extant animals from Africa (**Figure 2.3**). The two African samples taken from one individual are T1 and have the same haplotype as two other modern African sequences. The other positive sample is also T1, but this haplotype has not yet been described in the extant cattle studied.

2.4.4 *Genetic diversity analyses*

Linearised F_{ST} values can be used as an estimate of genetic distances between populations over shallow time depths (Slatkin 1995). In this study, these were calculated using published mtDNA sequences from seven geographical groups (Mannen *et al.* 1998; Troy *et al.* 2001), together with the 12 reliable ancient sequences, grouped into three regional populations: ancient Near East, Europe and Africa (**Table 2.5**). Both the ancient Near Eastern and ancient European populations show no discernable difference to the Anatolian, Middle Eastern or British populations, but the associated P-values are high so there is a possibility that these F_{ST} values may have been generated by a chance allocation of haplotypes. When compared to each other, the ancient and extant African populations have

a low pairwise F_{ST} value with a low associated P-value, implying a close affinity. The F_{ST} values between the ancient Near East population and both extant Mainland Europe and extant Africa are highly significant, as is the differentiation between extant Africa and ancient Europe.

The distance matrix was used to construct a neighbor-joining phylogeny, shown in **Figure 2.3**. The primary feature of this tree is the major branch separating the ancient and modern African populations from the others. Both the extant Japanese and the ancient Near Eastern populations have long branches, with the latter branching most closely with the modern Anatolian and Middle Eastern populations. The European samples, ancient and extant, cluster together.

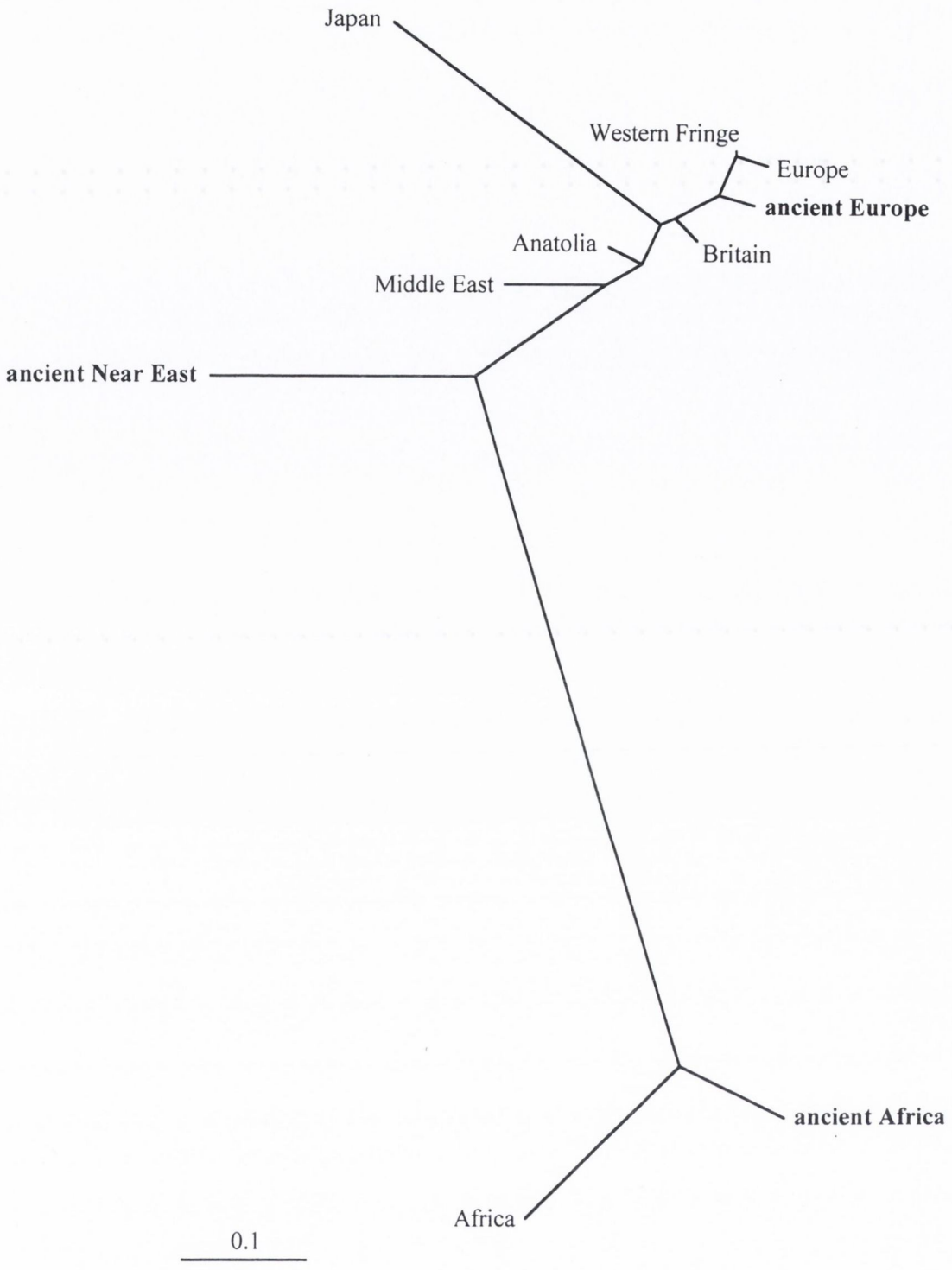


Figure 2.3 A neighbor-joining tree summarising *FST* genetic distances between seven extant cattle groupings and three groups comprising the 12 ancient sequences obtained (see text and Table 2.4). The scale represents units of genetic distance and is derived from the original comparison matrix.

	ancient Near East	ancient Europe	ancient Africa	Anatolia	Middle East	Mainland Europe	Britain	Western European Fringe	Africa	Japan
ancient Near East	-	0.008	0.381	0.852	0.733	0.352	0.792	0.399	< 0.001	0.080
ancient Europe	0.6377	-	0.875	0.933	0.725	0.525	0.720	0.308	< 0.001	0.023
ancient Africa	0.1172	0.0479	-	< 0.001	0.001	< 0.001	< 0.001	< 0.001	0.006	0.002
Anatolia	0.0000	0.0000	0.5427	-	0.749	< 0.001	0.012	< 0.001	< 0.001	< 0.001
Middle East	0.0000	0.0000	0.5388	0.0000	-	< 0.001	0.009	< 0.001	< 0.001	< 0.001
Mainland Europe	0.0337	0.0000	1.5395	0.0319	0.0616	-	0.018	0.031	1.253	< 0.001
Britain	0.0000	0.0000	0.5988	0.0162	0.0219	0.0000	-	0.020	0.788	< 0.001
Western European Fringe	0.0000	0.0218	1.3613	0.0440	0.0839	0.0000	0.0010	-	1.146	< 0.001
Africa	0.6216	0.9837	0.2709	0.7765	0.7179	0.0000	0.0000	0.0000	-	< 0.001
Japan	0.2209	0.2591	1.0127	0.2790	0.3049	0.4436	0.2397	0.3225	1.2098	-

Table 2.5 Population linearised *Fst* values (below diagonal), with P-values after 1023 permutations (above diagonal), calculated using Slatkin (1995). More information is given in the text. The ancient Near East group contains the sequences TB03, TB07 and CH11; the ancient Europe sample includes Ardèche 01 and 02, Bercy 09, Bercy 10 and Bercy 13; and the ancient African group contains Africa 05c, Africa 05d, Africa 06 and Africa 07.

2.5 Discussion

Here, analysis of 101 ancient cattle remains from 13 sites across the Near East, Europe and Africa is reported. Amplification products were generated from 30 samples, but reproducible results were only obtained from 12 of these. The sequences attained lie within the pattern of regional diversity seen in modern cattle.

Ancient DNA results are more readily accepted when they fit into an appropriate context. Here all reproducible sequences obtained fit within the phylogenetic framework of domesticated Near Eastern, European and African *B. taurus*. Also, data are consistent with regional continuity between ancient and modern cattle. Although studies of mtDNA in modern cattle have been informative in tracing the underlying patterns of European cattle domestication, the resolution of *B. taurus* mtDNA haplotypes imposes limitations on their use. The observed differences are not sufficient to differentiate between populations within regions (for example, Mannen *et al.* 1998).

As with modern cattle, there appears to be no definable maternal contribution from Western European *B. primigenius* in the ancient cattle studied here; that is when the sequences are compared to those from six British aurochs (Bailey *et al.* 1996; Troy *et al.* 2001). Of the 12 ancient mtDNA haplotypes obtained here, six correspond to the most common and putatively ancestral Middle Eastern and European haplotype, termed here T/T3. Apart from the three ancient samples from Africa with T1 haplotypes, which only differ from the ancestral African (T1) haplotype by up to three mutations, none of the remaining three ancient sequences differ by more than five transitions from this T/T3 consensus.

Modern *B. taurus* data show a pattern of elevated diversity in the Middle East (with haplogroups T, T2 and T3), a subset of which is transferred to Europe (T3); a different haplogroup (T1) predominates in Africa. This pattern is followed broadly in our ancient samples. Two haplogroups, T and T3, are encountered in the three samples from Çatalhöyük and Tell Brak; four of the five French samples also show T and T3 haplotypes, with the remaining one being T2, all of which are encountered in modern populations. On the other hand, the majority of the West African samples (three of four) fall in the modern haplogroup T1. The affinity of corresponding ancient and modern regional samples is also attested to by the proximity of these within the geographic phylogenies in **Figure 2.2**.

In addition, although still insufficiently supported from a quantitative perspective, data in Anatolia and Europe both fit chronological and geographical trends of the general scheme of domestic diffusion as described by archaeological models (for example, Guilaine 2000). Clusters on the neighbor-joining tree (**Figure 2.3**) are organised starting from ancient Near East toward extant Middle East on the one hand, and extant/ancient Europe on the other. The branch between these two groups bears extant haplotypes of Anatolia, which is thought to have been stocked with domestic cattle at an intermediate date between the Near East and Europe (Peters *et al.* 1999; Vigne & Buitenhuis 1999). In France, the two haplotypes obtained from samples dated to the very beginning of the Neolithic in the Mediterranean area (Baume d'Oullins and Combe Obscure), which is suspected to derive more or less directly from Anatolia and the Near East, are both identical to those of ancient Near East (**Table 2.4**). Conversely, the three haplotypes from Bercy are all of different types, in agreement with the date of this site (later by more than two millennia) and, overall, with the supposed mosaic origin of the Middle Neolithic in the Paris Basin, resulting from both Central and Mediterranean flows (Tresset 2000).

In the present survey, a minority of bones contained verifiable and reproducible DNA. Sporadic amplification in one replicate extraction and not in the other is most likely to be due to the excessive fragmentation and degradation seen in aDNA (Handt *et al.* 1994), particularly in poorly conserved remains, which can lead to jumping PCR (Pääbo *et al.* 1990). Here, poorly amplifying samples often yielded products which each gave a different haplotype. This was most likely due to the presence of either a small number of copies of endogenous and/or contaminating DNA, or to the accumulation of *Taq* polymerase errors following amplification from low-copy number DNA (Richards *et al.* 1995). The patterns of success and failure in these remains from various conditions and ages are typical, and increasing the number of extractions and amplifications could lead to spurious results as indicated by mosaicism, contradictions in sequence data obtained, and contamination problems.

It was possible to generate positive replicable amplification products from 12 of the 101 ancient samples. When a further 15 cattle remains from Çatalhöyük and five bones from Abu Gosh, containing no amplifiable DNA, are added to this total (D.E. MacHugh unpublished data), the overall success rate drops to 9.9%. The success rate varied from 0% in the sites from Israel, to two out of three in the French sites of Ardèche. This latter value is similar to success rates reported by Richards *et al.* (1995), Bailey *et al.* (1996),

MacHugh *et al.* (1999) and Troy *et al.* (2001), in a range of Western European cattle and aurochs specimens dating from 450 to 12,000 BP. It also confirms the good success rate (90 out of 110) obtained from rabbit bones from sites in Spain and Southern France (Hardy *et al.* 1995). The successes generated in the Bercy and African bones, 23.1% and 25.0% respectively, are mirrored in the 22.2% reproducible ancient cattle samples analysed by Turner *et al.* (1998).

Those samples taken from more arid areas gave considerably lower numbers of reproducible amplification, with Çatalhöyük having a success rate of 4.3%, the sites in Israel giving no reliable products at all, and Tell Brak giving 8.7% reproducibility. This disappointing trend has also been noted by Krings *et al.* (1999), where mtDNA amplification in 132 human mummies and skeletons from Egypt, dating to c. 3,000-2,500 BP, gave only two reproducible sequences. Other aDNA analyses undertaken in arid climes have reported more success. Fox (1997) amplified a 109 bp region of mtDNA from 15 out of 28 human Nubian bone and teeth remains from Sudan, dated 2,320-2,130 BP. Although this is a success rate of 53.6%, only seven samples (25%) were analysed twice. Bar-Gal *et al.* (2000) investigated mtDNA in 20 goat bones from Abu Gosh, Israel, dating to 9,500-5,500 BP. Although eight bones were said to have yielded product (40.0%), only a maximum of four (20.0%) would be counted as reproducible using our criteria.

It is generally accepted that dry depositional conditions are the most conducive to survival of aDNA and, on the whole, samples unearthed from arid environments are less promising for archaeological genetics, which is disappointing because Anatolia and the Middle East represent the most interesting areas concerning the origins of domesticated cattle. Kumar *et al.* (2000) reached similar conclusions when looking at human remains from open-air sites in India. Here, all the archaeological contexts from which the samples originated in arid or semi-arid areas studied were open field, and therefore strongly submitted to weather influences. In contrast, cave contexts in Mediterranean or temperate climatic areas seem to be especially favourable for DNA preservation. It would thus be necessary to test deep sedimentary contexts, such as wells or deep pits, in arid zones before being too pessimistic.

CHAPTER 3

Ancient DNA Analysis of European Aurochsen

3.1 Abstract

A total of 45 *Bos primigenius primigenius* bones, from nine European archaeological sites dated to between 3,720 to 12,380 years old, were assessed for the presence of verifiable mitochondrial sequences. It was possible to reproducibly amplify and sequence control region DNA extracted from 24 of the samples, and cytochrome *b* gene DNA from a subset of 19 samples. Comparisons between the aurochs control region sequences and published extant data from Near Eastern *Bos taurus* seems to suggest that the star-like patterns seem around the four main taurine haplotypes, previously explained as domestication-induced population expansion, may be due to a capture of a greater diversity of extinct *B. p. primigenius* individuals during the domestication process. Of the 24 aurochs control region sequences generated, seven (29.2%) exhibited taurine-like haplotypes, which may most plausibly be explained as cross-breeding between early domestic cows and wild aurochs bulls. The single mutation seen in the cytochrome *b* region of both the aurochs samples and representative extant taurine animals did not distinguish between those individuals carrying a taurine-like D-loop haplotype from those with an aurochs-like haplotype.

3.2 Introduction

3.2.1 *The origins of aurochsen*

The earliest form of aurochs are thought to have originated in the region stretching from southern Iran, through Pakistan, eastward to the Indian desert. During the interglacial periods, aurochsen extended their range over wide areas, but retreated south during subsequent glaciation. The latest form of aurochs (*Bos primigenius*) evolved during the Interglacial of the mid-Pleistocene, about 500,000 BC (Guintard 1999), and are thought to be descendents of *B. acutifrons* from Asia (Pilgrim 1947). In the Middle and Upper Pleistocene, *B. primigenius* migrated across Asia into Europe and Africa, and were widespread throughout Eurasia, except for the Himalayas, West China and the colder areas of the North, and over large areas of the Indian and Indo-Chinese subcontinents by the end of the Ice Age. Due to climate changes, post-glacial aurochsen were smaller than earlier aurochsen (Jarman 1969).

3.2.2 *Classification of the aurochs*

Aurochsen used to be called urus, from the Greek word *oros* meaning mountain. It is believed that sub-populations of *B. primigenius* arose during their radiation in the Pleistocene, and it is reasonable to assume that several local subspecies existed. Differences in horn shape and body size have led to the recognition of two species of aurochsen:

1. *Bos primigenius*, with two subtypes:
    i) the European *B. p. primigenius*
    ii) the North African *B. p. opisthonomus*
2. *Bos namadicus*

Analysis of mitochondrial DNA approximates that these two species diverged between 275,000 – 117,000 BP, with separation occurring between the European and African sub-types of *B. primigenius* around 26,000 – 22,000 BP (Bradley *et al.* 1996). Although such groupings have been rejected as too subjective (Epstein & Mason 1984), it is widely accepted that the aurochs was the wild ancestor of all extant cattle (Zeuner 1963; Grigson 1978; Payne 1991), and these two species are thought to have each provided the source population for the two types of modern domestic cattle, with *B. primigenius* giving rise to taurine cattle and *B. namidicus* giving rise to the Asian zebu cattle. Within the former case, there is sufficient evidence, from the disciplines of both archaeology (Clutton-Brock 1989;

Grigson 1991; Wendorf & Schild 1994) and genetics (Bradley *et al.* 1996; MacHugh *et al.* 1997; Troy *et al.* 2001), to suggest that *B. p. primigenius* gave rise to European *B. taurus*, and *B. p. opisthonomus* gave rise to African *B. taurus*.

3.2.3 *The aurochs in Europe*

During the Ice Age, the range of the aurochs expanded and contracted in response to interglacial and glacial stages. For example, the European aurochs (*B. p. primigenius*) is present only sporadically in the archaeological record of Britain from 250,000 BP onwards (Rackham 1994). After the last deglaciation, around 13,000 years ago, *B. p. primigenius* was found over almost the whole continent, apart from northern Scandinavia, northern parts of Russia and Ireland. In contrast with ovicaprids, this wide distribution makes it difficult to clearly assess the domestication process, as it is less obvious that domestic animals were introduced from the Near East (Clutton-Brock 1987). Also, *B. p. primigenius* remains were once considered as two species or sub-species, but this has since been shown merely to be the result of sexual dimorphism, and it is now suggested that *B. p. primigenius* and *B. taurus* form a continuous species (Bohlken 1962; Grigson 1978).

Bull aurochs were 157-200 cm at the withers, with an average height of 180 cm. It is estimated that their weight was between 600 and 1,000 kg. The cows were considerably smaller, averaging 160 cm in height, with a spread from 147-170 cm, and their weight has been estimated at 450-750 kg (Felius 1995). In cave-painting depictions of aurochs from Lascaux in France, and Altamira in Spain, dating to 17,000 and 14,000 years respectively, both the direction of the horns and the colouration of aurochsen is seen to vary. Studies of the paintings at Lascaux led Zeuner (1953) to conclude that aurochs calves were born wheat-coloured, with cows turning reddish-brown with a lighter-coloured saddle, and bulls becoming black with a light dorsal stripe.

As domestic livestock husbandry took hold over Europe, aurochs numbers diminished. Aurochsen were still to be found widely in Roman times but, by the 9th century AD, their range had been severely reduced to eastern Europe and southern Russia. It was here that aurochs herds lasted longest, where they were relatively protected by the introduction of game parks during the 15th century. However by 1557, due to a decrease in forest cover, competition for grazing by domestic animals, and poaching, a herd of approximately only 50 in number remained in the Jaktorow forest near Varsovia, Poland. In 1599, this number

had decreased further to 24 animals, with only four remaining in 1602. In 1620, there was only one cow left, which died in 1627 at the age of 17 (Szafer 1968).

3.2.4 *Genetic studies of aurochs mitochondrial DNA*

The first instance of amplifiable DNA from aurochs was reported by Bailey *et al.* (1996). They generated 214 bp of reproducible control region mtDNA from two out of four British Pleistocene *B. p. primigenius* cave remains, all of which were radiometrically dated to 12,380-10,860 years. The British aurochs sequences were found to cluster more closely with extant taurine samples than with zebu animals and, when a greater number of extant taurine cattle were included in the analysis, the aurochsen grouped more with European than African *B. taurus*. As extant *B. taurus* and *B. indicus* can be crossed to produce viable fertile offspring, it was seen as probable that *B. primigenius* and *B. taurus* would also be able to interbreed. This was seen as being confirmation of the status of *B. p. primigenius* and *B. taurus* as members of a single species, as proposed by Bohlken (1962) and Grigson (1978). The two sequences generated were very similar to each other, with only one base difference over 220 bp, but, due to the small sample size, it was impossible to determine if this represented a frequent Pleistocene lineage now extinct in modern taurine cattle populations. The novel sequences obtained from the Pleistocene samples, in conjunction with the extant data, led to the overall conclusion was that there was a diverse population of aurochsen in the Pleistocene, with the star-burst of lineages centred around the central European and African cattle haplotypes being the result of a more recent expansion superimposed upon the backdrop of a diverse ancient population.

Troy *et al.* (2001) analysed 201 bp of the control region mtDNA in four more British aurochs from cave deposits, this time radiocarbon dated to between 7,570 and 3,720 BP. The aurochs sequences were found to cluster tightly and were again distinct from modern cattle, despite their spatial and temporal dispersal. The difference between the root of the extant *B. taurus* and the putative root of the aurochsen was eight transitions, and one sample was found to have an identical sequence to one generated by Bailey *et al.* (1996).

3.3 Materials & Methods

40 *B. p. primigenius* bones from several locations across Europe (**Table 3.1**, **Figure 3.1**), were assessed for survival of ancient DNA. Apart from six bones, which were only tentatively labelled as aurochs, the other 34 bones were differentiated as wild aurochs, rather than domestic cattle, on the basis of size. A further five samples, D740 (Bailey *et al.* 1996), and TP65, CHWF, NORF and CPC98 (Troy *et al.* 2001), were also re-examined in this study, and a further aurochs sequence, D812 (Bailey *et al.* 1996) was included in the analyses.

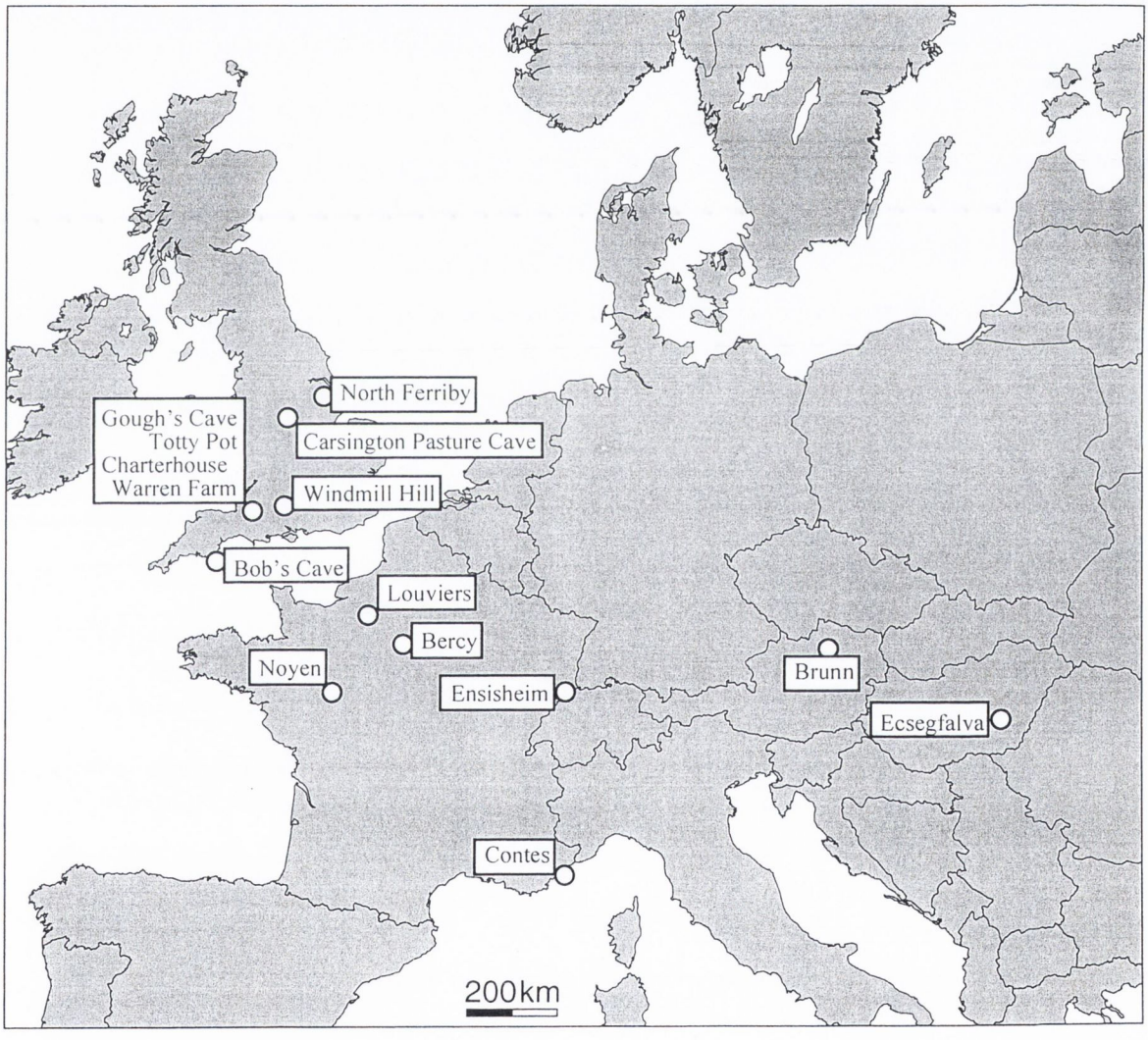


Figure 3.1 Approximate geographic location of aurochsen sampled.

The bone samples were prepared using the procedure modified from Yang *et al.* (1998) described in Section **2.3.2**. At least two DNA extractions were performed for each specimen.

The first region, analysed in all 40 samples, plus the previously analysed D740, TP65, CHWF, NORF and CPC98, was a defined, highly variable region of the mtDNA control region between bases 16,022 and 16,334 (Loftus *et al.* 1994a). Amplification was attempted for both a 157 base pair fragment, $AN2_{FOR}$ (16,022–16,041) – $AN1_{REV}$ (16,178–16,159); and a 176 bp region, $AN1_{FOR}$ (16,159–16,178) – $AN3_{REV}$ (16,334–16,314). In those samples that gave an amplification product with the D-loop primers, a 278 bp region of the cytochrome *b* mitochondrial gene was also targeted, with two pairs of overlapping primers: CY1F (14,643-14,664) – CY2R (14,827-14,808); and CY3F (14,759-14,778) – CY4R (14,920-14,901). The outer cytochrome primers (CY1F-CY4R) were also used to amplify product from three representative individuals from each of the taurine haplotypes (T, T1, T2 and T3), and one example of each of the zebu haplotypes (Z1 and Z2), previously described from amplification of the control region fragment in extant animals by Troy *et al.* (2001) and Magee (2002) respectively. All primers are species-specific so do not amplify human DNA. Ancient PCR set-up was conducted in a laboratory dedicated solely to pre-amplification ancient work, with the extant PCRs completed in the main lab. PCR conditions were as described in MacHugh *et al.* (1999), and the resulting amplification products were sent away for professional sequencing to either Lark (*http://www.lark.com/products/ sequencing.html*) or Oswel (*http://www.oswel.com/code/ en/prod_geno_dna.htm*).

Second-round PCR was not undertaken on any samples that did not amplify in the first-round. All non-amplifiable samples were tested for presence of inhibitors that may have been impeding PCR amplification of endogenous DNA. This involved spiking each negative sample extract with pig DNA (from 6,000 BP Carsington Pasture Cave) in a ratio of 1:4. PCR was continued as previously described in Section **2.3.2** with the pig primers discussed in Section **B.3**. Those samples that yielded a product from the spiked sample were then designated as containing little or no endogenous DNA.

The criteria for authenticating mitochondrial haplotypes were as described in Section **2.3.2**. Both control region and cytochrome *b* gene mtDNA sequences were aligned by eye. A reduced median network was constructed for the control region data, using the median algorithm of Bandelt *et al.* (1995). Two comparison networks were drawn from modern cattle and bison data. The first included modern data from 43 Anatolian and 37 Middle Eastern *B. taurus* cattle (Troy *et al.* 2001), and the second included modern data from 13 wood bison (*Bison b. athabascae*) and 40 plains bison (*B. b. bison*) individuals (Ward *et al.*

1999). These networks include sequence data between positions 16,023 and 16,262 (as denoted in the mitochondrial cattle sequence; Anderson *et al.* 1982) and, as such, only represent a truncated fragment of the control region analysed in the aurochs specimens. Although the cattle sequence length is 240 bp, the bison length between 16,023 and 16,262 is only 233 bp, as this region in bison includes one insertion event (16,057) and two deletion events (16,127-16,133 and 16,143).

Pairwise sequence mismatch distributions were generated according to Rogers & Harpending (1992) and Rogers (1995), using the ARLEQUIN computer program (version 2.000; Schneider *et al.* 1997). Calculations were made from a shortened fragment of the control region, between positions 16,023 and 16,262, so that the aurochs data could be directly compared to that generated from the 43 Anatolian plus 37 Middle Eastern cattle, 13 wood bison and 40 plains bison. It has been suggested that the demographic history of populations is reflected by analysis of frequency distributions of pairwise sequence mismatches in control region sequences. Populations that have recently experienced a rapid expansion tend to exhibit smooth unimodal distributions, due to the retention of newly formed lineages (Bradley *et al.* 1996). The time depth of the expansion can also be approximated from the position of the curve along the horizontal, sequence mismatch axis, with nearness to the vertical axis concurrent with a more recent expansion.

To determine which of the four taurine (T, T1, T2, T3 and T4) and two zebu (Z1 and Z2) cattle haplotypes the main aurochs sequence rooted to, the number of differences between this central aurochs haplotype and 381 taurine (207 European, 79 Near Eastern and 95 African; Troy *et al.* 2001) and 102 zebu (D.A. Magee unpublished data) extant cattle sequences were counted. The total aurochs sequence data considered were truncated to 240 bp so that a comparison could be made between the ancient and extant results.

The average mutational distance from the central sequence in a haplotype, ρ, was calculated from the truncated 240 bp control region data for the aurochs. This involved dividing the number of mutational steps from the central haplotype by the total number of samples in the haplogroup. The average age of the aurochsen was determined, and, using a mutation rate estimate of 38% per million years (Macaulay *et al.* 1999), the ρ value that might be expected in the aurochs population had it expanded up to the modern age was calculated. This value was then compared to ρ values generated from each of the four taurine haplotypes seen in extant cattle (Magee 2002).

3.4 Results

Control region DNA has previously been amplified from six aurochsen by Bailey *et al.* (1996) and Troy *et al.* (2001); however, these data comprised 214 bp and 201 bp non-contiguous regions respectively. Amplification was attempted for two overlapping fragments (157 and 176 bp) of the control region in 45 aurochs bones, leading to a total segment length of 272 bp. These 45 specimens included five of the six aurochsen mentioned above. Sample details are shown in **Table 3.1**. Amplification of the whole 272 bp product was successful in 21 samples, with a further three samples yielding the 157 bp product only (**Table 3.2**), a success rate of 53.3%. These data were used to construct a phylogenetic network (**Figure 3.2**), including the four central taurine haplotypes (T, T1, T2, T3 and T4) described by Troy *et al.* (2001). When the *B. p. primigenius* sequences were compared to *B. taurus*, six of the 24 (25.0%) exhibited taurine-like haplotypes (**Figures 3.2** and **3.3**).

To determine the position of the root between the taurine haplotypes and the central aurochs node, a comparison was made between a representative sequence of the central, and presumably ancestral, aurochs haplotype (TP65) and 483 modern taurine and zebu sequences (**Figure 3.4**). Haplotypes T and T1 contained sequences that were only six mutations away; T3 and T4 exhibited closest neighbours of seven mutations distant; the nearest sequence that had a T2 haplotype was eight mutations away; and the two closest zebu haplotypes of Z1 and Z2 were 21 and 22 mutations distant respectively. When modal values were considered, T and T3 were closest to the ancestral aurochs haplotype, with modal figures of eight mutations different. The most distant modes were Z1 and Z2, each with a figure of 24 mutations. The central T3 sequence was eight bases away, with both T1 and T2 being nine bases different, T4 being distinct by ten bases changes, and the two zebu central haplotypes (Z1 and Z2) being differentiated by 23 bases. From this calculation, it was determined that the aurochs central node should root to the taurine haplotype T, with seven changes seen between these nodes.

Sequences from the *B. p. primigenius* amplified here were compared to a corresponding 272 bp of the control region amplified from five of the six previously analysed aurochsen (Bailey *et al.* 1996; Troy *et al.* 2001), and 202 bp from a Pleistocene aurochs (D812; Bailey *et al.* 1996). Those sequences that resemble these aurochsen are termed as being *aurochs-like* (**Table 3.2**). As can be seen in **Figure 3.2**, all aurochs-like specimens are

lab code	archaeological code	skeletal part	sourced by	extra information	DNA present
Bob's Cave, Britain - 12,380-12,200 BP					
D740*	D740	n/a	Jill Bailey		yes
Gough's Cave, Britain - 12,040-11,760 BP					
D812*	D812	n/a	Jill Bailey		yes
Totty Pot, Britain - 7,570-7,320 BP					
TP65*	TP65	n/a	Andrew Chamberlain		yes
Charterhouse Warren Farm, Britain - 4,090-3,720 BP					
CHWF*	CHWF	n/a	Andrew Chamberlain		yes
North Ferriby, Britain - 3,990-3,720 BP					
NORF*	NORF	n/a	Andrew Chamberlain		yes
Conty, France - n/a					
P7366	P7366 Conty 23	n/a	Jill Bailey		no
P7367	P7367 Conty 27	n/a			no
P7368	P7368 Conty 27	n/a			no
P7369	P7369 Conty 28 #2464	n/a			no
Carsington Pasture Cave, Britain - 6,200-5,650 BP					
CPC98*	CPC98	n/a	Andrew Chamberlain		yes
CPC-01	CPC-98-1026	left pelvis		below 2nd chamber	no
CPC-02	CPC-98-1038	left tibia		below 2nd chamber	yes
CPC-03	CPC-98-1039	left tibia		below 2nd chamber	yes
CPC-04	CPC-98-1046	left radius		below 2nd chamber	yes
CPC-05	CPC-98-1093	right pubis		below 2nd chamber	yes
CPC-06†	CPC-98-2295	left ulna fragment		below 3rd chamber	yes
CPC-07†	CPC-98-2296	left ulna fragment		below 3rd chamber	yes
CPC-08†	CPC-98-2297	left ulna fragment		below 3rd chamber	yes
CPC-09	CPC-98-2299	left pelvis		below 3rd chamber	no
CPC-10	CPC-98-2300	right pelvis		below 3rd chamber	yes
CPC-11	CPC-98-2301	right pelvis		below 3rd chamber	yes
CPC-12	CPC-98-2306	right femur		below 3rd chamber	yes
CPC-13	CPC-98-2312	right centrotarsale		below 3rd chamber	yes
CPC-14	CPC-99	os centrotarsale		north chamber	yes
Windmill Hill, Britain - c. 4,800-4,600 BP					
WH06	9758; context 215; trench B	proximal phalange	Rick Schulting		yes
WH10	1483; context 103; trench A	femur			yes
Noyen/Seine, France - c. 9000 BP					
Noyen 01	NOY 84 XVI E 137 7d 33	bone	Jean-Denis Vigne		mosaic
Noyen 02	NOY 84 16148 10a I	tooth		aurochs based on size only	yes
Ecsegfalva, Hungary - early Neolithic					
H1	301 3089 SQ E2	carpal	László Bartosiewicz	aurochs based on size only	yes
H2	23B 314 4334 F3	distal phalanx		aurochs based on size only	inhibition
H3	23B 327 E7 5002	distal phalanx		aurochs based on size only	yes
H4	301 3792 D4	os sacrum		aurochs based on size only	no
H5	302 3028 01/08/99	os centrotarsale		aurochs based on size only	yes
Brunn, Austria - early Neolithic					
Wien01	16144; A 1994-12 Brunn 3	radius	Erich Pucher		no

lab code	archaeological code	skeletal part	sourced by	extra information	DNA present
Ensisheim, NE France - early Neolithic					
EF01	F.31.2	humerus	C. Jeunesse		no
EF02	F.31.3	talus			yes
Bercy, France - c. 5,000 BP					
Baur 01	B2000	radius	Anne Tresset	Layer LXV-15 c21	inhibition
Baur 02	B2100	metacarpal		Layer MXV	no
Baur 03	B2200	metacarpal		Layer LXII-13 c21	inhibition
Baur 04	B5500	sacrum		Layer MXV-8 c2n4	inhibition
Baur 05	B5600	metatarsal		Layer LXV-24 c21	inhibition
Baur 06	B6200	radius		Layer KXVIII-25	inhibition
Louviers, France - c. 5,000 BP					
Lou 01	Lo3600	phalange	Anne Tresset	Layer c18 34-80	inhibition
Lou 02	Lo3700	phalange		Layer c18 48-81	inhibition
Lou 03	Lo3800	metacarpal		Layer c18 58-81	inhibition
Lou 04	Lo3800	metacarpal		Layer c16 31-82	inhibition

Table 3.1 (on previous two pages) Table of archaeological *B. primigenius* samples studied, with associated codes, skeletal element used and extra information supplied. Site details and dates are given above each set of specimens. Samples signalled with an asterix (*) have previously been analysed by Bailey *et al.* (1996) and Troy *et al.* (2001), and have been radiocarbon dated as part of these analyses. In the case of all other samples, the archaeological dates are based on the context of, and evidence from, each excavation. The final column indicates which samples gave reproducible DNA with the first set of control region primers ($AN2_{FOR}$–$AN1_{REV}$). †, CPC-06, CPC-07 and CPC-08 are fragments from the same individual; *n/a*, not available.

forming a *star-burst* phylogeny, which is usually representative of past population expansions. Modern extant cattle exhibit star-like clusters centred around the four central taurine haplotypes (T, T1, T2, T3 and T4), and this has been suggested as a legacy of domestication expansion (Mannen *et al.* 1998; Troy *et al.* 2001). For comparison, networks were drawn for 221 bp from extant Middle Eastern and Anatolian domestic cattle, as well as wild populations of wood and plains bison (**Figure 3.5**). There does not appear to be any star-burst pattern to the bison network (**Figure 3.5c**), and the aurochs network (**Figure 3.5b**) more closely resembles the patterns seen in cattle (**Figure 3.5a**).

ρ values are useful as a device for dating expansions from founding sequences (Macaulay *et al.* 1999). The ρ value was calculated from the aurochs-like sequences by dividing the number of mutational steps from the central aurochs haplotype by the total number of samples in the haplogroup. This was determined over the control region positions 16,023 to 16,262, so that a direct comparison could be made with the ρ values generated from each of the four taurine haplotypes seen in extant cattle by Magee (2002). This gave a ρ

	16042	16049	16050	16051	16055	16057	16058	16074	16082	16084	16085	16093	16102	16109	16113	16116	16117	16119	16121	16122	16130	16137	16138	16141	16143	16147	16185	16196	16200i	16229	16231	16232	16247	16248	16255	16264	16278	16301	16302	16304	node	geographic location
TP65	T	T	C	C	T	G	T	C	G	C	C	G	G	T	T	T	G	T	G	C	T	T	T	T	A	T	G	G	-	A	T	C	C	C	C	A	C	C	G	T		
D740	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
D812	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	?	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
CHWF	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
NORF	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
CPC98	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	.	.	.	.	.	.	.	.	T	.	.	aurochs-like	Britain
CPC-02	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
CPC-03	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
CPC-04	.	C	.	T	.	.	C	T	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	-	.	C	.	.	.	T	G	T	.	.		T3-like	Britain
CPC-05	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
CPC-06	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	G	.	.	.	.	aurochs-like	Britain
CPC-07	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	G	.	.	.	.	aurochs-like	Britain
CPC-08	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	G	.	.	.	.	aurochs-like	Britain
CPC-10	.	C	.	T	.	.	C	T	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	-	.	C	.	.	.	T	G	.	.	.		T3-like	Britain
CPC-11	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
CPC-12	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
CPC-13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
CPC-14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	aurochs-like	Britain
WH06	.	C	.	T	.	.	C	T	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	-	.	C	.	.	.	T	G	.	.	.	A	T3-like	Britain
WH10	.	C	.	T	.	.	C	T	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	-	.	C	.	.	.	T	G	.	.	.	.	T3-like	Britain
Noyens 02	.	C	.	T	.	.	C	T	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	-	.	C	.	.	.	T	G	.	.	.	A	T3-like	France
H1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	?	?	?	?	?	?	?	?	?	?	aurochs-like	Hungary
H3	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	C	.	.	.	.	.	.	.	.	.	aurochs-like	Hungary
H5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	?	?	?	?	?	?	?	?	?	?	aurochs-like	Hungary
EF02	.	C	.	T	.	.	C	T	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	-	.	?	?	?	?	?	?	?	?	?	?	T3-like	France
T consensus	.	C	.	T	.	.	C	T	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	-	.	C	.	.	.	.	G	.	.	.	.		
T1 consensus	.	C	T	T	.	.	C	T	.	.	T	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	-	.	C	.	.	.	.	G	.	.	.	.		
T2 consensus	.	C	.	T	.	C	C	T	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	-	.	.	C	.	.	.	.	G	.	.	.	.		
T3 consensus	.	C	.	T	.	.	C	T	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	-	.	.	C	.	.	.	T	G	.	.	.	.		
T4 consensus	C	C	.	T	.	.	C	T	.	.	T	A	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	-	.	.	C	.	.	.	T	G	.	.	A	.		
Z1 consensus	.	.	.	T	.	A	.	.	A	.	.	.	A	C	C	C	A	C	A	.	C	C	C	C	-	C	.	A	A	G	C	T	T	T	T	?	?	?	?	?		
Z2 consensus	.	.	.	T	.	A	.	.	A	T	T	.	A	C	C	C	A	C	A	.	C	C	C	.	-	C	.	A	A	G	C	.	T	T	T	?	?	?	?	?	_	_

Table 3.2 The variable positions in control region sequences of archaeological *B. primigenius* samples aligned to the aurochsen consensus haplotype (TP65), for positions 16,042 to 16,313. Sample D812 includes the positions 16,042-16,157 and 16,228-16,313 only (Bailey *et al.* 1996). Differences are indicated with a period (.) denoting identity. Individual archaeological codes are given in the first column, with geographic locations in the final column. The sequence positions from the BOVMT GENBANK sequence are given above each column (accession number V00654; Anderson *et al.* 1982), with *i* denoting an insertion, and only variable sites are shown. The seven haplotypes seen in extant cattle (taurine types T, T1, T2, T3 and T4; and the zebu types Z1 and Z2) are included for comparison. All ancient sequences generated as part of this study are shown in bold type, and each sample has been assigned to be either aurochs-like or T3-like by means of its relative position in the median network.

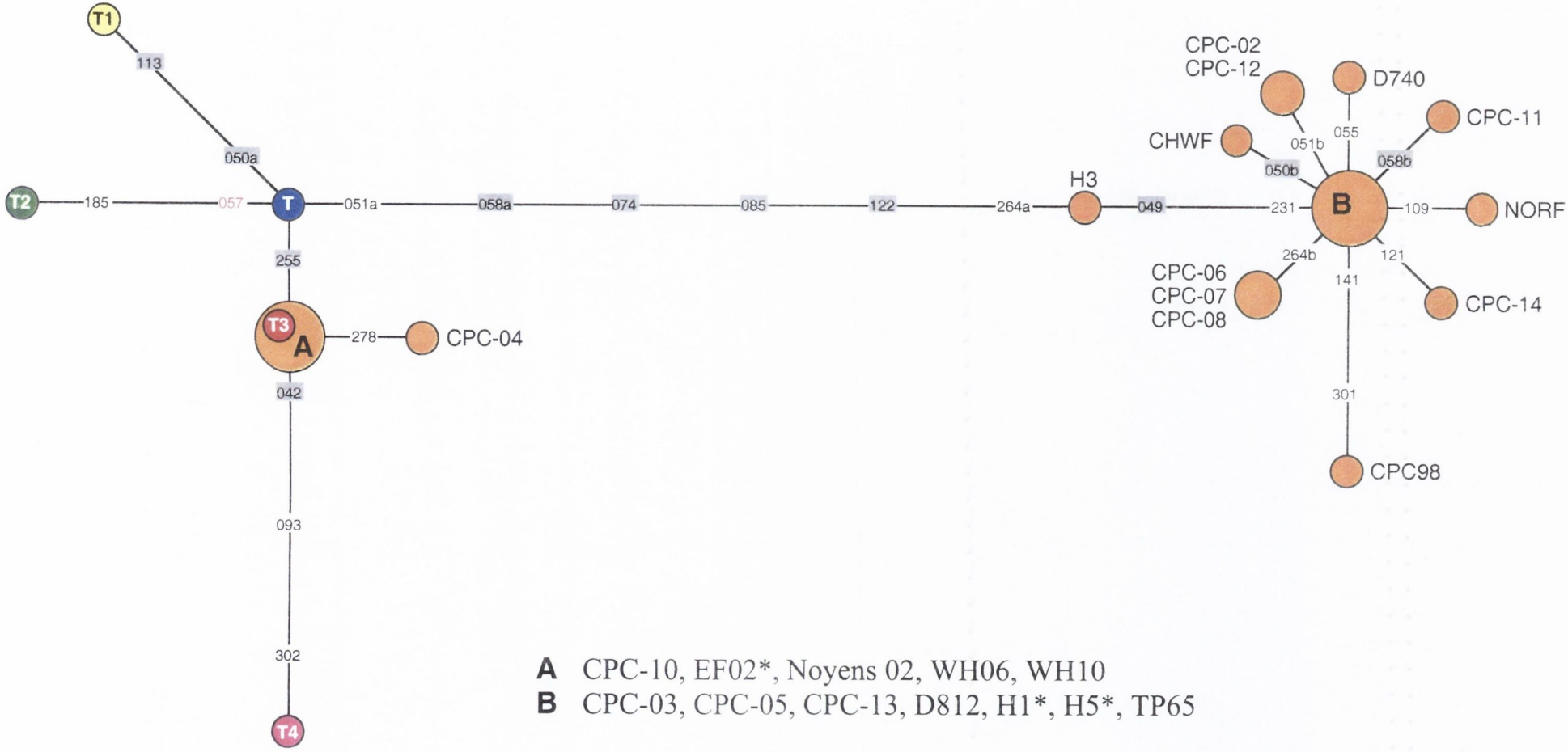


Figure 3.2 Reduced control region network of *B. primigenius* sequences obtained. The network is for positions 16,042 to 16,313 of the control region of the mitochondria; numbers represent the Anderson sequence minus 16,000, with the numbers in red indicating a transversion, and those shaded grey being sites with an increased substitution rate. All aurochs samples are coloured in orange. T = taurine (T = ancestral, T1 = A, T2 = ME+E+J, T3 = ME+E+J, T4 = J). *Samples H1, H5 and EF02 only generated the smaller 157 bp amplicon. D812 includes the positions 16,042-16,157 and 16,228-16,313 only (Bailey *et al.* 1996).

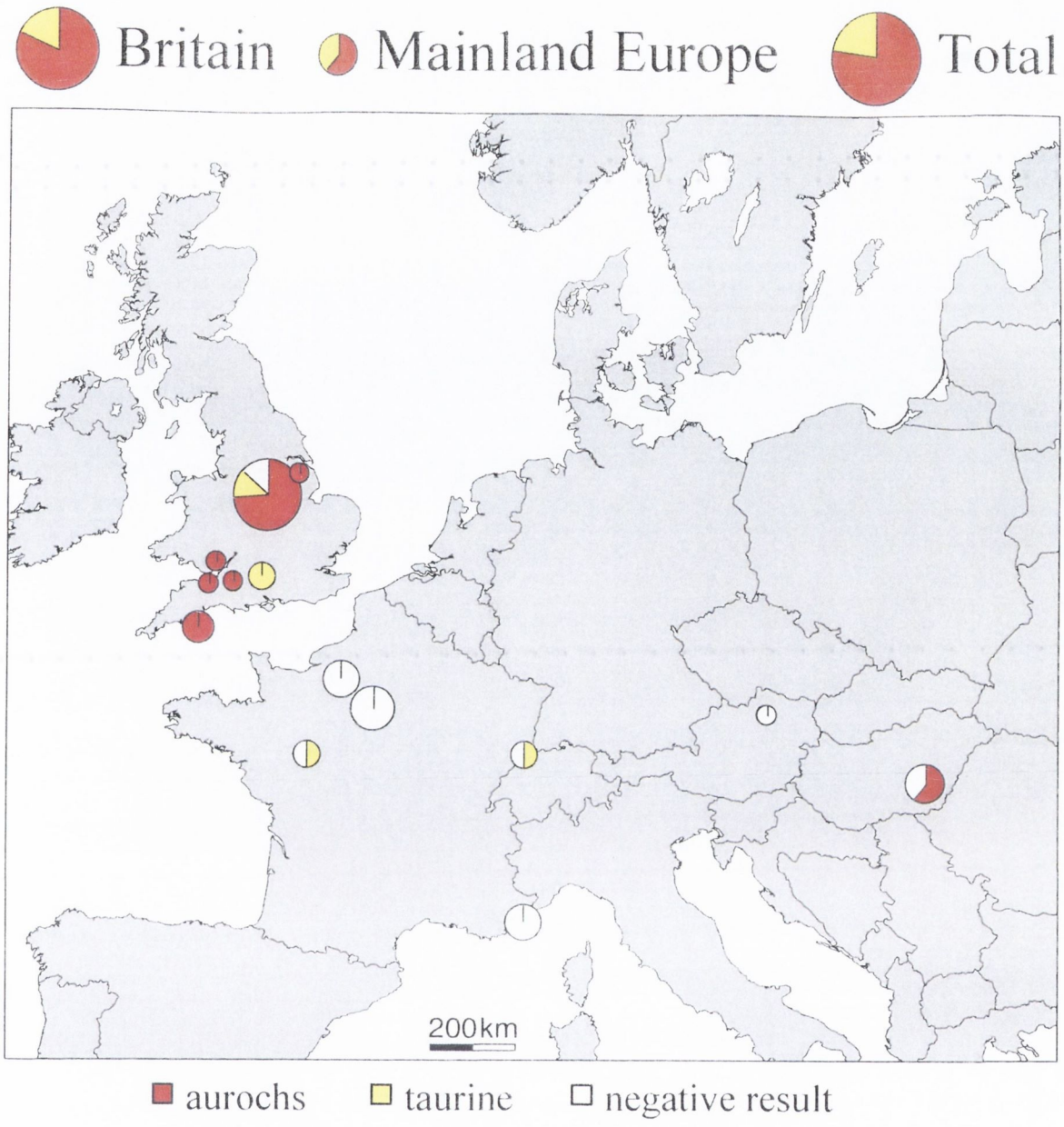


Figure 3.3 Map showing the relative proportions of aurochs and taurine mitotypes in aurochs remains from 14 European archaeological sites.

value for aurochs of 0.526. Using a mutation rate estimate of 38% per million years (Troy *et al.* 2001), it was determined that one mutation would occur every 10,965 years over the comparative 240 bp length of D-loop. The average archaeological age of the 19 aurochsen sequences used was 6,850 BP, and therefore approximately 0.625 mutations would be expected to have occurred over the time of deposition and the present. The ρ value that might be expected in the aurochs population had it expanded up to the modern age was therefore calculated to be 1.151, which is consistent with the values obtained for the taurine haplogroups, namely 1.258 for T, 0.968 for T1, 1.195 for T2, 0.937 for T3 and 0.235 for T4 (Magee 2002).

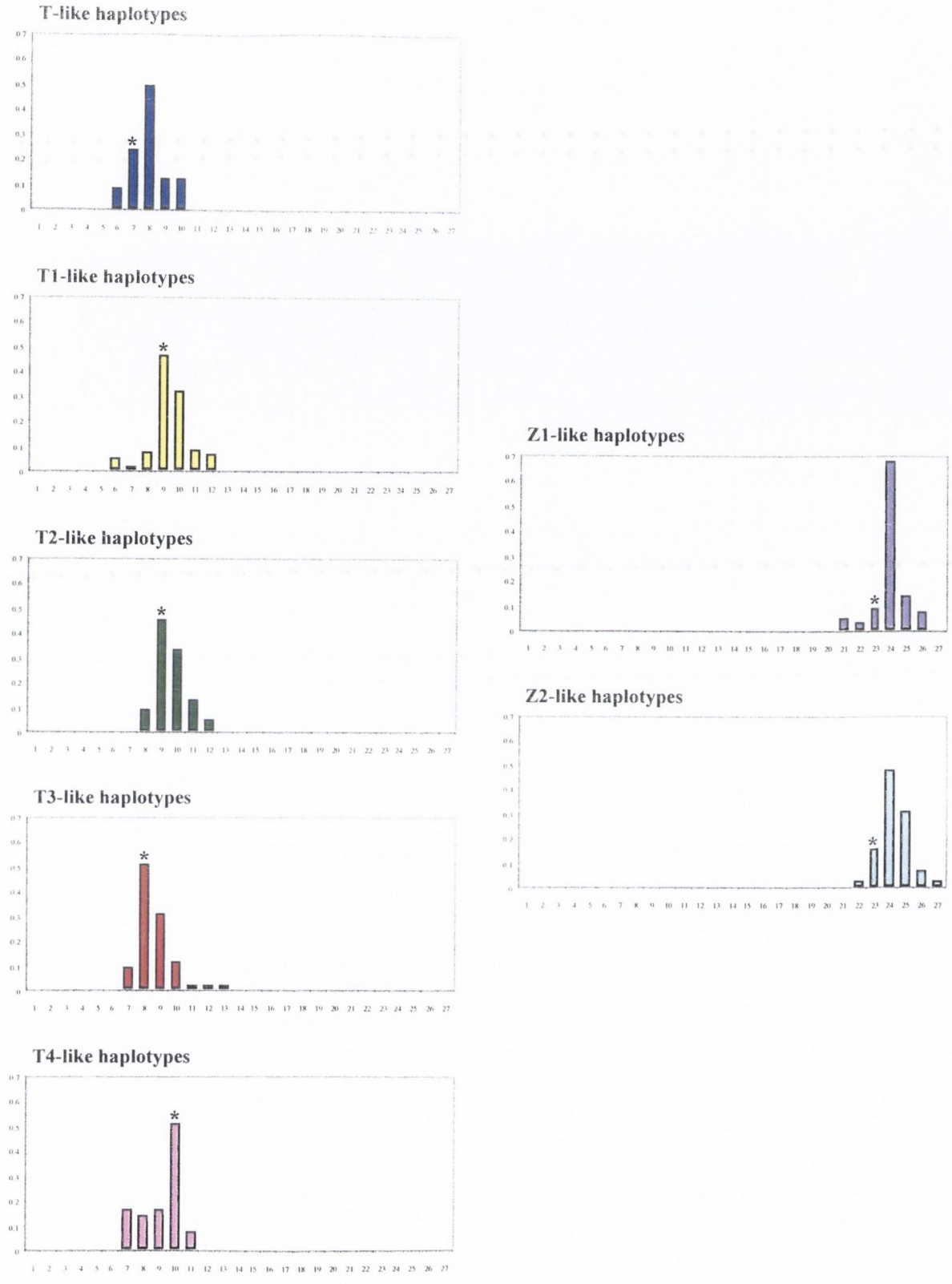


Figure 3.4 Comparison of central aurochs haplotype (TP65) to 381 taurine (T) (Troy *et al.* 2001) and 102 zebu (Z) (D.A. Magee unpublished data) cattle comprising all seven haplotypes seen in extant cattle populations. The horizontal axis shows the number of differences between TP65 and each individual in the seven groups, and the vertical axis is the frequency of samples at each category. The position marked with an asterix (*) on each histogram denotes the position of the central haplotype of each group.

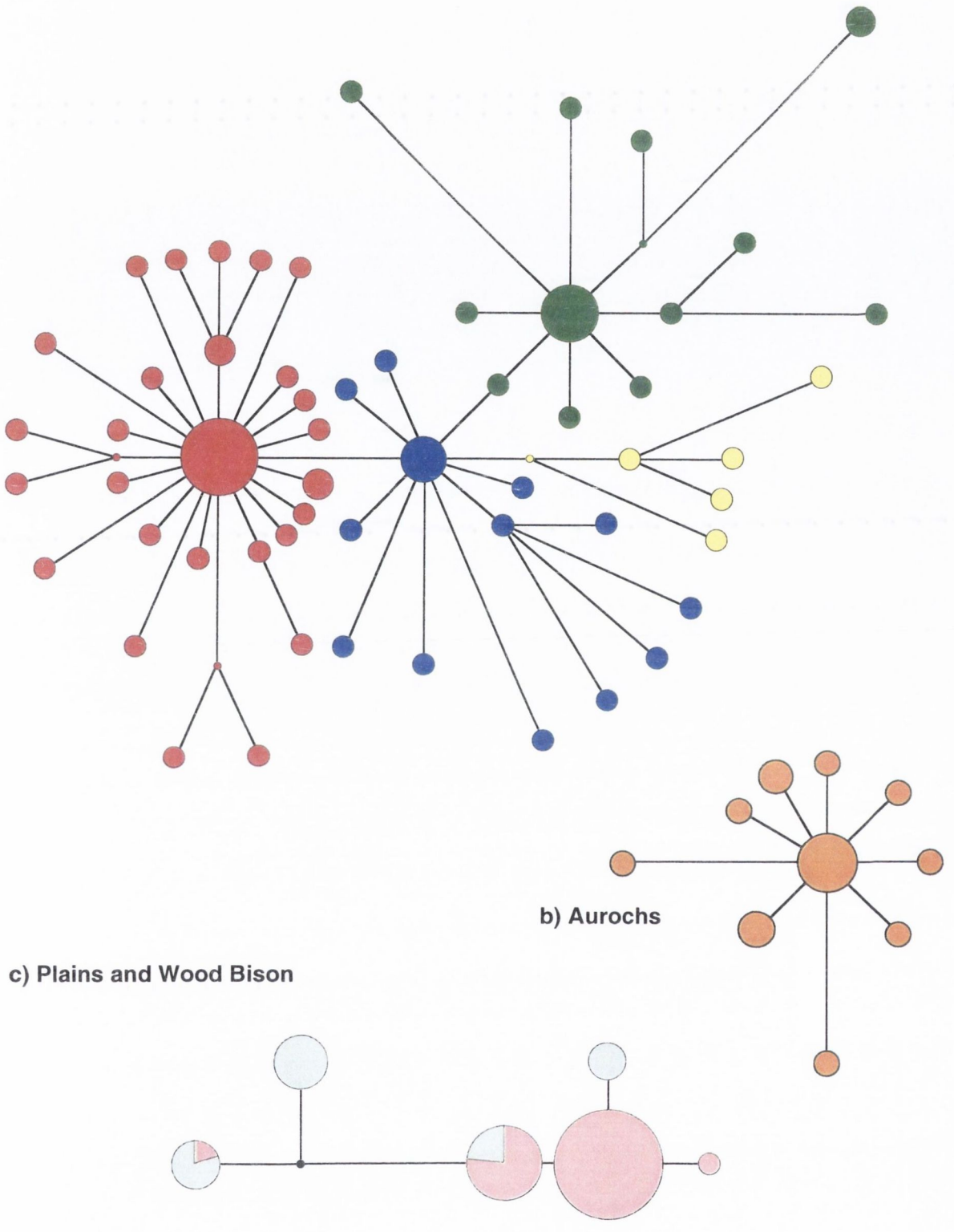


Figure 3.5 Reduced control region networks of cattle and bison for comparison with aurochs network. The network is for positions 16,042 to 16,262 of the control region of the mitochondria.

- **a)** Network of 37 Middle Eastern plus 43 Anatolian cattle, taken from Troy *et al.* (2001). Blue = haplotype T; yellow = T1; green = T2; and red = T3.
- **b)** Network of 19 aurochs (orange).
- **c)** Network of 40 plains bison (pink) and 13 wood bison (light green), taken from Ward *et al.* (1999).

Frequency distributions of pairwise sequence mismatches in control region sequences are thought to reflect demographic population history. Distributions were drawn from 240 bp (positions 16,023 to 16,262) from the 17 aurochs-like sequences that gave full amplification products. 37 Middle Eastern plus 43 Anatolian cattle, 13 wood bison and 40 plains bison sequences were also analysed for comparison (**Figure 3.6**). Both the aurochs and cattle distributions exhibit smooth unimodal distributions, characteristic of a recent rapid expansion. The time depth of the expansion can be approximated from the position of the curve along the horizontal, sequence mismatch axis, and here it appears that the aurochs expansion had occurred only shortly prior to the deposition of the remains. However, conclusions are difficult to draw from this as the aurochs samples differ both spatially, with remains from all across Europe, and temporally, with dates ranging from the Pleistocene to the early Neolithic.

Aurochs specimens that generated the complete 272 bp section of the control region were amplified with primers targeting the cytochrome *b* gene (**Table 3.3**). Discounting the zebu sequences added as reference, there is only one base change seen over this 236 bp region, a synonymous third-base position transition at 14,873. Unfortunately, this change does not distinguish between those individuals that have a taurine-like D-loop haplotype and those with an aurochs-like D-loop haplotype.

Taurine-like haplotypes are seen over a range of dates across Europe, from the 9,000 BP site of Noyens in France, through the two early Neolithic mainland European sites of Ecsegfalva and Ensisheim, to the two British sites of Carsington Pasture Cave, dated to around 6,000 BP, and Windmill Hill, dated to 4,750 BP. Over all the *B. p. primigenius* sequences, six of the 25 (24.0%) exhibited taurine-like haplotypes. From the cytochrome *b* data, plus that included in **Table 3.2** from the control region, the mean number of individuals sampled from Carsington Pasture Cave is 11, with nine aurochs-like and two taurine-like individuals when the control region is considered. Although these samples can not strictly be deemed to be a population due to possible temporal differences, the CPC group exhibit a taurine haplotype percentage of 18.2%.

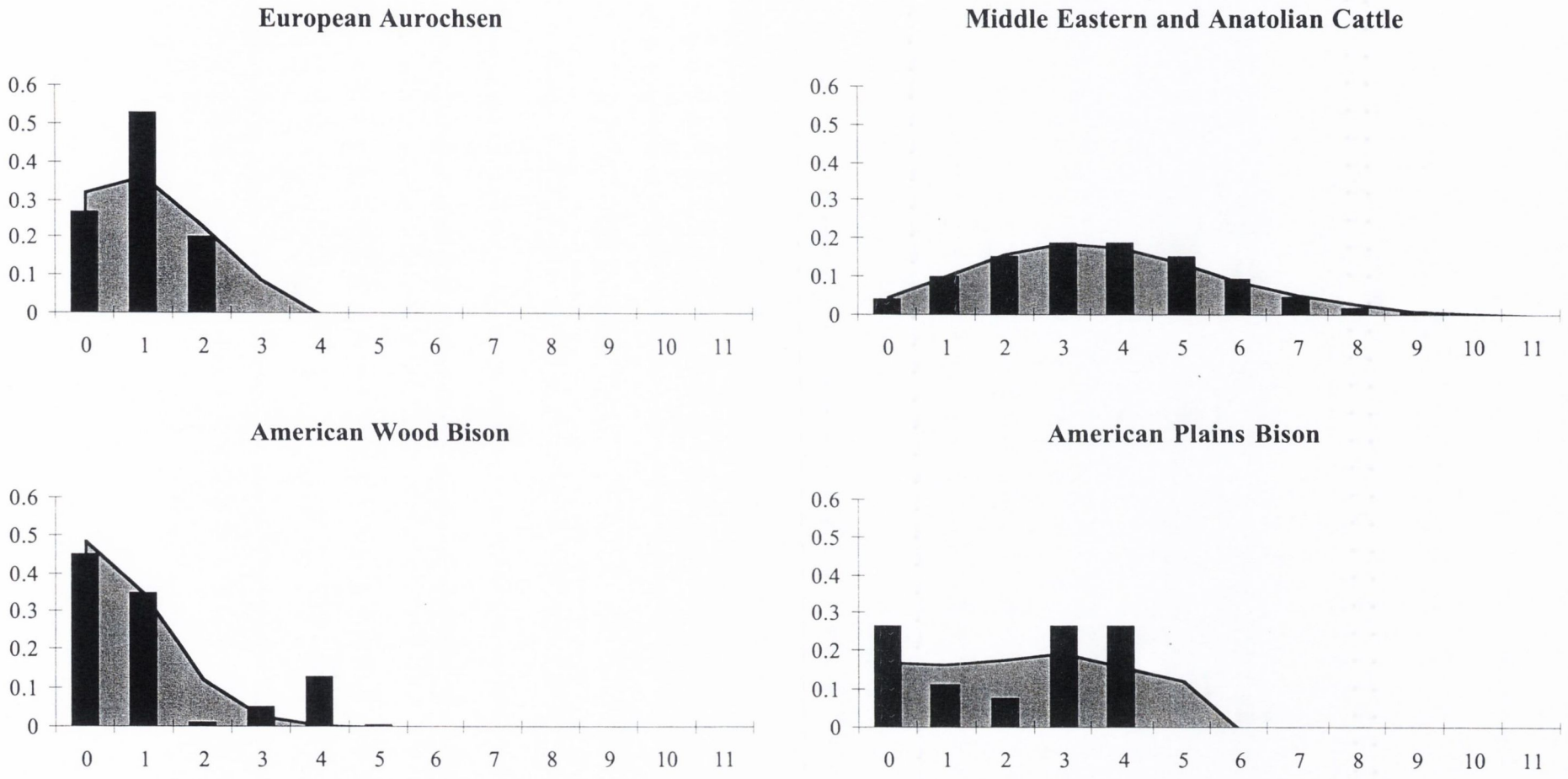


Figure 3.6 Pairwise differences graphs for 17 aurochsen, 37 Middle Eastern plus 43 Anatolian cattle, 13 wood bison and 40 plains bison sequences over the D-loop region from 16,023 to 16,262. The black bars represent observed differences, with the grey areas showing the simulated data generated by the ARLEQUIN algorithm. The *x*-axes show the number of differences seen between sequences, and the *y*-axes show the frequency with which this occurs.

	14825	14858	14873	14897
TP65	**C**	**T**	**G**	**T**
D740	.	.	**A**	.
CHWF	.	.	**A**	.
NORF	.	.	**A**	.
CPC98	.	.	**A**	.
CPC-02	.	.	.	.
CPC-03	.	.	**A**	.
CPC-04*	.	.	.	.
CPC-05	.	.	.	.
CPC-06	.	.	**A**	.
CPC-07	.	.	**A**	.
CPC-08	.	.	**A**	.
CPC-10*	.	.	.	.
CPC-11	.	.	.	.
CPC-12	.	.	.	.
CPC-13	.	.	.	.
CPC-14	.	.	**A**	.
WH06*	.	.	.	.
WH10*	.	.	.	.
T consensus	.	.	.	.
T1 consensus	.	.	.	.
T2 consensus	.	.	.	.
T3 consensus	.	.	.	.
Z1	T	C	.	C
Z2	T	C	.	C

Table 3.3 The variable positions in cytochrome *b* sequences of archaeological *B. primigenius* samples aligned to the aurochsen consensus haplotype (TP65), for positions 14,665 to 14,900. Individual archaeological codes are given in the first column and only variable sites are shown. The sequence positions from the BOVMT GENBANK sequence are given above each column (accession number V00654; Anderson *et al.* 1982). Three individuals each were sequenced from three of the four taurine haplotypes seen in extant cattle (T, T1, T2 and T3); and one representative of each of the zebu haplotypes (Z1 and Z2) was also sequenced for comparison. All ancient sequences generated as part of this study are shown in bold type, and all mutations are synonymous third position changes. Ancient sequences that had T-like control region sequences are denoted by an asterix (*).

Osteological data

To investigate the origins of two of the taurine-like sequences seen in the *B. p. primigenius* samples, measurement data from Carsington Pasture Cave were consulted. Osteological measurements of the Carsington Pasture Cave remains were undertaken by Paxton (1999) using the osteometric definitions given in Von den Driesch (1976). **Table 3.4** shows a summary of the metrical data from five of the CPC samples. Measurements were not obtained on seven CPC specimens: four were too fragmentary (CPC-01, CPC-02, CPC-05 and CPC-10); CPC-12 was a juvenile bone with unfused epiphyses; and no data were available for either CPC-14 or CPC98. The five measured CPC specimens have been

compared to measurements of aurochsen from Holocene sites in Denmark, France and Switzerland (Degerbøl & Fredskild 1970; Chaix & Arbogast 1999), and cattle from Neolithic sites in Denmark and Britain (Grigson 1965; Higham & Message 1969; Degerbøl & Fredskild 1970; Harcourt 1971). As can be seen in **Table 3.4**, all specimens are comparable in size with specimens confirmed metrically as *B. primigenius*, and it is clear that they are outside the measurements for prehistoric *B. taurus*.

				measurements (in millimetres)		
lab code	skeletal part	type of measurement or reason for non-measurement	CPC measurement	male aurochs (n)	female aurochs (n)	male + female cattle (n)
CPC-03	left tibia	distal breadth	86	82 - 93 (19)	68 - 83 (9)	43 - 76 (50)
CPC-04*	left radius	distal breadth	> 93	98 -114 (29)	78 - 93 (5)	57 - 88 (95)
CPC-06	left ulna fragments	olecranon length	> 127	not recorded	not recorded	98 -103 (3)
CPC-07	left ulna fragment	olecranon length	> 127	not recorded	not recorded	98 -103 (3)
CPC-08	left ulna fragment	olecranon length	> 127	not recorded	not recorded	98 -103 (3)
CPC-09	left pelvis	acetabulum length	97	not recorded	not recorded	56 - 70 (9)
CPC-11	right pelvis	acetabulum length	> 95	not recorded	not recorded	56 - 70 (9)
CPC-13	right centrotarsale	greatest breadth	75	78 - 88 (6)	61 - 67 (2)	51 - 62 (11)

Table 3.4 Measurements of *B. primigenius* specimens from Carsington Pasture Cave. All osteological measurements of CPC bones were taken by Paxton (1999) using the osteometric definitions given in Von den Driesch (1976). Values are in millimetres and are rounded to the nearest integer. Measurements were not obtained on seven CPC specimens: four of the specimens (CPC-01, CPC-02, CPC-05 and CPC-10*) were too fragmentary; CPC-12 was a juvenile bone with unfused epiphyses; and no data were available for CPC-14 or CPC98. However, all five of these specimens are described by Paxton (1999) as being comparable in size with specimens confirmed metrically as *B. primigenius*. The aurochs specimen that gave T-like mitotypes is denoted by an asterix (*). Measurements of aurochs are of males and female skeletons from Holocene sites in Denmark, France and Switzerland, published by Degerbøl & Fredskild (1970) and Chaix & Arbogast (1999). Measurements of cattle are from pooled-sex samples from Neolithic sites in Denmark and Britain (Grigson 1965; Higham & Message 1969; Degerbøl & Fredskild 1970; Harcourt 1971), apart from the olecranon lengths, which were measured from modern cattle. Data for aurochs and cattle are presented as ranges (minimum to maximum) with the sample sizes given in parentheses. Table adapted from A. Chamberlain.

Of the six bones that appear to have a taurine-like D-loop mitotype (**Table 3.2**), measurements were only available for CPC-04 (**Table 3.5**). CPC-04 is larger than the reference specimens held by the University of Sheffield (Paxton 1999), and is at the top range of variation for female *B. primigenius* (Degerbøl & Fredskild 1970). It is also larger than 20 Neolithic cattle remains from Windmill Hill (Grigson 1965), implying that the sample is either a female aurochs, or an exceptionally large male prehistoric *B. taurus*. However, it could still be the case that this animal represents a modern improved breed male *B. taurus* that has intruded into the archaeological assemblage. CPC-10 is substantially larger than the domestic cattle measured from Windmill Hill, but this survey only included two pelvis specimens, and both of these were from female *B. taurus*. It has

not been measured due to the fragmentary nature of the specimen, but is similar in size to another pelvis sample, CPC-09. Unfortunately, CPC-09 did not give reliable amplification, but its measurements are clearly greater than those seen in cattle. So, although it is still possible that CPC-10 originates from a large male *B. taurus*, this is less likely. Although, during the first part of the Neolithic it is difficult to differentiate between domesticated male cattle and female wild aurochs (Grigson 1978), the question of the origins of both of these bones, plus the other four remains that have taurine-like haplotypes, could be resolved by radiocarbon dating the remains and also by DNA sexing as, if female, then they would be more likely to be aurochsen.

taxon/specimen	period	site/country	size (mm)	no.	reference
radius - distal breadth					
CPC-04	**?Neolithic/Bronze Age**	**Carsington Pasture Cave, U.K.**	**>93**	**1**	**Paxton (1999)**
female aurochs	Neolithic	Denmark	78 - 93	6	Degerbøl & Fredskild (1970)
male aurochs	Neolithic	Denmark	100 - 114	25	Degerbøl & Fredskild (1970)
male and female cattle	Neolithic	Denmark	72 - 87	4	Degerbøl & Fredskild (1970)
male and female cattle	Neolithic	Windmill Hill, U.K.	70 - 85	11	Grigson (1965)
male and female cattle	Late Neolithic	Troldebjerg, Denmark	65.5 - 88.0	40	Higham & Message (1969)
male and female cattle	Late Neolithic	Durrington Walls, U.K.	57 - 78	37	Harcourt (1971)
male and female cattle	?Neolithic/Bronze Age	Carsington Pasture Cave, U.K.	58 - 69	3	Paxton (1999)
male and female cattle	Bronze Age	Bovenkarspel, Holland	51.1 - 70.8	33	Ijzereef (1981)
female cattle	Modern	Denmark	79.0 - 92.1	32	Higham & Message (1969)
male cattle	Modern	Denmark	91.5 - 93.6	2	Higham & Message (1969)
pelvis - length of acetabulum					
CPC-09*	**?Neolithic/Bronze Age**	**Carsington Pasture Cave, U.K.**	**97**	**1**	**Paxton (1999)**
male and female cattle	Early Neolithic	Windmill Hill, U.K.	56 - 70	7	Grigson (1965)
male and female cattle	?Neolithic/Bronze Age	Carsington Pasture Cave, U.K.	65 - 66	2	Paxton (1999)

Table 3.5 Osteological measurements from CPC-04, one of the *B. primigenius* specimens from Carsington Pasture Cave that exhibits a cattle-like haplotype. CPC-10 also exhibited a cattle-like haplotype but the bone was too fragmentary for measurements to be taken. Therefore CPC-09* was used as a comparison as it is similar in size to CPC-10. All osteological measurements of CPC bones were taken by Paxton (1999) using the osteometric definitions given in Von den Driesch (1976). Values are in millimetres and are rounded to the nearest integer. Data are presented as ranges (minimum to maximum). Table adapted from A. Chamberlain.

3.5 Discussion

Of the 24 control region sequences obtained from the 45 aurochs remains sampled, 17 were described as *aurochs-like* when compared to data from two British Pleistocene aurochs (**Table 3.2**; **Figure 3.2**). Previously only described in aurochs from Britain, the aurochs-like sequences are seen in three samples from Hungary, supporting the presence of the aurochs haplogroup across Europe in the Neolithic. The diversity seen in the aurochs shows some similarity to that present in modern taurine cattle (**Figure 3.6**). The star-burst structure of the aurochs sampled resembles the patterns seen in extant domestic cattle more closely than extant wild populations of bison (**Figure 3.5**). Also, when the ρ value (the average number of mutations from founding haplotype) calculated from these data was added to an estimate of the number of mutation events occurring between the deposition of the samples and the modern day, the resulting figure was consistent with the values obtained for the expansions seen around each of the four taurine haplogroups. These analyses raise the question that the star-burst patterns within haplotypes sampled in modern cattle may not solely be a result of the domestication process, but could incorporate sampling of the diversity seen in the extinct *B. primigenius* population that was domesticated in the Near East. A plausible explanation for this *B. p. primigenius* expansion is the climatic change of the end of the last glacial stage.

Several diverse haplotypes may have been present in aurochsen prior to the last glaciation, encompassing both the *B. p. primigenius* sampled here, the taurine haplotypes captured by domestication, and possibly other, as yet unsampled, haplotypes. The aurochs variants in the Near East would not have been affected by the ice ages of the Quaternary (2.4 million years ago to the present), but the size and distribution of the aurochsen in Northern Europe would have been influenced by the onset of glacial events. As populations were forced into different glacial refugia, followed by separate post-glacial colonisation routes, a situation that must have occurred repeatedly, genetic subdivision may have occurred (Hewitt 2000). Even considering one refugium only, populations at the northern limits of the refugial range would have expanded into large uncolonised areas quickly. The main refugial areas of Europe were Iberia, Italy, the Balkans and Caucasus, each of which would have contributed differently to the repopulation of northern parts. Taberlet *et al.* (1998) analysed the phylogeographies of ten taxa and concluded that the northern regions were colonised from Iberic and Balkanic refugia, with the Italian lineages generally isolated due to the Alpine barrier. Four main hybrid zones, termed *suture zones* by Remington (1968), have

been identified by comparison of these colonisation routes, formed when expanding populations from different refugia met (Taberlet *et al.* 1998). As the last glaciation occurred around 20,000 years ago, the lineages seen in the Pleistocene and early Neolithic aurochs specimens are those that survived many ice ages.

There are two hypotheses for the appearance of six modern taurine-like haplotypes in remains diagnosed as wild aurochs on the basis of osteological measurement (**Table 3.2**; **Figures 3.2** and **3.4**). Firstly, the aurochs-like haplotype, present in 19 of the samples analysed here, would have been present in Europe prior to the Neolithic, as the two Pleistocene aurochs remains, D740 and D812, have these sequences (**Table 3.2**). It is probable that these aurochs populations recolonised Europe from a refugia present in either the Balkans or Iberia. This aurochs-like haplotype is seen across Europe and exhibits a star-burst pattern indicative of a post-glacial expansion, but was not sampled as part of the Near Eastern domestication process as it is absent in modern cattle (Troy *et al.* 2001). In the situation under consideration, all T-like haplotypes would have originated from domestication and therefore the presence of taurine haplotypes in remains classified as aurochs would be most likely due to local introgression of domestic cattle into wild aurochs populations. This hypothesis is represented graphically in **Figures 3.7a** and **3.7b**.

It is possible that cross-breeding was occurring relatively often under the conditions of the time; a similar crossing of the wild boar and domesticated pig occurs sometimes even today. The female *B. taurus* may have escaped from captivity and been incorporated into *B. primigenius* herds, whereas escaping male *B. taurus* would have been unlikely to have competed with aurochs males for mating opportunities in the wild (A. Chamberlain personal communication). On the other hand, calves resulting from a cross between domestic cows kept in open locations and wild male aurochs may have been released into the wild, as the selection aim would have been for small-bodied animals (Bodø *et al.* 1996), and the calves would most probably have become unmanageable after a few weeks (Uerpmann 1996).

Five of the taurine-like haplotypes exhibited by the aurochs sampled here (**Table 3.2**; **Figure 3.2**) share the central T3 haplotype, with the sixth taurine like individual (CPC04) one mutation away from T3. The presence of T3-like, rather than T-like, haplotypes supports the hypothesis of local introgression. The six taurine haplotypes sequenced do not

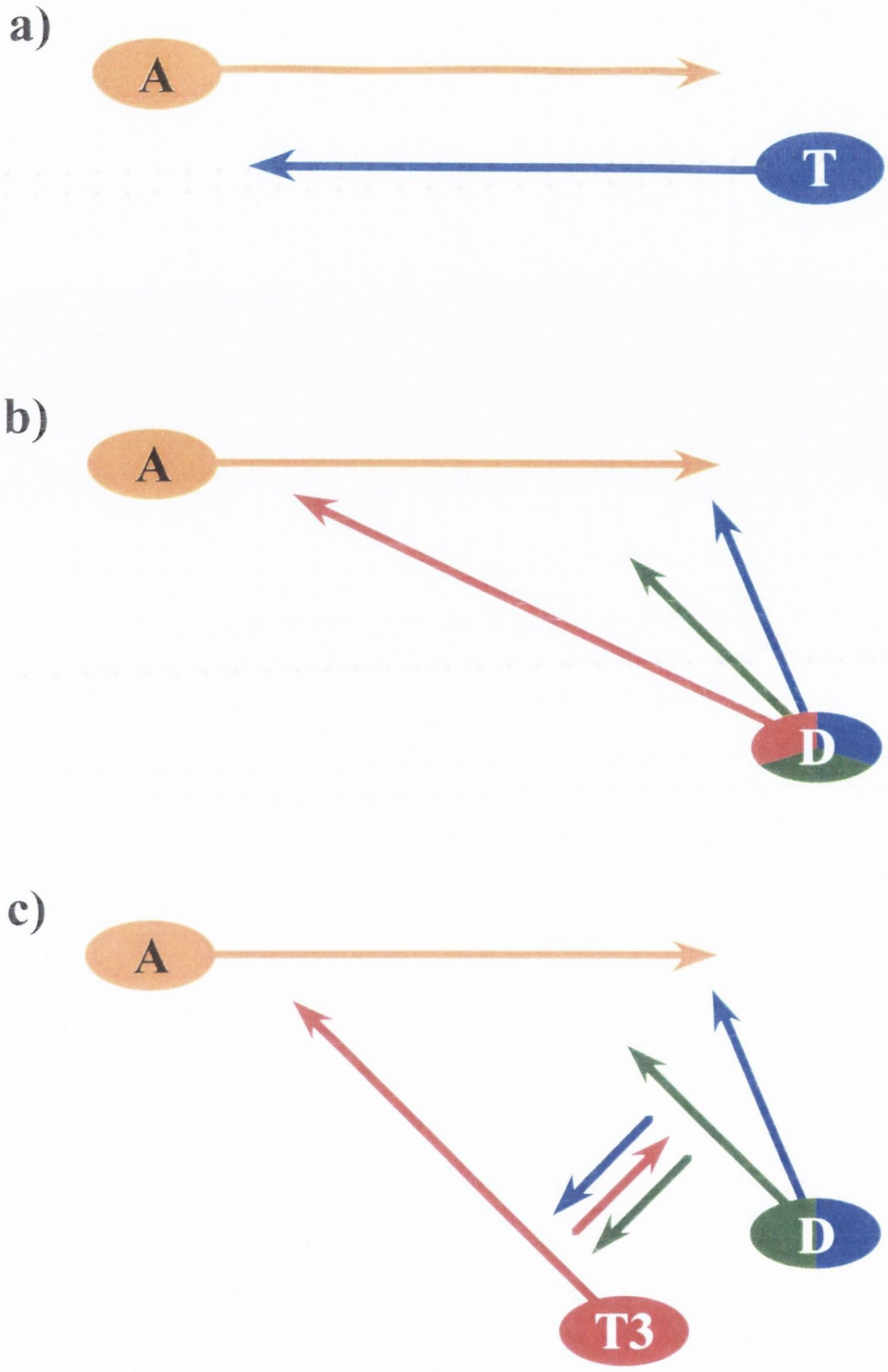


Figure 3.7 Hypothetical movements of aurochs and cattle mitotypes. Aurochs-like (A) haplotypes are coloured orange; cattle haplotypes are as follows: T = blue, T2 = green, T3 = red. D denotes a domestic population. **a)** Hypothetical movements of aurochs populations containing aurochs-like or T-like haplotypes pre-domestication. The A haplotype may have moved out of a refugia in either Iberia or the Balkans, and the T haplotype would have migrated from the Near East. This may have been followed by **b)**, where movements of domesticated T, T2 and T3 domestic cattle migrating from the domestication centre in the Near East would have mixed with A-like aurochs on route. This scenario would possibly lead to introgression of the T3 haplotype into wild aurochsen populations. Alternatively, **c)** may have occurred, where aurochs carrying the T3 haplotype recolonised Europe from a refugia in the Balkans, leading to mixture between A and T3 aurochsen, followed by migration of domesticated T and T2 haplotypes from the Near East. In this scenario, T3-carrying aurochs must have been domesticated in secondary events as the T and T2 cattle moved across Europe, which is one way of explaining the preponderance of the T3 haplotype in modern European cattle.

represent a random sampling of a wild T-like population, which might be expected to be similar to that existing in the Near East, but rather are all T3 or T3-related; a point that is consistent with an origin from the domestic haplotypes captured in the Near East. Also, the majority of sites containing taurine-like aurochs (Ensisheim in France, and Carsington Pasture Cave and Windmill Hill in Britain) have been dated to the early Neolithic. However, care must be taken in the absence of radiocarbon dates for these remains, especially as the sample from Noyens was sourced from a site dating to 9,000 BP.

The second scenario argues the presence of an aurochs-like haplotype in the Iberian refugium and a T3-like refugium in the Balkans, both present in pre-Neolithic aurochs populations. The Near East would have contained diverse taurine haplotypes, of which T and T2 were domesticated most successfully. The incidence of the T3 haplotype in modern European cattle would then have arisen due to introgression of T3-carrying aurochs into the domestic cattle populations as they spread westwards out of the Near East. This is represented graphically in **Figure 3.7c**. Support for this hypothesis lies in the fact that T3 is the most common haplotype in Europe (**Figure 2.1**; Troy *et al.* 2001), and is present in appreciable frequencies in both the Near East and Britain. Also, the T3 haplotype is found in 24.0% of the 25 aurochs analysed. However, it is unlikely that T3 would have also become a major Near Eastern domesticated sequence, especially as this haplotype is found in modern cattle from Iraq and East Asia. Also, there is no clinal decrease associated with the T3 haplotype as sampling moves eastwards through Anatolia toward Iraq (Magee 2002), which is usually indicative of a founding effect.

On balance, the first hypothesis stated is more probable, with the presence of taurine mitotypes in remains classified as morphologically aurochs explained by local introgression of domestic cows into the aurochs gene pool. However, before concrete conclusions can be stated, more extensive samplings of aurochs remains are required from a wider geographic range in Europe, as well as from sites dating from the Pleistocene to the Neolithic.

CHAPTER 4

Y-specific Microsatellite Polymorphism Analyses

4.1 Abstract

Several dinucleotide $(CA)_n$ microsatellite repeat arrays have previously been assigned to the bovine Y-chromosome, with one marker (*INRA124*) shown to be polymorphic. In this new study, a panel of four Y-specific microsatellite markers was assessed for polymorphism in a range of cattle and related species is described here. It was possible to amplify all the markers in the animals sampled and all showed variation. Three of the microsatellite loci (*INRA124*, *INRA189* and *BM861*) displayed putative taurine- and zebu-specific alleles, which can be useful indicators of male-mediated gene flow in hybrid populations. Using the polymorphism at the four loci, genetic relationships between 73 European, African, Asian and Creole breeds were investigated. Within-breed variation was estimated from average genetic diversity values and mean number of alleles. Breeds were grouped into twelve geographic regions, and reduced networks, MDS plots, mismatch and AMOVA analyses generated.

Polymorphism at the *INRA124* marker differentiated between zebu and taurine Y-chromosomes, and the results support the independent domestications of zebu and taurine cattle. European cattle breeds exhibited exclusively taurine Y-chromosomal haplotypes, with Creole, Africa, Near East and South Russia containing a mixture of both zebu and taurine. European cattle groups share the majority of their haplotypes with the Near Eastern group, with general homogeneity across central Europe. Although little substructuring was determined within Europe, there appeared to be a loss of both haplotypes and diversity in relation to the distance of the geographic group from the Near East. This supports the idea that European cattle are a subset of an initially diverse Near Eastern domestic population. In contrast to the diversity patterns seen in Europe, the African group shared few haplotypes with the Near East, evidence pointing toward an independent origin for African domestic cattle. There was also evidence for introgression in Scandinavia, a group that contained two highly diverse haplotypes that have yet to be found in any other extant cattle. The presence of these haplotypes could possibly originate from introgression of local aurochs or wisent populations.

Analysis of the diversity and distribution of haplotypes in the Near East supports the suggestion that this region was a major centre of cattle domestication. The presence of both Near Eastern and putative African Y-haplotypes in the Iberian group may account for the

increased levels of diversity. The British Isles group also has increased diversity when compared to the rest of Europe, which may either be a result of a fast migratory route out of the Near East, whereby there was less time for loss of genetic diversity, or alternatively may point toward repeated colonisation by sea.

The results shown here appear to offer evidence of the two migratory routes of domesticated Near Eastern cattle into Europe: the *Mediterranean* and *Danubian* routes. The speed of migration - slow across the Danubian route, quick along the Mediterranean route – is highlighted by the elevated diversity values in the southern and north-western groups when compared to the eastern and central groups.

4.2 Introduction

4.2.1 *Simple tandem repeats: microsatellites*

Microsatellites, also known as simple sequence loci, are DNA repeat regions with numerous copies of units between two and five nucleotides in length that are distributed randomly throughout the nuclear genome. Due to their rapid rate of evolution, ease of scoring and general abundance, these genetic elements have been widely used for kinship studies, population genetic analyses and to infer evolutionary history. In addition, microsatellites have been essential tools for the construction of genetic linkage maps, which can then be used to search for loci associated with economically important traits. It is not yet clear what role, if any, microsatellite repeats play in the genome function, although it has been hypothesised that they act either as hot spots for gene recombination, or are associated with regulation of gene expression at the transcriptional level (Hamada *et al.* 1982). Comparison of protein marker systems and microsatellite loci by Arranz *et al.* (1996) found microsatellites to be more useful for the estimation of genetic distance among closely related populations due to their increased variability.

Microsatellite evolution

The mutation rate at microsatellite loci has been estimated to be of the order of 1×10^{-3} per locus per generation (Weber & Wong 1993). These short simple sequence stretches are thought to mutate by polymerase slippage events during DNA replication (Schlötterer & Tautz 1993), which increase or decrease the number of repeat units incrementally, although larger mutational steps are also possible. This is supported by the observation that allele sizes at a dinucleotide locus differ by even numbers of bases. Two possible models of mutation are as follows: firstly, the Infinite Alleles Model (IAM) of neutral selection, where each mutation creates a new allele that is not dependent upon the original allele size (Kimura & Crow 1964); and secondly, the Stepwise Mutation Model (SMM), where length changes occur via the insertion or deletion of one repeat unit per mutation (Ohta & Kimura 1973). Mutational events to any of the possible alleles are equally likely in IAM, whereas under the SMM the alleles are different in terms of mutational steps. Evidence has been presented that most Y-chromosomal microsatellites follow the SMM (Cooper *et al.* 1996), leading to the assumption that any given haplotype may arise independently on more than one Y-chromosome lineage. This would make interpretation of small samples sizes (less than 30) difficult, but using multiple loci to define haplotypes could reduce the extent of

this haplotypic homoplasy. Two alleles are identical-by-descent (IBD) if they descend from the same ancestral allele without mutation. If two alleles have the same size (that is, are identical-in-state, IIS) this does not necessarily infer that they are IBD. They may have shared a common allele state in the past but have since followed their own history. Allelic homoplasy is defined as alleles that are IIS but not IBD. This can cause problems when it comes to applying the SMM as the actual divergence between populations can be underestimated.

4.2.2 *Review of the Y-chromosome literature*

Brief description of the Y-chromosome

In mammals the male is the heterogametic sex, possessing the male-specific, effectively haploid Y-chromosome. Maleness is determined in a dominant way through the inheritance of a single gene, *SRY*, present on this chromosome, which causes testis differentiation. The Y-chromosome shares homology with the X at its tip, and undergoes recombination here, in an area termed the pseudoautosomal region. Except for this region the whole of the Y-chromosome acts as a single non-recombining unit. This ensures that original combinations of mutational events along male lineages are preserved and this genetic diversity can be traced both within and among populations. Because of the increased variability of male reproductive success and sexual differences in population mixing and migration patterns, male lineages are likely to reflect different evolutionary patterns to female lineages. Research targeting the Y-chromosome can provide additional but complementary data to those generated from the maternally-inherited mitochondria.

Y-chromosome studies in humans

Y-chromosome-specific microsatellites have been used in humans for discrimination of closely-related populations and to estimate relationships between these populations (for example, diversity within populations has been shown to be greater than that between populations, Cooper *et al.* 1996; and geographic clustering is evident, Linares *et al.* 1996). In the past, the human Y-chromosome had been reported as being the least variable chromosome with a general paucity of polymorphisms, but research from the last decade has not found this to be the case. For example, Roewer *et al.* (1996) found high levels of polymorphism in Dutch and German populations, with little haplotype sharing when compared over seven Y-specific loci; and Cooper *et al.* (1996) and Kittles *et al.* (1998)

also found high levels of haplotype diversity within populations in East Anglia and Finland respectively.

The effective population size of Y-haplotypes is ~ 25% that of the autosomes (Kittles *et al.* 1998). However, due to the lack of recombination compared with possible episodes of selective sweep, and unequal variance in the reproductive success of males and females, the effective size may be lower than this (Petit *et al.* 2002). This small effective size makes the Y-chromosome sensitive to genetic drift, and it is therefore affected more by recent demographic events; for example, by population bottle-necks, founder effects and population expansions.

Y-chromosomal haplogroups constructed using Y-specific loci are extremely informative for population genetic studies, as well as for analyses of settlement, mating structure and migration of populations in historic times. A combinatorial approach, utilising both microsatellite and SNP markers, has been adopted to study isolation and admixture in human populations; for example, paternal inheritance within the Jewish Cohanim priesthood (Thomas *et al.* 1998) and mass migration of Anglo-Saxons into Britain (Weale *et al.* 2002).

Sex chromosomes in cattle

The karyotype of cattle was determined in the 1950's to be 2n = 60 (Melander 1959) and this number is almost uniform across all bovids. With regard to the sex chromosomes, there is one form of the X-chromosome in cattle, which is a large submetacentric in both taurine, and zebu cattle. A single X comprises 2.6 to 3.0% of the diploid genome by length (Halnan 1989). In contrast there are two forms of the Y-chromosome, with the typical *B. taurus* Y-chromosome morphology being submetacentric and *B. indicus* having small acrocentric Y-chromosomes. It is believed that the size differences between these two types are due to a Y-to-autosome translocation (Rendel 1980). Halnan & Watson (1982) demonstrated that the short (*p*) arm of pure taurine Y-chromosomes is equivalent to the long (*q*) arm in pure-bred zebu, a pericentric inversion. The submetacentric Y represents about 1.5% of the total genome by length and the acrocentric about 0.9% (Halnan 1989).

The *taurine-type* of the Y-chromosome is present in European bison, Banteng cattle, gaur and yak, and the *indicine-type* is present in American bison (Halnan 1989). The fact that European bison have taurine Y-chromosomes may possibly be due to admixture with

cattle. Y-chromosome morphology varies in different cattle breeds but no comprehensive study has been undertaken. However, it has been suggested that the Y-chromosome in Jersey cattle is metacentric rather than submetacentric, and Charolais and Simmental have larger Y-chromosomes than other breeds (see Halnan 1989 for a review). The difference in Y-morphology has been useful for detecting genetic erosion of taurine populations due to zebu migration and introgression.

Y-chromosome work in cattle

Based on RFLP analysis of a polymorphic Y-specific DNA probe (*btDYZ-1*), Bradley *et al.* (1994) found two Y-chromosomal haplotypes, one zebu and the other taurine. This is consistent with the two forms of Y-chromosome determined by cytogenetic studies. The probe was used to assess the pattern of gene flow observed in African cattle, and significant sexually-mediated zebu introgression was discovered. When compared to autosomal variation (MacHugh *et al.* 1997) and mtDNA sequences (for example, Bradley *et al.* 1996), the work by both Bradley *et al.* (1994) and Teale *et al.* (1995) (who used RAPD probe *ILO1065*) suggests that the introgression of Y-chromosomes into populations occurs more extensively than other genomic features. This can be explained by movement of male zebu cattle into Africa, but may also have occurred because of preference exercised for zebu bulls within both that continent and Asia.

A thorough search of the available literature and databases uncovered nine non-pseudoautosomal Y-specific markers (Bishop *et al.* 1994; Vaiman *et al.* 1994; Mezzelani *et al.* 1995; Kappes *et al.* 1997; Ward *et al.* 2001). Prior to this study, the only Y-specific marker to be typed as polymorphic had been *INRA124* (Hanotte *et al.* 1997), which showed two length variants; a 130 bp zebu allele and a 132 bp taurine allele. This marker was used to type sub-Saharan African *zebu* cattle and it was discovered that 30% of those animals studied had the taurine Y-chromosome (Hanotte *et al.* 2000).

Microsatellite analyses of the Y-chromosome are preferable to cytogenetic and other molecular techniques as this type of approach has the potential to provide a higher resolution perspective on evolutionary history and recent population movements. Microsatellite loci can be typed more rapidly than other methods, and it is highly likely that male lineages will reflect different evolutionary histories than those of female lineages. These Y-specific microsatellite polymorphisms should provide a starting point

for an informative paternal analogue to mitochondrial DNA for studies of genetic variation in cattle and related species.

4.3 Materials & Methods

4.3.1 *Polymorphism screening in a range of bovid species*

In the first part of the present study, primer and sequence data pertaining to Y-chromosome-specific microsatellites from scientific literature and GENBANK were consulted. Four markers were located (Bishop *et al.* 1994; Vaiman *et al.* 1994; Kappes *et al.* 1997), each comprising at least one dinucleotide repeat region of nine or more repeat units. All four repeat segments were of the $(dC\text{-}dA)_n\cdot(dG\text{-}dT)_n$-type (this array being the most common dinucleotide repeat motif found in vertebrate genomes; Hamada *et al.* 1982); two are perfect microsatellites and the other two imperfect. Of the four, only one marker (*INRA124*) has been previously shown to be variable (Hanotte *et al.* 1997). **Table 4.1** shows information on these four microsatellites.

The four bovine Y-specific microsatellites were assessed for polymorphism in a panel of male animals from the following breeds and species: Hereford (European *Bos taurus*, n = 6); N'Dama (African *Bos taurus*, n = 6); Ongole (Indian *Bos indicus*, n = 3); Nellore (Indian *Bos indicus*, n = 12); plains bison (*Bison bison*, n = 6); mithan (*Bos [Bibos] gaurus*, n = 11); swamp buffalo (*Bubalus arnee*, n = 6); and yak (*Bos* [*Pöephagus*] *grunniens*, n = 6). Three samples of male sheep (African *Ovis aries*, n = 3) were also assayed to assess the specificity of the primers to the *Bovini* family. Six female Jersey animals were also included to check marker specificity for the cattle Y-chromosome. In addition, a small number of females (1-6) from the other species and subspecies were screened to ensure Y-chromosome specificity for each marker.

20-30 ng of DNA was used as template in an 11 µl reaction volume. 0.5 Units of PLATINUM® *Taq* polymerase (GIBCOBRL®) were used with the buffers supplied (1.5 mM $MgCl_2$; 50 mM KCl; 20 mM Tris·HCl [pH 8.4]). The concentration of dNTPs was 200 µM, with the exception of dCTP, which was reduced to 20 µM. [α-^{32}P] dCTP radionucleotide was randomly incorporated during the PCR amplification by the addition of 0.5 µCi [α-^{32}P] dCTP at 3,000 Ci mmol^{-1} per reaction. 3.0 pmoles of both the forward and the reverse primers were also added. The reactions were carried out using a HYBAID© *TouchDown*™ thermal cycling machine with the following cycling parameters: 93°C for 3 minutes, followed by 40 cycles of 93°C for 40 seconds, 65 → 55°C for 40 seconds and

Marker	Primer sequences (5'-3')	Repeat motif	No. of alleles	Range of alleles (bp)	Reference
INRA124 (DYS6)	for-GATCTTTGCAACTGGTTTG rev-CAGGACACAGGTCTGACAATG	$(GT)^4(A)(TG)^9(N)^{43}(TG)^9$	2	130-132	Vaiman *et al.* (1994)
INRA126 (DYS7)	for-GTTGTTGCCTCTGCAGAGTAGG rev-GACACTCTTTCTATTTTCAAGG	$(TG)^{11}$	3	182-186	Vaiman *et al.* (1994)[1]
INRA189	for-TACACGCATGTCCTTGTTTCGG rev-CTCTGCATCTGTCCTGGACTGG	$(TG)^{22}$	9	68-124	Kappes *et al.* (1997)[2]
BM861	for-TTGAGCCACCTGGAAAGC rev-CAAGCGGTTGGTTCAGATG	$(GT)^6(C)(TG)^{10}$	6	144-158	Bishop *et al. (*1994)

Table 4.1 Y-specific microsatellite loci used to assay haplotypic variation in the cattle surveyed.

[1] Published amplification primers differ from those presented here.
[2] No published primers available so those presented here were designed from sequences submitted to GENBANK.

72°C for 40 seconds. Annealing temperatures were decreased from 65 to 55°C successively for each of the first 21 cycles (that is, decrements of 0.5°C per cycle: *TouchDown* PCR). A final extension of 5 minutes at 72°C was also carried out. The radiolabelled products were then assessed for length variation using standard 6% polyacrylamide sequencing gel electrophoresis and autoradiography. DNA sequenced from the phage vector M13mp18 was run with the amplified test samples to size the allelic variation. This technique was precise enough to reproducibly resolve single repeat unit differences.

4.3.2 *Four-marker analysis of cattle breeds and related species*

Once polymorphism had been established in the four markers, DNA samples collected from 577 male cattle were typed. In addition, a small number of females (42) from certain breeds were screened to further ensure Y-chromosome specificity. Unrelated animals were chosen, as far as could be ascertained. Hair or blood samples were collected from between 3 and 46 male individuals from each of the 38 breeds, and genomic DNA was extracted as described previously (Loftus *et al.* 1999). In addition, a further 297 samples, comprising 35 breeds, were included (J. Kantanen unpublished data). Details for all breeds are shown in **Table 4.2**. The total 73 breeds were split into twelve regional groups dependent upon geographical location, as shown in **Figure 4.1**.

To assess the levels of variability at the DNA level in the breed groups, unbiased estimates of expected gene diversity, with associated standard errors, were calculated according to Nei (1987). Gene diversity can be defined as the probability that two randomly chosen alleles from a population are non-identical, and is not affected by non-random mating or by the number of loci or alleles present. Allele frequencies, mean number of alleles (MNA), and haplotype frequencies were determined by direct counting. Graphical representations of both the diversity and MNA values were drawn.

Analysis of frequency distributions of pairwise mismatches of haplotypic systems can be indicative of population demographic history. The FACTOR program (Meacham 1993), part of the PHYLIP package (Felsenstein 1989), was used to recode the multistate allele counts for the four markers to binary characters for mismatch analysis. Mean pairwise differences and mismatch distributions were generated according to Rogers & Harpending (1992) and Rogers (1995), using the ARLEQUIN computer program (version 2.000; Schneider *et al.* 1997).

Geographic grouping	Breed name	Country of origin	Number of females sampled	Total number of males sampled	Missing male data	Complete male data
India (IND)	Ongole	India		5	1	**4**
Creole (CRE)	Antigua	Antigua		46	26	20
	Guadaloupe	Guadaloupe		27	16	11
	Nellore	Brazil	7	13	1	12
	St. Lucia	St. Lucia		3	1	2
						45
Africa (AFR)	Benin Somba	Benin		15	7	8
	Egyptian	Eygpt	4	9	4	5
	Gobra	Senegal		15	5	10
	Kuri	Chad		15	7	8
	N'Dama	Guinea		22	0	22
	North Togolese Somba	Togo		30	7	23
	Pure Togolese Somba	Togo		15	0	15
	Sokoto Gudali	Nigeria		15	10	5
	South Togolese Somba	Togo		15	6	9
	White Fulani	Nigeria		8	4	4
						109
Near East (NEA)	Anatolian Black	Turkey		14	2	12
	Damascus	Syria		14	6	8
	East Anatolian Red	Turkey		6	1	5
	Kurdi	Iraq		5	0	5
	Middle Iraqi	Iraq		10	1	9
	North Iraqi	Iraq		12	2	10
	South Anatolian Red	Turkey		8	3	5
	South Iraqi	Iraq		7	0	7
	Turkish Grey	Turkey		21	12	9
						70
South Russia (SRU)	Alatauskaya*	Russia		6	0	6
	Bushuevskaya*	Russia		4	0	4
	Red Byelorus*	Russia		23	0	23
	Red Gorbatovskay*	Russia		2	0	2
	Ukrainian Grey*	Ukraine		7	0	7
	Ukrainian Whiteheaded*	Ukraine		11	0	11
	Yurinskaya*	Russia		6	0	6
						59
North-central Eurasia (NCE)	Eastern Finncattle*	Finland		14	3	11
	Istobenskaya*	Russia		9	0	9
	Kholomogorskaya*	Russia		8	0	8
	Northern Finncattle*	Finland		3	0	3
	Western Finncattle*	Finland		9	0	9
	Yakutian*	Russia		17	0	17
	Yaroslavskaya*	Russia		4	0	4
						61
Iberia (IBE)	Alentejana	Portugal	3	25	6	19
	Arouquesa	Portugal		28	11	17
	Barrosa	Spain	7	19	15	4
	Mertolenga	Portugal	8	19	1	18
						58
Western-central Europe (WCE)	Charolais	France		15	1	14
	Limousin	France	7	7	4	3
	Pustertaler	Italy/Germany		14	0	14
	Simmental	Switzerland		10	3	7
						38
North-western Europe (NEU)	Friesian	Holland		15	2	13
	Jutland (old native)*	Denmark		6	0	6
	Red Danish (old type, native)*	Denmark		15	0	15
						34

Geographic grouping	Breed name	Country of origin	Number of females sampled	Total number of males sampled	Missing male data	Complete male data
British Isles (BRT)	Aberdeen Angus	Scotland		15	1	14
	Galloway	Scotland		15	8	7
	Hereford	England		15	2	13
	Highland	Scotland		15	5	10
	Jersey	Jersey	6	15	4	11
	Jersey (old type)*	Denmark		9	0	9
						64
Baltic (BAL)	Estonian Native*	Estonia		10	0	10
	Estonian Red*	Estonia		10	0	10
	Latvian Blue (native)*	Latvia		9	0	9
	Latvian Brown (commercial)*	Latvia		8	0	8
	Lithuanian Grey*	Lithuania		10	0	10
	Lithuanian Red*	Lithuania		9	0	9
	Lithuanian White-Back (native)*	Lithuania		10	0	10
						66
Scandinavia (SCN)	Blacksided Troender*	Norway		7	0	7
	Doela*	Norway		4	0	4
	Eastern Red Polled*	Norway		6	0	6
	Icelandic*	Iceland		12	1	11
	Norwegian (commercial, hybrid)*	Norway		12	0	12
	Swedish Mountain*	Sweden		11	2	9
	Swedish Old Mountain (unimproved)*	Sweden		4	0	4
	Swedish Red Polled*	Sweden		5	0	5
	Telemark*	Norway		2	0	2
	Western Fjord*	Norway		10	4	6
	Western Red Polled*	Norway		5	1	4
						70
		TOTAL	42	874	196	**678**

Table 4.2 (on previous two pages) Breed information, including geographical grouping (with the respective code in parentheses) for the 38 cattle breeds sampled as part of this study, plus 35 breeds analysed by J. Kantanen (denoted by an asterix *). Details include the number of female individuals sampled as part of the verification that the four markers exhibit Y-chromosome specificity; the total number of male samples typed, with the missing male data column relating to those individuals that were genotyped with less than four of the markers; and the complete male data set, with both breed totals and geographic group totals.

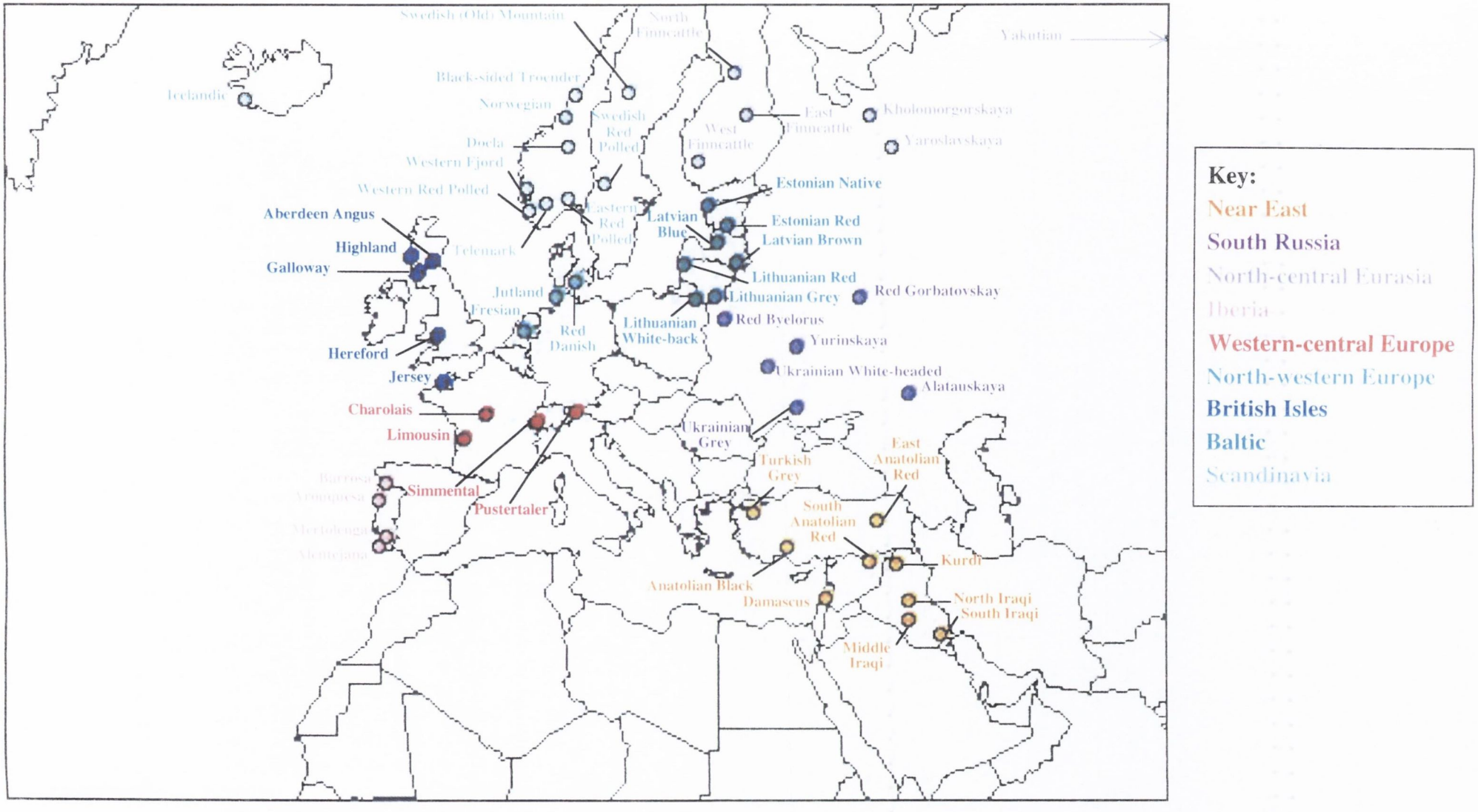


Figure 4.1 Locations of all the breeds sampled as part of this study, with breed names and points coloured depending upon geographic group (Table 4.2).

A reduced skeleton median network was constructed using the median algorithm of Bandelt *et al.* (1995). Reduced networks were also drawn using four marker haplotype data from each geographical group. These networks display the principal character relationships among the haplotypes present in the data. Using the marker *INRA124*, the Y-haplotypes can be partitioned into zebu (presence of the 130 bp allele) and taurine (presence of the 132 bp allele). The reliability of using this demarcation was examined by comparing haplotypes generated from 100 African individuals using the four markers here, plus two extra markers, *DYZ1* and *BYM-1*, typed in samples by Dr. Juha Kantanen.

An attempt was made to determine the root of the skeleton network using the four marker results from the four bovid species, plains bison, mithan, swamp buffalo and yak. For each unique bovid haplotype, the allele size exhibited at each marker was compared to all those alleles seen in the cattle breeds at the same marker. Zero denoted identity between the bovid allele and the marker allele under consideration, with an extra 1 added to this score for each 2 bp mutation exhibited with relation to the marker allele. The overall score for each cattle allele at each marker was then tallied to reconstruct the root haplotype.

Population genetic distances (pairwise F_{ST} values) were generated from sum of size differences using the ARLEQUIN computer program (version 2.000; Schneider *et al.* 1997). This was carried out using all geographic groups, groups truncated to contain only breeds with zebu haplotypes, and groups truncated to contain only breeds with taurine haplotypes. The pairwise F_{ST} genetic distance is considered the most appropriate measure for surveys of genetic variation among populations where genetic drift has been the primary stimulus of differentiation (Reynolds *et al.* 1983; Slatkin 1995). Multidimensional scaling (MDS; Kruskal 1964) was performed, based on the F_{ST} matrices obtained, using the SPSS© program (Standard Version, release 10.0.5).

Analysis of molecular variation (AMOVA; Excoffier *et al.* 1992) was performed on all data, and on data partitioned into groups containing only zebu or only taurine haplotypes, using ARLEQUIN. Locus-by-locus AMOVA was also calculated. AMOVA assesses the partitioning of molecular variance between individuals within populations/breeds, between populations within groups, and among groups.

Taurine haplotypes were grouped into eight Eurasian geographic groups (Near East, South Russia, North-central Eurasia, Iberia, Western-central Europe, North-western Europe,

British Isles, Baltic and Scandinavia), excluding the two outlying populations: the Yakutian breed of North-central Eurasia and the Icelandic breed. Geographic coordinates for the central location of each group were estimated. The first principal component (PC) was calculated using SPSS©, and the PC values for the nine groups used to construct an interpolation map illustrating continent-wide genetic trends. Similar maps were also constructed from regenerated values of genetic diversity and MNA, as well as mean number of mutations calculated from mismatch distributions. The interpolation maps were drawn using the *Spatial Analyst* extension of ArcView (version 3.2; available at *www.esri.com/software/arcview/*). The inverse distance weighted (IDW) option, with a power of two, was selected for the interpolation of the surface. IDW assumes that each input point has a local influence that diminishes with distance.

4.4 Results

4.4.1 *Polymorphism screening in bovids*

Figure 4.2 shows sample autoradiograms for a panel of animals screened with each marker. **Table 4.3** shows the allele sizes and frequencies for the various populations screened with these markers. Using the four microsatellite loci, a total of eleven alleles were detected in various populations of cattle. This number rose to twenty alleles when individuals from closely-related species were also surveyed with the same genetic markers. One Y-specific microsatellite used in this study (*INRA124*) had already been shown to be polymorphic by Hanotte *et al.* (1997). They found two length variants in several cattle populations, but failed to obtain amplification in samples from swamp buffalo, yak or mithan (domesticated gaur). This was not found to be the case here, as all four loci gave PCR products when amplified with DNA from American plains bison (*Bison bison*), mithan (*Bos [Bibos] gaurus*), swamp buffalo (*Bubalus arnee*), yak (*Bos [Pöephagus] grunniens*) and sheep (*Ovis*). In all but one case, the amplification of these Y-chromosome-derived markers was restricted to the male samples. However, *INRA126* displayed an amplification product in six female yaks of exactly the same size as that detected in the panel of male yaks (182 bp). This would indicate that the yak X-chromosome has retained an homologous sequence to the Y-chromosomal segment containing the *INRA126* microsatellite.

As the second part of this study, 577 male cattle were typed with four microsatellite markers, *INRA124*, *INRA126*, *INRA189* and *BM861*. Full haplotypic data were obtained from 329 of these animals, with incomplete typings from the remaining 185 samples due to failure of one or more markers. No results were obtained from the 42 females included in this data set, verifying the Y-chromosome specificity of the four markers in cattle. A further 286 samples were contributed by kind permission of J. Kantanen (unpublished data). Here, only eleven animals failed to give complete haplotype data when amplified with the same four markers used here. Therefore, the complete data set available for further analyses included 678 fully-typed Y-chromosomes from 73 breeds, classified into twelve groups on the basis of geography (**Table 4.2**; **Figure 4.1**).

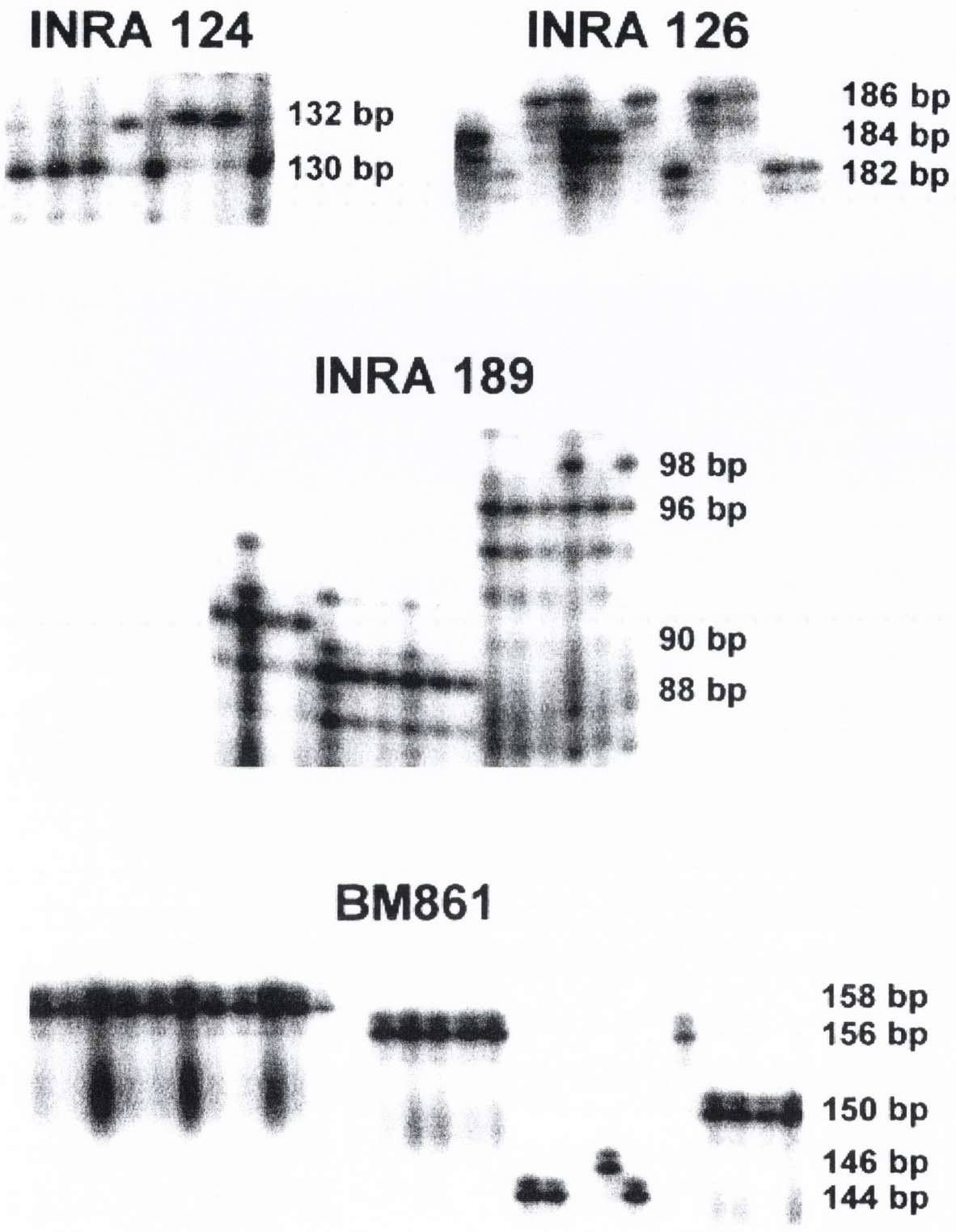


Figure 4.2 Electronic scans of autoradiograms showing each of the four microsatellite markers amplified in a range of animals. Allele sizes are also indicated in base pairs.

Marker	Length (bp)	Hereford (6)	N'Dama (6)	Ongole (3)	Nellore (12)	Plains Bison (6)	Mithan (11)	Swamp Buffalo (6)	Yak (6)	Togo Sheep (3)
INRA 124	130	0	0	100	100	67	100	0	0	0
(DYS6)	132	100	100	0	0	33	0	100	100	100
INRA 126	182	17	17	100	100	100	100	0	100	0
(DYS7)	184	67	17	0	0	0	0	100	0	33
	186	17	67	0	0	0	0	0	0	67
INRA 189	68	0	0	0	0	0	100	0	0	0
	84	0	0	0	0	0	0	83	0	0
	88	0	0	100	100	0	0	0	0	0
	90	0	83	0	0	0	0	0	0	0
	96	0	0	0	0	67	0	17	0	0
	98	100	0	0	0	33	0	0	0	0
	102	0	0	0	0	0	0	0	100	0
	104	0	17	0	0	0	0	0	0	0
	124	0	0	0	0	0	0	0	0	100
BM861	144	0	0	0	0	33	0	0	0	0
	146	0	0	0	0	67	0	0	0	0
	148	0	0	0	0	0	0	0	100	0
	150	0	0	0	0	0	82	0	0	0
	156	0	0	100	100	0	18	0	0	0
	158	100	100	0	0	0	0	100[1]	0	100

Table 4.3 Allele size ranges and frequencies for *INRA124*, *INRA126*, *INRA189* and *BM861* in a variety of cattle types and related species.
[1] Of the six animals sampled, only one gave an amplification product with marker *BM861*.

4.4.2 *Diversity within cattle geographic groups*

Allele frequencies

A total of 21 alleles were detected across the four loci when they were screened in 73 breeds grouped into twelve biogeographical regions. The mean number of alleles per locus was therefore 5.25, although the actual number of observable alleles at each locus varied from two alleles at the *INRA124* locus, to ten alleles at the *INRA189* locus. The frequency of each allele in each geographic group is given in **Table 4.4**.

Locus *INRA124* exhibits two alleles, which differentiate between taurine and zebu individuals (Hanotte *et al.* 1997). Each geographic group was analysed for presence of taurine and zebu haplotypes depending upon which allele was present at this locus; 130 bp is zebu and 132 bp is taurine. The Indian group contains only zebu alleles at this locus, with Creole showing 60% zebu, Africa 45%, the Near East 24.3% and the Southern Russian group showing 5.1%. Studying these groups leads to the putative diagnosis of the 88 bp allele at *INRA189* also being a zebu allele, as similar allele frequencies for the 130 bp allele in *INRA124* are exhibited by this allele. *INRA189* also exhibits a possible breed-specific structure, with the 82 bp allele only present in the Norwegian Doela (4 individuals) and Telemark (2) breeds from Scandinavia, 92 bp only present in the Alentejana breed (1) from Iberia, and 96 bp only present in the Egyptian breed (1) from Africa. The 190 bp allele in *INRA126* is only present in the Iberian group.

The number of zebu, taurine and unique haplotypes observed over the twelve geographic groups is shown in **Table 4.5**. Only four zebu haplotypes were unique to one group, and three of these are seen in the Near East. The Near East exhibits the greatest number of taurine alleles (17), but none of these are unique to this area alone. Both Iberia and the British Isles have a comparable number of taurine haplotypes to the Near East (16 and 15 respectively), with five and four unique haplotypes each respectively.

Mean number of alleles (MNA) and genetic diversity

When populations are at mutation-drift equilibrium and the sample size is approximately the same for each population, the mean number of alleles (MNA) observed over a range of loci for different populations is considered to be a reliable indicator of genetic variation (Nei 1987). Higher MNA values are seen in the mixed zebu/taurine groups rather than

	India	Creole	Africa	Near East	South Russia	North-central Eurasia	Iberia	Western-central Europe	North-western Europe	British Isles	Baltic	Scandinavia
INRA124												
130	100.00	60.00	44.95	24.29	5.08							
132		40.00	55.05	75.71	94.92	100.00	100.00	100.00	100.00	100.00	100.00	100.00
INRA126												
180		2.22		15.71			6.90					
182	75.00	33.33	24.77	14.29	15.25	9.84	8.62	21.05	11.76	6.25	4.55	4.29
184		40.00	5.50	40.00	27.12	49.18	24.14	36.84	47.06	57.81	65.15	71.43
186		20.00	53.21	30.00	47.46	40.98	51.72	36.84	32.35	32.81	25.76	24.29
188	25.00	4.44	16.51		10.17		3.45	5.26	8.82	3.13	4.55	
190							5.17					
INRA189												
82												8.57
88	100.00	57.78	47.71	20.00	5.08							
90		4.44	16.51	4.29	10.17							
92							1.72					
96			0.92									
98		24.44	0.92	28.57	44.07	59.02	5.17	2.63	100.00	21.88	96.97	75.71
100		6.67		4.29	1.69		1.72			20.31	1.52	10.00
102		2.22	11.01	7.14	35.59		31.03	97.37		26.56	1.52	5.71
104		4.44	17.43	17.14	3.39	40.98	60.34			31.25		
106			5.50	18.57								
BM861												
156	75.00	62.22	45.87	21.43	5.08		1.72			9.38		
158	25.00	37.78	52.29	77.14	94.92	100.00	98.28	92.11	97.06	79.69	100.00	100.00
160			1.83	1.43				7.89	2.94	10.94		

Table 4.4 Allele size ranges and frequencies for *INRA124*, *INRA126*, *INRA189* and *BM861* in each of the twelve geographic groupings

	number of zebu chromosomes	zebu haplotypes	unique zebu haplotypes	number of taurine chromosomes	taurine haplotypes	unique taurine haplotypes
India	4	2	0	0	0	n/a
Creole	27	6	1	18	8	1
Africa	49	4	0	60	18	8
Near East	17	7	3	53	17	0
South Russia	3	2	0	56	14	4
North-central Eurasia	0	0	n/a	61	6	0
Iberia	0	0	n/a	58	16	5
Western-central Europe	0	0	n/a	38	7	1
North-western Europe	0	0	n/a	34	5	0
British Isles	0	0	n/a	64	15	4
Baltic	0	0	n/a	66	6	0
Scandinavia	0	0	n/a	70	9	2

Table 4.5 Direct haplotype counts of all geographic groups, detailing the number of zebu and taurine chromosomes present in each group, and the total number of zebu and taurine haplotypes encountered in each group, with the number of haplotypes unique to each.

zebu and taurine groups. MNA values vary from 1.25 in the South Russian zebu group to 4.00 in the Near Eastern and African groups containing both zebu and taurine data. Among the different geographical groupings, the MNA values are highest in the Near East, with 4.00 over both zebu and taurine haplotypes, 2.25 for zebu haplotypes alone, and 3.00 for taurine haplotypes alone (**Table 4.6**). This has previously been noted from mitochondrial DNA (Troy *et al.* 2001) and nuclear microsatellite (Loftus *et al.* 1999) analyses. Although little store may be attached to these MNA values in a linked haplotypic system, the British Isles group exhibits a similarly high MNA (3.00), with the Iberian group having an even higher value of 3.50. Geographic groups from northern and western Eurasia (North-central Eurasia, North-western Europe, Baltic and Scandinavia) have low MNA values, from 1.75 to 2.25 when taurine haplotypes alone are considered. Values for all data, zebu data and taurine data are 5.25, 2.50 and 5.00 respectively.

A similar pattern is seen in the unbiased estimates of genetic diversity for zebu and taurine haplotypes (**Table 4.6**). Gene diversity is the probability that two randomly chosen haplotypes are different in a population. The values vary from 0.589 in Africa, containing both zebu and taurine haplotypes, to 0.143 in the taurine Baltic group. The values obtained for mixed zebu-taurine groups would be expected to show higher diversity due to the presence of zebu and taurine alleles, and this seems particularly evident in the values obtained from the African group, which contains a hybrid Togolese Somba population (**Table 4.6**). Looking at taurine haplotypes alone, the British Isles group shows the greatest

	zebu and taurine haplotypes			zebu haplotypes only			taurine haplotypes only			
	genetic diversity	MNA	mean mismatch	genetic diversity	MNA	mean mismatch	genetic diversity	MNA	mean mismatch	PC1
India	n/a	n/a	n/a	0.250 +/- 0.144	1.50	nd	n/a	n/a	n/a	n/a
Creole	0.571 +/- 0.053	3.75	4.82	0.195 +/- 0.151	2.00	1.16	0.307 +/- 0.149	2.75	2.29	nd
Africa	0.589 +/- 0.048	4.00	5.84	0.178 +/- 0.164	2.00	1.33	0.365 +/- 0.181	3.75	4.59	nd
Near East	0.567 +/- 0.117	4.00	5.46	0.329 +/- 0.166	2.25	1.75	0.367 +/- 0.202	3.00	2.95	-0.490
South Russia	0.388 +/- 0.167	3.50	3.33	0.167 +/- 0.167	1.25	nd	0.332 +/- 0.192	2.75	2.81	0.030
North-central Eurasia	n/a	n/a	n/a	n/a	n/a	n/a	0.270 +/- 0.157	1.75	2.15	nd
North-central Eurasia (minus Yakutian)	n/a	n/a	n/a	n/a	n/a	n/a	0.211 +/- 0.131	1.75	1.51	0.622
Iberia	n/a	n/a	n/a	n/a	n/a	n/a	0.312 +/- 0.172	3.50	1.96	-1.367
Western-central Europe	n/a	n/a	n/a	n/a	n/a	n/a	0.225 +/- 0.161	2.25	1.20	-1.501
North-western Europe	n/a	n/a	n/a	n/a	n/a	n/a	0.183 +/- 0.164	2.00	0.94	1.191
British Isles	n/a	n/a	n/a	n/a	n/a	n/a	0.417 +/- 0.162	3.00	2.30	-0.613
Baltic	n/a	n/a	n/a	n/a	n/a	n/a	0.143 +/- 0.124	2.25	0.70	1.372
Scandinavia	n/a	n/a	n/a	n/a	n/a	n/a	0.212 +/- 0.122	2.25	2.11	nd
Scandinavia (minus Icelandic)	n/a	n/a	n/a	n/a	n/a	n/a	0.218 +/- 0.126	2.25	2.35	0.757

Table 4.6 Genetic diversity and mean number of alleles (MNA) calculated for each geographic group, plus the mean value generated from the pairwise mismatch distribution analysis, for all data (zebu and taurine), zebu individuals and taurine individuals. For the taurine data, the first principal component (PC1) values were calculated for the geographic groups shown, from the *Fst* values shown in red in Table 4.8c. n/a = not applicable; nd = not done.

value (0.417), closely followed by the Near East (0.367). The mean genetic diversity value for all chromosomes was estimated as 0.499 ± 0.131, with the zebu and taurine averages being 0.223 ± 0.167 and 0.358 ± 0.185 respectively.

Although the MNA values appear to have little structure, trends in the genetic diversity data are more noticeable when plotted in graphical representations, shown in **Figure 4.3**. When all the data are considered (**Figure 4.3a**), the groups with the greatest diversity are those that contain both zebu and taurine chromosomes. Again, the zebu data (**Figure 4.3b**) suffer from a poor sample size, with the greatest diversity seen in the Near East. The purely zebu Indian group includes only four individuals from one breed, leading to the, perhaps unexpected, low diversity value seen. The taurine data (**Figure 4.3c**) are interesting as, although there is high diversity in the African, Near Eastern, and South Russian groups, the greatest diversity is in the British Isles.

Pairwise mismatch distributions

Rogers & Harpending (1992) have suggested that the demographic history of populations is reflected by analysis of frequency distributions of pairwise mismatches of haplotypic systems. Until recently, mismatch distributions have only been studied at the mitochondrial level. With this system, populations that have expanded rapidly in the recent past, leading to the high probability of survival of numerous lineages, tend to exhibit smooth unimodal distribution curves. The time depth of the expansion has also been estimated from the position of the mode of the distribution, with an older expansion more likely to have a distribution skewed towards greater values of mismatches between individuals. Pereira *et al.* (2001) have applied mismatch analysis to Y-chromosomal bialleleic markers, with their resulting distributions being bimodal in appearance, which they explain as evidence for the different demographic histories for maternally and paternally transmitted loci. Therefore, it is suggested that pairwise mismatch analysis can be a useful addition to Y-chromosomal studies, with the proviso that haplotypic homoplasy may obscure the true picture. **Figure 4.4** shows the mismatch distributions for all individuals typed, for zebu haplotypes, and for taurine haplotypes. As would be expected, the spread of all data (**Figure 4.4a**) exhibits a bimodal distribution, due to the presence of both zebu and taurine haplotypes in this data set. By comparison, both the zebu (**Figure 4.4b**) and taurine (**Figure 4.4c**) mismatch curves are smooth unimodal distributions. The centre of the taurine mismatch is shifted to the right of that seen in zebu, indicating a deeper expansion time for taurine individuals, but the small sample size of the zebu data set (100 individual chromosomes) when

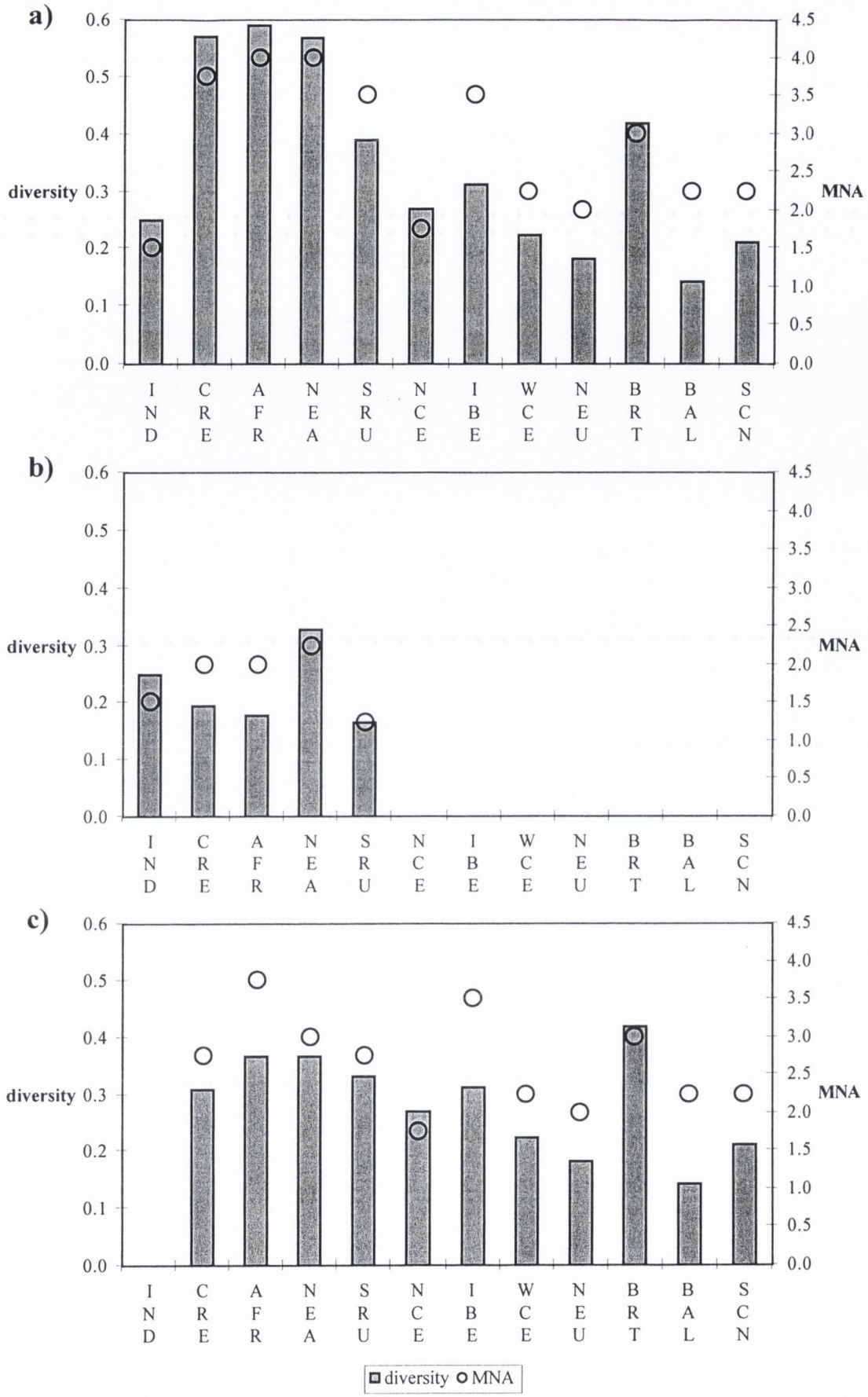


Figure 4.3 Graphical representations of the statistical data included in Table 4.6, showing diversity and mean number of alleles (MNA) values for the twelve geographic groupings, including: **a)** all individuals typed; **b)** zebu haplotypes; and **c)** taurine haplotypes. The three-letter codes are as shown in Table 4.2.

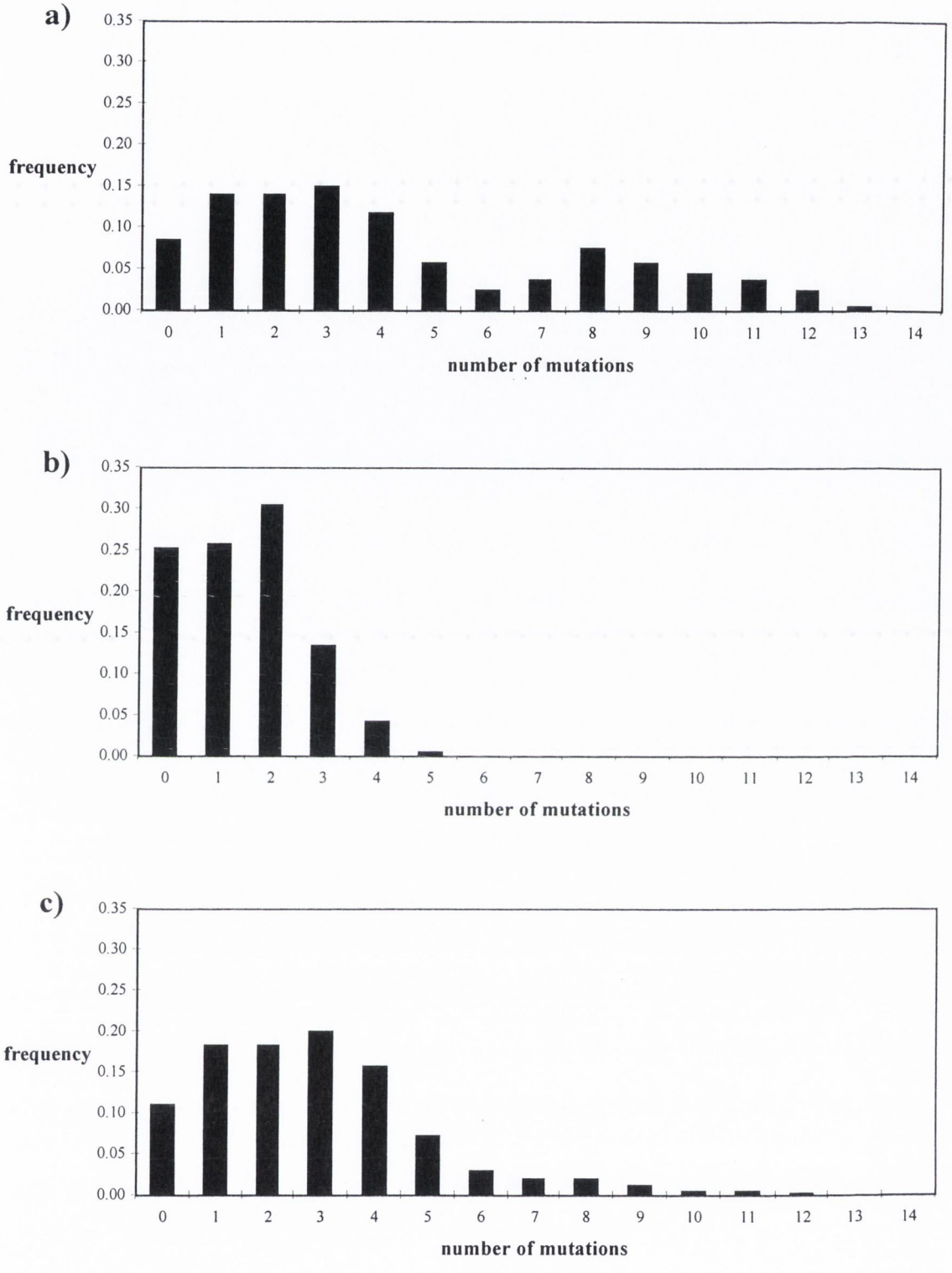


Figure 4.4 Frequency distributions of the number of mismatches observed in pairwise comparisons within geographic groups, for: **a)** all individuals typed; **b)** zebu haplotypes; and **c)** taurine haplotypes.

compared to that of the taurine data set (578) makes it difficult to make conclusions based upon this. The taurine mismatch is very dispersed, with a slight bimodal appearance, the causes of which can more easily be discussed with reference to the network drawings, as discussed in the next section.

As may be expected for groups with a mixture of zebu and taurine chromosomes, mismatches calculated using all data from the Creole, African, Near East and South Russian groups (**Figure 4.5**) exhibit bimodal distributions. The South Russia histogram has less of a bimodal structure, but this group has three zebu to 56 taurine chromosomes, whereas the other three mixed groups have more equal numbers of zebu and taurine chromosomes (**Table 4.5**). Mismatch curves were also constructed for the zebu individuals from the mixed groups (**Figure 4.5**), except the three South Russian zebu individuals as the sample size for this exercise was deemed to be too small. The mismatch of the four Indian zebu individuals was also not drawn. The curves for zebu haplotypes seen in both the Creole and Near Eastern groups are smooth and unimodal, with the Near East curve shifted along the horizontal axis suggesting an older or more diverse introgression of zebu animals into this area in comparison to the distribution of the Creole group. The African zebu curve appears to be bimodal in structure.

The Creole taurine distribution (**Figure 4.5**) is highly uneven in nature. Of the taurine mismatches shown in **Figure 4.6**, the Near East, Iberia, Western-central Europe, North-western Europe, the British Isles and the Baltic exhibit smooth unimodal curves, with the African, North-central Eurasian and Scandinavian groups being bimodal in nature. The South Russian distribution is also bimodal to some extent. The outlying individuals causing the bimodal distributions are more easily described by network construction using haplotypic data from each group. The modal values are distributed along the horizontal axis for the various groups presenting unimodal distributions, with the Near East appearing to have the greatest spread of data and a modal value of 3-4, suggestive of an earlier expansion than the other taurine groups. The mode of the South Russian and British Isles groups is 2, with Iberia, Western-central Europe and North-western Europe having a mode of 1. The Baltic group has a modal value of zero.

The mean mismatches of the distributions varied among the groups (**Table 4.6**). On the whole, when the mean number of differences is greater, the population is considered to be older and more diverse, coinciding with a shift in the curve away from the vertical axis. As

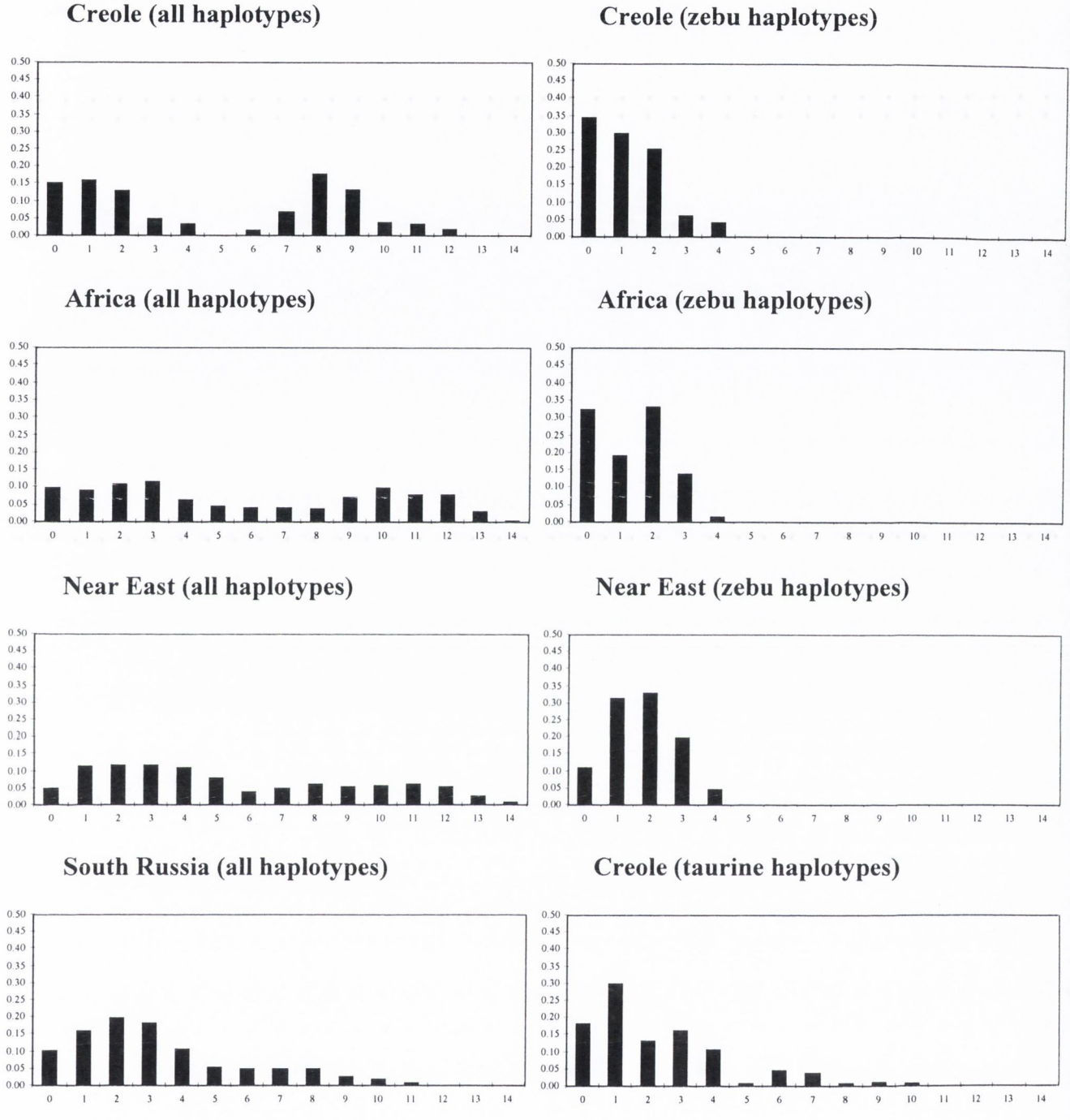


Figure 4.5 Frequency distributions of the number of mismatches observed in pairwise comparisons within those geographic groupings that contain both zebu and taurine haplotypes. The numbers of pairwise differences are given on the horizontal axis, with the frequency of each category represented on the vertical scale.

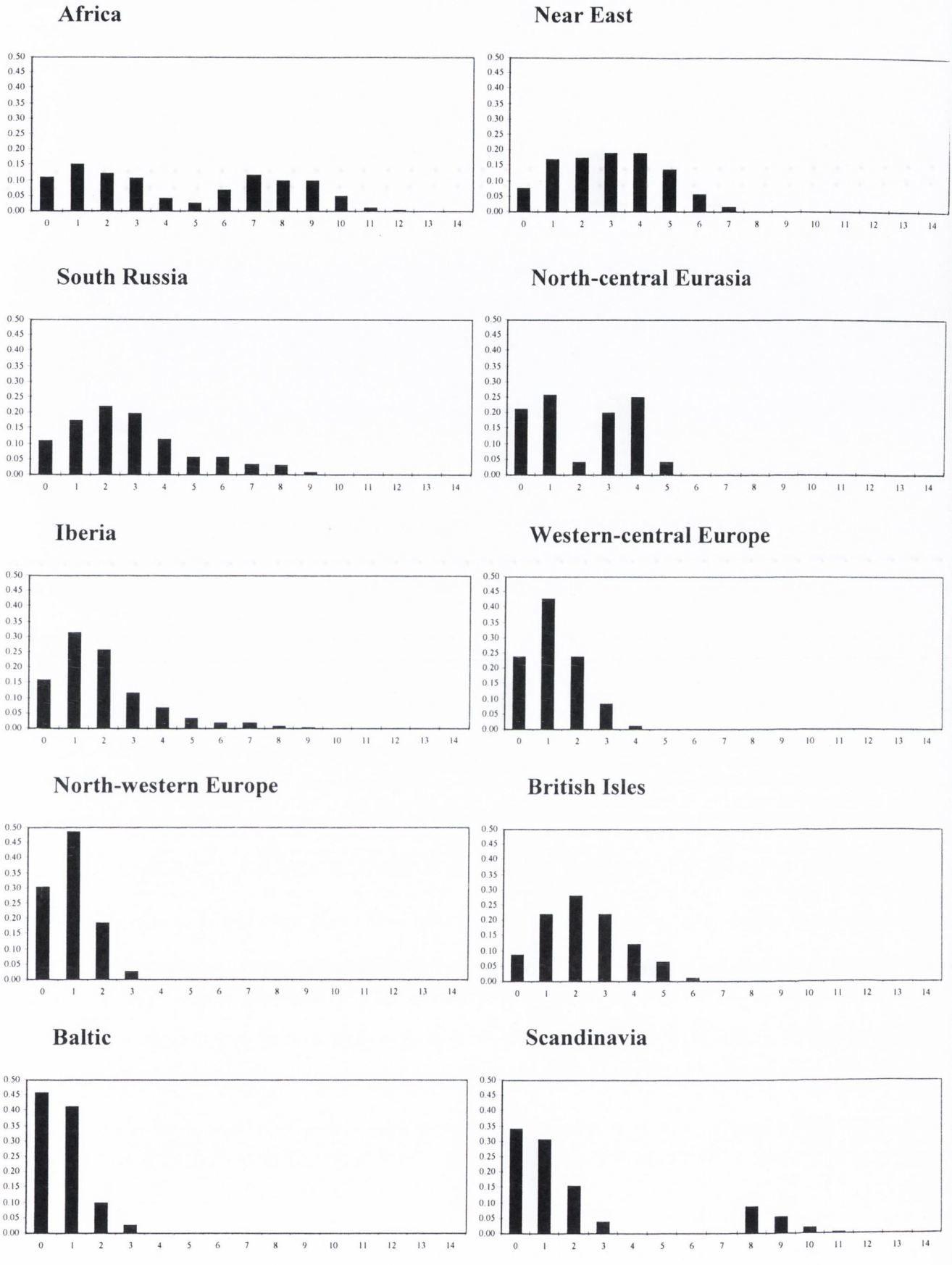


Figure 4.6 Frequency distributions of the number of mismatches observed in pairwise comparisons using taurine haplotypes from ten geographic groupings. The numbers of pairwise differences are given on the horizontal axis, with the frequency of each category represented on the vertical scale.

may be expected, the mean of the mixed zebu/taurine groups are generally higher than those seen when the data are partitioned into zebu and taurine groups. The zebu means are low, with the highest value seen in the Near East (1.75), which is reflected in the histograms shown in **Figure 4.5**. The bimodal taurine African mismatch mean (4.59) is the highest of the taurine means calculated. Of the other taurine groups, the Near East has the highest mean value (2.95), with the lowest mean in the Baltic group (0.70).

4.4.3 Network analyses

A total of 54 four-marker haplotypes were observed on the 678 chromosomes studied. 21 haplotypes were unique to individuals and were not shared between or within groups. Therefore, of all the 54 haplotypes, 39% were encountered only once. A reduced skeleton network was constructed to show all non-unique haplotypes (**Figure 4.7**). *INRA126* was reduced as this locus appears to be fast-mutating, a point that is confirmed by the high percentage of variation seen within populations (81%), and low differentiation among geographical groups (8%), as calculated by AMOVA analysis (**Table 4.7**). This can be explained thus: as the greatest diversity is seen within populations, the same mutation may have occurred more than once in several groups, resulting in extensive homoplasy. The other three loci (*INRA124*, *INRA189* and *BM861*) have not been reduced. Similar variation is exhibited among groups (over 20% in each case), a larger value than that seen in *INRA126*, indicating that there are larger genetic differences at each locus when compared across groups and each mutation is more likely to have arisen once only.

Source of variation	Percentage of variation						
	all data	zebu only	taurine only	*INRA124*	*INRA126*	*INRA189*	*BM861*
Among groups	22.84	10.11	15.14	26.31	7.68	26.66	21.03
Among populations within groups	42.79	6.48	42.15	54.78	11.67	45.91	50.96
Within populations	34.37	83.41	42.71	18.91	80.65	27.44	28.01

Table 4.7 Hierarchical AMOVA analysis, showing the percentage of variation at each of three levels of population hierarchy.

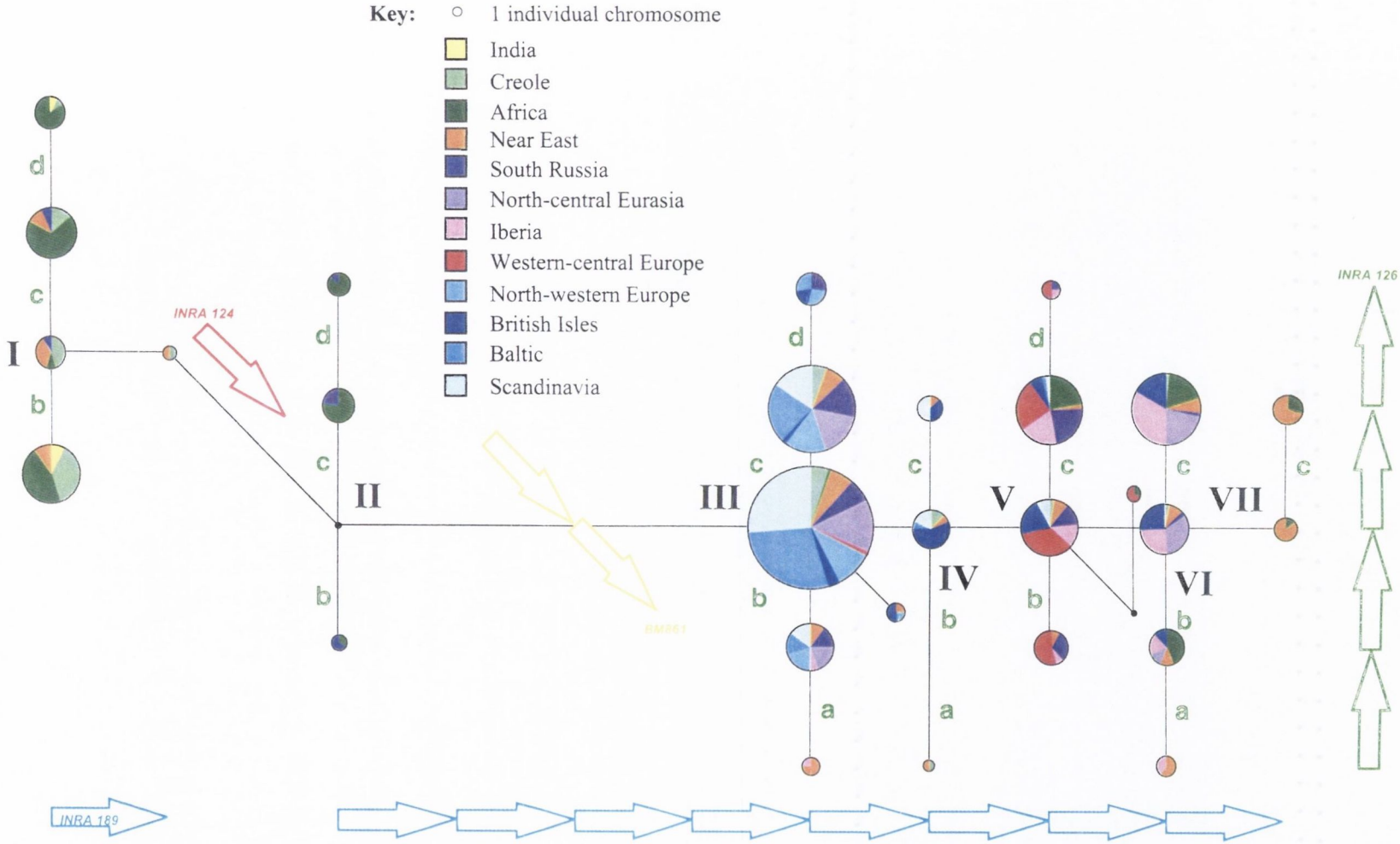


Figure 4.7 Reduced skeleton network of Y-chromosome haplotypes using four Y-specific microsatellites markers. The arrows represent the mutations seen in these markers, with the direction notifying an increase of two base pairs. The single *INRA124* mutation is shown as a red arrow, with *INRA126* in green, *INRA189* in turquoise and *BM861* in yellow. The network has been reduced at *INRA126*, and recurrent mutations are labelled a to d in green. Only haplotypes that are non-unique are shown and the size of each circle represents the number of chromosomes present in each haplotype. The pie charts show the number of individuals from each geographic group present at each haplotype. Roman numerals are used to indicate haplogroups defined by the allele present at *INRA189*, with I = 88 bp; II = 90 bp; III = 98 bp; IV = 100 bp; V = 102 bp; VI = 104 bp; and VII = 106 bp.

Over all geographic groups, the main taurine haplotype (148 individuals) was *INRA124* (132 bp): *INRA126* (184): *INRA189* (98): *BM861* (158), with the main zebu haplotype being 130:156:88:182 (40 individuals). With relationship to the skeleton network shown in **Figure 4.7**, these two major haplotypes are indicated by III and Ib respectively. Each of the twelve groups was assigned a colour and the network has been shaded accordingly, an approach which shows the differences seen across the network and strengthens the choice of reducing marker *INRA126*. Mutations seen in this marker are represented vertically (as shown by the green arrows to the right of the diagram).

Haplotypes with mutations in marker *INRA189* have been assigned haplogroup numbers I to VII, where I is a zebu haplogroup, containing the 130 bp allele at marker *INRA124* (shown by the red arrow), and haplogroups II to VII are taurine, with the 132 bp allele at *INRA124*. Zebu haplogroup I includes mainly individuals from Creole and Africa, but also has animals from India, the Near East and South Russia. Haplogroup II has individuals from Africa and, to a lesser extent, South Russia. Haplogroup III contains all groups with taurine chromosomes, with the main groups being South Russia, North-central Eurasia, North-western Europe, the Baltic and Scandinavia. Haplogroup IV is quite small in relation to III, V and VI, and has a similar distribution as haplogroup II; its main contributors are the British Isles and Scandinavia. Haplogroup V shows a sudden change from III and IV, with African, Iberian and Western-central European chromosomes exhibited. Africa, North-central Eurasia, Iberia and the British Isles are represented in haplogroup VI, with the small haplogroup VII having individuals from Africa and the Near East only. Chromosomes from the Near East are present in all haplogroups except II.

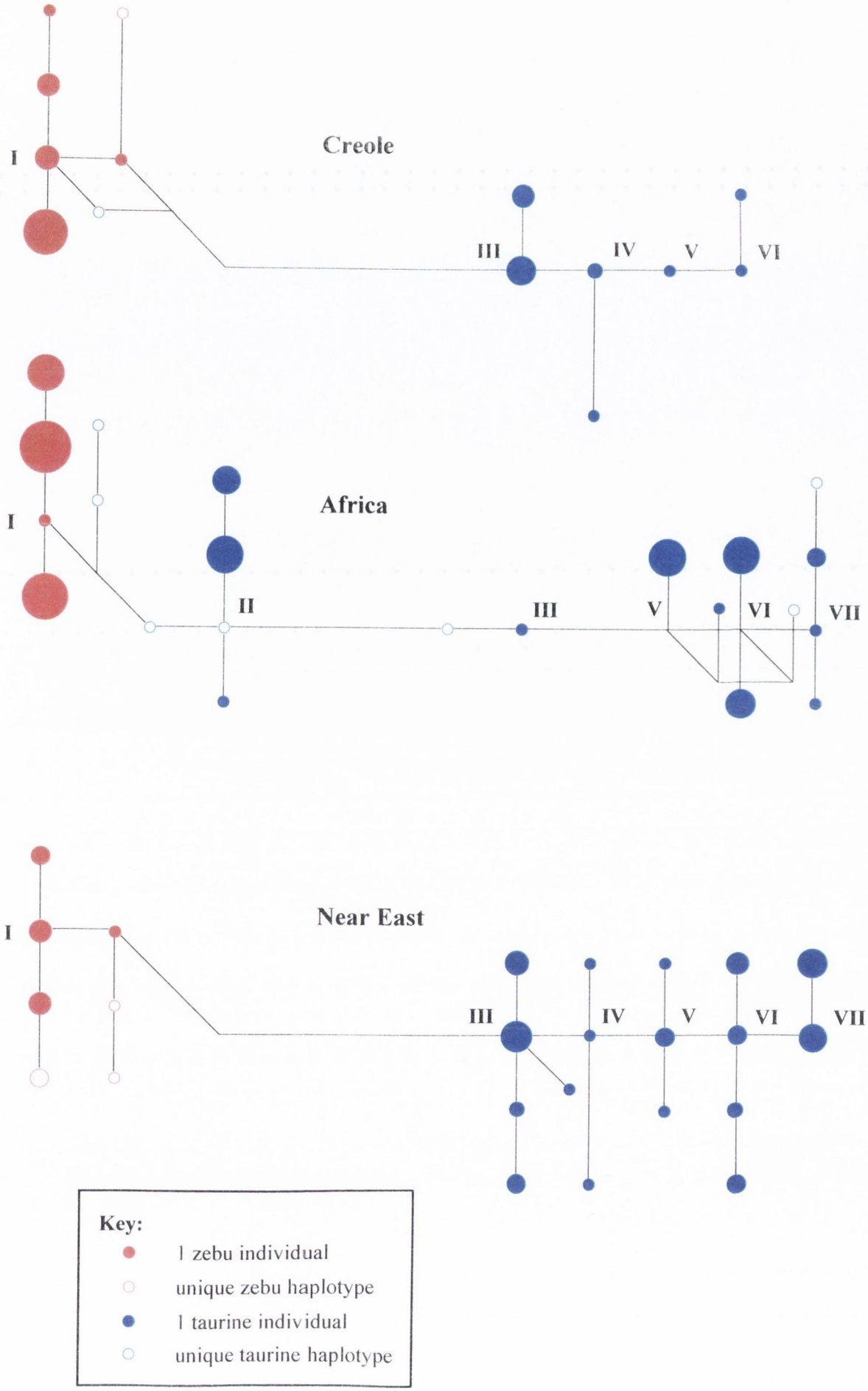
Rooting of the skeleton network was attempted by analysis of 29 bovid chromosomes taken from plains bison, mithan, swamp buffalo and yak. Eight unique haplotypes were found in these 29 individuals: three in plains bison, two each with mithan and swamp buffalo, and one in yak. With the alleles in base pairs and in the marker order [*INRA124*:*INRA126*:*INRA189*:*BM861*], the haplotypes are as follows: plains bison – [130:182:96:144], [130:182:98:146] and [132:182:96:144]; mithan – [130:182:68:150] and [130:182:68:156]; swamp buffalo – [132:184:84:158] and [132:184:96:158]; yak – [132:182:102:148]. These haplotypes were compared, marker-by-marker, against all alleles seen in the cattle breeds. In cattle, *INRA124* has two alleles, *INRA126* six, *INRA189* ten and *BM861* three. A score of zero denoted identity between the bovid allele and the marker allele under consideration, with an extra 1 added for each 2 bp mutation seen between the

two alleles. The cattle allele closest in size to each bovid allele was then compared for each of the unique bovid haplotypes to determine the haplotype of the root. However, although completion of this pointed toward the root allele at *INRA126* being 182 bp (closest allele in six of the eight bovid haplotypes), and 156 bp at *BM861* (again, six out of eight), there was no overall consensus for either the two alleles at *INRA124* or the ten alleles at *INRA189*.

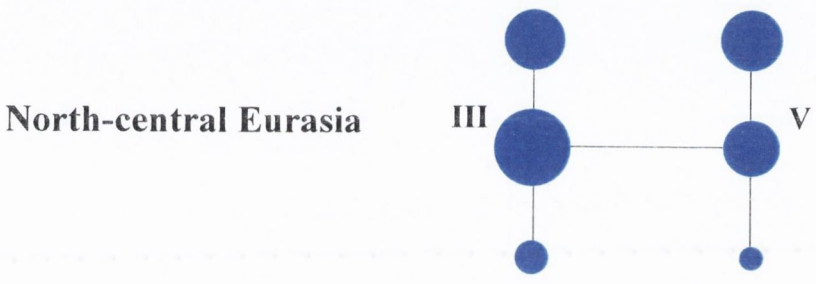
As the four Indian individuals are expected to be zebu in origin, the positioning of these animals in the skeleton network should indicate zebu haplotypes. Correspondingly, individuals from central Europe would be expected to exhibit taurine haplotypes. Without accounting for the zebu/taurine split said to be explained by the presence of either allele 130 bp or 132 bp at *INRA124* (Hanotte *et al.* 1997), when positions of the haplogroups are looked at in more detail, haplogroup II appears to be closer to the zebu haplogroup I than the taurine haplogroups III to VII. This is due to this haplogroup having allele 90 bp at *INRA189*. The *correct* positioning of this haplogroup, in relation to whether it should be classified as zebu or taurine, was tested using a similar method to that used to determine the root. The majority of the chromosomes in this haplogroup belong to the African group, and therefore six marker data from 100 African individuals were consulted. However, the extra two markers (*DYZ1* and *BYM-1*, typed by J. Kantanen) did not lend any additional information to those four already analysed as part of this study. The African individuals in haplogroup II all belong to the N'Dama breed, and it has previously been assumed that this breed, located in Guinea, West Africa, is a pure taurine breed (for example, Loftus *et al.* 1999).

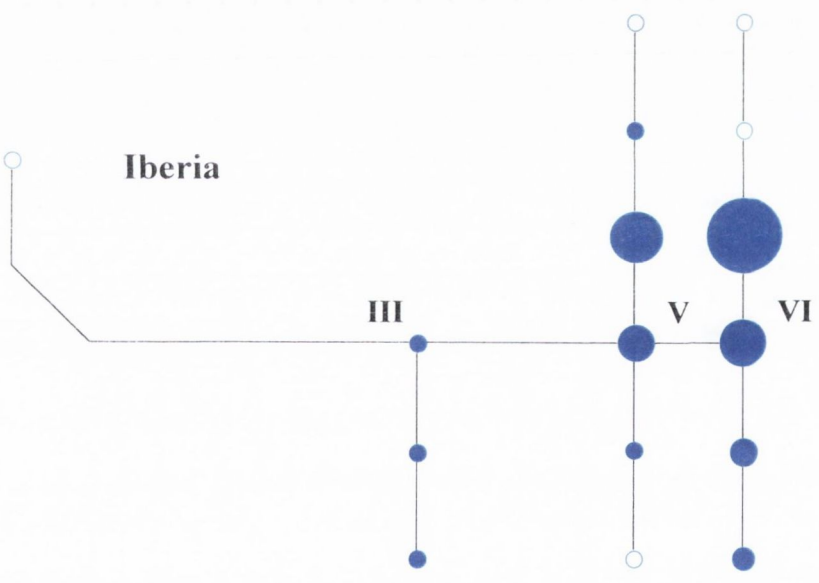
Networks were constructed for each geographic grouping separately (**Figure 4.8**), with unique haplotypes denoted by open circles. Using the *INRA124* discrimination between zebu and taurine haplotypes, with the presence of the 130 bp or 132 bp allele respectively, zebu haplotypes were coloured red and taurine haplotypes blue. The network of the four Indian zebu individuals was not drawn, but the location of these animals can be seen in **Figure 4.7**.

The Creole, African, Near Eastern and South Russian groups contain both zebu and taurine haplotypes. The most common zebu haplotype in the Creole group is haplogroup I plus mutation b at *INRA126*, with the most common taurine haplotype at the central haplogroup III. Two Creole individuals exhibit unique haplotypes, one of which is zebu, the other taurine. The most frequent zebu and taurine African haplotypes are Ic and Vc/Vic, with


Creole
I
III
IV
V
VI
Africa
I
II
III
V
VI
VII
Near East
I
III
IV
V
VI
VII
Key:
1 zebu individual
unique zebu haplotype
1 taurine individual
unique taurine haplotype


South Russia
I
II
III
IV
V
VI


North-central Eurasia
III
V


Iberia
III
V
VI

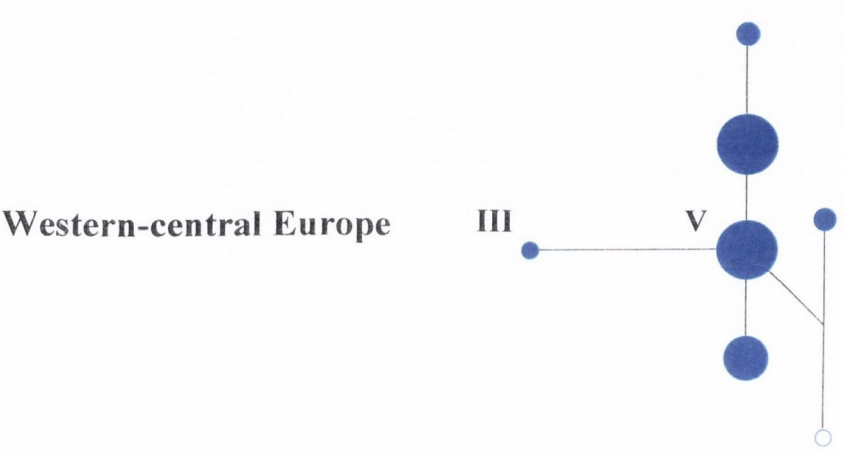

Western-central Europe
III
V



Figure 4.8 (previous three pages) Reduced networks of eleven geographical groups (see Table 4.2 for details). Using the marker *INRA124*, the haplotypes can be partitioned into zebu (red) and taurine (blue). Solid circles are haplotypes present in more than one group, with open circles denoting unique haplotypes. Roman numerals are used to indicate haplotypes dependent upon the allele present at *INRA189*, as used in Figure 4.7.

seven unique taurine haplotypes present. The Near East group has two most common zebu haplotypes, I and Ib, with the most common taurine haplotype being III. This group exhibits three unique haplotypes, all zebu, with three individuals typing as Ia. The South Russian group has two zebu haplotypes, from three individuals of the Bushuevskaya breed. This breed was historically founded from herds of Uzbekistan zebu cattle crossed with Friesian and Swiss Brown bulls, which were interbred to produce a recognised breed in the 1930's (Felius 1995). There is one unique taurine haplotype (IVd) and the most numerous taurine haplotype is Vc. The South Russian group contains animals at haplogroup II, as seen in the N'Dama individuals of the African network. These haplogroup II South Russian individuals belong to the Ukrainian Grey breed, and six out of seven of the chromosomes from this breed are haplogroup II. In a similar way to the presence of the predominantly African mitotype T1 in the Near East (Troy *et al.* 2001), it is not improbable to find the presence of a Y-chromosome haplogroup prevalent in the African N'Dama taurine breed in the Ukraine breed.

The remaining seven groups only contain taurine haplotypes. North-central Eurasia has two haplogroups only, III and V, with an equal partition between the two. The majority of the individuals in haplogroup V (17 out of 25) are from the peripheral Yakutian breed. Iberia displays more diversity in haplotypes, with five unique chromosomes present. The most common haplotype is VIc. The outlying individual belongs to the Alentejana breed and has allele 92 bp at marker *INRA189*. Western-central Europe has two haplogroups present (III and V), with one unique haplotype and one other individual showing allele 160 bp at locus *BM861*. The chromosome at haplogroup III belongs to an individual from the Pustertaler breed. North-western Europe contains haplogroup III only. The diversity of haplotypes in the British Isles group is greater than that seen in other taurine groups. The main haplotype is present at VIc, and of the four unique haplotypes, one contains four, and the other six, individuals. This group exhibits the most coherent breed structure, with haplogroup III containing Aberdeen Angus individuals only, haplogroup IV with Hereford animals alone, haplogroup VI including only Jersey chromosomes, and haplogroup V with a mixture of Highland and Galloway individuals. The majority of the Baltic group chromosomes are haplogroup III, with two individuals as outliers, at haplogroup IV and Vc. Scandinavia shows a similar distribution of haplotypes as the Baltic group, but with two outlying unique haplotypes, which have allele 82 bp at marker *INRA189*. These two haplotypes consist of the four Doela and two Telemark cattle sampled. The Doela breed is

locally derived from major crossing with the neighbouring Telemark, which is an old authentic breed from the region (Felius 1995).

The bimodal taurine mismatch distributions seen in **Figure 4.6** can be explained by consultation of the geographic networks of **Figure 4.8**. The African mismatch is caused by the presence of haplogroup II, and it is also this haplogroup that makes the South Russian distribution slightly bimodal in nature. The curves seen with the North-central Eurasia mismatch are caused by the presence of two haplogroups (III and V) both at high frequencies. The unlinked bimodal structure of the Scandinavian group is caused by the presence of the six individuals with allele 82 bp at *INRA189*.

4.4.4 *Relationships between geographic groups*

FST values and multidimensional scaling (MDS)

The fixation index, *FST*, is a measure of the genetic differentiation of subpopulations, and is the standardised variance in allele frequencies among populations. *FST* values were calculated for all haplotypic data, zebu haplotypes and taurine haplotypes (**Table 4.8**). For the matrix calculated from all the data (**Table 4.8a**), the greatest distance, of 0.903, is found between the Indian group (four zebu individuals from the Ongole breed) and the Baltic (containing taurine haplotypes). The majority of large *FST* values are between groups containing both zebu and taurine haplotypes and those with taurine individuals alone. All the *FST* values are low when the zebu data are considered (**Table 4.8b**), indicating close relationships between the zebu individuals in each of the five groups, although the small sample numbers (as shown in **Table 4.5**) may influence the data. When the taurine data are considered (**Table 4.8c**), the two groups with the closest relationship are the Baltic and North-western Europe group, with a value of 0.005, which may be expected due to the geographic location of these two groups (**Figure 4.1**). On the whole, the taurine genetic diversities are low, with the highest values seen between Western-central Europe and the Baltic and North-western European groups (0.677 and 0.636 respectively), and between Iberia and the Baltic (0.677). The taurine Near East haplotypes most resemble the British Isles (0.047) and North-central Eurasia (0.048), with Africa's lowest genetic diversity value exhibited when compared to South Russia (0.109), which may in part be explained by the sharing of haplotypes with allele 90 bp at *INRA189* (**Figure 4.7**).

a)

	India	Creole	Africa	Near East	South Russia	North-central Eurasia	Iberia	Western-central Europe	North-western Europe	British Isles	Baltic
Creole	0.100										
Africa	0.169	0.061									
Near East	0.427	0.202	0.153								
South Russia	0.629	0.340	0.231	0.068							
North-central Eurasia	0.773	0.477	0.349	0.101	0.080						
Iberia	0.809	0.570	0.399	0.197	0.254	0.217					
Western central Europe	0.873	0.567	0.405	0.212	0.230	0.306	0.183				
North-western Europe	0.870	0.457	0.357	0.190	0.137	0.216	0.557	0.636			
British Isles	0.772	0.499	0.365	0.120	0.120	0.069	0.114	0.132	0.359		
Baltic	0.903	0.534	0.411	0.233	0.186	0.249	0.609	0.677	0.005	0.398	
Scandinavia	0.743	0.411	0.337	0.152	0.105	0.163	0.473	0.472	0.050	0.272	0.041

b)

	India	Creole	Africa	Near East
Creole	-0.034			
Africa	0.097	0.120		
Near East	-0.032	0.015	0.171	
South Russia	0.228	0.216	-0.041	0.156

c)

	Creole	Africa	Near East	South Russia	North-central Eurasia	North-central Eurasia*	Iberia	Western-central Europe	North-western Europe	British Isles	Baltic
Africa	0.175										
Near East	0.122	0.153									
South Russia	0.037	0.109	0.120								
North-central Eurasia	0.051	0.171	0.048	0.070							
North-central Eurasia*	nd	nd	0.158*	0.084*	nd						
Iberia	0.366	0.169	0.120	0.255	0.217	0.408*					
Western central Europe	0.413	0.235	0.199	0.231	0.306	0.468*	0.183				
North-western Europe	0.093	0.314	0.311	0.152	0.216	0.075*	0.557	0.636			
British Isles	0.138	0.169	0.047	0.110	0.069	0.193*	0.114	0.132	0.359		
Baltic	0.101	0.384	0.362	0.203	0.249	0.076*	0.609	0.677	0.005	0.398	
Scandinavia	0.007	0.282	0.253	0.119	0.163	nd	0.473	0.472	0.050	0.272	0.041
Scandinavia†	nd	nd	0.241†	0.112†	nd	0.051*†	0.459†	0.455†	0.059†	0.261†	0.054†

Table 4.8 Genetic distance estimates between the geographic groupings using: **a)** all the data; **b)** geographic groups containing individuals with zebu haplotypes; and **c)** geographic groups containing individuals with taurine haplotypes. In Table **c)**, * = North-central Eurasia minus the Yakutian breed, and † = Scandinavia minus the Icelandic breed. Those values in red were used to calculate the PC1 values shown in Table 4.6, while those values not suffixed with a symbol were used to draw the MDS plot shown in Figure 4.11c.

Calculation of genetic distance is usually followed by the construction of phylogenetic trees that show the main divergences seen between different populations or species. However, this approach does not consider the effects of admixture upon populations residing on different branches. Multidimensional scaling reconstructs a *map* representing the relationships between populations, which may have been under the influence of gene flow from external sources, from a table of differences between them. Two-dimensional

MDS plots were drawn from each pairwise *FST* matrix (**Figure 4.9**). When all the data are considered (**Figure 4.9a**), variable 1, along the *x*-axis, appears to differentiate between those groups with a mixture of zebu and taurine haplotypes (right-hand side of plot), with the pure zebu Indian group furthest to the right, and the pure taurine groups (left-hand side of the plot). The second variable (*y*-axis) seems to distribute the European taurine data, with the British Isles, Iberia and Western-central Europe to the bottom of the plot, the western fringe groups (North-western Europe, the Baltic and Scandinavia) clustering to the top of the plot, and Near Eastern and Eurasian groups between the two. The r^2 value calculated for this plot, that is the amount of variance described by the data, is 97.8%. The first two variables of the zebu plot account for 97.8% of the diversity seen, but there is no real structure here (**Figure 4.9b**), most probably due to the small sample sizes involved (**Table 4.5**). The first two variables of the taurine plot account for 98.6% of the diversity seen. The first variable in **Figure 4.9c** places the Near East group between Western-central Europe, Iberia and the British Isles on the left-hand side of the plot, and the cluster of northern and eastern European groups on the far-right of the plot. The British Isles group closely with the Near East, and the Western Fringe groups cluster together, as seen in the plot containing all data. Variable 2 appears to distinguish the African taurine from the European taurine groups, with the deviation of Iberia toward the African group and away from the zero *y*-axis due to possible admixture of African alleles into the Iberian population, as has been reported for mitochondrial data (Cymbron *et al.* 1999).

AMOVA analysis

AMOVA is based upon the mutational difference between pairs of different haplotypes and yields values equating to the haplotypic diversity within and between populations. Keeping the geographic groupings demarked in **Table 4.2**, AMOVA values were calculated for all data, zebu haplotypes and taurine haplotypes (**Table 4.7**). When applied to the complete data, a larger mean proportion of variance is seen among breeds within the twelve groups (42.79%), with the smallest proportion of the data (22.84%) accounted for by differences seen between haplotypes present between the diverse groups. This is mirrored in the taurine calculations, with 42.15% and 15.14% respectively. In contrast, when the zebu haplotypes are considered, the majority of the difference seen in the data (83.41%) occur between breeds within geographic groups. The difference seen among groups is, however, again a low value, at 10.11%. All variance components were significantly greater than random allocations of haplotypes, with $P < 0.001$, $P < 0.03$ and $P < 0.001$ for all data, zebu data and taurine data respectively.

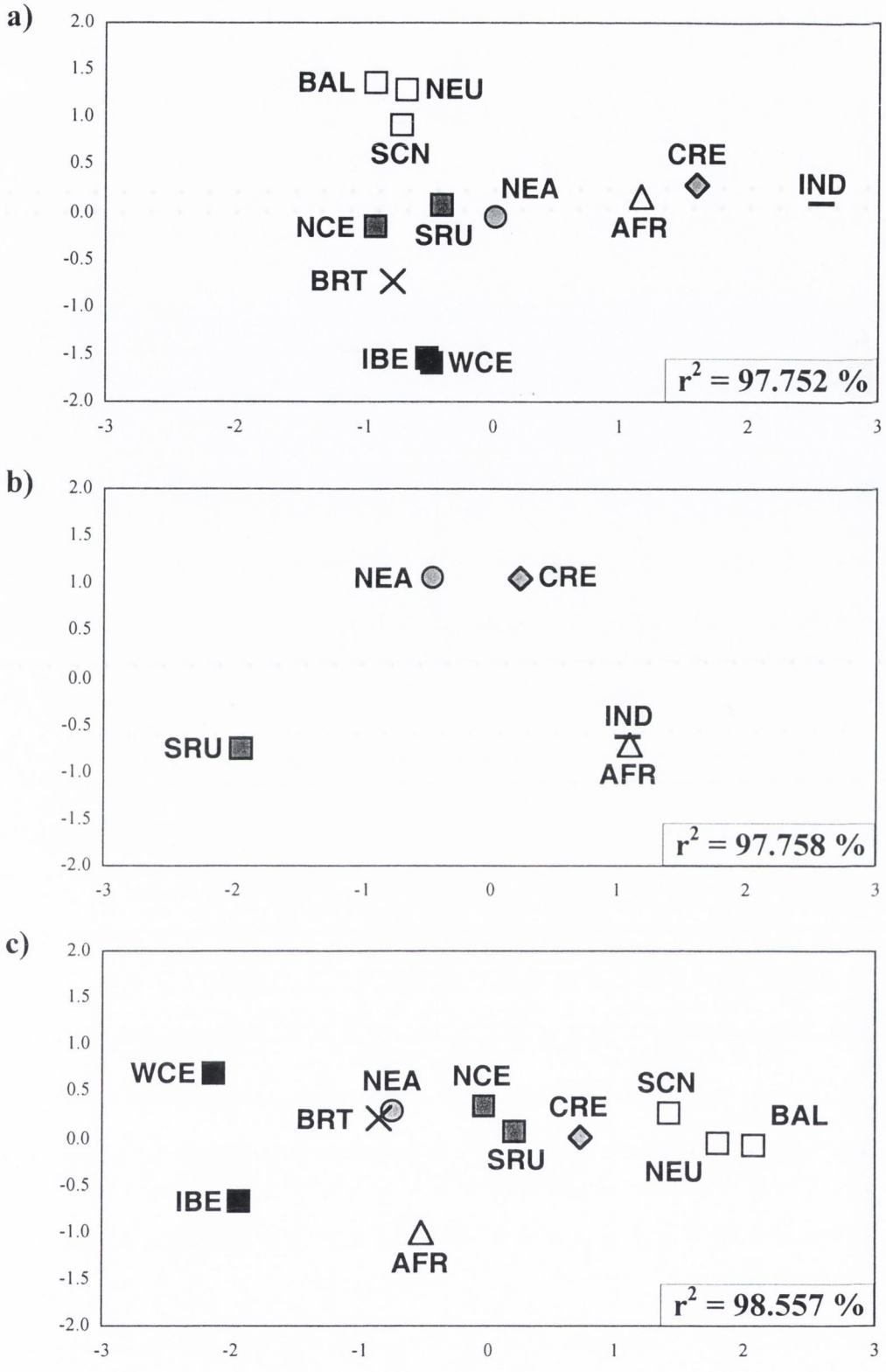


Figure 4.9 Two-dimensional MDS plots, drawn using data from four microsatellites, summarising sum of size differences among groups of modern cattle breeds, including: **a)** all individuals typed; **b)** zebu haplotypes; and **c)** taurine haplotypes. Squares represent Continental European groups. The three-letter codes are as follows: IND, India; CRE, Creole; AFR, Africa; NEA, Near East; SRU, South Russia; NCE, North-central Eurasia; IBE, Iberia; WCE, Western-central Europe; NEU, North-western Europe; BRT, British Isles; BAL, Baltic; SCN, Scandinavia. Breeds included in these groupings are detailed in Table 4.2. r^2 values are shown for each plot.

Geographic analysis using synthetic map drawing

The taurine data were analysed using the synthetic map drawing program, ArcView, with the following groups under consideration: Near East, South Russia, North-central Eurasia, Iberia, Western-central Europe, North-western Europe, British Isles, Baltic and Scandinavia. Central locations for each group were estimated (**Figure 4.10a**) by consideration of the breeds present in each group (**Figure 4.1**) and, for this reason, the outlying populations of the Yakutian breed, in North-central Eurasia, and the Icelandic breed, in Scandinavia, were removed from subsequent investigations. The pairwise *FST* genetic distances shown in red in **Table 4.8c** were used to calculate the first principal coordinate (PC1) using MDS for the nine groups (**Table 4.6**). The ranking of the PC1 values shows the Near Eastern group with a value mid-way between the Iberian, West-central European and British Isles groups at one extreme, and the remaining groups at the other, a distribution that is mirrored in the taurine MDS plot (**Figure 4.9c**). The amount of variance described by PC1, the r^2 value, was calculated at 96%. A synthetic map was constructed using interpolated PC1 values to illustrate the spread of data (**Figure 4.10b**). Genetic diversity and MNA values, as well as the mean number of mutations calculated from mismatch distributions, were regenerated for North-central Eurasia and Scandinavia, minus the Yakutian and Icelandic breeds respectively (**Table 4.6**). Although the MNA values for these newly constructed groups did not alter, both the genetic diversity and mean mismatch values decreased in the North-central Eurasian group once the Yakutian breed had been removed (0.270 to 0.211, and 2.15 to 1.51, respectively), whereas the opposite was true for the Scandinavian group minus the Icelandic breed (0.212 to 0.218, and 2.11 to 2.35). Synthetic maps were constructed using the genetic diversity, MNA and mean mismatch figures shown in **Table 4.6**. As can be seen from **Figure 4.10c**, genetic diversity values are greater along the southern and western edge of Europe, in comparison to the low values in north-eastern and central Eurasia. A similar pattern is seen when the MNA values are superimposed upon the map of Eurasia (**Figure 4.10d**). Although the outliers seen in the Scandinavian network (**Figure 4.8**) produce a high mismatch mean (2.35), with an associated peak in colour intensity, the pattern seen when the mean mismatch figures are interpolated is also comparable to the genetic diversity map (**Figure 4.10e**).

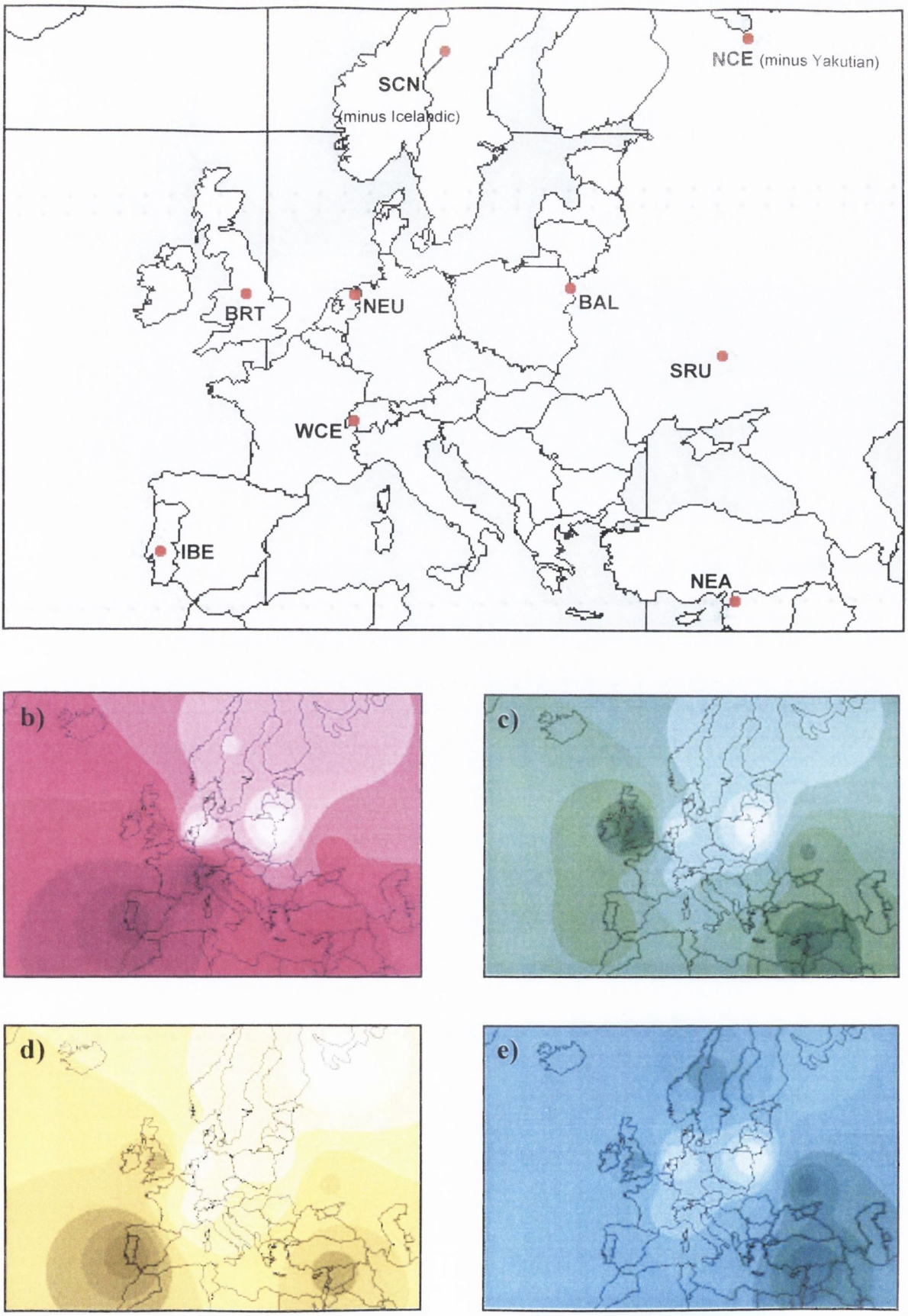


Figure 4.10 a) Central locations of the nine geographic groups of cattle breeds. Three letter codes are as for Table 4.2. Synthetic maps illustrating the taurine geographic variation of the following statistical variants (taken from Table 4.6): **b)** the first principal coordinate, which accounts for 96% of the total allelic variation; **c)** genetic diversity; **d)** mean number of alleles (MNA); and **e)** the mean number of mutations calculated from mismatch distributions. The colour gradient increases for maps **c)**, **d)** and **e)** with increase in relative value used.

4.5 Discussion

4.5.1 *Distribution of zebu and taurine haplotypes*

As Hanotte *et al.* (1997) showed, *INRA124* displays taurine- and zebu-diagnostic alleles. The genetic dichotomy often observed between zebu and taurine cattle is useful for detecting gene flow and introgression between the two taxa (MacHugh *et al.* 1997; Bradley *et al.* 1998). Both *BM861* (a similar marker to *INRA124*) and *INRA189* seem to show allele size variants specific to taurine and zebu in the cattle breeds studied (**Tables 4.3** and **4.4**). Although these loci have zebu- and taurine-diagnostic alleles, the presence of these alleles in related bovid species cannot be taken to be meaningful and care must be taken before making assumptions about inter-specific relationships based on these microsatellite data. There are at least three reasons why two samples can possess the same length variant at any one locus. These are: common descent from the same ancestor; introgressive hybridisation; and convergent evolution (or independent mutation; that is, identical in state but not identical by descent).

Previous genetic studies have demonstrated the depth of separation between zebu and taurine cattle, using mitochondrial (Loftus *et al.* 1994a, 1994b; Bradley *et al.* 1996), Y-chromosome (Bradley *et al.* 1994; Teale *et al.* 1995) and nuclear (MacHugh *et al.* 1997) markers. Using the *INRA124* marker to differentiate between zebu and taurine Y-chromosomes, the results presented here support the independent domestications of zebu and taurine cattle, as well as the geographic component to the distribution of these two subspecies noted by other researchers. It has been proposed that zebu cattle may have had two centres of domestication (Allchin 1969), and analysis of mitochondrial DNA from a range of zebu breeds has lent support to this (Troy 1998; Magee 2002), with the presence of two mitotypes, called Z1 and Z2, distributed clinally across India. Unfortunately, results generated here are not sufficient to make any concrete conclusions about two culturally-linked domestication processes in zebu as only four individuals from one Indian breed were analysed.

European cattle breeds exhibited exclusively taurine Y-chromosomal haplotypes, with Creole, Africa, Near East and South Russia containing a mixture of both zebu and taurine haplotypes (**Table 4.5**). The presence of zebu mitotypes has been shown previously in the Creole breeds by Magee *et al.* (2002), with five out of 78 individuals (6.4%) exhibiting

zebu ancestry. The same three Creole breeds (Antigua, Guadeloupe and St. Lucia), plus one Brazilian breed (the Nellore, imported from India in the 19th century; Felius 1995), were analysed here, with 27 out 45 (60%) having zebu chromosomes (**Table 4.5**). The presence of zebu haplotypes in the Creole cattle can be attributed to the importations of West African zebu cattle to the region subsequent to the 17th and 18th century introductions of taurine European and African cattle (Felius 1995). Zebu cattle are recent migrants to Africa, with two separate introgression events: the first around 1,500 BC and the second linked to the Arab invasion in 699 AD (Epstein 1971). The distribution of zebu Y-chromosomes has been studied using locus *INRA124* by Hanotte *et al.* (2000), with human migration, habitat and disease adaptation, and phenotypic preferences by pastoralists explaining the geographic dispersal of zebu alleles across Africa, with the cline observed running in an east-west direction across Africa, and in a north-south direction in West Africa. Further analysis of nuclear microsatellites by Hanotte and colleagues (2002) has detailed the initial expansion of *B. taurus* individuals from a single region of origin, most probably in the north-east of the continent, with *B. indicus* animals entering Africa through the Horn and the East Coast. Zebu and taurine haplotypes have previously been noted in the mitochondrial and nuclear DNA of the Near East (Troy 1998; Loftus *et al.* 1999), and their distribution patterns indicate an introduction of zebu animals from their domestication centre in the east. Zebu individuals in the South Russian group all originate from the Bushuevskaya breed, which is recognised as a zeboid breed due to their origin from a zebu breed of Uzbekistan (Felius 1995).

4.5.2 *Trends in diversity among groups*

When taurine haplotypes are considered, the Near East emerges as the most diverse group sampled (with the exception of the British Isles; **Table 4.6**); a feature that also holds true for both nuclear (Loftus *et al.* 1999) and mitochondrial (Troy *et al.* 2001) analyses. When compared to the other groups, analysis of the distribution of haplotypes in the Near East supports the suggestion that this region was a major centre of cattle domestication (**Figure 4.10**). Analyses of the Y-chromosome have added information to data collected from studies of the mitochondrial (Loftus *et al.* 1994a, 1994b; Bradley *et al.* 1996) and nuclear (MacHugh *et al.* 1994, 1997, 1998) components of the genome.

Africa

Taurine animals in Africa are thought to represent descendants of the herds maintained by the first agriculturalists in the continent, having originated from the Near East and arriving in the Nile Delta around 6,000 BC (Epstein 1971). The prehistory of cattle herding in Africa indicates that the earliest domesticates were taurine, with the presence of zebu animals due to later migrations (Epstein 1971; Payne 1991). However, this viewpoint has recently been in question and a local independent domestication of African aurochs (*B. primigenius opisthonomus*) has been proposed from both archaeological (Bökönyi 1976; Grigson 1991; Meadow 1993; Wendorf & Schild 1994) and genetic (Bradley *et al.* 1996; Hanotte *et al.* 2002) evidence. Using mitochondrial data, it has been estimated that at least 22,000 years separates the taurine African and European domestic cattle strains (Bradley *et al.* 1996), and this divergence increases to between 150,000 and 250,000 years when nuclear microsatellites are considered (MacHugh *et al.* 1997). Although the genetic data support an autochthonous domestication of African cattle, the pattern observed could also be due to genetic exchange between domestic populations emerging from the Near East and indigenous African aurochs.

In contrast to the diversity patterns seen in Europe, the African group shares few haplotypes with the Near East, and contains seven unique haplotypes (**Figure 4.10**; **Table 4.5**). Only the African group, and individuals from the Ukrainian Grey breed of the South Russian group, exhibit chromosomes in haplogroup II. African animals in this haplogroup belong to the West African N'Dama breed from Guinea (**Figure 4.9**), and may point towards an independent origin for African domestic cattle. However, it was also considered that the proximity of this haplogroup to haplogroup I (**Figure 4.7**), which includes known zebu and zebu-admixed individuals, may reflect a zebu ancestry of these animals. The 22 N'Dama chromosomes typed here with four microsatellite loci have previously been analysed by Southern blotting techniques using the Y-specific probe *btDYZ-1* (Bradley *et al.* 1994). In a system that does not rely on the presence of putative taurine and indicine alleles, all Guinean N'Dama animals typed contained taurine Y-chromosomes. Autosomal and mitochondrial analyses of representative individuals from this breed have also failed to discover zebu admixture (MacHugh *et al.* 1997 and Troy *et al.* 2001 respectively).

Europe

European cattle groups share the majority of their haplotypes with the Near Eastern group (**Figure 4.10**), which relates to the idea that European cattle are a subset of an initially diverse Near Eastern domestic population. On the whole, the haplotypes observed show a general homogeneity across central Europe, and, although there is little substructuring within European cattle, there appears to be a loss of both haplotypes and diversity with relation to the distance of the geographic group from the Near East. Exceptions to this general rule include the Iberian and British Isles groups (**Table 4.6**; **Figures 4.3** and **4.12**).

The indigenous cattle of the Iberian peninsula are thought to share a common Near Eastern origin with the rest of Europe, although there is conjecture that local domestication also contributed to the gene pool (Davidson 1989). Although this hypothesis can not be supported by the data presented here, the Iberian group does exhibit a diverse network structure (**Figure 4.10**), with high diversity values (**Table 4.6**). The presence of a unique Y-haplotype containing allele 92 at *INRA189* may be linked to the N'Dama haplogroup II, with allele 90 bp at this marker. Analysis of Iberian mitochondrial DNA by Cymbron *et al.* (1999) has exposed introgression of African animals into the peninsula, with 11 out of 49 individuals (22.4%) having African mitotypes, which most likely occurred during the Arabian Moor occupation of the 8th to 15th centuries. The presence of putative African and European haplotypes in the Iberian group may account for the increased levels of diversity with comparison to the Near East.

The British Isles group also has increased diversity when compared to the rest of Europe (**Table 4.6**; **Figure 4.3**). This might be explained as possible local introgression from the autochthonous aurochs of the region, however, the haplotypes exhibited in the British Isles group are similar to those seen in other European groups (**Figure 4.7**) and this scenario is improbable. The high diversity is more likely to be the result of a fast migratory route out of the Near East, whereby there was less time for loss of genetic diversity. The Scandinavian group contains two highly diverse haplotypes (**Figure 4.8**), which are yet to be found in any extant cattle, and it is more probable that these are present as a result of local aurochs or wisent introgression, especially when the small allele sizes seen at locus *INRA189* in mithan and swamp buffalo (**Table 4.3**) are considered.

4.5.3 *Movement of cattle across Europe in the Neolithic*

Archaeological evidence indicates that agriculture spread out of the Near East, with the first farming settlements appearing in Greece around 7,000 BC, followed at about 5,000 BC by habitation of the plains of south-eastern Europe by groups of the Starcevo-Körös tradition (Waterbolk 1968) and then by the expansion of the Bandkeramik culture after 6,400 BP (Zilhão 2001). Genetic evidence from humans suggests that the spread of farming into Europe was demic, rather than cultural (Sokal *et al.* 1991; Cavalli-Sforza *et al.* 1993). European cattle are derived from aurochsen domesticated in the Near East between 10,000 and 8,000 BP, and similar protein polymorphism studies in cattle have also indicated large-scale cattle population movements in the early Neolithic (Medjugorac *et al.* 1994). In the south-east of Europe, environmental conditions were similar to those of the Near East, and domestic crops and livestock would have migrated easily. The spread of farming into temperate Europe, however, would not have been a steady progression, as farming practices had to be adapted to the northern climate, with development of new strains of cereal and greater reliance upon cattle and pig, which were to be found wild in the temperate forests, unlike sheep and goats. This distinction between the warm southern coast of Europe and colder central Europe led to differences in the speed of transfer of farming practices across the continent, and it is generally accepted that agriculture reached Central and Western Europe by two routes: the *Mediterranean* (also known as the *Western Neolithic* or *coastal*) route and the *Danubian* (or *continental*) route.

Waterbolk (1968) estimated that it may have taken as much as 3,000 years for village farming to travel by way of diffusion from the Aegean to the North Sea along the coastal route. However, Zilhão (2001) has dated the first appearance of the Neolithic package in central Italy and Portugal to 5,400 BC, with an even earlier date of 6,000 BC for southern Italy. He believes the rapid diffusion to be due to what he describes as *maritime pioneer colonisation*. Although the shorter continental route was not faster due to geographic barriers (for example, mountains) and expansion into hunter-gatherer territory, agricultural communities had reached the edge of the German lowland plain by 4,400 BC (Waterbolk 1968).

The position of Britain, with regard to the two colonisation routes, appears to mirror the east-west split seen in the Paris Basin, evidence of the mixing of communities travelling via both routes, with cattle dominating the faunal remains in the west (Tresset 2000). This possible suture zone in Britain between Neolithic cultures migrating via the Danube and

Mediterranean may explain the higher diversity values seen in modern cattle populations. The Jersey breed has previously been described as belonging to breeds that migrated via the Mediterranean and south-west Europe to the western part of the continent and to the offshore isles (Astolfi *et al.* 1983), and their distinctness has been reported from nuclear analysis (MacHugh *et al.* 1994). However, this nuclear microsatellite analysis found the British Isles to display reduced levels of genetic variation, which shows a disparity with the Y-chromosome data presented here.

The results shown here appear to offer evidence of the two migratory routes of domesticated Near Eastern cattle into Europe, not least in the diversity patterns shown in **Figure 4.10**. The speed of migration - slow across the Danubian route, quick along the Mediterranean route – is highlighted by the elevated diversity values in the southern and north-western groups (Iberia, British Isles and Western-central Europe) when compared to the eastern and central groups (North-central Eurasia, North-western Europe, the Baltic and Scandinavia) (**Table 4.6**; **Figure 4.3**).

4.5.4 *Future work*

At the time of designing this Y-chromosome project, at least eight dinucleotide $(CA)_n$ microsatellite repeats arrays had been assigned to the bovine Y-chromosome. As well as the four markers used here, *INRA008*, *INRA057* and *INRA062* (Vaiman *et al.* 1994), and *BTIGA50* (Mezzelani *et al.* 1995) have also been assessed for polymorphism (Edwards 1998), but were found to be present in multiple copies on the Y-chromosome. Subsequent to concluding analyses of the breeds presented here, J. Kantanen evaluated a further two markers, *DYZ1* (Perret *et al.* 1990) and *BYM-1* (Ward *et al.* 2001). In 2002, Liu *et al.* published a radiation hybrid map for the bovine Y-chromosome and, as a result, 28 new microsatellite markers located outside the pseudoautosomal region are now available.

CHAPTER 5

Relationships between the endangered Pustertaler-Sprinzen and three related European cattle breeds as analysed with 20 microsatellite loci

5.1 Abstract

Genetic relationships between the endangered German Pustertaler Sprinzen cattle breed and the Pinzgauer, Vosges and Simmental breeds were estimated - decided upon after consultation of the available historical literature. Within-breed diversity of the four breeds was also assessed. 20 microsatellite markers were amplified in 27-50 unrelated individuals from populations of each breed. Within-breed variation was estimated from average heterozygosity values and mean number of alleles. Breed relationships were evaluated by genetic distance and a neighbour-joining tree was calculated from these estimates. The robustness of the tree topology obtained was tested by bootstrap resampling of loci. A tree was also constructed from distance matrices using individual animals as operational taxonomic units. From both the average heterozygosity values and mean number of alleles calculated, the Pustertaler breed appears to be no more genetically impoverished than the other breeds analysed. The breed tree showed an 85% support for the Pustertaler-Pinzgauer grouping, and this result is echoed in the genetic distance values and allele-sharing individual tree.

5.2 Introduction

To determine which breeds should be selected for conservation, it is necessary to devise objective criteria with which to assess variability, especially as breed conservation is a complex problem and involves assessing several contributing factors; for example, cost of implementation, and cultural and economic importance of the breeds involved. Although criticism has recently been levelled at the reliability and relevance of microsatellite markers for conservation programmes (Ruane 1999), the results obtained from such surveys can provide useful insights into population structures and may help address specific conservation questions. When using microsatellite data, it is important not only to look at *taxonomic distinctiveness* (Hall & Bradley 1995) and between population variation, but also at within-population diversity, as measured by the average heterozygosities and mean numbers of alleles calculated from microsatellite analyses. Worldwide there are more than 780 cattle breeds, with 305 of these being native to Europe (Scherf 1995). It has been shown by microsatellite methods that, despite gene flow between populations, European cattle represent separate gene pools and possess unique combinations of alleles (MacHugh *et al.* 1994, 1998). This is most likely due to the introduction of herd books and selective breeding in the early 19th century, and more recent breeding technology.

The Pustertaler cattle are an authentic Shortheaded breed of cattle originating in the Puster Valley of the South Tyrol. The breed declined in the late 19th century and were subsequently outcrossed with both Simmental and Pinzgauer cattle (Felius 1995) with a view to increasing breed size. The total population size was estimated at 60 individuals in 1994, and the last Pustertaler animals are now seriously endangered and enjoy protected status under the *critically-maintained* World Watch List of the Food and Agriculture Organisation (Scherf 1995). Animals from the Vosges breed were included in this study as it has been suggested that this breed originated from cattle introduced during the Napoleonic era (Sambraus 1992), but others believe Vosges to represent the ancestral Pustertaler population (J. Knox personal communication).

5.3 Materials & Methods

Genotype data from 20 microsatellite DNA markers typed in 138 animals have been analysed here to assess the population structure of these European cattle breeds. Hair or blood samples were collected from between 27 and 50 individuals from each of the four breeds: Pustertaler, Pinzgauer, Vosges and Simmental. As far as could be ascertained, unrelated animals were chosen, and genomic DNA was extracted as described previously (Loftus *et al.* 1999). The following number of animals were sampled for each breed: 31 Pustertaler (24 from Italy and seven from Germany), 30 Pinzgauer (20 from Germany and ten from Austria), and 27 Vosges and 50 Simmental, both from France. The 20 microsatellites chosen for this analysis were: *BM1818*, *BM2113*, *BM1824*, *CSSM66*, *CSRM60*, *ETH3*, *ETH10*, *ETH152*, *ETH225*, *HEL1*, *HEL5*, *HEL9*, *HEL13*, *ILSTS005*, *ILSTS006*, *INRA005*, *INRA023*, *INRA032*, *INRA035*, and *INRA063*. This represents a panel of 20 loci from 30 markers agreed by the International Society for Animal Genetics (ISAG, Tours France, July 1996) from a collaborative study of variability in European cattle (*http://www.ri.bbsrc.ac.uk/cdiv_www/homepage.htm*). The data from the Simmental breed have been published previously (Schmid *et al.* 1999).

To assess the levels of variability at the DNA level in the four breeds, both the mean number of alleles (MNA) and gene diversity (average heterozygosity) values were measured. Unbiased estimates of expected gene diversity, with associated standard errors, were calculated according to Nei (1987), and observed heterozygosity values were calculated from the data. Allele frequencies were determined by direct counting. Interbreed genetic distances (D_A) were estimated as described in Nei *et al.* (1983) using the DISPAN computer program (T. Ota, Center for Human Genetics, Boston University). D_A distances are considered the most appropriate with which to obtain correct tree topology from microsatellite data (Takezaki & Nei 1996), especially when drift is an issue. The neighbour-joining method (Saitou & Nei 1987) was used to construct a dendrogram of breed relationships from the genetic distances using the PHYLIP package (Felsenstein 1989), and a bootstrap value was generated by resampling loci over 1,000 replicates. Another neighbour-joining tree was constructed using the allele-sharing statistic suggested by Bowcock *et al.* (1994), with individual animals acting as operational taxonomic units (OTUs).

5.4 Results & Discussion

The average heterozygosities, MNA and genetic distances computed here (**Tables 5.1** and **5.2**) are similar to those obtained with other European cattle breeds (MacHugh *et al.* 1994). From both the former values it appears that the Pustertaler breed exhibits a similar level of genetic variability to other breeds. The average heterozygosities and MNA values for the Simmental and Vosges cattle are comparable to those obtained from similar microsatellite methods by MacHugh *et al.* (1997) and Moazami-Gouldarzi *et al.* (1997) respectively. The topology of the breed tree (**Figure 5.1a**) indicates strong support for the Pustertaler breed being more closely-related to the Pinzgauer breed (a bootstrap value of 85% with respect to the loci sampled); a grouping that is consistent with historical breed data (Felius 1995).

Breed	**Origin of breed**	**Average Heterozygosity**		**MNA**
	(No. of animals)	**Expected**	**Observed**	
Pustertaler	Italy (24) Germany (7)	0.69 ± 0.03	0.63 ± 0.03	5.3
Pinzgauer	Germany (20) Austria (10)	0.71 ± 0.03	0.67 ± 0.04	6.0
Vosges	France (27)	0.68 ± 0.03	0.68 ± 0.04	5.4
Simmental	France (50)	0.58 ± 0.04	0.56 ± 0.04	5.2

Table 5.1 Breeds included in the analysis and their country of origin, including the number of individuals sampled (in parentheses). The average heterozygosities and mean number of alleles (MNA) are also included.

Breed Grouping	**Genetic Distance (D_A)**
Pustertaler - Pinzgauer	0.1202
Pustertaler - Vosges	0.1240
Pustertaler - Simmental	0.1974

Table 5.2 Genetic distance (D_A) estimates between Pustertaler and each of the other three breeds.

The results from both neighbour-joining trees (**Figure 5.1a** and **5.1b**) are concordant with the conventional view that the Pinzgauer contributed to the gene pool of the Pustertaler breed in the past. However, a remarkable level of tree structure can be seen from the allele-sharing measure tree (**Figure 5.1b**); with 48 of the 50 Simmental animals forming a tight cluster. Of the other three breeds, definite clusters can be distinguished; 21 of the 31

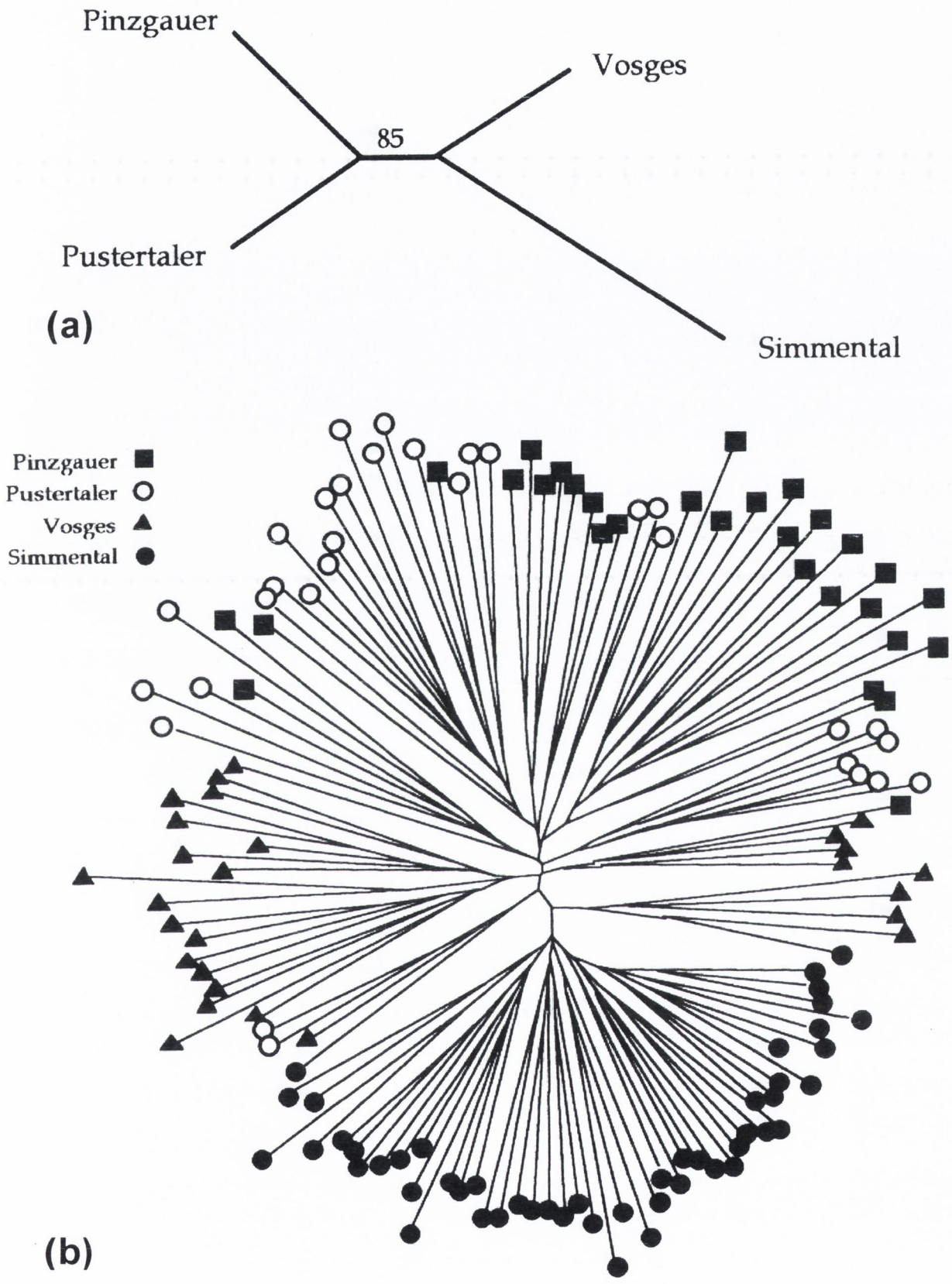


Figure 5.1 Neighbor-joining dendrograms constructed from: **(a)** genetic distances among the four cattle breeds; **(b)** allele-sharing distances among the 138 individual animals. Reproduced from Edwards *et al.* (2000).

Pustertaler animals form two exclusive groups, but by concentrating on the internal nodes it can be seen that, on the whole, this breed segregates with Pinzgauer individuals, as would be hypothesised from both the breed tree (**Figure 5.1a**) and historical information. MacHugh *et al.* (1998) found a similar level of clustering among seven European cattle breeds using this microsatellite allele-sharing distance. The data were also analysed with principal component analysis (PCA). The results are not shown here but they were consistent with those obtained from the neighbour-joining trees.

The overall level of genetic variability in the Pustertaler breed is not significantly reduced compared to the three breeds, or indeed when compared to other European breeds (MacHugh *et al.* 1994). This may be expected due to the known outcrossing with Pinzgauer and Simmental in the past. Although there is a difference between the expected and observed heterozygosity values for Pustertaler, this does not appear to be significant when compared to the corresponding values from the other three breeds. From the analysis of the Pustertaler breed and the three breeds considered to be most closely associated, it clusters most closely with the Pinzgauer breed.

The genotypes obtained in this study have been submitted to the Roslin Cattle Diversity Database (*http://www.ri.bbsrc.ac.uk/cdiv_www/homepage.htm*).

CHAPTER 6

Feasibility and utility of microsatellite markers in archaeological cattle remains from a Viking Age settlement in Dublin

6.1 Abstract

A panel of nineteen cattle bones, from the Viking tenth and early eleventh century levels in Dublin, was assessed for presence of reliable genotypes from three autosomal markers. The samples were chosen because previous analyses indicated that these remains were extremely well preserved. It was therefore possible to amplify and type the microsatellite loci (*CSRM60*, *HEL1* and *ILSTS001*) in eleven specimens. A comparative analysis was performed using data from the same three markers in eleven extant British, Irish and Nordic cattle breeds. The medieval remains displayed lower levels of diversity than the modern European breeds, but represent a normal population. The results indicate a possible origin for the ancient Irish cattle as the remains group more closely with breeds from the British Isles than with those from Scandinavia. The data collected suggest that microsatellites may be useful for the further study of ancient cattle. However, certain criteria need to be considered before embarking on such studies. These include the use of well preserved remains, informative loci, short amplification products and, crucially, ample scope for reproducibility.

6.2 Introduction

aDNA research in livestock has focused primarily on the analysis of mtDNA variation (Bailey *et al.* 1996; MacHugh *et al.* 1999; Troy *et al.* 2001), and these studies have been highly informative in tracing underlying patterns of cattle domestication. However, the information content of cattle mtDNA haplotypes imposes limitations on their use, particularly among Western European cattle populations that display little or no mtDNA population genetic structure. On the other hand, microsatellite loci, also known as short tandem repeat (STR) polymorphisms, have been shown to provide a breed-level genetic resolution within cattle (MacHugh *et al.* 1997, 1998). Therefore, it is desirable to extend microsatellite-based studies from extant populations to archaeological specimens.

The first evidence of Vikings in Ireland dates to the last part of the eighth century, when raiding parties landed to plunder the rich monastic settlements along the coasts. However, during the 840s AD, Scandinavian *longphorts*, or shipping ports, became established in Dublin and Annagassan, County Louth, serving as both permanent trading stations and shelter for the Norse invaders. The original Viking settlement in Dublin, founded in 841, was traditionally believed to have been located in the vicinity of the ninth century Scandinavian cemeteries located at Kilmainham and Islandbridge at the western edge of the present city (O'Brien 1998), although recent research indicates that there was further ninth century settlement in the Parliament Street area further to the east (Simpson 2000).

In 902, the Dublin Vikings were banished to northern England by a coalition of Irish kings, but returned in 917 to establish a real town (called a *dún*) at Dublin. The Viking settlement of Dublin, or *Dubh Linn* (the black pool), was situated near the natural harbour formed by the boundaries of the River Liffey to the north, and the River Poddle to the east, downstream from the original *longphort*. Extensive excavations indicate that the tenth and eleventh century town was located in the area surrounding Dublin castle, which includes the Parliament Street area. This resettled site evolved into an enclosed town with an extensive network of streets, buildings and plots (Wallace 1992). Although Scandinavian and English evidence places the end of the Viking Age in the middle of the eleventh century, the Scandinavians in Ireland blended with the indigenous population to produce the Hiberno-Norse culture. These people remained unchallenged until the arrival of the Anglo-Normans in 1169 AD (Wallace 1981).

One of the best-preserved areas of the Viking settlement was uncovered in Fishamble Street during excavations commenced in 1962 by The National Museum or Ireland. The work yielded detritus and thousands of artefacts from the early town. Due to the damp climate and waterlogged state of the site, most of the material had been conserved under anoxic conditions and the quality of preservation was exceptional (Wallace 1984). The deposits represented a maximum of thirteen property plots containing up to fourteen levels of occupation (Wallace 1992). The typical well-preserved plot level consisted of a house and a *garden* area, with post-and-wattle boundary fences separating it from its neighbours. The house floors and garden areas contained large amounts of food refuse, notably butchered animal bones.

The status of the domesticated cattle present in and around medieval Dublin is unclear. Although it is more widely accepted that they were indigenous cattle taken from the surrounding areas (Wallace 1987), it is possible that they may have included imported cattle from England or Scandinavia. An examination of domestic animal bones from Plots 2 and 3 shows that beef was the main contributor to the diet, although pig were the dominant species in terms of minimum numbers of individuals (55%; McCormick 1983, in prep.). This evidence indicates that the large urban population of Viking Dublin were, most likely, dependent on outside producers for their meat supplies, and the absence of calf remains at the site, coupled with the finding that 57% of the cattle present were over four years old (cited in Wallace 1985), suggests that mature cattle were being driven into the town from the surrounding countryside and were rarely kept within the settlement itself.

However, documentary evidence suggests that there was an area of Scandinavian influence in the rural hinterland of Dublin, an area known as the *Dyflinarskiri*. Bradley (1988) has argued that this area would have been settled by Scandinavian farmers, while Valente (2000) argues that it was settled by native people who were clients to the rulers of Viking Dublin. While it might be assumed that the beef consumed in Dublin represented indigenous livestock types, there is a possibility that the cattle may have included imported cattle from abroad. Early written evidence indicates that the Irish were not unaccustomed to importing domesticated animals. Welsh horses are referred to in the *Annals of Ulster* in 1029, the contemporary *Book of Rights* mentions Scottish horses (Dillon 1962), while, in the early laws, it is written that British horses were imported in order to improve the breeding stock in Ireland (Kelly 1997). The ninth century *Cormac's Glossary* suggests the importation of dogs to Ireland from Britain (Stokes 1862), while philological evidence

implies that cats may have been imported from Wales (Thurneysen 1949). Further examples for the importation of livestock from overseas can be seen in 1105, when the *Annals of Innisfallen* note that King Murthertach O'Brien came into the possession of a camel, as well as in the discovery of a Barbary ape in an second century BC context at the Royal site at Navan, Co. Armagh (Raftery 1997).

There is no evidence for the importation of cattle into Ireland at this time, and it seems more probable that the cattle from which the town derived its milk and meat were obtained from nearby farmlands rather than introduced from Scandinavia or Britain. However it should be noted that the original Neolithic cattle of Ireland must have been imported, as aurochs never existed in the country. Wilson (1909) argued that the hornless cattle of Ireland were originally imported from Scandinavia during the Viking period, his argument based on the observation that hornless breeds were a feature of lands settled by Scandinavians, such as the north of Scotland, the Hebrides and Iceland. More recent zooarchaeological studies, however, have indicated the presence of hornless cattle in Ireland prior to the Viking presence (McCormick 1987). It is certain that the Vikings transferred cattle during raiding overseas and their introduction of cattle to Greenland (McGovern 1992) indicates that they were capable of moving the species over long distances. The movement of cattle from Ireland by ship is recorded in a thirteenth century account of the tenth century King Olav Trygvason, contained in the *Heimskrinkla* of Snorre Sturlason (Monsen 1932). The incident describes a raid in which "*many hundred*" cattle, including cows, were taken on board the king's ships during a "*great shore raid*". So, there was at least the potential for cattle of imported Scandinavian stock to be present in Dublin during the Viking period. The aim of this study was to test this possibility.

Microsatellite analysis was performed on a panel of 19 cattle bones recovered from Plots 1 and 2 in Fishamble Street. Previous examination of mtDNA variation in a subset of these bones could not determine the origins of the domesticated animals used by the Dublin Vikings (MacHugh *et al.* 1999), but it was anticipated that the increased information content offered by microsatellite markers could unravel breed associations between the ancient cattle and relevant extant populations. The modern breeds studied included Kerry cattle, one of the few remaining indigenous cattle breeds left in Ireland, three Norwegian breeds and one breed from Iceland. The present study involves three microsatellites (*CSRM60*, *HEL1* and *ILSTS001*) selected from several loci previously used to study genetic variation in modern cattle populations (MacHugh *et al.* 1997, 1998; Loftus *et al.*

1999). All markers were chosen primarily for their small allele size ranges: *CSRM60*, 83-111 base pairs (bp); *HEL1*, 101-117 bp; and *ILSTS001*, 77-97 bp, as it has been demonstrated that ancient DNA is inevitably fragmented and degraded (Handt *et al.* 1994).

6.3 Materials & Methods

6.3.1 *Archaeological material collection from the Fishamble Street excavations*

Nineteen cattle calcaneus (ankle) bones were retrieved from the Fishamble Street site. Using morphometric criteria, these skeletal elements were ascertained to include a minimum number of sixteen individuals. The preservation environment was essentially anoxic, as demonstrated by the excellent preservation of timber used for construction of dwellings on the site (Wallace 1992). Coins and dendrochronological evidence indicates that the material dates to between about 920 and 1020 AD.

Previous studies of ancient DNA survival have found a good correlation between relatively subjective measures of bone preservation and the presence of amplifiable endogenous DNA (Richards *et al.* 1995; Colson *et al.* 1997). The Fishamble Street bovine bone material was dense, with undamaged external surfaces, and could therefore be considered well preserved using the gross morphological criteria outlined by Hagelberg *et al.* (1991a). These samples include the subset of bones used in a previous mitochondrial study conducted by MacHugh *et al.* (1999).

6.3.2 *DNA extraction and amplification of archaeological specimens*

Bone samples were prepared using the modified Yang *et al.* (1998) procedure described in Section **2.3.2**. Varying numbers of DNA extractions were performed for each specimen. PCR set-up was conducted in a laboratory dedicated solely to pre-amplification ancient work. A first-round of PCR was carried out in a dedicated PCR machine, and PCR amplifications were in 20 µl reaction volumes containing 10x PCR buffer (10 mM Tris·HCl [pH 8.8]; 50 mM KCl; 0.1% Triton X-100), 2.5 mM $MgCl_2$, 200 µM dNTPs, 4 µM of both the forward and the reverse primers, and 0.5 units of PLATINUM® *Taq* polymerase (GIBCO BRL®). Primer details for *CSRM60* (*D10S5*), *HEL1* (*D15S10*) and *ILSTS001* (*D7S13*) were taken from Barendse *et al.* (1994), Kaukinen & Varvio (1993), and Kemp *et al.* (1995) respectively. A 2 minute denaturation step at 94°C was followed by 20 cycles of 1 minute denaturation at 94°C, 1 minute of annealing at 55°C, and 1 minute of extension at 72°C, followed by a final 4 minute extension step at 72°C. These faint products were then visualized on 1.5% agarose mini-gels. Extraction and PCR blanks were always included and remained negative throughout.

Positive samples, obtained from 11 of the 19 bones, were cleaned using a Concert™ rapid purification kit (GIBCO BRL®; used according to the manufacturers instructions with DNA eluted in 50 µl of TE). A second-round, radioactively labelled PCR was then performed in the main analytical laboratory. Reactions took place in 96-well microtitre plates with 5.5 µl of cleaned DNA in 10 µl reaction volumes. The reaction conditions were as before, except that the concentration of dCTP in the dNTP mix was reduced to 10 µM, and the concentration of magnesium chloride was varied depending upon the marker used (*CSRM60* at 1.5 mM; *HEL1* and *ILSTS001* at 1.0 mM). ^{32}P radionucleotide was randomly incorporated during the PCR amplification by the addition of 0.5 µCi [α-^{32}P] dCTP at 3,000 Ci mmol^{-1} per reaction. The reactions were carried out with the following cycling parameters: 94°C for 2 minutes, followed by 30 cycles of 94°C for 30 seconds, 54-58°C for 30 seconds (*CSRM60* at 55°C; *HEL1* at 54°C; *ILSTS001* at 58°C) and 72°C for 30 seconds. A final extension of 4 minutes at 72°C was also carried out. The radiolabelled products were then assessed for length variation using standard 6% polyacrylamide sequencing gel electrophoresis and autoradiography. M13mp18 phage vector was used to provide sequence ladders for allele size calibration. This technique was precise enough to reproducibly resolve single repeat unit differences.

The criteria for authenticating microsatellite genotypes were those described by Burger *et al.* (1999). In brief, DNA was extracted from each sample at least twice, and each extract was amplified by PCR at least five times. Only if at least three of the five amplifications gave consistent allele readings, and this was reproduced in a second extract, were the results considered reproducible and thus authentic. To counteract the possibility of null alleles or the generation of false chimeric alleles, a modification of the approach advocated by Taberlet *et al.* (1996) was also used. If highly ambiguous results were obtained, an allele was only recorded if it was observed at least three times from the five positive PCR products, and an individual was only scored as homozygous if the allele was seen in all five independent amplifications. The analysis of five products from each sample-locus combination reduces the probability of false-homozygous typing to 0.008 (cited in Schmerer *et al.* 1999).

6.3.3 *Amplification of modern specimens*

For the purposes of this study, the positive aDNA results have been collated into a single population, even though the medieval samples are not strictly a single panmictic population as they came from deposits ranging in age from 920 to 1020 AD, covering

approximately 20 cattle generations. The ancient microsatellite data obtained were compared with 12 extant breeds. The 13 populations studied comprise the ancient group from Fishamble Street, eleven European breeds (5 British Isles, 1 indigenous Irish, 4 Scandinavian, and 1 Continental) and an African *Bos taurus* outgroup. Apart from the synthetic derived Norwegian Red breed, all of the extant populations have the status of *authentic breed.* The Kerry, Telemark and Western Fjord breeds are on the endangered list (Scherf 1995). Data for these populations were taken from previously analysed results (MacHugh *et al.* 1997, 1998; Loftus *et al.* 1999; Cymbron *et al.* in prep.), except in the case of the Galloway, Iceland, Norwegian Red, Telemark and Western Fjord breeds, which were amplified with all three markers as part of this study. Also, the Aberdeen Angus and Kerry cattle were amplified with *CSRM60*, and the Highland samples with *ILSTS001*. The reaction conditions for these modern samples were as described in Loftus *et al.* (1999) for *CSRM60* and *HEL1*, and MacHugh *et al.* (1997) for *ILSTS001*.

6.3.4 *Statistical and phylogenetic analysis*

To assess the levels of genetic variability in the 13 populations, both the mean number of alleles per locus (MNA) and gene diversity (average expected heterozygosity) values were estimated. Unbiased estimates of gene diversity, and observed heterozygosity values, with associated standard errors, were calculated according to Nei (1987). Allele frequencies were determined by direct counting. MNA values have been shown to be inadequate estimates of genetic variation if the sample sizes for each population are notably different (Valsecchi *et al.* 1997). To compensate for the large differences between the sample sizes, the MNA values were compared between the populations using a resampling approach. The smallest number of alleles equalled 14 for marker *CSRM60* in the ancient Dublin population. Fourteen random alleles were selected without replacement from each of the breeds, and this approach was repeated 1,000 times. This calculation was performed using a Microsoft Excel©-based microsatellite program (Park 2002). An exact test for deviations from Hardy-Weinberg equilibrium (HWE) was performed using the ARLEQUIN computer program (version 2.000, Schneider *et al.* 1997), with a Markov chain set to 10,000 steps with 1,000 steps of dememorisation.

Interbreed genetic distances (D_A) were estimated as described in Nei *et al.* (1983) using the DISPAN computer program (T. Ota, Center for Human Genetics, Boston University). Multidimensional scaling (MDS; Kruskal 1964) was performed based on the D_A matrix obtained using the SPSS© program (Standard Version, release 10.0.5). The SPSS program

was also used to perform a non-parametric Mann-Whitney rank-sum test (Mann & Whitney 1947).

The model-based Bayesian clustering method, implemented in the STRUCTURE program (Pritchard *et al.* 2000), was used to infer population structure from the three microsatellite loci, and to assign individuals to populations. This program identifies clusters of genetically similar individuals from genotypes without prior knowledge of their population affinities. It assumes HWE and linkage disequilibrium (LE) among the unlinked marker loci. Departures from HWE and LE lead the population to be split into subpopulations, to which individuals are assigned, and those with admixed ancestries are assigned to more than one source population. The admixture model was used with a burn-in period of 30,000 steps and a run length of 10^5 iterations. Two populations were assumed for the run. Individuals were assigned probabilistically to these populations, or jointly to the two populations if their genotypes indicated admixture. These values were plotted as a scatter graph.

A neighbor-joining tree (Saitou & Nei 1987) was constructed using the allele-sharing measure (ASM) suggested by Bowcock *et al.* (1994), with individual animals acting as operational taxonomic units (OTUs). Samples were counted as OTUs only if they had given amplification products with all three microsatellite loci, and therefore the tree contains 242 out of the 302 animals sampled, including 7 Fishamble Street individuals. The ASM matrix was calculated (Park 2002), and the dendrogram derived using the PHYLIP software package (version 3.572, Felsenstein 1993).

6.4 Results

6.3.1 *Authenticity of archaeological results*

As detailed in **Table 6.1**, it was possible to generate positive amplification products from at least two of the three microsatellites in 11 of the 19 cattle bones from the Wood Quay site. This success rate of approximately 58% is in general agreement with Richards *et al.* (1995), who suggest that more than 50% of skeletal remains from the past 2000 years are likely to contain endogenous DNA.

The critical issue was to verify that DNA extracted from the ancient material was not chemically altered to such an extent that analysis of the polymorphism of the microsatellite loci would not be reproducible. To fully tackle this problem, multiple extractions were performed on all of the bones and verified genotypes for each sample were obtained from at least two independent DNA extractions. Each extract was amplified by PCR at least five times, and only if three of the five amplifications gave consistent allele readings in both extracts were the results considered authentic. All of the eleven ancient samples that gave amplification products fulfilled these rigorous criteria. In addition, extraction and PCR controls were used for all DNA isolation, purification and amplification steps, and were blank in all cases.

To further evaluate the possibility of contamination, Thomas & Pääbo (1993) suggest that a PCR should be run using an extract from an unrelated ancient specimen for which the primers should not work. In this case, a *Sus scrofa* (wild boar; CPC-98) vertebra from the 5,000 year-old site of Carsington Pasture Cave in Derbyshire was extracted and amplified in tandem with the Fishamble Street samples. Aurochs bones from this site amplify extraordinarily well with mtDNA primers (Troy *et al.* 2001), but no results were obtained from any of the three microsatellite primer pairs with this boar specimen. Whereas *CSRM60* amplifies in red deer (Tate 1997) and *ILSTS001* amplifies in sheep and goat (Kemp *et al.* 1995), the locus *HEL1* has only been found to amplify in bovine material. Therefore, another positive indication that the ancient results are indeed authentic, and not due to cross-species contamination, is that seven of the 11 positive Fishamble Street samples gave amplification products with all three markers. Eleven of the 19 medieval cattle bones gave a positive amplification with at least two of the three microsatellites (*HEL1* and *ILSTS001*), but WQ01a, WQ04a, WQ12 and WQ15 did not amplify with the

Laboratory code	Number of extractions	PCR product	Microsatellite locus *CSRM60* (bp)	*HEL1* (bp)	*ILSTS001* (bp)
WQ01a	2	yes	no	(111, 111)	(93, 93)
WQ01b	2	no	n/a	n/a	n/a
WQ02	2	no	n/a	n/a	n/a
WQ03	5	yes	(101, 101)	(113, 113)	(93, 93)
WQ04a	3	yes	no	(113, 113)	(93, 93)
WQ04b	5	yes	(103, 105)	(113, 113)	(93, 95)
WQ05	2	no	n/a	n/a	n/a
WQ06	2	no	n/a	n/a	n/a
WQ07a	2	no	n/a	n/a	n/a
WQ07b	4	yes	(103, 105)	(103, 113)	(93, 93)
WQ08	3	no	n/a	n/a	n/a
WQ09	4	yes	(103, 105)	(105, 113)	(93, 93)
WQ10	2	no	n/a	n/a	n/a
WQ11	4	yes	(103, 105)	(113, 113)	(93, 93)
WQ12	4	yes	no	(105, 113)	(93, 93)
WQ13	4	yes	(103, 105)	(113, 113)	(93, 93)
WQ14	2	no	n/a	n/a	n/a
WQ15	4	yes	no	(113, 113)	(93, 93)
WQ16	4	yes	(103, 103)	(113, 113)	(93, 93)

Table 6.1 Cattle calcaneus bones analysed from the Fishamble Street excavations, with associated laboratory codes and PCR amplification results for each of the three microsatellites in base pairs (bp).

CSRM60 locus (**Table 6.1**). Two of the 7 positive results from *CSRM60* were homozygotes, with this value being 8 out of 11 for *HEL1*, and 10 out of 11 for *ILSTS001*. Apart from WQ11 and WQ13, which generated the same genotype with the three loci, the results are different for each fully-typed ancient individual, further emphasising the authenticity of the results.

6.3.2 *Genetic diversity and phylogenetic analyses*

Even when sample size is taken into account, the Fishamble Street population has a low MNA value (**Table 6.2**). The ancient samples also have the lowest value when the observed heterozygosity values are compared among breeds. Four populations significantly deviated from HWE at locus *CSRM60*, with $P < 0.01$ for Galloway and Highland, and $P < 0.05$ for Fishamble Street and N'Dama. The N'Dama also have a P-value of < 0.01 for marker *ILSTS001*. The two-dimensional MDS plot constructed using the 13 cattle populations (**Figure 6.1**) places the Fishamble Street population among the British Isles breeds studied, nearest the Hereford breed.

To test the assumption of a close association between the Fishamble Street and British Isles breeds, the D_A distances between the Fishamble Street population and the European breeds (**Table 6.3**, column 1) were ranked. The British breeds (Aberdeen Angus, Galloway, Hereford, Highland, Jersey) exhibit the closest affinity to the Fishamble Street samples (ranging from 0.0922 to 0.1652), with the Hereford breed having the smallest genetic distance of 0.0922. The Irish Kerry breed (0.2065) is separated from this grouping by the Norwegian Red (0.1694), with distances for the remaining Scandinavian breeds (Iceland, Telemark, Western Fjord) ranging from 0.2367 to 0.4745. This order of grouping with the Fishamble Street is evident to some extent in the MDS plot (**Figure 6.1**). The ranked list was analysed using the Mann-Whitney test, and the significance value of the similarity between the ancient Dublin population and modern British Isles breeds was calculated to be $P < 0.02$.

To further test the assumption that the ancient cattle are more closely related to the British Isles breeds than those from Scandinavia, population structure was assessed by using a model-based clustering approach implemented in the STRUCTURE program. Briefly, this model assumes K populations, each characterized by a set of allele frequencies at each locus, and individuals are assigned to these populations on the basis of their genotypes. The assignment of the individual cattle were inferred by STRUCTURE without using prior

Breed	Code	Origin of breed (No. of animals)	Average Heterozygosity		MNA (full sample)[a]	MNA (uniform sample)[b]
			Expected	Observed		
Ancient Group						
Fishamble Street	WQ	920-1020 AD Dublin, Ireland (11)	0.38 ± 0.16	0.36 ± 0.11	3.0 ± 0.58	2.7 ± 0.26
Britain						
Aberdeen Angus	AN	Northeast Scotland (33)	0.41 ± 0.17	0.41 ± 0.09	3.7 ± 1.20	2.7 ± 0.40
Galloway	GA	Galloway, Scotland (23)	0.43 ± 0.22	0.50 ± 0.15	4.0 ± 1.53	2.9 ± 0.38
Hereford	HE	Western England (34)	0.42 ± 0.14	0.43 ± 0.09	4.3 ± 0.88	2.8 ± 0.39
Highland	HI	West Highlands, Scotland (35)	0.42 ± 0.20	0.46 ± 0.14	4.3 ± 1.45	2.9 ± 0.37
Jersey	JE	Jersey (34)	0.49 ± 0.24	0.48 ± 0.14	4.0 ± 1.15	3.2 ± 0.31
Ireland						
Kerry	KE	Southwest Ireland (40)	0.48 ± 0.10	0.49 ± 0.07	2.7 ± 0.33	2.5 ± 0.18
Scandinavia						
Iceland	IC	Iceland (15)	0.68 ± 0.02	0.58 ± 0.06	3.7 ± 0.33	3.5 ± 0.20
Norwegian Red	NO	Norway (15)	0.50 ± 0.18	0.46 ± 0.09	3.7 ± 0.88	3.1 ± 0.30
Telemark	TK	Telemark, Norway (15)	0.52 ± 0.11	0.43 ± 0.03	3.7 ± 0.67	3.2 ± 0.30
Western Fjord	WF	Southwest Norway (15)	0.65 ± 0.13	0.63 ± 0.11	4.3 ± 0.33	4.0 ± 0.28
Continental						
Charolais	CH	Southeast France (36)	0.63 ± 0.11	0.64 ± 0.05	4.7 ± 1.20	3.9 ± 0.37
West African Outgroup						
N'Dama	ND	Fouta Djalon, Guinea (44)	0.67 ± 0.05	0.61 ± 0.04	5.3 ± 1.20	4.1 ± 0.43

Table 6.2 Populations included in the analysis and their country of origin, including the total number of individuals sampled (in parentheses) and the associated two letter codes. The average heterozygosities and mean number of alleles (MNA) are also included, with associated standard deviations.

[a] The mean number of alleles (MNA) was calculated using all the animals from each population sampled.
[b] Uniform population sizes were used for the second column of MNA values. A random sample of 7 animals was used for each population. (see text for a more detailed explanation). Associated standard errors were negligible.

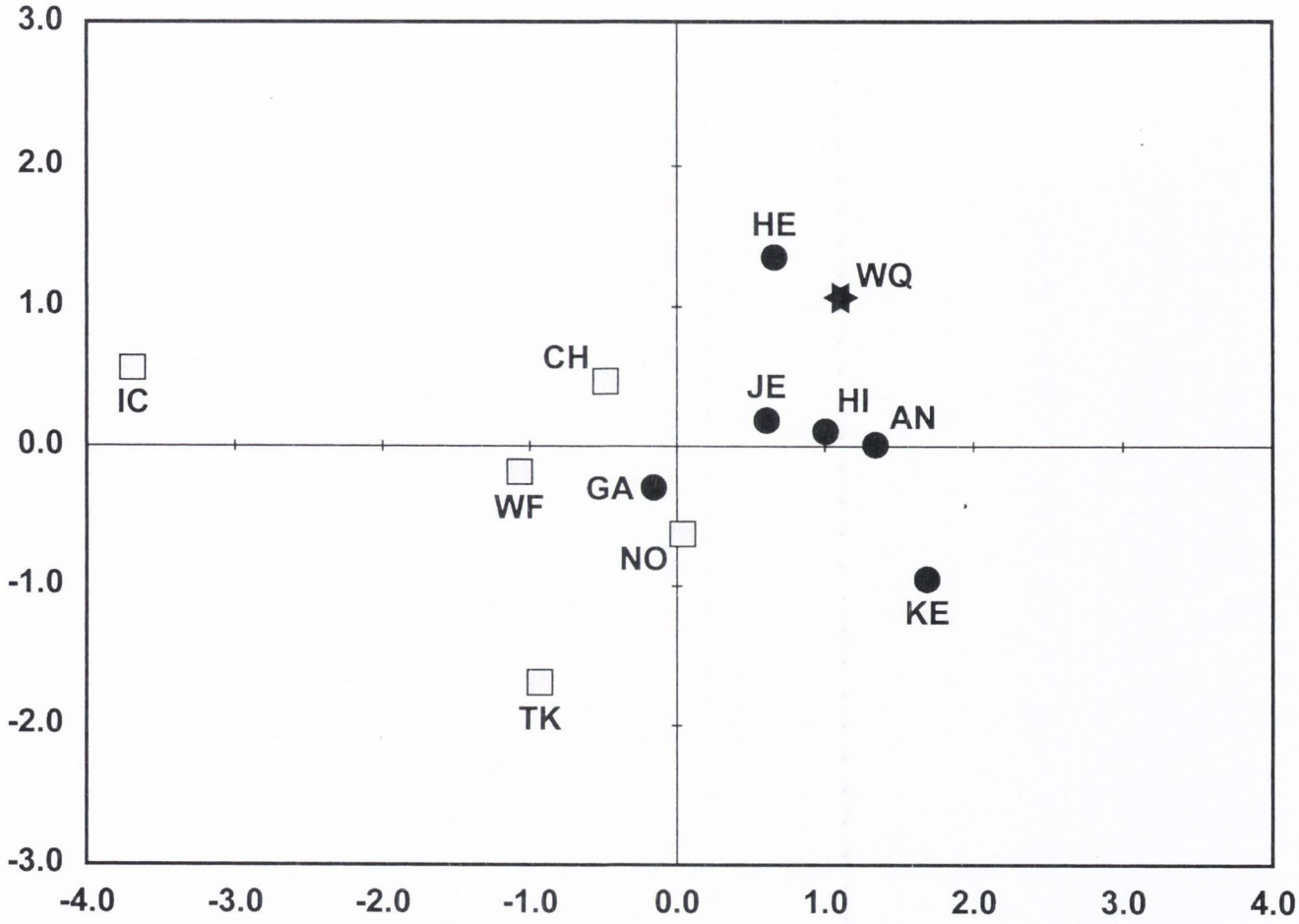


Figure 6.1 A two-dimensional MDS plot, drawn using data from three microsatellites, summarising D_A genetic distances among 11 extant European breeds and the medieval Fishamble Street population. Black circles represent the British and Irish breeds, grey squares denote the Scandinavian and Continental breeds, and a black star corresponds to the Fishamble Street population. The two-letter codes are as in Table 6.2. The ancient population groups more closely with the British Isles breeds than those from Scandinavia, and more specifically with the Hereford breed from western England. Reproduced from Edwards *et al.* (in prep.).

Population	WQ	AN	GA	HE	HI	JE	KE	IC	NO	TK	WF	CH	ND
Fishamble Street	0.0000												
Aberdeen Angus	0.1652	0.0000											
Galloway	0.1284	0.1623	0.0000										
Hereford	0.0922	0.2121	0.1733	0.0000									
Highland	0.1353	0.1440	0.1079	0.1136	0.0000								
Jersey	0.1232	0.0736	0.0821	0.1231	0.1107	0.0000							
Kerry	0.2065	0.2909	0.2329	0.2079	0.1926	0.2114	0.0000						
Iceland	0.4745	0.5329	0.3198	0.4005	0.4289	0.4030	0.5667	0.0000					
Norwegian Red	0.1694	0.1586	0.0808	0.1622	0.0821	0.0893	0.1783	0.3321	0.0000				
Telemark	0.2970	0.2802	0.0974	0.3201	0.2353	0.1905	0.2789	0.3253	0.0906	0.0000			
Western Fjord	0.2367	0.2180	0.1457	0.1803	0.1525	0.1487	0.2767	0.2303	0.0992	0.1508	0.0000		
Charolais	0.1258	0.1984	0.1358	0.1331	0.1730	0.1357	0.2457	0.2867	0.1179	0.1634	0.0961	0.0000	
N'Dama	0.5154	0.3767	0.3286	0.5114	0.4256	0.3847	0.5514	0.2730	0.3516	0.2741	0.2860	0.2660	0.0000

Table 6.3 Genetic distance (D_A) matrix obtained from the frequencies of 3 microsatellite markers among 13 breeds. The two-letter codes are as in Table 6.2.

population information, with a K value of 2. **Figure 6.2** is a representation of this computation, showing the spread of values for each individual along an axis, with proximity to a probability of 0 indicating closeness to the British Isles cluster, and proximity to a probability of 1 indicating closeness to a Scandinavian cluster. Nine of the Fishamble Street individuals have a probability value of less than 0.5, which groups them more closely to the British Isles end of the graph.

Figure 6.3 shows a neighbor-joining phylogenetic tree constructed from an allele-sharing measure. Of the 302 animals sampled from the 12 European populations, 242 gave results from all three microsatellite markers and were therefore considered as OTUs. The most common reason for exclusion from the ASM tree was non-amplification of locus *CSRM60*. Although 12 of the 30 Kerry samples form a clade, and 17 of the 36 Charolais group together, overall the tree shows little breed structure, most probably due to the small amount of genetic variation that can be measured from three microsatellites. However, the British Isles breeds do, in general, exhibit some level of clustering. Of the 7 aDNA specimens that gave reproducible results from all 3 markers (**Table 6.1**), WQ11 and WQ13 generated the same genotype, WQ03 matched with one Scandinavian animal (NO), WQ07b corresponded to 1 Charolais animal, WQ09 matched with 4 British (1 AN, 1 HE, 1 HI and 1 JE) and 1 Charolais animal, and WQ16 had the same genotype as 6 British animals (1 GA, 3 HE, 1 HI and 1 JE). WQ04b had a novel haplotype not seen in any of the other 11 populations studied. When the percentages of number of individuals shared with animals from the same breed against total number of OTUs from each breed are calculated, the value of 28.6% for the ancient population is low compared with the average value of 53.2% for the British Isles breed group, and 50.0% for the Scandinavian group. This dissimilarity within the Fishamble Street samples is a further intimation that each individual result is indeed from endogenous ancient DNA.

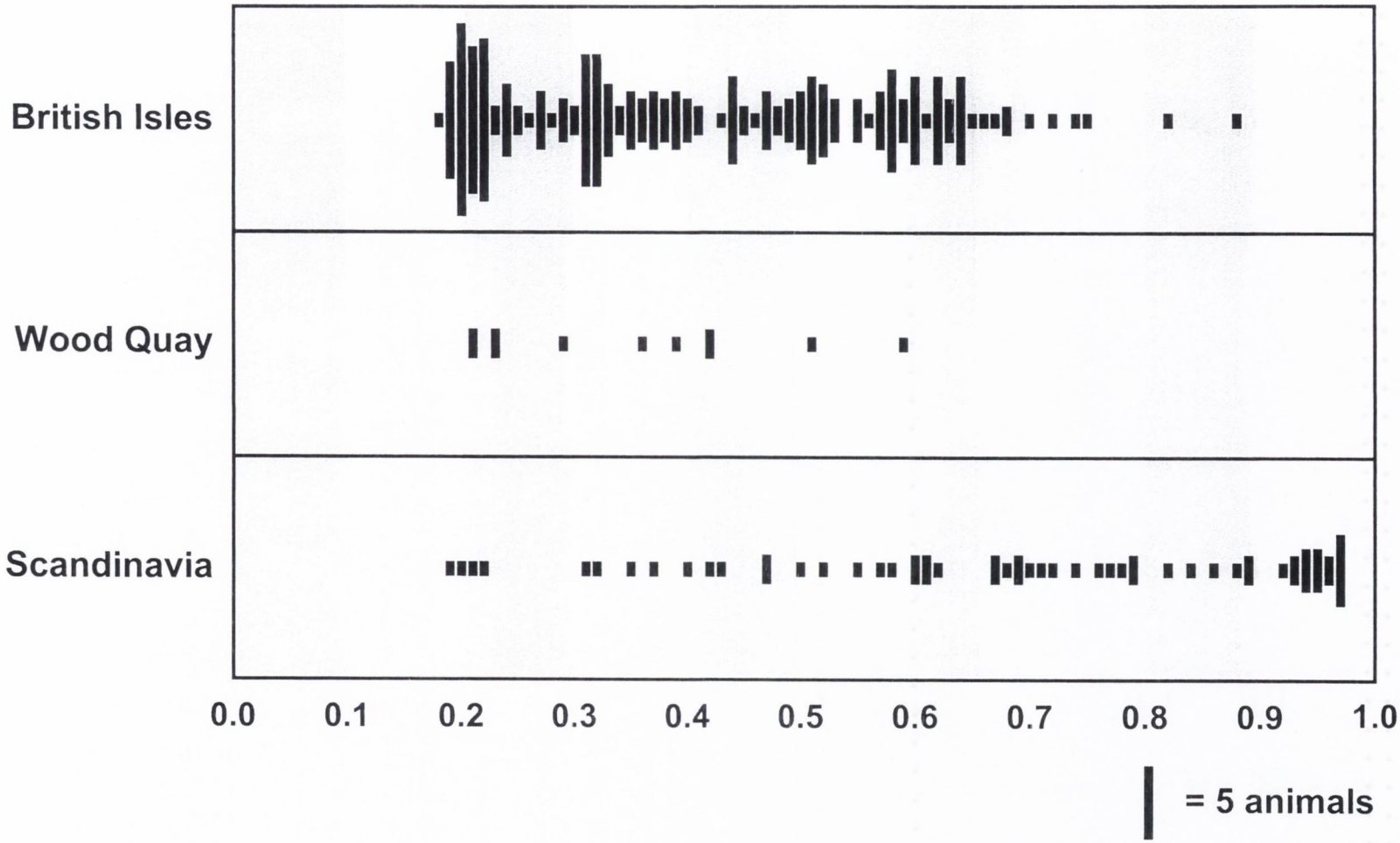


Figure 6.2 Summary of the clustering results for the cattle data, assuming two populations. Each bar of unit one shows the mean estimated ancestry for an individual in the sample. The fraction of the individual's ancestry is represented by its distance along the line. After the clustering was performed, the points were labelled according to origin, with the British Isles breeds at the top and the Scandinavian samples below. The Fishamble Street individuals are shown between these. Reproduced from Edwards *et al.* (in prep.).

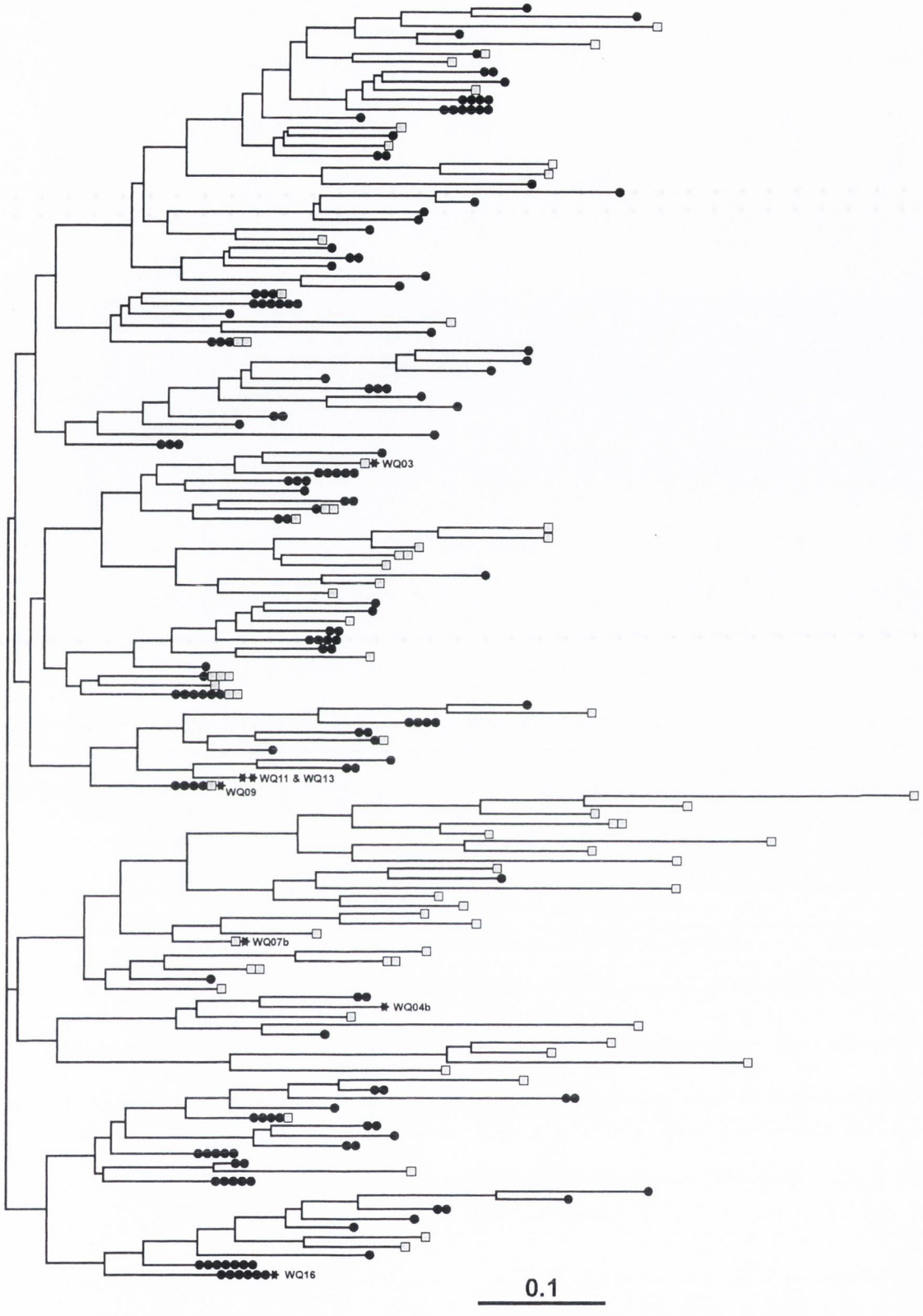


Figure 6.3 Neighbor-joining dendrogram constructed from allele-sharing distances among 242 individual animals from 11 extant European breeds and the medieval Fishamble Street population. Black circles represent the British and Irish breeds, grey squares denote the Scandinavian and Continental breeds, and a black star corresponds to the Fishamble Street population. The 7 aDNA individual codes are shown in Table 6.1. The scale represents arbitrary units of genetic distance and is derived from the original ASM matrix. Five of the seven ancient individuals (WQ03, WQ09, WQ11, WQ13 and WQ16) fall within clusters that predominantly confine British Isles individuals. Reproduced from Edwards *et al.* (in prep.).

6.5 Discussion

Genuine ancient results are more readily accepted when they fit into an appropriate phylogenetic context. The isolation of breeds located on the edge of Europe, especially those from the British Isles and Scandinavia, has had an effect on their genetic structure, leading to low variation. As can be seen from **Table 6.2**, although the Fishamble Street population has a low normalised MNA value (of 2.7), this figure fits within the ranges of all the European breeds. The ancient population has the lowest observed heterozygosity value of all the 13 breeds, however, this is not surprising given the similarity between the expected and observed values. Overall, the Fishamble Street medieval cattle exhibit patterns of microsatellite variation that place them well within the normal European cattle radiation.

In molecular studies using ancient specimens, the amount of DNA available for genetic typing is often in the picogram range. When the template has a low-copy number, or the sample quality is poor, selective amplification of one allele often occurs, leading to the incorrect typing of a heterozygous individual as a homozygote. The ancient Fishamble Street population deviated from HWE at locus *CSRM60*, with a P value of < 0.05, but this was caused by an excess of heterozygotes. Although it is possible that false alleles may have occurred from slippage artefacts generated during the first cycles of the first-round PCR, rigorous precautions were taken to ensure that each genotype was reliable and reproducible. As illustrated in the tree constructed from an ASM distance matrix using individual animals as OTUs (**Figure 6.3**), apart from WQ11 and WQ13, the Fishamble Street individuals are all different from each other, further supporting the authentication and reliability of the results. The authenticity of the ancient data is additionally supported by the low value of number of individuals shared with animals from same breed against total number of OTUs from each breed. Only 28.6% (2 out of 7) of the Fishamble Street samples included in the ASM matrix grouped with each other, whereas the values for the British Isles and Scandinavian breed groups are 53.2% and 50.0% respectively.

When using microsatellites for archaeogenetic studies, it is essential that strict ancient DNA procedures be adhered to, as described in MacHugh *et al.* (2000). The conclusions drawn from authentication of these results lead us to suggest strict criteria for studying microsatellite DNA in ancient samples.

These are:

a) well preserved samples;
b) short amplification lengths (primers can be redesigned, where necessary, to amplify small products of between 50 and 150 bp, making them amenable for use with damaged and degraded DNA);
c) robust and non-ambiguous markers;
d) relatively polymorphic loci;
e) sufficient modern data to provide an adequate modern genetic context; and
f) internally consistent reproducibility.

The positive results obtained here demonstrate that autosomal markers are informative to the study of ancient cattle, and further reinforce the feasibility of using microsatellite analyses for archaeogenetic applications. There have been similar studies looking at the usefulness of microsatellites as markers for ancient material. Ramos *et al.* (1995) claim that microsatellites are not reliable enough for the analysis of ancient DNA extracted from bone or teeth; their main reason being that the loci chosen did not yield reproducible results for the human samples they typed. Our results suggest that this is not so for all markers or all bones, and similar conclusions were reached by Zierdt *et al.* (1996) and Burger *et al.* (1999). The reproducibility of the results in this study reflects the strongly preserved nature of the Fishamble Street bones, which is due to the anoxic conditions of the waterlogged site in which they were found.

Microsatellite analysis of archaeological remains provides information concerning autosomal genetic variation, therefore allowing more general questions to be addressed than studies using mitochondrial DNA alone. For small populations, microsatellite markers allow a greater degree of structural analysis than mtDNA due to their increased level of polymorphism and high heterozygosity. MtDNA analysis is of limited use in studies of cattle diversity in Europe as resolution is afforded only to the regional level. However, a large number of autosomal microsatellites have been typed in extant cattle, and this database acts as a good reference for comparison with archaeological material, especially when relationships among modern populations have become well understood (MacHugh *et al.* 1997, 1998; Loftus *et al.* 1999; Edwards *et al.* 2000; Kantanen *et al.* 2000).

Whereas older ancient DNA may allow macroevolutionary problems to be investigated with as little as one specimen, more recent remains will inevitably demand a population genetic approach. This can require many specimens to give an indication of variation, but

this type of aDNA is less rare and more readily available. However, the highly informative nature of microsatellite data can facilitate a credible study as, although it is often quite difficult to acquire a large range of specimens from a site, microsatellite typing has the capability to make the analysis of a single animal as instructive as that of a population. Burger *et al.* (2000) have reviewed the utility of STR-genotyping for analysing a large range of sample types, including leather, parchment, pot contents, and glues. They conclude that microsatellite-profiling could be useful in authenticating data obtained from less variable mitochondrial sequences, and even that genotyping differences could be used to match specimens to individual animals.

When the normalised MNA value for the Fishamble Street samples (2.7 alleles; **Table 6.2**) is compared between populations, the value is more similar to the Island breeds sampled (Aberdeen Angus, Galloway, Hereford, Highland, Jersey and Kerry: mean = 2.8 alleles) than the Scandinavian group (Iceland, Norwegian Red, Telemark and Western Fjord: mean = 3.5 alleles). Overall, there is little variation among the observed heterozygosity values of the populations, although the low Fishamble Street value of 0.36 is closer to the British Isles group mean of 0.46 than the Scandinavian group mean of 0.53. From the MDS plot (**Figure 6.1**), the ancient Irish population most closely groups with the British breeds, especially the Hereford breed. MacHugh *et al.* (1999), who carried out a similar archaeological study using mtDNA sequence data, also found Fishamble Street grouping near Hereford.

As the data seem to comprise little geographic distinction between individual breeds, another approach had to be used to assess whether an overlying structure could be determined. The Mann-Whitney statistic shows to a significant level ($P < 0.02$) that the ancient Fishamble Street population exhibits clear relationships to the British Isles breeds, rather than to the Scandinavian group. This grouping is reinforced by the STRUCTURE analysis (**Figure 6.2**), where the genotypes from the individual cattle were split into two distinct populations. The STRUCTURE algorithm constructs genetic clusters from a collection of individual multilocus genotypes, estimating for each individual the fractions of its genome that belong to each cluster. With extensive simulations, Pritchard *et al.* (2000) demonstrated that this genetic cluster analysis method could accurately infer individual ancestries. On average, the Fishamble Street samples exhibit a probability of 0.35, with a spread of 0.21 to 0.59. This places the ancient individuals closer to the British

Isles area (average = 0.41; spread = 0.18-0.88) of the graph, than the Scandinavian end (average = 0.69; spread = 0.19-0.97).

Under normal circumstances, three microsatellite markers would be considered a sparse data source for population genetic analyses. However, five of the 7 ancient individuals (WQ03, WQ09, WQ11, WQ13 and WQ16) fall within clusters that predominantly confine British Isles individuals (**Figure 6.3**), therefore providing extra support for an association between Fishamble Street and the British and Irish breeds.

The results reported here are in concordance with the mitochondrial study of MacHugh *et al.* (1999), suggesting that the cattle used in early medieval Dublin were comparable to modern cattle populations found in Europe. However, due to the increased depth of information that can be gathered from microsatellite typing (even from only three loci), this analysis further shows that the ancient Fishamble Street population display more affinity with the extant breeds from the British Isles rather than with Scandinavian breeds. From these results it can be inferred that the ancient Viking Dublin cattle did not originate from Scandinavia, but rather represented indigenous cattle taken from the surrounding areas or, less likely, animals imported from England. There was a period of intense contact between Dublin and York in England during the early 10th century (Wallace 1981), and Dublin traded extensively with Chester until 980 AD, and then afterwards with Bristol (Wallace 1984, 1985). Also, some urban influences at the medieval Wood Quay site are more directly attributable to northern and western England than to Scandinavia (Wallace 1986). However, the Vikings at Dublin dominated the surrounding areas and, on balance, it is more probable that the cattle were supplied from nearby Irish farmlands, rather than being stock introduced from Britain.

In conclusion, this study suggests an Irish origin for these early medieval cattle. However, due to the quality of preservation and volume of animal remains excavated at Fishamble Street, future analyses using additional microsatellites may give a more accurate profile of their genetic origins. It could also provide additional insights into the overall Hiberno-Norse culture, especially as the Viking settlement at this site represents a centre where the three ethnic groupings of Irish, English and Scandinavian have, to varying degrees, played a role.

LITERATURE CITED

Literature Cited

Aldhous P. (1995) 'Jurassic DNA' looks distinctly human. *New Scientist* **11th February**:5.

Allaby R.G., Jones M.K. & Brown T.A. (1994) DNA in charred wheat grains from the Iron Age hillfort at Danebury, England. *Antiquity* **68**:126-132.

Allaby R.G., O'Donoghue K., Sallares R., Jones M.K. & Brown T.A. (1997) Evidence for the survival of ancient DNA in charred wheat seeds from European archaeological sites. *Ancient Biomolecules* **1**:119-129.

Allard M.W., Miyamoto M., Jarecki L., Kraus F. & Tennant M.R. (1992) DNA systematics and evolution of the artiodactyl family Bovidae. *Proceedings of the National Academy of Sciences of the U.S.A.* **89**:3972-3976.

Allard M.W., Young D. & Huyen Y. (1995) Detecting dinosaur DNA. *Science* **268**:1192.

Allchin F.R. (1969) Early domestic animals in India and Pakistan. In: *The Domestication and Exploitation of Plants and Animals* (ed. P.J. Ucko and G.W. Dimbleby), pp. 317-322. Duckworth, London.

Anderson S., de Bruijn M.H., Coulson A.R., Eperon I.C., Sanger F. & Young I.G. (1982) Complete sequence of bovine mitochondrial DNA. *Journal of Molecular Biology* **156**:683-717.

Archibald J.D. (1996) Fossil evidence for a Late Cretaceous origin of hoofed mammals. *Science* **272**:1150-1152.

Arranz J.J., Bayón Y. & Primitivo F.S. (1996) Comparison of protein markers and microsatellites in differentiation of cattle populations. *Animal Genetics* **27**:415-419.

Astolfi P., Pagnacco G. & Guglielmino-Matessi C.R. (1983) Phylogenetic analysis of native Italian cattle breeds. *Journal of Animal Breeding and Genetics* **100**:84-100.

Bailey J.F. (1998) *Analysis of Population Structure using Ancient and Modern Mitochondrial DNA.* Ph.D. thesis, University of Oxford.

Bailey J.F., Richards M.B., Macaulay V.A., Colson I.B., James I.T., Bradley D.G., Hedges R.E.M. & Sykes, B.C. (1996) Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proceedings of the Royal Society of London Series B Biological Sciences* **263**:1467-1473.

Baker C.M.A. & Manwell C. (1980) Chemical classification of cattle: 1. Breed groups. *Genetics* **11**:127-150.

Baker C.M.A. & Manwell C. (1991) Population genetics, molecular markers and gene conservation of bovine breeds. In: *Cattle Genetic Resources* (ed. C.G. Hickman), pp. 221-304. Elsevier, Amsterdam.

Balasse M., Bocherens H., Tresset A., Mariotti A. & Vigne J.-D. (1997) Émergence de la production laitière au Néolithique? Contribution de l'analyse isotopique d'ossements de bovins archéologiques. *Sciences de la Terre et des Planétes* **325**:1005-1010.

Bandelt H.-J., Forster P., Sykes B.C. & Richards M.B. (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**:743-753.

Barendse W., Armitage S.M., Kossarek L.M., Shalom A., Kirkpatrick B.W., Ryan A.M., Clayton D., Li L., Neibergs H.L., Zhang N., Grosse W.M., Weiss J., Creighton P., McCarthy F., Ron M., Teale A.J., Fries H.R., McGraw R.A., Moore S.S., Georges M., Soller M., Womack J.E. & Hetzel D.J.S. (1994) A genetic linkage map of the bovine genome. *Nature Genetics* **6**:227-235.

Bar-Gal G.K., Khalaily H., Marder O., Ducos P. & Horwitz L.K. (2000) Ancient DNA evidence for the transition from wild to domestic status in Neolithic goats: a case study from the site of Abu-Gosh, Israel. In: *5th International Ancient DNA Conference Proceedings, 12th-14th July 2000*, p. 21. UMIST, Manchester, England.

Bar-Gal G.K., Khalaily H., Mader O., Ducos P. & Horwitz L.K. (2002) Ancient DNA evidence for the transition from wild to domestic status in Neolithic goats: a case study from the site of Abu Gosh, Israel. *Ancient Biomolecules* **4**:9-17.

Bar-Yosef O. & Meadow R.H. (1995) The origins of agriculture in the Near East. In: *Last Hunters – First Farmers. New Perspectives on the Prehistoric Transition to Agriculture* (ed. T.D. Price and A.B. Gebauer), pp. 39-94. School of American Research, New Mexico.

Barnosky A.D. (1986) 'Big Game' extinction caused by late Pleistocene climatic change: Irish elk (*Megaloceros giganteus*) in Ireland. *Quaternary Research* **25**:128-135.

Bishop M.D., Kappes S.M., Keele J.W., Stone R.T., Sunden S.L.F., Hawkins G.A., Toldo S.S., Fries R., Grosz M.D., Yoo J. & Beattie C.W. (1994) A genetic linkage map for cattle. *Genetics* **136**:619-639.

Blumenbach J. (1799) *Hanbuch der Naturgeschichte* (6th edition) **16**:697.

Bocherens H., Tresset A., Wiedemann F., Giligny F., Lafage F., Lanchon Y. & Mariotti A. (1997) Diagenetic evolution of mammal bones in two French Neolithic sites. *Bulletin de la Société Geologique de France* **168**:555-564.

Bodø I., Gera I & Koppány G. (1996) *The Hungarian Grey Cattle Breed* (translated B. Borsos). Association of the Hungarian Grey Cattle Breeders, Budapest.

Bohlken K. (1962) Probleme der Merkmalsbewertung am Säugetierschadel, dargestellt am Beispiel des *B. primigenius* Bojanus 1827. *Mörphologisches Jahrbuch* **103**:509-661.

Bökönyi S. (1969) Archaeological problems and methods of recognizing animal domestication. In: *The Domestication and Exploitation of Plants and Animals* (ed. P.J. Ucko and G.W. Dimbleby), pp. 219-229. Duckworth, London.

Bökönyi S. (1976) Development of early stock rearing in the Near East. *Nature* **264**:19-23.

Boom R., Sol C.J.A., Salimans M.M.M., Jansen C.L., Wertheim van Dillen P.M.E. & van der Noordaa J. (1990) Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* **28**:495-503.

Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R. & Cavalli-Sforza L.L. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**:455-457.

Bradley D.G., MacHugh D.E., Loftus R.T., Sow R.S., Hoste C.H. & Cunningham E.P. (1994) Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations. *Animal Genetics* **25**:7-12.

Bradley D.G., MacHugh D.E., Cunningham P. & Loftus R.T. (1996) Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the U.S.A.* **93**:5131-5135.

Bradley D.G., Loftus R.T., Cunningham P. & MacHugh D.E. (1998) Genetics and domestic cattle origins. *Evolutionary Anthropology* **6**:79-86.

Bradley J. (1988) The interpretation of Scandinavian settlement in Ireland. In: *Settlement and Society in Medieval Ireland* (ed. J. Bradley), pp. 49-78. Boethius Press, Kilkenny.

Brown K.A. (1998) Gender and sex – what can ancient DNA tell us? *Ancient Biomolecules* **2**:3-16.

Brown T.A. (1999) How ancient DNA may help in understanding the origins and spread of agriculture. *Philosophical Transactions of the Royal Society of London Series B* **354**:89-98.

Brown T.A., Allaby R.G., Brown K.A. & Jones M.K. (1993) Biomolecular archaeology of wheat: past, present and future. *World Archaeology* **25**:64-73.

Brown T.A., Allaby R.G., Brown K.A., O'Donoghue K. & Sallares R. (1994) DNA in wheat seeds from European archaeological sites. *Experientia* **50**:571-575.

Brown W.M., George Jr. M. & Wilson A.C. (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the U.S.A.* **76**:1967-1971.

Buntjer J.B., Otsen M., Nijman I.J., Kuiper M.T.R. & Lenstra J.A. (2002) Phylogeny of bovine species based on AFLP fingerprinting. *Heredity* **88**:46-51.

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

Burger J., Hummel S. & Herrmann B. (2000) Palaeogenetics and cultural heritage. Species determination and STR-genotyping from ancient DNA in art and artefacts. *Thermochimica Acta* **365**:141-146.

Byrne R. (1987) Climatic change and the origins of agriculture. In: *Studies in the Neolithic and Urban Revolutions: The V. Gordon Childe Colloquium, Mexico, 1986* (ed. L. Manzanilla). British Archaeological Reports International Series, Oxford **349**:21-34.

Cano R.J., Poinar H.N., Pieniazek N.J., Acra A. & Poinar Jr. G.O. (1993) Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature* **363**:536-538.

Cavalli-Sforza L.L., Manozzi P. & Piazza A. (1993) Demic expansions and human evolution. *Science* **259**:639-646.

Chaix L. & Arbogast R.-M. (1999) Holocene aurochs from Western Europe. Osteometrical data. In: *Archäologie und Biologie des Auerochsen* (ed. G.-C. Weninger), pp.35-48. Neanderthal Museum, Germany.

Chamberlain A.T. (1999) Carsington Pasture Cave, Brassington, Derbyshire: a Prehistoric Burial Site. *Cave Archaeology and Palaeontology Research Archive* **1** [available at *http://www.shef.ac.uk/~capra/1/carsing.html*].

Chamberlain A.T. (2001) Radiocarbon Dates from Carsington Pasture Cave, Brassington, Derbyshire. *Cave Archaeology and Palaeontology Research Archive* **3** [available at *http://www.shef.ac.uk/~capra/3/carsdates.html*].

Chikuni K., Mori Y., Tabata T., Saito M., Monma M. & Kosugiyama M. (1995) Molecular phylogeny based on the kappa-casein and cytochrome *b* sequences in the mammalian suborder Ruminantia. *Journal of Molecular Evolution* **41**:859-866.

Clutton-Brock J. (1981) *Domestic Animals from Early Times*. Heinemann, London.

Clutton-Brock J. (1987) *A Natural History of Domesticated Mammals*. Cambridge University Press, Cambridge.

Clutton-Brock J. (1989) Cattle in ancient North Africa. In: *The Walking Larder: Patterns of Domestication, Pastoralism and Predation* (ed. J. Clutton-Brock), pp. 201-206. Unwin Hyman, London.

Clutton-Brock J. (1999) *A Natural History of Domesticated Mammals*. 2nd edition. University Press, Cambridge.

Cole S. (1967) *The Neolithic Revolution*. 4th edition. British Museum, London.

Colson I.B., Bailey J.F., Vercauteren M., Sykes B.C. & Hedges R.E.M. (1997) The preservation of ancient DNA and diagenesis. *Ancient Biomolecules* **1**:109-117.

Cooper A. & Poinar H.N. (2000) Ancient DNA: do it right or not at all. *Science* **289**:1139.

Cooper A., Mourer-Chauviré C., Chambers G.K., von Haeseler A., Wilson A.C. & Pääbo S. (1992) Independent origins of New Zealand moas and kiwis. *Proceedings of the National Academy of Sciences of the U.S.A.* **89**:8741-8744.

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

Cooper G., Amos W., Hoffman D., & Rubinsztein D.C. (1996) Network analysis of human Y microsatellite haplotypes. *Human Molecular Genetics* **5**:1759-1766.

Cymbron T., Loftus R.T., Malheiro M.I. & Bradley D.G. (1999) Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proceedings of the Royal Society of London Series B Biological Sciences* **266**:597-603.

Cymbron T., Loftus R.T., Malheiro M.I. & Bradley D.G. (in prep.) Microsatellite variation and patterns of cattle migration in Europe.

Davidson I. (1989) Escaped domestic animals and the introduction of agriculture to Spain. In: *The Walking Larder: Patterns of Domestication, Pastoralism and Predation* (ed. J. Clutton-Brock), pp. 59-71. Unwin Hyman, London.

Davies K. (1991) The past comes alive. *Nature* **352**:381.

Deakin W.J., Rowley-Conwy P. & Shaw C.H. (1998) Amplification and sequencing of DNA from preserved sorghum of up to 2800 years antiquity found at Qasr Ibrim. *Ancient Biomolecules* **2**:27-43.

Degerbøl M. (1962) Ur und hausrind. *Zeitschrift für Tierzüchtung und Züchtungbiologie* **76**:243-251.

Degerbøl M. & Fredskild B. (1970) *The Urus (Bos primigenius Bojanus) and Neolithic Domesticated Cattle (Bos domesticus Linné) in Denmark.* Munksgaard, København.

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.

Diamond J. (1997) *Guns, Germs, and Steel: the Fates of Human Societies.* Norton, New York.

Diamond J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature* **418**:700-707.

Dillon M. (1962) *Lebor na Cert: The Book of Rights.* Irish Texts Society, Dublin.

Douzery E. & Randi E. (1997) The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. *Molecular Biology and Evolution* **14**:1154-1166.

Edwards C.J. (1998) *Y-chromosome Microsatellite Variation in Domesticated Cattle and Related Species*. M.Sc. thesis, University of Sheffield.

Edwards C.J., Dolf G., Looft R.T. & Bradley D.G. (2000) Relationships between the endangered Pustertaler-Sprinzen and three related European cattle breeds as analysed with 20 microsatellite loci. *Animal Genetics* **31**:329-332.

Edwards C.J., Connellan J., Wallace P.F., Park S.D.E., McCormick F.M., Olsaker I., Eythrósdóttir E., MacHugh D.E., Bailey J.F. & Bradley D.G. (in prep.) Feasibility and utility of microsatellite markers in archaeological cattle remains from a Viking Age settlement in Dublin.

Epstein H. (1971) *The Origin of the Domestic Animals of Africa, Volume 1*. Africana Publishing Corporation, New York.

Epstein H. & Mason I.L. (1984) Cattle. In: *Evolution of Domesticated Animals* (ed. I.L. Mason), pp.6-27. Longman Inc., New York.

Excoffier L., Smouse P.E. & Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**:479-491.

Fairbanks R.G. (1989) A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**:637-642.

Felius M. (1995) *Cattle Breeds - An Encyclopaedia*. Misset, The Netherlands.

Felsenstein J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**:164-166.

Felsenstein J. (1993) PHYLIP version 3.572: distributed by the author. Department of Genetics, University of Washington, Seattle.

Fox C.L. (1997) mtDNA analysis in ancient Nubians supports the existence of gene flow between sub-Sahara and North Africa in the Nile Valley. *Annals of Human Biology* **24**:217-227.

Galton F. (1865) The first steps towards the domestication of animals. *Transactions of the Ethnological Society of London* **3**:122-138.

Gautier A. (1987) Prehistoric men and cattle in North Africa: a dearth of data and a surfeit of models. In: *Prehistory of Arid North Africa: Essays in Honor of Fred Wendorf* (ed. A. Close), pp.163-187. Southern Methodist University Press, Dallas.

Gentry A.W. (1992) The subfamilies and tribes of the family Bovidae. *Mammalian Review* **22**:1-32.

Gill A. & West A. (2001) *Extinct*. Channel 4 Books, Pan Macmillan Ltd., London.

Gill P., Ivanov P.L., Kimpton C., Piercy R., Benson N., Tully G., Evett I., Hagelberg E. & Sullivan K. (1994) Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics* **6**:130-135.

Giuffra E., Kijas J.M., Amarger V., Carlborg O., Jeon J.T. & Andersson L. (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**:1785-1791.

Goloubinoff P., Pääbo S. & Wilson A.C. (1993) Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proceedings of the National Academy of Sciences of the U.S.A.* **90**:1997-2001.

Golenberg E.M., Giannasi D.E., Clegg M.T., Smiley C.J., Durbin M., Henderson D. & Zurawski G. (1990) Chloroplast DNA sequence from a Miocene Magnolia species. *Nature* **344**:656-658.

Gonzalez S., Kitchener A.C & Lister A.M. (2000) Survival of the Irish elk into the Holocene. *Nature* **405**:753-754.

Grigson C. (1965) Faunal remains: measurements of bones, horncores, antlers and teeth. In: *Windmill Hill and Avebury: Excavations by Alexander Keiller 1923-1939* (ed. I.F. Smith), pp.145-167. Claredon Press, Oxford.

Grigson C. (1978) The craniology and relationships of four species of Bos. 4. The relationship between *Bos primigenius* Boj. and *B. taurus* L. and its implications for the phylogeny of the domestic breeds. *Journal of Archaeological Science* **5**:123-152.

Grigson C. (1980) The craniology and relationships of four species of Bos. 5. *Bos indicus L. Journal of Archaeological Science* 7:3-32.

Grigson C. (1982) Porridge and pannage: pig husbandry in Neolithic England. In: *Archaeological Aspects of Woodland Ecology* (ed. M. Bell and S. Limbrey). British Archaeological Reports International Series, Oxford **146**: 297-314.

Grigson C. (1985) *Bos indicus* and *Bos namadicus* and the problem of autochthonous domestication in India. In: *Recent Advances in Indo-Pacific Prehistory* (ed. V.N. Mistra and P. Bellwood), pp. 425-428. Oxford and IBH Publishing Company, New Delhi.

Grigson C. (1991) An African origin for African cattle? – some archaeological evidence. *The African Archaeological Review* **9**:119-144.

Guilaine J. (2000) De l'Orient à l'Occident: la Néolithisation de la Méditerranée. Questions ouvertes. In: *La Neolitizzazione tra Oriente e Occidente, Udine, April 1999. Museo Friulino di Storia Naturale – Accademia di Udine di Scienze, Lettere e Arti*, pp.11-21. Udine, Italy.

Guintard C. (1999) On the size of the ure-ox or aurochs (*Bos primigenius* Bojanus, 1827). In: *Archäologie und Biologie des Auerochsen* (ed. G.-C. Weniger), pp.7-21. Neanderthal Museum, Germany.

Hagelberg E. (1994) Mitochondrial DNA from ancient bones. In: *Ancient DNA* (ed. B. Herrmann and S. Hummel), p.195. Springer-Verlag, New York.

Hagelberg E. & Clegg J.B. (1991) Isolation and characterization of DNA from archaeological bone. *Proceedings of the Royal Society of London Series B Biological Sciences* **244**:45-50.

Hagelberg E., Sykes B. & Hedges R. (1989) Ancient bone DNA amplified. *Nature* **342**:485.

Hagelberg E., Bell L.S., Allen T., Boyde A., Jones S.J. & Clegg J.B. (1991a) Analysis of ancient bone DNA: techniques and applications. *Philosophical Transactions of the Royal Society of London Series B* **333**:399-407.

Hagelberg E., Gray I.C. & Jeffreys A.J. (1991b) Identification of the skeletal remains of a murder victim by DNA analysis. *Nature* **352**:427-429.

Hagelberg E., Quevedo S., Turbon D. & Clegg J.B. (1994) DNA from ancient Easter Islanders. *Nature* **369**:25-26.

Hall S.J.G. & Bradley D.G. (1995) Conserving livestock breed biodiversity. *Trends in Ecology and Evolution* **10**:267-270.

Halnan C.R.E. (1989) Karyotype and phenotype in cattle and hybrids of the genus. In: *Cytogenetics of Animals* (ed. C.R.E. Halnan), pp. 235-256. C.A.B. International, Wallingford, U.K..

Halnan C.R.E. & Watson J.I. (1982) Y chromosome variants in cattle *Bos taurus* and *Bos indicus*. *Annales de Genetique et de Selection Animale* **14**:1-16.

Hamada H., Petrino M.G. & Kakunaga T. (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Sciences of the U.S.A.* **79**:6465-6469.

Handt O., Höss M., Krings M. & Pääbo S. (1994) Ancient DNA: methodological challenges. *Experientia* **50**:524-529.

Hanotte O., Okomo M., Verjee Y., Rege E. & Teale A. (1997) A polymorphic Y chromosome microsatellite locus in cattle. *Animal Genetics* **28**:308-322.

Hanotte O., Tawah C.L., Bradley D.G., Okomo M., Verjee Y., Ochieng J. & Rege J.E. (2000) Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Molecular Ecology* **9**:387-396.

Harcourt R.A. (1971) Animal bones from Durrington Walls. In: *Durrington Walls: Excavations 1966-1968 Appendix II* (ed. G.J. Wainwright and I.H. Longworth), pp. 338-350. The Society of Antiquaries, London.

Hardy C., Casare D., Vigne J.-D., Callou C., Dennebouy N., Mounolou J.C. & Monnerot M. (1994) Ancient DNA from Bronze Age bones of European rabbit (*Oryctolagus cuniculus*). *Experientia* **50**:564-570.

Hardy C., Callou C., Vigne J.-D., Casañe D., Dennebouy N., Mounolou J.-C. & Monnerot M. (1995) Rabbit mitochondrial DNA diversity from prehistoric to modern time. *Journal of Molecular Evolution* **40**:227-237.

Hartl G.B., Göltenboth R., Grillitsch M. & Willing R. (1988) On the biochemical systematics of the Bovini. *Biochemical Systematics and Ecology* **16**:575-579.

Hauswirth W.W., Dickel C.D., Rowold D.J. & Hauswirth M.A. (1994) Inter- and intrapopulation studies of ancient humans. *Experientia* **50**:585-591.

Hayashi J., Tagashira Y. & Yoshida M.C. (1985) Absence of extensive recombination between inter- and intraspecies mitochondrial DNA in mammalian cells. *Experimental Cell Research* **160**:387-395.

Hedges R.E.M., Housley R.A., Bronk C.R. & van Klinken G.J. (1992) Radiocarbon dates from the Oxford AMS system: date list 14. *Archaeometry* **34**:141-159.

Hedges R.E.M., Millard A.R. & Pike A. (1995) Measurements and relationships of diagenetically altered bone from three archaeological sites. *Journal of Archaeological Science* **22**:201-210.

Hedges S.B. & Schweitzer M.H. (1995) Detecting dinosaur DNA. *Science* **268**:1191.

Henikoff S. (1995) Detecting dinosaur DNA. *Science* **268**, 1192.

Heun M., Schäfer-Pregl R., Klawan D., Castagna R., Accerbi M., Borghi B. & Salamini F. (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. *Science* **278**:1312-1314.

Hewitt G. (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**:907-913.

Hiendleder S., Mainz K., Plante Y. & Lewalski H. (1998) Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from Ural and Argali sheep. *Journal of Heredity* **89**:113-120.

Hiendleder S., Kaupe B., Wassmuth R. & Janke H. (2002) Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society of London Series B Biological Sciences* **269**:893-904.

Higham C. & Message M. (1969) An assessment of a prehistoric technique of bovine husbandry. In: *Science in Archaeology* (ed. D. Brothwell and E. Higgs), pp.315-330. Thames and Hudson, London.

Higuchi R., Bowman B., Freiberger M., Ryder O.A. & Wilson A.C. (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**:282-284.

Horai S., Hayasaka K., Murayama K., Wate N., Koike H. & Nakai N. (1989) DNA amplification from ancient human skeletal remains and their sequence analysis. *Proceedings of the Japanese Academy of Science Series B* **65**:229-233.

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

Höss M., Dilling A., Currant A. & Pääbo S. (1996) Molecular phylogeny of the extinct ground sloth *Mylodon darwinii*. *Proceedings of the National Academy of Sciences of the U.S.A.* **93**:181-185.

Hummel S. & Herrmann B. (1991) Y-chromosome-specific DNA amplified in ancient human bone. *Naturwissenschaften* **78**:266-267.

Ijzereef G.F. (1981) *Bronze Age Animal Bones from Bovenkarspel. The Excavation at Het Valkje*. Rijksdienst voor het Oudheidkundig Bodemonderzoek, Utrecht.

Irwin D.M., Kocher T.D. & Wilson A.C. (1991) Evolution of the cytochrome *b* gene of mammals. *Journal of Molecular Evolution* **32**:128-144.

Janecek L.L., Honeycutt R.L., Adkins R.M. & Davis S.K. (1996) Mitochondrial gene sequences and the molecular systematics of the Artiodactyl subfamily Bovinae. *Molecular Phylogenetics and Evolution* **6**:107-119.

Jarman M.R. (1969) The prehistory of Upper Pleistocene and recent cattle. Part 1: East Mediterranean, with reference to North-West Europe. *Proceedings of the Prehistoric Society* **35**:236-266.

Jeffreys A., Allen M., Hagelberg E. & Sonnberg A. (1992) Identification of the skeletal remains of Josef Mengele by DNA analysis. *Forensic Science International* **56**:65-76.

Jones G.F. (1998) Genetic aspects of domestication, common breeds and their origin. In: *The Genetics of the Pig* (ed. A. Ruvinsky and M.F. Rothschild), pp.17-50. C.A.B. International, Oxon.

Kantanen J., Olsaker I., Holm L.-E., Lien S., Vilkki J., Brusgaard K., Danell B., Eythrósdóttir E. & Adalsteinsson S. (2000) Genetic diversity and population structure of 20 North European cattle breeds. *Journal of Heredity* **91**:446-57.

Kappes S.M., Keele J.W., Stone R.T., McGraw R.A., Sonstegard T.S., Smith T.P.L., Lopez-Corrales N.L. & Beattie C.W. (1997) A second-generation linkage map of the bovine genome. *Genome Research* 7:235-249.

Kaukinen J. & Varvio S.L. (1993) Eight polymorphic bovine microsatellites. *Animal Genetics* **24**:148.

Kelly F. (1997) *Early Irish Farming: A Study Based Mainly on the Law-Texts of the 7th and 8th Centuries AD.* School of Celtic Studies, Dublin Institute for Advanced Studies, Dublin.

Kemp S.J., Hishida O., Wambugu J., Rink A., Longeri M.L., Ma R.Z., Da Y., Lewin H.A., Barendse W. & Teale A.J. (1995) A panel of polymorphic bovine, ovine and caprine microsatellite markers. *Animal Genetics* **26**:299-306.

Kijas J.M.H & Andersson L. (2001) A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *Journal of Molecular Evolution* **52**:302-308.

Kim K.-I., Lee J.-H., Li K., Zhang Y.-P., Lee S.-S., Gongora J. & Moran C. (2002) Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism. *Animal Genetics* **33**:19-25.

Kimura M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**:111-120.

Kimura M. & Crow J.F. (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**:725-738.

Kittles R.A., Perola M., Peltonen L., Bergen A.W., Aragon R.A., Virkkunen M., Linnoila M., Goldman D. & Long J.C. (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. *American Journal of Human Genetics* **62**:1171-1179.

Krings M., Halim Salem A., Bauer K., Geisert H., Malek A. K., Chaix L., Simon C., Welsby D., Di Rienzo A., Utermann G., Sajantila A., Pääbo S. & Stoneking M. (1999) mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? *American Journal of Human Genetics* **64**:1166-1176.

Kruskal J.B. (1964) Multidimensional scaling by optimising goodness to fit to a nonmetric hypothesis. *Pyschometrika* **29**:1-27.

Kumar S. S., Nasidze I., Walimbe S. R. & Stoneking M. (2000) Brief communication: discouraging prospects for ancient DNA from India. *American Journal of Physical Anthropology* **113**:129-133.

Kuwayama R. & Ozawa T. (2000) Phylogenetic relationships among European red deer, wapiti, and sika deer inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* **15**:115-123.

Lau C.H., Drinkwater R.D., Yusoff K., Tan S.G., Hetzel D.J. & Barker J.S. (1998) Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial DNA D-loop and cytochrome *b* sequence variation. *Animal Genetics* **29**:253-264.

Leonard J.A., Wayne R.K., Wheeler J., Valadez R., Guillén S. & Vilà C (2002) Ancient DNA evidence for Old World origin of New World dogs. *Science* **298**:1613-1616.

Lev-Yadun S., Gopher A. & Abbo S. (2000) The cradle of agriculture. *Science* **288**:1602-1603.

Lien S., Kantanen J., Olsaker I., Holm L.-E., Eythorsdottir E., Sandberg K., Dalsgard B. & Adalsteinsson S. (1999) Comparison of milk protein allele frequencies in Nordic cattle breeds. *Animal Genetics* **30**:85-91.

Linares A., Nayar K., Goldstein D.B., Hebert J.M., Seielstad M.T., Underhill P.A., Lin A.A., Feldman M.W. & Cavalli-Sforza L.L. (1996) Geographic clustering of human Y-chromosome haplotypes. *Annals of Human Genetics* **60**:401-408.

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

Lindahl T. (1997) Facts and artefacts of ancient DNA. *Cell* **90**:1-3.

Lister A.M. (1981) *Evolutionary Studies on Pleistocene Deer*. Ph.D. thesis, University of Cambridge.

Lister A.M. (1994a) Ancient DNA: not quite Jurassic Park. *Trends in Ecology and Evolution* **9**:82-84.

Lister A.M. (1994b) The evolution of the giant deer, *Megaloceros giganteus* (Blumenbach). *Zoological Journal of the Linnean Society* **112**:65-100.

Liu W.S., Mariani P., Beattie C.W., Alexander L.J. & Ponce de Leon F.A. (2002) A radiation hybrid map for the bovine Y-chromosome. *Mammalian Genome* **13**:320-326.

Loftus R.T., MacHugh D.E., Bradley D.G., Sharp P.M. & Cunningham P. (1994a) Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the U.S.A.* **91**:2757-2761.

Loftus R.T., MacHugh D.E., Ngere L.O., Balain D.S., Badi A.M., Bradley D.G. & Cunningham E.P. (1994b) Mitochondrial genetic variation in European, African and Indian cattle populations. *Animal Genetics* **25**:265-271.

Loftus R.T., Ertugrul O., Harba A.H., El-Barody M.A., MacHugh D.E., Park S.D. & Bradley D.G. (1999) A microsatellite survey of cattle from a centre in the Near East. *Molecular Ecology* **8**:2015-2022.

Loreille O., Vigne J.-D., Hardy C., Callou C., Treinen-Claustre F., Dennebouy N. & Monnerot M. (1997) First distinction of sheep and goat archaeological bones by the means of their fossil mtDNA. *Journal of Archaeological Science* **24**:33-37.

Luikart G., Gielly L., Excoffier L., Vigne J.-D., Bouvet J. & Taberlet P. (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences of the U.S.A.* **98**:5927-5932.

Macaulay V., Richards M., Hickey E., Vega E., Cruciani F., Guida V., Scozzari R., Bonne-Tamir B., Sykes B. & Torroni A. (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *American Journal of Human Genetics* **64**:232-249.

MacHugh D.E. (1996) *Molecular Biogeography and Genetic Structure of Domesticated Cattle*. Ph.D. thesis, University of Dublin.

MacHugh D.E. & Bradley D.G. (2001) Livestock genetic origins: goats buck the trend. *Proceedings of the National Academy of Sciences of the U.S.A.* **98**:5382-5384.

MacHugh D.E., Loftus R.T., Bradley D.G., Sharp P.M. & Cunningham E.P. (1994) Microsatellite DNA variation within and among European cattle breeds. *Proceedings of the Royal Society of London Series B Biological Sciences* **256**:25-31.

MacHugh D.E., Loftus R.T., Cunningham P. & Bradley D.G. (1996) Molecular biogeography of domesticated cattle revealed with microsatellite DNA polymorphisms. *Animal Genetics* **27** Supplement 2:23.

MacHugh D.E., Shriver M.D., Loftus R.T., Cunningham P. & Bradley D.G. (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146**:1071-1086.

MacHugh D.E., Loftus R.T., Cunningham P. & Bradley D.G. (1998) Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Animal Genetics* **29**:333-340.

MacHugh D.E., Troy C.S., McCormick F., Olsaker I., Eythrósdóttir E. & Bradley D.G. (1999) Early medieval cattle remains from a Scandinavian settlement in Dublin: genetic analysis and comparison with extant breeds. *Philosophical Transactions of the Royal Society of London Series B* **354**:99-109.

MacHugh D.E., Edwards C.J., Bailey J.F., Bancroft D.R. & Bradley D.G. (2000) The extraction and analysis of ancient DNA from bone and teeth: a survey of current methodologies. *Ancient Biomolecules* **3**: 81-102.

Maddison W.P. & Maddison D.R. (1992) MacClade version 3.05, software for the analysis of phylogeny and character evolution. Sinaeuer Associates Inc. Publishers, Massachusetts.

Magee D.A. (2002) *Molecular Genetic Investigations of the Diversity and Origins of Old and New World Cattle Populations*. Ph.D. thesis, University of Dublin.

Magee D.A., Meghen C., Harrison S., Troy C.S., Cymbron T., Gaillard C., Morrow A., Maillard J.C. & Bradley D.G. (2002) A partial African ancestry for the Creole cattle populations of the Caribbean. *Journal of Heredity* in press.

Mann H.B. & Whitney D.R. (1947) On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **18**:50-60.

Mannen H., Tsuji S., Loftus R.T. & Bradley D.G. (1998) Mitochondrial DNA variation and evolution of Japanese Black cattle (*Bos taurus*). *Genetics* **150**:1169-1175.

Mannucci A., Sullivan K M., Ivanov P.L. & Gill P. (1994) Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin. *International Journal of Legal Medicine* **106**:190-193.

Marshall F. (1989) Rethinking the role of *Bos indicus* in Sub-Saharan Africa. *Current Anthropology* **30**:235-240.

McCormick F. (1983) Dairying and beef production in early Christian Ireland: the faunal evidence. In: *Landscape Archaeology in Ireland* (ed. T. Reeves-Smyth and F. Hamond). British Archaeological Reports International Series, Oxford **116**:253-267.

McCormick F. (1987) *Stockrearing in Early Christian Ireland.* Ph.D. thesis, Queen's University Belfast.

McCormick F. (in prep.) The provision of meat in Viking Dublin.

McGovern T.H. (1992) Bones, buildings and boundaries: palaeoeconomic approaches to Norse Greenland. In: *Norse and Later Settlement and Subsistence in the North Atlantic* (ed. C.D. Morris and D.J. Rackham), pp. 193-230. University of Glasgow.

Meacham C.A. (1993) FACTOR program: distributed by J. Felsenstein as part of PHYLIP version 3.572. Department of Botany, University of Georgia, Athens.

Meadow R.H. (1984) Animal domestication in the Middle East: a view from the Eastern Margin. In: *Animals and Archaeology: 3. Early Herders and Their Flocks* (ed. J. Clutton-Brock and C. Grigson), pp. 309-337. Oxford University Press, Oxford.

Meadow R.H. (1989) Osteological evidence for the process of animal domestication. *In: Patterns of Domestication, Pastoralism and Predation* (ed. J. Clutton-Brock), pp. 80-90. Unwin Hyman, London.

Meadow R.H. (1993) Animal domestication in the Middle East: a revised view from the Eastern Margin. In: *Harappan Civilisation* (ed. G. Possehl), pp. 295-320. Oxford & I.B.H., New Dehli.

Medjugorac I., Kustermann W., Lazar P., Russ I. & Pirchner F. (1994) Marker-derived phylogeny of European cattle supports demic expansion of agriculture. *Animal Genetics* **25 Supplement 1**:19-27.

Melander Y. (1959) The mitotic chromosomes of some cavicorn mammals *Bos taurus* L., *B. bonasus* L., *O. aries* L.. *Hereditas* **45**:649-664.

Merriwether D.A., Rothhammer F. & Ferrell R.E. (1994) Genetic variation in the New World: ancient teeth, bone, and tissue as sources of DNA. *Experientia* **50**:592-601.

Mezzelani A., Zhang Y., Redaelli L., Castiglioni B., Leone P., Williams J. L., Toldo S. S., Wigger G., Fries R. & Ferretti L. (1995) Chromosomal localisation and molecular characterisation of 53 cosmid-derived bovine microsatellites. *Mammalian Genome* **6**:629-635.

Moazami-Goudarzi K., Laloë D., Furet J.P. & Grosclaude F. (1997) Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites. *Animal Genetics* **28**:338-345.

Moen R.A., Pastor J. & Cohen Y. (1999) Antler growth and extinction of Irish elk. *Evolutionary Ecology Research* **1**:235-249.

Molyneux T. (1697) A discourse concerning the large horns frequently found under ground in Ireland, concluding from them that the great American deer, call'd a moose, was formerly common in that island: with remarks on some other things natural to that country. *Philosophical Transactions of the Royal Society of London* **29**:489-512.

Monaghan N.T. (1995) The Irish giant deer or 'Irish elk'. *The Geological Curator* **6**:171-173.

Monsen E. (1932) *Heimskringla by Snorre Sturlason*. Heffer, Cambridge.

Moore A.M.T. & Hillman G.C. (1992) The Pleistocene to Holocene transition and human economy in southwest Asia: the impact of the Younger Dryas. *American Antiquity* **57**:482-494.

Nei M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M., Tajima F. & Tateno Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* **19**:153-170.

Nijman I.J., Bradley D.G., Hanotte O., Otsen M & Lenstra J.A. (1999) Satellite DNA polymorphisms and AFLP correlate with *Bos indicus-taurus* hybridization. *Animal Genetics* **30**:265-273.

Nowak R.M. (1991) *Walker's Mammals of the World.* John Hopkins University Press, Baltimore.

O'Brien E. (1998) The location and context of Viking burials at Kilmainham and Islandbridge, Dublin. In: *Ireland and Scandinavia in the Early Viking Age* (ed. H.B. Clarke, M. Ní Mhaonaigh and F. Ó Floinn), pp. 203-221. Four Courts Press, Dublin.

Ohta T. & Kimura M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**:201-204.

Okumura N., Ishiguro N., Nakano M., Hirai K., Matsui A. & Sahara M. (1996) Geographic population structure and sequence divergence in the mitochondrial DNA control region of the Japanese wild boar (*Sus scrofa leucomystax*), with reference to those of domestic pigs. *Biochemical Genetics* **34**:179-189.

Oliver W.L.R. (ed.) (1993) *Pigs, Peccaries, and Hippos.* International Union for Conservation of Nature and Natural Resources, Switzerland.

Pääbo S. (1993) Ancient DNA. *Scientific American* **November**:60-66.

Pääbo S., Higuchi R.G. & Wilson A.C. (1989) Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *Journal of Biological Chemistry* **264**:9709-9712.

Pääbo S., Irwin D.M. & Wilson A.C. (1990) DNA damage promotes jumping between templates during enzymatic amplification. *Journal of Biological Chemistry* **265**:4718-4721.

Palmer D. (1992) The DNA that stays forever amber. *New Scientist* **17th October**:15.

Park S.D.E. (2002) *Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection.* Ph.D. thesis, University of Dublin.

Paxton J. (1999) *A Study of the Macrofaunal Assemblage from Carsington Pasture Cave.* Unpublished dissertation, Department of Archaeology and Prehistory, University of Sheffield.

Payne W.J.A. (1964) The origin of domestic cattle in Africa. *Empire Journal of Experimental Agriculture* **32**:97-113.

Payne W.J.A. (1991) Domestication: a forward step in civilisation. In: *Cattle Genetic Resources* (ed. C.G. Hickman). 1st edition. Elsevier, Amsterdam.

Payne W. & Bull G. (1988) Components of variation in measurements of pig bones and teeth, and the use of measurements to distinguish wild from domestic pig remains. *Archaeozoologia* 2:27-66.

Pereira L., Dupanloup I., Rosser Z.H., Jobling M.A. & Barbujani G. (2001) Y-chromosome mismatch distributions in Europe. *Molecular Biology and Evolution* **18**:1259-1271.

Perkins D. (1973) The beginnings of animal domestication in the Near East. *American Journal of Archaeology* **77**:278–282.

Perret J., Shia Y.-C., Fries R., Vassart G. & Georges M. (1990) A polymorphic satellite sequence maps to the pericentric region of the bovine Y chromosome. *Genomics* **6**:482-490.

Peters J., Helmer D., von den Drisch A. & Saña-Segui M. (1999) Animal domestication and husbandry in PPN Northern Syria and South-Eastern Turkey. *Paléorient* **25**:27-47.

Petit E., Balloux F. & Excoffier L. (2002) Mammalian population genetics: why not Y? *Trends in Ecology and Evolution* **17**:28-33.

Pilgrim G.E. (1947) The evolution of the buffaloes, oxen, sheep and goats. *Journal of the Linnean Society of London Zoo* **41**:272-286.

Poinar H.N., Cano R.J. & Poinar Jr. G.O. (1993) DNA from an extinct plant. *Nature* **363**:677.

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

Polziehn R.O. & Strobeck C. (1998) Phylogeny of wapiti, red deer, sika deer, and other North American cervids as determined from mitochondrial DNA. *Molecular Phylogenetics and Evolution* **10**:249-258.

Polziehn R.O., Hamr J., Mallory F.F. & Strobeck C. (1998) Phylogenetic status of North American wapiti (*Cervus elaphus*) subspecies. *Canadian Journal of Zoology* **76**:998-1010.

Pringle H. (1998a) The slow birth of agriculture. *Science* **282**:1446.

Pringle H. (1998b) Reading the signs of ancient animal domestication. *Science* **282**:1448.

Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**:945-959.

Rackham J. (1994) *Interpreting the Past: Animal Bones.* British Museum Press, London.

Raftery B. (1997) Implications of the monkey bones from Site B. In: *Excavations at Navan Fort 1961-71* (ed. C.J. Lynn), pp. 121-125. The Stationary Office, Belfast.

Ramos M.D., Lalueza C., Girbau E., Perez-Perez A., Quevedo S., Turbon D. & Estivill X. (1995) Amplifying dinucleotide microsatellite loci from bone and tooth samples of up to 5000 years of age: more inconsistency than usefulness. *Human Genetics* **96**:205-212.

Randi E., Mucci N., Pierpaoli M. & Douzery E. (1998) New phylogenetic perspectives on the Cervidae (Artiodactyla) are provided by the mitochondrial cytochrome *b* gene. *Proceedings of the Royal Society of London Series B Biological Sciences* **265**: 793-801.

Randi E., Mucci N., Claro-Hergueta F., Bonnet A. & Douzery E.J.P. (2001) A mitochondrial DNA control region phylogeny of the Cervinae: speciation in *Cervus* and implications for conservation. *Animal Conservation* **4**:1-11.

Remington C.L. (1968) Suture-zones of hybrid interaction between recently joined biotas. *Evolutionary Biology* **2**:321-428.

Rendel J.M. (1980) Low calving rates in Brahman cross cattle. *Theoretical and Applied Genetics* **58**:207-210.

Reynolds J., Weir B.S. & Cokerham C.C. (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105**: 767-779.

Richards M.B., Hedges R.E.M. & Sykes B.C. (1995) Authenticating DNA extracted from ancient skeletal remains. *Journal of Archaeological Science* **22**:291-299.

Ritz L.R., Glowatzki-Mullis M.L., MacHugh D.E. & Gaillard C. (2000) Phylogenetic analysis of the tribe Bovini using microsatellites. *Animal Genetics* **31**:178-185.

Robin E.D. & Wong R. (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of Cellular Physiology* **136**:507-513.

Roewer L., Kayser M., Dieltjes P., Nagy M., Bakker E., Krawczak M. & de Knijff P. (1996) Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Human Molecular Genetics* **5**:1029-1033.

Rogers A.R. (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* **49**:608-615.

Rogers A.R. and Harpending H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**:552-569.

Rollo F., Venanzi F.M. & Amici A. (1991) Nucleic acids in mummified plant seeds: biochemistry and molecular genetics of pre-Columbian maize. *Genetical Research* **58**:193-201.

Ruane J. (1999) A critical review of the value of genetic distance studies in conservation of animal genetic resources. *Journal of Animal Breeding and Genetics* **116**:317-323.

Saitou N. & Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.

Salo W.L., Aufderheide A.C., Buikstra J. & Holcomb T.A. (1994) Identification of *Mycobacterium tuberculosis* DNA in a Pre-Columbian Peruvian Mummy. *Proceedings of the National Academy of Sciences of the U.S.A.* **91**:2091-2094.

Sambraus H.H. (1992) *A Colour Atlas of Livestock Breeds*. Wolfe Publishing Ltd., London.

Scherf B.D. (ed.) (1995) *World Watch List for Domestic Animal Diversity*. Food and Agriculture Organization (F.A.O.) of the United Nations, Rome.

Schlötterer C. & Tautz D. (1993) Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* **20**:211-215.

Schmerer W.M., Hummel S. & Herrmann B. (1999) Optimized DNA extraction to improve reproducibility of short tandem repeat genotyping with highly degraded DNA as target. *Electrophoresis* **20**:1712-1716.

Schmid M., Saitbekova N., Gaillard C. & Dolf G. (1999) Genetic diversity in Swiss cattle breeds. *Journal of Animal Breeding and Genetics* **116**:1-8.

Schneider S., Kueffer J.-M., Roessli D. & Excoffier L. (1997) ARLEQUIN version 2.000, a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sidow A., Wilson A.C. & Pääbo S. (1991) Bacterial DNA in Clarkia fossils. *Philosophical Transactions of the Royal Society of London Series B* **333**:429-433.

Simpson L. (2000) Forty years a-digging: a preliminary synthesis of archaeological investigations in medieval Dublin. In: *Medieval Dublin I* (ed. S. Duffy), pp. 11-68. Four Courts Press, Dublin.

Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**:457-462.

Smith B.D. (1995) *The Emergence of Agriculture*. Scientific American Library, New York.

Sokal R.R., Oden N.L. & Wilson C. (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* **351**:143-145.

Stokes W. (1862) *Three Irish Glossaries*. Norgate, London.

Stone A.C., Milner G.R., Pääbo S. & Stoneking M. (1996) Sex determination of ancient human skeletons using DNA. *American Journal of Physical Anthropology* **99**:231-238.

Stuart A.J. (1991) Mammalian extinctions in the late Pleistocene of northern Eurasia and north America. *Biological Review* **66**: 453-563.

Stuart A.J. & van Wijngaarden-Bakker L.H. (1985) Quaternary vertebrates. In: *The Quaternary History of Ireland.* (ed. K.J. Edwards and W. Warren), pp. 221-249. Academic Press, London.

Sykes B.C. (1991) Ancient DNA: the past comes alive. *Nature* **352**:381-382.

Szafer W. (1968) The ure-ox, extinct in Europe since the XVIIth century: and early attempt at conservation that failed. *Biological Conservation* **1**:45-47.

Taberlet P., Griffin S., Goossens B., Questiau S., Manceau V., Escaravage N., Waits L.P. & Bouvet J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**:3189-3194.

Taberlet P., Fumagalli L., Wust-Saucy A.-G. & Cosson J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7:453-464.

Takezaki N. & Nei M. (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**:389-399.

Tanaka K., Solis C.D., Masangkay J.S., Maeda K., Kawamoto Y. & Namikawa T. (1996) Phylogenetic relationship among all living species of the genus *Bubalus* based on DNA sequences of the cytochrome *b* gene. *Biochemical Genetics* **34**:443-452.

Tate M. (1997) *Evolution of Ruminant Chromosomes.* Ph.D. thesis, University of Otago, New Zealand.

Teale A.J., Wambugu J., Gwakisa P.S., Stranzinger G., Bradley D. & Kemp S.J. (1995) A polymorphism in randomly amplified DNA that differentiates the Y chromosomes of *Bos indicus* and *Bos taurus*. *Animal Genetics* **26**:243-248.

Thomas M.G., Skorecki K., Ben-Ami H., Parfitt T., Bradman N. & Goldstein D.B. (1998) Origins of Old Testament priests. *Nature* **394**: 138-139.

Thomas W.K. & Pääbo S. (1993) DNA sequences from old tissue remains. *Methods in Enzymology* **224**:406-419.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F. & Higgins D.G. (1997) The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**:4876-4882.

Tresset A. (2000) Early husbandry in Atlantic areas. Animal introductions, diffusions of techniques and native acculturation at the North-Western Fringe of Europe. In: *The Prehistory and Early History of Atlantic Europe, Götebord, 1998* (ed. J.C. Henderson). British Archaeological Reports International Series, Oxford **861**:17-32.

Troy C.S. (1998) *Mitochondrial Diversity and Biogeography in Domestic Cattle.* Ph.D. thesis, University of Dublin.

Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P., Chamberlain A.T., Sykes B.C. & Bradley D.G. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**:1088-1091.

Tudge C. (1998) *Neanderthals, Bandits and Farmers. How Agriculture Really Began.* Weidenfeld & Nicolson, London.

Turner C.L., Grant A., Bailey J.F., Dover G.A. & Barker G.W.W. (1998) Patterns of genetic diversity in extant and extinct cattle populations: evidence from sequence analysis of mitochondrial coding regions. *Ancient Biomolecules* **2**:235-249.

Uerpmann H.-P. (1996) Animal domestication – accident or intention? In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia* (ed. D.R. Harris), pp. 227-237. U.C.L. Press Limited, London.

Ursing B.M. & Arnason U. (1998) The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). *Journal of Molecular Evolution* **47**:302-306.

Vaiman D., Mercier D., Moazami-Goudarzi K., Eggen A., Ciampolini R., Lepingle A., Velmala R., Kaukinen J., Varvio S.L., Martin P., Levéziel H. & Guérin G. (1994) A set of 99 cattle microsatellites: characterisation, synteny mapping, and polymorphism. *Mammalian Genome* **5**:288-297.

Valente M. (2000) Dublin's economic relations with hinterland and periphery in the later Viking age. In: *Medieval Dublin I* (ed. S. Duffy), pp. 69-83. Four Courts Press, Dublin.

Valsecchi E., Palsböll P., Hale P., Glocker-Ferrari D., Ferrari M., Clapham P., Larsen F., Mattila D., Sears R., Sigurjonsson J., Brown M., Corkeron P & Amos B. (1997) Microsatellite genetic distances between oceanic populations of the Humpback whale (*Megaptera novaeangliae*). *Molecular Biology and Evolution* **14**:355-362.

Vigne J.-D. (1998) The large 'true' Mediterranean islands as a model for the Holocene human impact on the European vertebrate fauna? Recent data and new reflections. In: *The Holocene History of the European Vertebrate Fauna. Modern Aspects of Research* (ed. N. Benecke), pp. 295-322. Verlag Marie Leidorf, Germany.

Vilà C., Savolainen P., Maldonado J.E., Amorim I.R., Rice J.E., Honeycutt R.L., Crandall K.A., Lundeberg J. & Wayne R.K. (1997) Multiple and ancient origins of the domestic dog. *Science* **276**:1687-1689.

Vilà C., Leonard J.A., Gotherstrom A., Marklund S., Sandberg K., Liden K., Wayne R.K. & Ellegren H. (2001) Widespread origins of domestic horse lineages. *Science* **291**:474-477.

Von den Driesch A. (1976) *A Guide to the Measurement of Animal Bones from Archaeological Sites*. Peabody Museum, Harvard.

Wallace P.F. (1981) Dublin's waterfront at Wood Quay: 900-1317. In: *Waterfront Archaeology in Britain and Northern Europe* (ed. G. Milne and B. Hobley). C.B.A. Research Reports **41**:109-118.

Wallace P.F. (1984) A reappraisal of the archaeological significance of Wood Quay. In: *Viking Dublin Exposed* (ed. J. Bradley), pp. 112-133. The O'Brien Press, Dublin.

Wallace P. (1985) The archaeology of Viking Dublin. In: *Comparative History of Urban Origins in Non-Roman Europe* (ed. H.B. Clarke and A. Simms). British Archaeological Reports International Series, Oxford **255**:103-145.

Wallace P.F. (1986) The English presence in Viking Dublin. In: *Anglo-Saxon Monetary History* (ed. M.A.S. Blackburn), pp. 201-222. Leicester University Press, U.K..

Wallace P.F. (1987) The economy and commerce of Viking Age Dublin. In: *Untersuchungen zu Handel und Verkehr der von- und Frühgeschichtlichen Zeit in Mittel- und Nordeuropa* (ed. K. Duwel, H. Jankuhn, H. Siems and D. Timpe), pp. 200-245. Vanderhoeck & Ruprecht, Gottingen.

Wallace P.F. (1992) *The Viking Age Buildings of Dublin: Medieval Dublin Excavations 1962-81*. National Museum of Ireland, Royal Irish Academy, Dublin.

Ward T.J., Bielawski J.P., Davis S.K., Templeton J.W & Derr J.N (1999) Identification of domestic cattle hybrids in wild cattle and bison species: a general approach using mtDNA markers and the parametric bootstrap. *Animal Conservation* **2**:51-57.

Ward T.J., Skow L.C., Gallagher D.S., Schnabel R.D., Nall C.A., Kolenda C.E., Davis S.K., Taylor J.F. & Derr J.N. (2001) Differential introgression of uniparentally inherited markers in bison populations with hybrid ancestries. *Animal Genetics* **32**:89-91.

Watanobe T., Okumura N., Ishiguro N., Nakano M., Matsui A., Sahara M. & Komatsu M. (1999) Genetic relationship and distribution of the Japanese wild boar (*Sus scrofa leucomystax*) and Ryukyu wild boar (*Sus scrofa riukiuanus*) analysed by mitochondrial DNA. *Molecular Ecology* **8**:1509-1512.

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. *Journal of Molecular Evolution* **52**:281-289.

Watanobe T., Ishiguro N., Nakano M., Takamiya H., Matsui A. & Hongo H. (2002) Prehistoric introduction of domestic pigs onto the Okinawa Islands: ancient mitochondrial DNA evidence. *Journal of Molecular Evolution* **55**:222-231.

Waterbolk H.T. (1968) Food production in prehistoric Europe. *Science* **162**:1093-1102.

Weale M.E., Weiss D.A., Jager R.F., Bradman N. & Thomas M.G. (2002) Y chromosome evidence for Anglo-Saxon mass migration. *Molecular Biology and Evolution* **19**:1008-1021.

Weber J.L. & Wong C. (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* **2**:1123-1128.

Wendorf F. & Schild R. (1994) Are the early Holocene cattle in the Eastern Saharan domestic or wild? *Evolutionary Anthropology* **3**:118-128.

Wenke R.J. (1999) *Patterns in Prehistory: Humankind's First Three Million Years*. 4th edition. Oxford University Press, Oxford.

Wilson J. (1909) The Scandinavian origin of the hornless cattle of the British Isles. *Scientific Proceedings of the Royal Dublin Society* **12**:145-164.

Wood N.J. & Phua S.H. (1996) Variation in the control region sequence of the sheep mitochondrial genome. *Animal Genetics* **27**:25-33.

Woodman P., McCarthy M. & Monaghan N. (1997) The Irish Quaternary Fauna Project. *Quaternary Science Reviews* **16**:129-159.

Woodward S.R., Weyand N.J. & Bunnell M. (1994) DNA Sequence from Cretaceous period bone fragments. *Science* **266**:1229-1232.

Yalden D. (2001) The return of the prodigal swine. *Biologist* **48**:259-262.

Yang D.Y., Eng B., Waye J.S., Dudar J.C. & Saunders S.R. (1998) Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology* **105**:539-543.

Zeuner F.E. (1953) The colour of the wild cattle at Lascaux. *Man* **53**:68-69.

Zeuner F.E. (1955) The goats of early Jericho. *Palestine Exploration Quarterly* **April**:70-86.

Zeuner F.E. (1963) *A History of Domesticated Animals*. Hutchinson, London.

Zierdt H., Hummel S. & Herrmann B. (1996) Amplification of human short tandem repeats from medieval teeth and bone samples. *Human Biology* **68**:185-199.

Zilhão J. (2001) Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proceedings of the National Academy of Sciences of the U.S.A.* **98**:14180-14185.

Zimmerman S., Zehner R. & Herzog A. (1998) Cytochrom *b* – sequenzvergleich zwischen Rothirsch (*Cervus elaphus hippelaphus*), Damhirsch (*Dama dama dama*) und Reh (*Capreolus capreolus capreolus*). *Zeitschrift fur Jagdwissenschaft* **44**:16-25.

Zohary D. (1988) *Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe and the Nile Valley*. Clarendon Press, Oxford.

APPENDICES

Appendix A - Publications arising from this thesis

Edwards C.J., Bradley D.G., MacHugh D.E., Dobney K., Martin L., Horwitz L.K., McIntosh S.K., Tresset A., Vigne J.-D. & Helmer D. (in prep.) Ancient DNA analysis of 101 domestic cattle remains: limits and prospects.

Edwards C.J., Connellan J., Wallace P.F., Park S.D.E., McCormick F.M., Olsaker I., Eythrósdóttir E., MacHugh D.E., Bailey J.F. & Bradley D.G. (in prep.) Feasibility and utility of microsatellite markers in archaeological cattle remains from a Viking Age settlement in Dublin.

Kantanen J., Miceikiené I., Holm L.-E., Viinalass H., **Edwards C.J.**, Eythrósdóttir E. & Bradley D.G. (2001) Y-chromosomal haplotypes in North-European cattle breeds. In: *Book of Abstracts of the 52nd Annual Meeting of the European Association for Animal Production* (ed. Y. van der Honing), p. 7.

MacHugh D.E., **Edwards C.J.**, Bailey J.F., Bancroft D.R. & Bradley D.G. (2000) The extraction and analysis of ancient DNA from bone and teeth: a survey of current methodologies. *Ancient Biomolecules* **3**: 81-102.

Edwards C.J., Dolf G., Looft R.T. & Bradley D.G. (2000) Relationships between the endangered Pustertaler-Sprinzen and three related European cattle breeds as analysed with 20 microsatellite loci. *Animal Genetics* **31**:329-332.

Edwards C.J., Gaillard C., Bradley D.G. & MacHugh D.E. (2000) Y-specific microsatellite polymorphisms in a range of bovid species. *Animal Genetics* **31**:127-130.

Appendix B - Ancient DNA Analysis of Wild Boar and Domestic Pig

B.1 Abstract

418 base pairs, of a 440 bp area of the hypervariable mitochondrial D-loop region studied by Giuffra *et al.* (2000), was amplified in six out of seven attempted ancient *Sus scrofa* remains from Carsington Pasture Cave in Derbyshire. Dating to around 6,000 BP, these bones included four domestic animals, two wild boar specimens and one bone of undefined domestic status. This study is the first known ancient DNA analysis of British *S. scrofa* remains.

B.2 Introduction

B.2.1 *Classification of pigs and related species*

The hog (pig) has existed for at least 45 million years as a stocky, round-bodied, flat-snouted creature. Pigs are part of the order Artiodactyla (even-toed, hoofed animals) and all belong to the family Suidae, which comprises three subfamilies and five genera. The subfamily Suinae, the *true* pigs, contains the *Sus* genus, which includes eight species, including the Eurasian wild boar (*Sus scrofa*), a popular game animal during medieval times, which still persists over many parts of Europe, Asia, and North Africa. In its native form, *S. scrofa* has one of the largest ranges of any wild ungulate, and, together with its domestic and feral derivatives, it has one of the most extensive distributions of all mammals. They are markedly variable in size, pelage and other characters. The 16 subspecies of Eurasian pig have been clustered into four subspecies groupings: 1) *Western races* – Europe (including *S. s. scrofa*), North Africa and Middle East; 2) *Indian races*; 3) *Eastern races* – Mongolia and the Soviet Far East, Japan (*S. s. leucomystax* and *S. s. riukiuanus*), Taiwan, southeast China and Vietnam; and 4) *Indonesian races* – Malay Peninsula, Sumatra, Java, Bali and other offshore islands in the archipelago (Oliver *et al.* 1993).

B.2.2 *The origins of domesticated pigs*

While most early livestock were utilised by nomadic peoples, swine are more indicative of a settled farming community and domestication of the pig coincided with the formation of the first permanent human settlements. In many respects, wild pigs are well suited to domestication. Rather than forming large herds, they form smaller family groups where

one animal is dominant, making it easier for humans to manipulate behaviour by taking control (Smith 1995). Another advantage to domesticating pigs is the relative ease with which younger animals can be tamed, as well as the ability of pigs to grow quickly, produce large litters and survive on waste food (Grigson 1982).

The earliest positively domesticated pigs were found in the Near East at Qalat Jarmo, dated to around 6,750 BC (Bökönyi 1976). Management of pigs has also been found in Jericho, at 7,000 BC (Clutton-Brock 1999). However, these remains are from fully developed domestic animals and there are no traces of local domestication at either site. One possible candidate for the first occurrence of domestication in the Near East is the site of Cayönü in Anatolia, dating to 8,500 BP, with pigs accounting for 45% of animal remains (Perkins 1973; Smith 1995). This site, plus the others generating remains of domestic pig prior to 7,000 BP (Assoud, Gritille, Hayaz and Cafer), all fall into the northern range of the Fertile Crescent (Smith 1995).

Taxonomic relationships between wild boar and domestic pigs have long been uncertain. Although European wild boar are ancestral to the original domestic pigs of the Middle East and Europe, archaeologically it is less certain whether Chinese domesticates share the same ancestry, have a separate ancestry from an eastern race of the same species, or an ancestry from one of the related Oriental species. The first domestic pigs in China existed about 6,900 years ago, although there is evidence for domestication dating to 10,000 BP from the Zengpiyan Caves. Domestic pigs have been present in Britain from the early Neolithic, and the Vikings, plus Spanish and English colonists, brought swine to the Americas. Today wild and domestic pigs can be found on every continent except Antarctica. By means of selection and controlled mating, over 300 breeds of pigs have been raised, including: Berkshire, Chester White, Duroc, Hampshire, Landrace, Large White, Tamworth and Yorkshire. These are all termed as *S. s. domesticus*.

During the 18th and early 19th centuries, a transformation occurred in the European breeds with the introduction of domestic swine from China and south-east Asia to improve stock (Jones 1998). This cross-breeding proved so popular that it is now thought that few pigs in Europe are pure-bred descendants of pigs domesticated in western Europe (Clutton-Brock 1999).

B.2.3 *History of pigs in Britain*

Wild boar were absent from Britain during the late glacial, but moved quickly back into the country in the Flandrian period (Yalden 2001). Boar were being hunted, along with red deer, roe deer, elk, aurochs and beaver, by Mesolithic hunters at Starr Carr in Yorkshire around 9,500 BP. During the Neolithic around 5,500 BP, domestic sheep, goats and cattle were introduced to Britain; but archaeological evidence for domesticated pigs also being brought in at this time is less certain. It is thought, rather, that domestic pigs have been present in Britain since 6000 BP (Clutton-Brock 1999). Most of the bones found at Windmill Hill are certain to have belonged to domestic animals (based on measurements by Grigson 1965), and three such *S. scrofa* remains have been radiocarbon dated to 4,730 ± 70, 4,690 ± 70 and 4,650 ± 70 bp uncalibrated (approximately 3,500 BC; Hedges *et al.* 1992). The earliest record of domestic pig in Britain is at Hazelton North, dated to 4,915 bp (A. Chamberlain personal communication).

In the early British Neolithic, between 5,000 and 4,500 bp, cattle were the predominant domestic animals. However, pigs comprise the major amount of ungulate bones present in the late Neolithic (4,000 to 3,500 bp), with their numbers varying from 42 to 68%. Almost all pig bones and teeth found during this time are from domestic and not wild animals, and the two populations are thought to have been distinct with no suggestion of interbreeding (Grigson 1982).

B.2.4 *Carsington Pasture Cave (CPC) site details*

CPC is formed in dolomitised limestone of the Caboniferous series, near the southern margin of the Derbyshire lead-zinc orefield (Chamberlain 1999). The faunal remains are dominated by domestic species, principally cattle (*Bos taurus*) and pig, with dog, horse, sheep and goat occurring less frequently. There are a small number of bones of wild species, including red deer, roe deer, aurochs, badger, red fox and bear. These bones show few signs of butchery or other secondary product utilisation, but there is copious evidence of gnawing by medium sized carnivores. *S. scrofa* remains occur throughout the cave system, and surviving long bone elements suggest that whole carcasses rather than individual skeletal elements were entering the cave system. Chamberlain (1999) inferred from this, and the low frequency of butchery on the bones (just 2 out of 139 pig bones exhibit cut marks), that the pigs may have fallen into the cave through vertical fissures, or have been brought in by carnivores as partial or complete carcasses.

Although radiocarbon dating has not been done on any of the pig remains, an aurochs bone (CPC-98-1000) has been dated to the early Neolithic (5145 ± 70 uncalibrated bp or 4,250 – 3,700 cal. BC) (Chamberlain 2001; Troy *et al.* 2001), and there is no reason to suspect that the *Sus* bones would not date to a similar time. Although exploration of the cave has led to partial disturbance of the deposits, the sediments have been left *in situ* wherever possible.

B.2.5 *Genetic studies of modern pig mitochondrial DNA*

The complete *S. scrofa* mitochondrion genome (16,679 bases) was sequenced in 1998 by Ursing & Arnason (GENBANK accession number AJ002189). Giuffra *et al.* (2000) analysed the entire mitochondrial cytochrome *b* gene plus a 440 bp area of the hypervariable D-loop region. They found clear evidence for independent domestication of two subspecies of wild boar, alternately in Europe and Asia, with the time since divergence of the two forms estimated to be of the order of 500,000 years. There are 27 polymorphic sites in the 440 bp control region, and the 130 modern sequences analysed fell into three distinct mtDNA clades: one Asian and two European.

1) The Asian clade (A) includes Japanese wild boars, domestic Chinese Meishan pigs, and some European domestic animals.

2a) European clade 1 (EI) is composed of the majority of wild boars from Europe and Israel, most European domestic pigs, and a pig from Cook Island.

2b) European clade 2 (EII) includes only three wild boars from southern Italy.

All mtDNA haplotypes having a presumed European origin and found in domestic pigs belonged to EI. These three clades have since been confirmed by analysis of 15,997 bp of the mtDNA genome in three domestic pigs and two European wild boars (Kijas & Andersson 2001). Using the functional component of the mitochondrial genome, Guiffra *et al.*'s study predicted the time since divergence for the three clades, resulting with estimates of 669,000 years between clade EI and clade EII; 908,000 years between clade EI and clade A; and 807,000 years between clade EII and clade A. However, Kim *et al.* (2002) have estimated a shorter divergence time of 58,000 BP between the European clades and clade A using 1036 bp of the control region. In either case, these large divergence estimates predate the domestication of pigs.

Ancient DNA analyses in *S. scrofa* have concentrated on 574 bp of the 3' end of the control region in Japanese island populations dating to between the 5th and 12th centuries AD (Watanobe *et al.* 2001, 2002).

B.3 Materials & Methods

Seven *S. scrofa* bones from Carsington Pasture Cave (CPC), in Derbyshire, were sourced by Andrew Chamberlain, University of Sheffield, for ancient DNA analysis. Details of the samples are shown in **Table 4.1**. Apart from P7, which is only tentatively labelled as wild boar, measurements taken from the other six bones (P1 to P6) by Andrew Chamberlain were compared to reference data (Payne & Bull 1988), enabling the differentiation of wild and domestic forms of European pig. These six specimens were taken from different locations in the chamber of the cave, and therefore each sample was taken to represent a separate individual animal. The specimen P7 is a juvenile or subadult, and therefore attribution to either wild or domestic pig on the basis of size was insecure.

Lab code	CPC code	Skeletal element	Designation	Location
A211 - P1	CPC-98-1112	left pelvis	wild boar	surface of 2nd chamber
A212 - P2	CPC-98-2218	right mandible	wild boar	below 2nd chamber
A213 - P3	CPC-98-2230	right tibia	wild boar	below 3rd chamber
A214 - P4	CPC-98-1109	left humerus	domestic pig	surface of 2nd chamber
A215 - P5	CPC-98-2280	left radius	domestic pig	surface of 3rd chamber
A216 - P6	CPC-98-2325	right scapula	domestic pig	below 3rd chamber
A217 - P7	CPC-98-?	vertebra	unknown	below 2nd chamber

Table B.1 Details for the seven *Sus scrofa* samples from Carsington Pasture Cave (CPC).

The bone samples were prepared using the procedure modified from Yang *et al.* (1998) described in Section **2.3.2**. Before extraction, each bone was sandblasted to remove external contamination that can outcompete endogenous DNA (Richards *et al.* 1995). At least two DNA extractions were performed for each specimen.

The control region sequences of the following boars/pigs were available from GENBANK: Asian wild boar (AWB), European wild boar (EWB), Okinawa, Satsuma, Yucatan (Okumura *et al.* 1996: accession numbers AB015084-95); Swedish (Ursing & Arnason 1998: AJ002189); AWB, EWB, Hampshire, Landrace, Large White, Mangalica, Meishan, unknown (Giuffra *et al.* 2000: AF136555-68, AF182446); AWB, Berkshire, Duroc, Hampshire, JinHua, Landrace, Large White, Meishan, Moncai, Pietrain, Potbelly, Yontsuan (Watanobe *et al.* 2001:AB041467-99); AWB, EWB, Berkshire (Okumura *et al.* unpublished: AB059650-2). The sequences were aligned by eye and CLUSTALX (Thompson *et al.* 1997), and nested primers were designed to amplify the 5' end of the mtDNA control

region of all *S. scrofa* sub-species using the guidelines suggested by MacHugh *et al.* (2000).

Amplification was attempted for the following fragments: 198 bp from $PIG1_{FOR}$ (15,448–15,468) – $PIG2_{REV}$ (15,645–15,626); 345 bp from $PIG1_{FOR}$ – $PIG2_{REV}$ (15,792–15,773); and 418 bp from $PIG1_{FOR}$ – $PIG3_{REV}$ (15,865–15,846) (numbers designated by comparison to the whole pig mitochondria; Ursing & Arnason 1998). Primers were shown not to amplify either cattle or human DNA. PCR set-up was conducted in a laboratory dedicated solely to pre-amplification ancient work. PCR conditions were as described in MacHugh *et al.* (1999), and the resulting amplification products were sent away for professional sequencing to MWG-Biotech (*http://www.mwg-biotech.com/html/services/sequencing/service_custseq.shtml*).

The criteria for authenticating mitochondrial haplotypes were as described in Section **2.3.2**. MtDNA sequences were aligned by eye against those sequences above taken from GENBANK, and a reduced median network was constructed using the median algorithm of Bandelt *et al.* (1995). The network includes the following published data, truncated to the 377 bp presented here: AB015093-5, AB041484-6, AB041488-9, AB041491-9, AF136556-63, AF182446, D42170 and D42179 (Okumara *et al.* 1996; Giuffra *et al.* 2000; Watanobe *et al.* 2001). The neighbor-joining method (Saitou & Nei 1987) was used to construct a dendrogram from genetic distances generated by the DNADIST program (Felsenstein 1989). All sequences used for primer design were included here, as well as the following animals sequenced by Kim *et al.* (2002): Berkshire, Cheju, Erhualian, Gandonghei, Jinghua, Neijang, Putian, Tongcheng, Wan'an, Wanhua, Wannanhua, Welsh, Westran and Yanxin (accession numbers AF276921-38). A peccary sequence (AF276938: Kim *et al.* 2002) was used as an outgroup, and bootstraps were calculated from 1,000 random replications of the data.

B.4 Results

Amplification was attempted for three fragments from the 5' end of the control region, sized 198 bp (positions 15,448–15,645), 345 bp (15,448–15,792), and 418 bp (15,448–15,865), of seven pig bones dating to around 6,000 BP. Due to the excellent preservation condition of the specimens, amplification of the 418 bp product was successful in six samples, with the seventh giving no PCR product with any primer combination (a success rate of 85.7%). Unfortunately, this present study was already in

progress before the publication of several ancient pig sequences from Japanese island populations (Watanobe *et al.* 2001, 2002). In the latter studies, the 3' end of the D-loop was targeted, and the primers designed here fall within the 5' end of the control region. Although the sequences obtained could not be compared, similar aDNA amplification success was generated. Watanobe *et al.* (2001) amplified a 574 bp fragment in 39 out of 45 pig bones and teeth from shell middens on Rebun Island, a success rate of 87.7%. In their further study (2002) of 161 bones from 12 sites on the Okinawa islands, only 33 (20.5%) gave reproducible amplification results. Working samples were taken from shell middens, with the poorly amplifying specimens located under cliffs or on hilly areas. The reason for successful amplification in such a high number of remains from both CPC and Rebun Island must be due to the excellent preservation afforded by the cool dry conditions of the underground cave and the alkaline conditions of the shell middens, respectively.

Analysis of the mtDNA D-loop sequences of 100 extant and six ancient pig sequences showed 41 polymorphic sites (**Table B.2**), representing 10.9% of the total DNA sequence analysed (which, minus both primer sequences, was 377 bp). Three of the 41 variable positions represented insertion/deletion events. The remaining 38 variable positions were single nucleotide substitutions, all but one of which were transitions. A highly variable region between positions 15,500 and 15,600, previously described by Kim *et al.* (2002), represented 39.0% of the total number of polymorphic sites. The total number of sites in the 377 bp region was 18 (16 mutations and two indel sites) for clade EI (36 individuals, including the six ancient samples), one for clade EII (three individuals), and 24 (22 mutations and two indel sites) for clade A (67 individuals).

Figure B.1 shows a network drawn from 377 bp of control region mtDNA, using modern data from 26 domestic pig and seven wild boar, plus the six ancient sequences (**Table B.2**). As the European and Asian clades diverged from each other between 900,000 and 58,000 years (Giuffra *et al.* 2000 and Kim *et al.* 2002 respectively), the network consists only of those animals that belonged to the two European clades identified by Giuffra *et al.* (2000). These clades can be clearly distinguished, with three Italian wild boar representative of EII (haplotypes 18 and 19) and clade EI consisting of a combination of domestic and wild animals (haplotypes 1 to 17). Indel sites were excluded from the network, and therefore the two indel sites found within the European clades were excluded, leading to two haplogroups being amalgamated (1+9 and 13+14).

| Clade | wild/ domestic | Animals | 1 5 4 9 1 | 1 5 5 0 4 | 1 5 5 1 1 | 1 5 5 2 4 | 1 5 5 4 3 | 1 5 5 4 4 | 1 5 5 5 6 | 1 5 5 8 2 | 1 5 5 6 5 | 1 5 5 6 6 | 1 5 5 7 1 | 1 5 5 7 5 | 1 5 5 7 7 | 1 5 5 7 9 | 1 5 5 7 7 | 1 5 5 8 2 | 1 5 5 9 7 | 1 5 5 9 9 | 1 5 5 9 7 | 1 5 5 0 3 | 1 5 6 6 0 | 1 5 6 6 1 | 1 5 6 6 5 | 1 5 6 6 6 | 1 5 6 6 9 | 1 5 6 6 1 | 1 5 6 6 4 | 1 5 6 6 8 | 1 5 6 6 5 | 1 5 6 6 7 | 1 5 6 8 8 | 1 5 6 8 2 | 1 5 7 7 9 | 1 5 7 7 4 | 1 5 7 7 0 | 1 5 7 7 1 | 1 5 7 7 1 | 1 5 7 7 4 | 1 5 7 7 3 | 1 5 7 7 6 | 1 5 7 7 7 | 1 5 7 7 1 | 1 5 7 7 0 | 1 5 7 8 8 | 1 5 7 8 2 | 1 5 8 8 2 | 1 5 8 8 3 | 1 5 8 8 5 | 1 5 8 8 4 | 1 5 8 8 0 | 1 5 8 8 4 | Haplotype number |
|---|---|---

The neighbor-joining tree, shown in **Figure B.2**, includes reference sequences from each of the 51 extant *S. scrofa* haplotypes, which include wild and domestic pigs, originating in Europe and Asia, plus six ancient sequences (**Table B.2**). This tree shows the division seen between the three clades, with the Asian clade (A) and European clade I (EI) including both wild and domestic animals. As with the network (**Figure B.1**) and sequence table (**Table B.2**), it can clearly be seen that all of the ancient specimens cluster with clade EI.

Domestic animals P4, P5 and P6, as well as undefined P7, have the same haplotype as one Duroc, one Landrace, one Large White and one Pietrain domestic animal (AB041485, AB041496, AB041494, AB041489: Watanobe *et al.* 2001), one Hampshire domesticate (AF136558: Giuffra *et al.* 2000), and three Israeli wild boars (Giuffra *et al.* 2000). This haplotype is the most numerous and appears to be the central European pig sequence (**Figure B.1**). It is most probable that the specimen P7 is an example of a domestic pig, especially as only a small proportion (less than 10%; A. Chamberlain personal communication) of the pig remains from CPC are definitely wild boar. The wild boar P3 exhibits a unique haplotype not previously described, but it is most similar to wild boar P2 and its extant counterpart, a domestic Berkshire animal (AB041484: Watanobe *et al.* 2001).

B.5 Discussion

The three domestic ancient samples (P4, P5 and P6), plus the undefined bone (P7), have sequences that are the same as the most numerous central haplotype (**Figure B.1**). This haplotype (1+9) includes five domestic pig and three Israeli wild boars (**Table B.2**). The two wild boar sequences differ from the ancient domestic pig, with one (P2) grouping with a domestic modern sample (haplotype 11) and the other (P3) exhibiting an undescribed haplotype (12). These two haplotypes are two and four mutations from the central haplotype respectively. Although these results are consistent with a differentiation between wild and domestic pig remains, the differences seen between them are not diagnostic when compared to the spread of modern domestic and wild animals.

When the overall pattern of haplotype diversity in both Asian and European pigs is considered (**Figure B.2**), it is apparent that there are three clades present: one Asian (A) and two European (EI and EII). All six of the ancient sequences fall within the EI clade. Clade EII is very divergent from either clade, and contains two haplotypes (18 and 19) with three individuals present, all of which are European wild boar from Umbria in Italy

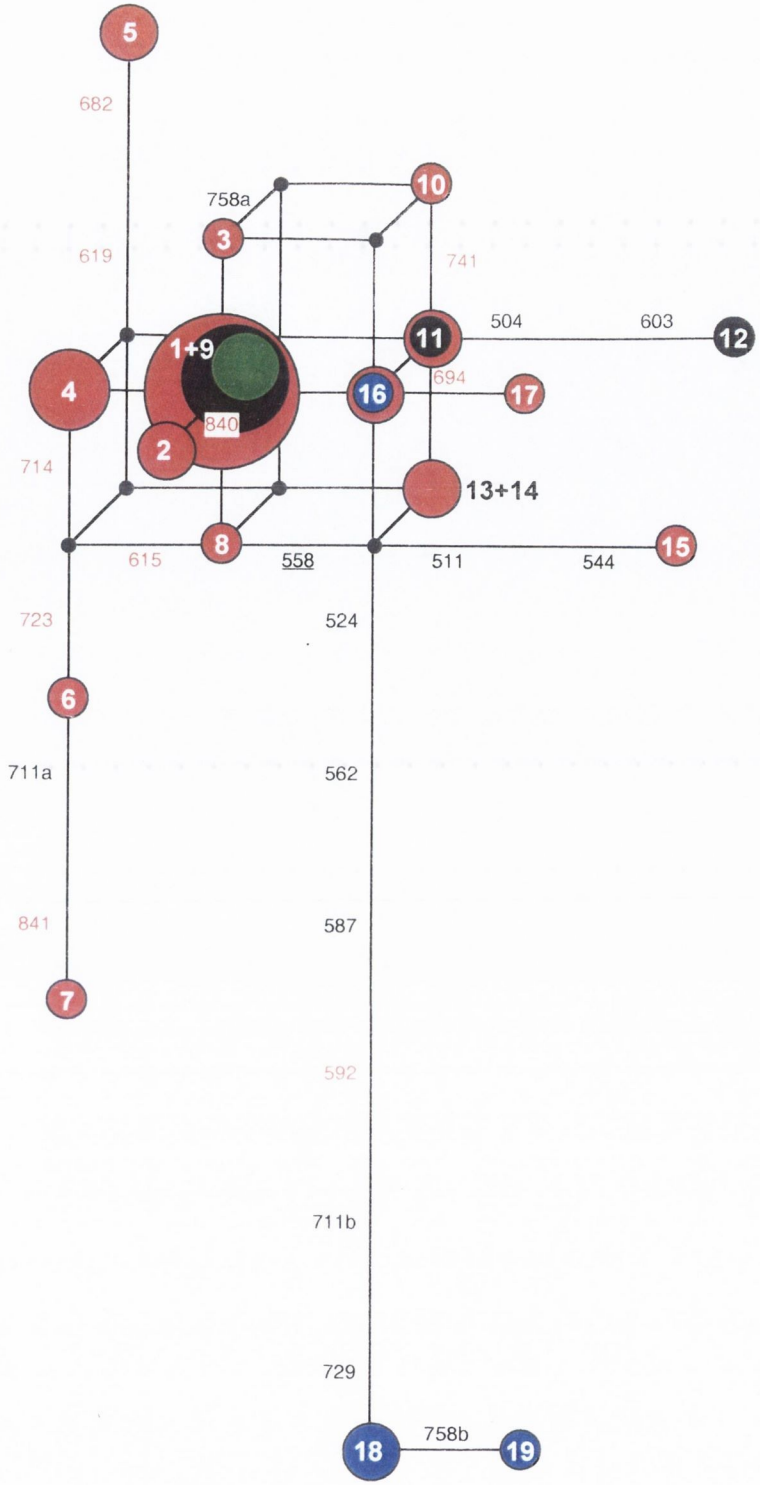


Figure B.1 Reduced median network of the two European clades seen in wild boar and domestic pig, putting ancient sequences in a modern context. The network includes both extant lineages (taken from GENBANK, see text) and the six ancient sequences obtained. Circles represent sequence haplotypes, the area being proportional to the frequency of the haplotypes. Circle-associated numbers denote codes with respect to haplotype identification (Table B.2); here haplotypes 1 + 9, and 13 + 14 have been amalgamated as the network excludes indel sites. Points are theoretical intermediate nodes introduced by the median algorithm, and branches between haplotypes represent mutations. The single transversion site is underlined, and mutations in red occur within the Asian clade also (see Table B.2). The colour key is as follows: blue, European wild boar; green, Israeli wild boar; red, domestic pig; black, ancient sequences. The three ancient domestic sequences, plus that from the undefined bone, are included in haplotype 1+9, with the two wild sequences falling within haplotypes 11 and 12 respectively.

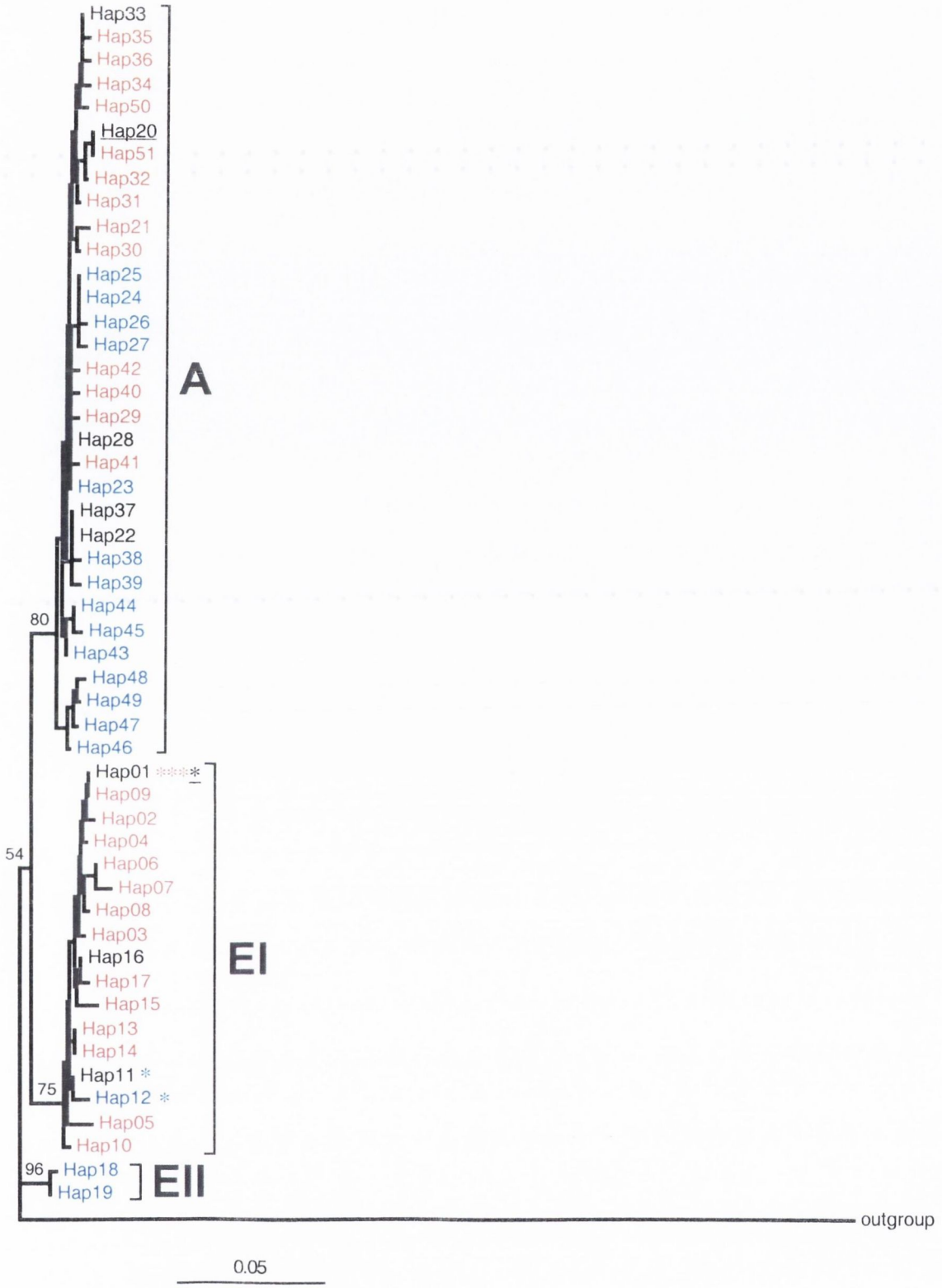


Figure B.2 Neighbor-joining trees of extant and ancient wild boar and domestic pig mitochondrial D-loop haplotypes based on 377 bp of the control region. Bootstrap values (threshold of 50% after 1000 replicates) are reported on the nodes. Haplotype numbers are as for Table B.2. Haplotypes containing only wild boar are coloured blue; those with only domestic pig are coloured red; and those containing a mixture of wild and domestic pigs are coloured black. Unknown designations are underlined. Positions of the six ancient sequences are denoted by asterixes, with the three ancient domestic sequences, plus that from the undefined bone, included in haplotype 1, and the two wild sequences falling within haplotypes 11 and 12 respectively. The three phylogenetic clades identified by Giuffra *et al.* (2000) are shown as A, EI and EII. The outgroup sequence is from a peccary (see text).

(Watanobe *et al.* 1999). Clade EI consists of 26 modern domestic pigs from a variety of breeds, three wild boars from Israel plus one from Poland, and the six ancient sequences from Britain (**Table B.2**). The wide disparity between the wild sequences present in clades EI and EII (**Figure B.1**) may be explained by recolonisation of Europe from two distinct refugia after the last glacial maxim.

Although it could reasonably be considered that the genetic constitution of Britain's wild boar population has remained more or less constant through time, and has not been unduly affected by introgression of domestic pigs (A. Chamberlain personal communication), the pattern seen in both the network (**Figure B.1**) and the neighbor-joining tree (**Figure B.2**) seems to refute this, with introgression clouding the picture. The date of other faunal remains from CPC (at around 6,000 BP) place the site in the very earliest part of the British Neolithic, and therefore near to the initial introduction of domestic pigs to this island.

As far as known, this is the first analysis of ancient DNA from British pigs, both wild and domestic forms. The results presented here would benefit from a more thorough sampling of both wild and domestic pig remains from archaeological sites across Europe.

Appendix C - Ancient DNA Analysis of Giant Irish Deer

C.1 Abstract

Two *Megaloceros giganteus* (giant Irish deer) bone samples from Irish sites dating between 30,000 and 10,600 years BP were assessed for the presence of reliable mitochondrial haplotypes. It was possible to reproducibly amplify and sequence an 80 base pair fragment of the control region, plus a 186bp region of the cytochrome *b* gene, from one of the samples, although sequence data obtained from the other bone were in agreement. This result was compared with published extant data by direct alignment and construction of neighbor-joining trees, and appears to be authentically from a previously unsequenced cervid. In the absence of genetic analysis, the supposed close relationship between giant deer and fallow deer (*Cervus dama*) has been based on similar morphology, most notably antler design. However, in the preliminary data shown here, *Megaloceros* clusters more closely with red deer (*C. elaphus*), although caution needs to be exercised with regard to the reliability of this statement. This is the first phylogenetic study of the giant Irish deer.

C.2 Introduction

Megaloceros giganteus (Blumenbach 1799 [cited in Lister 1994b]), the giant Irish deer, more commonly known as the *Irish elk* (Molyneux 1697 [cited in Monaghan 1995]), stood almost two metres high at the shoulders, was approximately three metres long, and is estimated to have weighed around 700 kilograms (kg) (Stuart 1991). The most striking aspect of the animal was the huge antlers carried by the males. These could measure up to 4 metres along the curvature and weigh up to 40 kg, about a third heavier than modern moose, the largest extant cervid. The antlers varied considerably between individuals, but all were outspread at right angles to the head and broadened into a flat *palm*, which faced slightly backwards and upwards when the head was held horizontally, with pointed prong-like projections, or *tines* (Lister 1994b). Only extant fallow deer have a similar *palmation* (Gill & West 2001). It is now believed that these gigantic antlers were the ultimate cause of the disappearance of *Megaloceros*, as, during the Younger Dryas, the dietary intake of minerals would have been insufficient to meet requirements for antler mineralisation, leading to osteoporosis, malnutrition, enfeeblement and, eventually, extinction (Barnosky 1986; Moen *et al.* 1999).

Although the giant deer is associated with Ireland due to the sheer concentration of numbers found there (close to 200 find spots of giant deer have been recorded in Ireland; Woodman *et al.* 1997), it has a history across Europe, western Asia, and even northern Africa (Barnosky 1986), spanning 300,000 to 400,000 years. It is thought that the deer originated in Siberia and gradually migrated westwards in search of a more favourable climate. Adapted to flat country, covered with grasses and scattered copses, the lowlands of central and eastern Ireland were ideally suited to the animal. The species is thought to have inhabited Ireland from 37,000 to 32,000 BP, and again from 11,750 to 10,950 BP. In the intervening period, the Gleneavy stadial, Ireland would have been in the grip of ice and conditions would have been difficult for survival of the giant deer. Recolonisation occurred during the Woodgrange interstadial, either from the south or from Britain via Scotland or the Isle of Man (Gill & West 2001), and the deer finally vanished from the country during the Nahanagan stadial (which correlates to the Younger Dryas in Europe; Woodman *et al.* 1997). The latest date reported for the Irish deer in Ireland is 10,610 ± 495 BP in an individual from Ballybetagh Co. Dublin (UB-2699; Stuart & van Wijngaarden-Bakker 1985). Man did not arrive in Ireland until approximately 1,500 years after the demise of the Irish deer, and therefore deposits in this country offer the ideal opportunity to examine the giant deer in an environment where they had no natural predators.

Although giant deer were believed to have died out across Europe during the megafaunal extinction of the late Pleistocene, recently radiocarbon dated finds from the Isle of Man date to 9,200 BP (Gonzalez *et al.* 2000) and it is reasonable to assume that the deer endured into the Holocene on the Isle of Man because Man kept a moderate climate longer and had similar vegetation to Ireland prior to 10,500 BP.

M. giganteus is usually given tribal status within the Cervidae as Megalocerini, and the possibility has been mooted that the fallow deer (*Cervus dama*, also known as *Dama dama*) is the closest extant relative of the giant deer (Lister 1981; Monaghan 1995). Recognised characteristics of the Megalocerini include palmation of the antlers and thickening of the lower jaw bone. Within the Cervidae, the Megalocerini lie with the Cervinae (*Old World deer*), rather than the Odocoileinae (*New World deer*) or related taxa such as the Alcinae or Rangiferinae (Lister 1994b), but the origins of the Megalocerini are unclear. Ancient genetic analysis of skeletal remains from *M. giganteus* would offer a unique opportunity to unravel the relationship between the giant Irish deer and the extant Cervinae.

C.3 Materials & Methods

Two giant Irish deer bones were sourced by Nigel Monaghan, from the National Museum of Ireland, for ancient DNA analysis. ELK 1 (NMING: F21174) is a left femur taken from Liskelly Stream, Co. Cork, dating to the late glacial age (approx. 12,100-10,600 BP). ELK 2 (NMING: F21175) is a phalanx taken from Ballynamintra Cave, Co. Waterford, dating to the glacial age (30,000-10,600 BP). The bone samples were prepared using the procedure modified from Yang *et al.* (1998) described in Section **2.3.2**. Before extraction, each bone was sandblasted to remove external contamination that can outcompete endogenous DNA (Richards *et al.* 1995). Varying numbers of DNA extractions were performed for each specimen.

The control region sequences of the following taxa were available from GENBANK: *Cervus alfredi*, *C. canadensis*, *C. dama*, *C. elaphus*, *C. eldi*, *C. nippon*, *C. timorensis* and *C. unicolor* (Randi *et al.* 2001: accession numbers AF291879-88, AF291890-3, AF291895); *C. elaphus* (Polziehn *et al.* 1998: AF016968-9, AF016971); *C. elaphus* and *C. nippon* (Polziehn & Strobeck 1998: AF016972-5; Feng *et al.* unpublished: U12867-8); and *Capreolus capreolus* (Douzery & Randi 1997; Z70318). Cytochrome *b* sequences were available for the following taxa: *C. albirostris*, *C. elaphus*, *C. nippon*, *C. timorensis* and *C. unicolor* (Ludt *et al.* unpublished: AF423195-2, AF489279-81, AY035871, AY035875-6, AY044856-63, AY070221-6); *C. dama* (Irwin *et al.* 1991: X56290); *C. dama* and *C. elaphus* (Randi *et al.* 1998: AJ000021-2); *C. elaphus* (Matsunaga *et al.* unpublished: AB001612); *C. elaphus* and *C. nippon* (Kuwayama & Ozawa 2000: AB021090-99); *C. nippon* (Chikuni *et al.* 1995: D32192); and *Capreolus capreolus* (Zimmerman *et al.* 1998: Y14951). The complete mitochondrial sequence from *Bos taurus* (Anderson *et al.* 1982: NC_001567) was used as an outgroup. The sequences were aligned using CLUSTALX (Thompson *et al.* 1997), and degenerate primers were designed to amplify both the mtDNA control region and a fragment of the cytochrome *b* gene of all cervid species using the guidelines suggested by MacHugh *et al.* (2000).

Amplification was attempted for two non-contiguous segments of the control region: a 161 base pair fragment, $DEER1_{FOR}$ (16,019–16,036) – $DEER2_{REV}$ (16,188–16,167); and a 168 bp region, $DEER5_{FOR}$ (136–154) – $DEER6_{REV}$ (279–260) (numbers designated by comparison to the bovine mitochondria; Anderson *et al.* 1982); and for a stretch of the cytochrome *b* gene: a 122 base pair fragment, $ECB2_{FOR}$ (499-520) – $ECB1_{REV}$ (620-599); and a 153 bp region, $ECB1_{FOR}$ (599-620) – $ECB3_{REV}$ (751-730) (numbers designated as

position in 1140 bp cytochrome *b* gene, with start codon being number 1). Primer sequences are shown in **Table C.1**, and were shown not to amplify either cattle, pig or human DNA. PCR set-up was conducted in a laboratory dedicated solely to pre-amplification ancient work. PCR conditions were as described in MacHugh *et al.* (1999), and direct sequencing conditions were described in Section **2.3.2**.

The degenerate D-loop primers did not produce clean amplification products. However, a small number of targeted products were sequenced successfully and alignment of these preliminary data against the modern sequences showed an affinity between the control region of Irish deer and *Cervus alfredi* (AF291891: Randi *et al.* 2001). The primers were therefore modified to be non-degenerate using the sequence data from *C. alfredi* (**Table C.1**).

Name of primer	Sequence
$DEER1_{FOR}$	5'-T**M**T**R**CCCCATGC**W**TATAA-3'
$DEER1_{FOR}$ (modified)	5'-T**C**T**G**CCCCATGC**T**TATAA-3'
$DEER2_{REV}$	5'-TTGT**Y**TGGTTTCACGCGGCATGG-3'
$DEER2_{REV}$ (modified)	5'-TTGT**C**TGGTTTCACGCGGCATGG-3'
$DEER5_{FOR}$	5'-AACTGCATCTTGAGCATCC-3'
$DEER6_{REV}$	5'-TCTAGGG**RR**AAG**Y**GTGTTAA-3'
$DEER6_{REV}$ (modified)	5'-TCTAGGG**AA**AAG**T**GTGTTAA-3'
$ECB2_{FOR}$	5'-GGCTTTTCAGTAGA**Y**AAAGCAA-3'
$ECB1_{REV}$	5'-TTATT**R**GATCCTGT**Y**TCGTGAA-3'
$ECB1_{FOR}$	5'-TTCACGA**R**ACAGGATC**Y**AATAA-3'
$ECB3_{REV}$	5'-CAAG**Y**AGGTCTGGTGCGAATAA-3'

Table C.1 Primer details for amplification of the mtDNA control region (DEER) and cytochrome *b* gene (ECB) of Irish deer. Degenerate sites are coloured blue (M = C/A; R = A/G; W = T/A; Y = C/T), and bases changed between original and modified primers are coloured red.

All non-amplifiable samples were tested for presence of inhibitors that may have been impeding PCR amplification of endogenous DNA. This involved spiking each negative sample extract with aurochs DNA (from 6,000 BP Carsington Pasture Cave) in a ratio of 1:4. PCR was continued as previously described, but using cattle-specific primers that amplify a 157 bp fragment of the control region (Section **2.3.2**). Those sample extracts that yielded a product from the spiked sample were then designated as containing unknown substances causing inhibition of PCR and no further analysis undertaken.

The criteria for authenticating mitochondrial haplotypes were as described in Section **2.3.2**. MtDNA sequences were aligned by eye against representative animals of the *Cervus* species, and outgroups of roe deer (*Capreolus capreolus*) and domestic cattle (*Bos taurus*). Distance matrices were generated according to Kimura's *2-parameter* model of nucleotide substitution (1980), using the DNADIST program of PHYLIP (Felsenstein 1989). The Kimura model assumes that the likelihood of a base change at any one site is independent of a change at any other site, while also allowing for a difference between the frequency rates of transitions and transversions. Each distance thus generated is an estimate of the divergence time between a particular pair of species. The neighbor-joining method (Saitou & Nei 1987) was used to construct dendrograms of breed relationships from these genetic distances using the PHYLIP package (Felsenstein 1989), and a bootstrap value was generated by resampling loci over 1,000 replicates.

C.4 Results

Amplification was attempted for two non-contiguous sections of the control region, sized 161 bp (positions 16,019–16,188) and 168 bp (136–279), and two fragments of the cytochrome *b* gene, sized 122 bp (499-620) and 153 bp (599–751). Only the second section of the D-loop targeted gave reproducible, reliable results. Sequences were obtained from both giant deer samples for the first segment (16,019–16,188), but the majority of bases were ambiguous, which, because of the degenerate nature of both $DEER1_{FOR}$ and $DEER2_{REV}$, is most likely due to amplification of both the target fragment as well as other areas of the giant deer genome. For the second segment, amplification was obtained from the non-degenerate primer $DEER5_{FOR}$ and the degenerate primer $DEER6_{REV}$ (**Table C.1**), but again some bases were ambiguous. In this case, however, there was enough sequence homogeny between the ancient sequences obtained and the extant cervid data to modify the reverse primer. Using this new primer, only one of the Irish deer samples, ELK2, gave reproducible sequence data, but the single sequence generated from amplification products of ELK1 was identical to those obtained from ELK2. Amplification was successful for both primer sets targeting a 252 bp section of the cytochrome *b* gene, and sequences were obtained from both giant deer samples. However, with ELK1, reproducible, reliable results were only obtained from the first region, positions 499 to 620 in the 1140 bp cytochrome gene. The success rates of each targeted fragment with each of the two giant deer samples are shown in **Table C.2**.

Sample #	Number of extractions	Prescence of inhibition	122 bp cytb fragment		153 bp cytb fragment		161 bp D-loop fragment		168 bp D-loop fragment	
			# of PCR products	readable sequences	# of PCR products	readable sequences	# of PCR products	readable sequences	# of PCR products	readable sequences
ELK 1	10	6 out of 10	10 out of 16 attempts	7 out of 12 attempts	1 out of 8 attempts	0 out of 0 attempts	14 out of 24 attempts	1 out of 4 attempts	9 out of 32 attempts	2 out of 4 attempts
ELK 2	3	0 out of 3	12 out of 12 attempts	7 out of 12 attempts	2 out of 6 attempts	4 out of 10 attempts	15 out of 18 attempts	1 out of 3 attempts	16 out of 18 attempts	5 out of 13 attempts

Table C.2 Information about extraction from giant Irish deer.

The ability of ELK2 to give reproducible amplifications and sequence data is taken to be due to the survival conditions. Bone ELK1 was dark-brown in colour, which is most likely due to presence of tannins and humic acids that can inhibit PCR, and this can be seen in the proportion of extracts that exhibited inhibition (**Table C.2**). Bone ELK2 was off-white in colour and overall looked better preserved. This might be expected as this sample was found in a cave, and a recent comparison of the effect of environmental factors upon DNA preservation identified that cool cave conditions, even with the presence of water, gave the highest yield of DNA (Burger *et al.* 1999).

The table of mutations for the D-loop and cytochrome *b* regions are shown in **Tables C.3** and **C.4** respectively. Translation of the cytochrome *b* sequence was performed on 186 bp, between bases 523 and 597, and bases 619 and 729, and the residues numbered from 1 to 62 accordingly (**Table C.5**). None of the amino acid changes between the giant deer and the cervids were seen to occur only in giant deer, and the majority of substitutions were between amino acids of similar properties; for example, between isoleucine (I) and valine (V).

Distance matrices are shown in **Table C.6**. These were generated from 80 bp of the control region and 186 bp of the cytochrome *b* (positions 523-597 plus 619-729; that is, minus the central degenerate primer CY1F/R) using representative cervid species as comparison, and *B. taurus* as an outgroup. Individual taxa that included both regions were also joined together and a matrix generated from the total 266 bp sequence. When the D-loop sequence is considered, *M. giganteus* groups most closely to the Philippine spotted deer (*Cervus alfredi*) than any of the other representative extant cervids considered, with a bootstrap of 88%, and this is reflected in the neighbor-joining tree generated from the matrix (**Figure C.1a**). The branching structure of this tree is the same as when the complete control region is considered for the modern taxa (Randi *et al.* 2001).

	1 5 5	1 6 3	1 6 6	1 6 9	1 7 0	1 7 1	1 7 3	1 7 3 i	1 7 7	1 7 8	1 7 9	1 8 1	1 8 3	1 9 0	1 9 1	2 0 1	2 0 2	2 0 3	2 0 4	2 0 5	2 0 8	2 1 0	2 1 3	2 1 4	2 1 5	2 1 6	2 1 7	2 1 9	2 2 0	2 2 3	2 2 4	2 2 6	2 2 7	2 2 9	2 3 0	2 3 1	2 3 2
M. giganteus	C	G	G	A	T	-	G	G	G	G	C	G	C	C	A	A	C	T	A	T	C	T	C	T	G	G	A	T	A	A	C	C	T	A	G	C	C	T	A
Cervus alfredi	.	A	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
C. dama	.	.	.	.	C	-	A	.	C	A	T	A	T	.	.	.	T	.	.	.	T	.	.	C	A	-	-	.	C	.	.	.	.	C	A	G	T	.	T
C. elaphus	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	T	C	.	.	T	.	.	.	.	A	G	.	.	.	.	.	.	.	T	.	.	.
C. eldi	.	A	.	.	.	-	.	.	T	.	.	.	.	.	.	G	T	.	.	.	T	.	.	.	.	A	.	.	.	.	.	.	.	.	T	.	C	T
C. nippon	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	G	T	.	.	.	T	.	.	.	.	A	G	.	.	.	.	.	.	.	T	.	.	.
C. timorensis	.	A	.	G	C	A	.	.	T	.	.	.	.	.	.	T	T	C	.	.	T	.	.	.	.	A	G	.	.	.	.	.	.	.	T	.	.	T
C. unicolor	.	.	.	.	.	-	.	.	C	.	.	.	.	.	.	.	T	C	.	.	T	.	.	.	.	A	G	.	.	.	.	.	.	.	A	.	.	.
Capreolus capreolus	.	.	.	.	.	-	A	.	.	A	.	.	.	A	G	.	T	.	.	.	T	.	.	C	.	A	C	.	.	.	.	.	.	C	T	T	T	C	T
Bos taurus	G	A	A	.	.	G	A	-	T	A	.	.	.	.	.	.	A	.	T	A	T	A	T	.	A	T	C	C	C	C	T	T	C	T	A	A	A	A	T

Table C.3 An 80 base pair portion of the mitochondrial control region from *Megaloceros giganteus*, more commonly known as the giant Irish deer, or Irish elk, compared with a panel of *Cervus* (deer) consensus sequences taken from GENBANK (see text). The deer samples were chosen as it is believed, based on morphological conditions, that Irish elk were closely related to fallow deer (*C. dama*). However, this seems not to be the case, with *M. giganteus* showing closer affinity with *C. alfredi* (Philippine deer), and *C. elaphus* (red deer). The *Capreolus* (roe deer) and *Bos taurus* (cattle) sequences were included as outgroups. Only variable sites are shown, with differences given and a full stop (.) denoting identity. The sequence positions from the BOVMT GENBANK sequence are given above each column (Anderson *et al.* 1982). *i* = insertion.

	523	525	528	531	534	537	540	546	552	555	561	564	567	568	569	570	574	576	579	585	588	589	594
M. giganteus	**T**	**A**	**C**	**A**	**C**	**C**	**C**	**C**	**C**	**T**	**C**	**T**	**T**	**G**	**C**	**A**	**C**	**T**	**C**	**A**	**T**	**T**	**C**
C. albirostris	C	G	.	G	T	.	T	.	T	C	.	C	C	.	.	.	.	.	.	.	C	.	.
C. dama	.	.	T	.	.	.	T	.	T	A	T	C	.	.	.	.	.	.	T	.	.	.	A
C. elaphus 1	C	.	.	.	T	.	T	.	T	C	T	C	C	.	.	.	.	C	T	.	C	.	.
C. elaphus 2	C	.	.	.	.	.	T	.	T	C	T	C	C	.	.	.	.	C	T	.	C	.	.
C. nippon	C	.	.	.	T	.	T	.	T	.	T	C	C	.	.	.	.	.	T	.	C	.	.
C. timorensis	C	.	.	.	.	T	T	.	.	C	T	C	C	.	.	G	.	C	T	.	C	.	.
C. unicolor	C	.	.	.	T	T	T	.	T	C	T	C	C	.	.	G	.	C	T	.	.	.	.
Capreolus capreolus	C	G	.	.	T	.	T	.	T	C	T	C	.	.	.	.	.	.	T	C	.	.	T
Bos taurus	C	T	.	.	.	.	T	T	.	.	T	C	C	A	T	.	A	.	.	C	C	C	A

	624	626	633	634	636	639	640	642	644	645	648	654	657	660	663	666	669	672	675	678	681	684	687	688	689	693	694	695	696	699	700	702	703	705	706	708	711	712	714	715	719	721	729
M. giganteus	**A**	**C**	**C**	**C**	**A**	**A**	**G**	**T**	**C**	**A**	**C**	**T**	**T**	**C**	**C**	**C**	**C**	**T**	**G**	**C**	**A**	**C**	**T**	**C**	**T**	**A**	**T**	**T**	**C**	**A**	**C**	**T**	**C**	**A**	**A**	**T**	**C**	**C**	**T**	**T**	**C**	**T**	**T**	**A**
C. albirostris	.	.	.	.	.	.	.	A	C	.	.	.	C	C	.	T	T	.	C	C	T	.	T	C	T	C	.	.	C	.	.	.	.	.	G	G	A	.	.	.	.	.	C	.	.
C. dama	.	.	.	.	.	.	.	.	.	.	.	T	C	A	T	.	.	.	C	C	.	.	T	C	T	G	G	C	A	.	.	.	.	A	.	.	C	T	A	G	.	.	C	A	.
C. elaphus 1	.	T	T	.	.	.	.	C	.	.	.	C	C	T	T	T	T	.	C	T	.	T	C	T	C	.	.	C	.	.	.	.	.	T	G	A	.	.	.	.	.	.	.	.	.
C. elaphus 2	.	.	.	.	.	.	.	C	.	.	.	C	C	.	T	T	.	.	C	T	.	T	C	T	.	.	.	C	T	.	.	.	.	.	G	A	.	.	.	.	.	.	C	.	.
C. nippon	.	.	.	.	.	.	G	.	C	.	.	.	C	C	.	T	T	.	C	C	T	.	T	C	T	C	.	.	C	T	.	.	.	.	.	G	A	.	.	.	.	.	.	C	.	.
C. timorensis	.	.	.	.	.	.	G	.	C	.	.	T	C	C	.	T	T	.	C	C	T	.	T	C	T	C	G	.	C	T	.	.	.	A	.	G	A	.	.	.	.	.	.	C	.	.
C. unicolor	.	.	.	.	.	.	G	.	C	.	.	T	C	C	.	T	T	.	C	C	T	.	T	C	T	C	.	.	C	T	.	.	.	A	.	G	A	.	.	.	.	.	.	C	.	.
Capreolus capreolus	G	.	.	.	.	.	A	C	T	G	.	.	A	T	.	.	.	C	C	T	.	T	C	.	G	G	.	.	.	.	.	C	T	.	.	.	.	T	.	C	.	C	.	C
Bos taurus	.	.	T	T	C	.	.	C	T	.	.	C	A	.	.	.	.	.	C	T	G	.	C	T	G	G	C	C	.	C	T	A	.	.	.	.	.	A	G	C	T	C	C	.

Table C.4 An 186 base pair fragment of the cytochrome *b* gene from *M. giganteus*, compared with a panel of *Cervus* (deer) consensus sequences taken from GENBANK (see text). Only variable sites are shown, with differences given and a full stop (.) denoting identity. The sequence positions given above each column are the relative positions of the changes in the 1140 bp cytochrome gene. The upper table includes changes seen over the first fragment (523-597) and the lower table shows changes seen over the second fragment amplified (619-729).

	16	18	28	31	33	34	51	54	55	57	60
M. giganteus	**A**	**L**	**T**	**P**	**D**	**A**	**I**	**L**	**I**	**F**	**L**
C. albirostris	.	.	.	.	N	.	.	.	V	.	.
C. dama	.	.	.	.	.	.	A	I	.	V	I
C. elaphus 1	.	.	I	.	.	.	.	.	V	.	.
C. elaphus 2	.	.	.	.	.	.	.	.	V	.	.
C. nippon	.	.	.	.	.	.	.	.	V	.	.
C. timorensis	.	.	.	.	.	.	V	I	V	.	.
C. unicolor	.	.	.	.	.	.	.	I	V	.	.
Capreolus capreolus	.	.	.	.	N	V	V	.	.	S	.
Bos taurus	I	I	.	S	.	V	A	.	.	A	.

Table C.5 Protein information for 186 bp of the cervid cytochrome *b* region, including base positions 523-597, and 619-729. The amino acid residues are numbered from 1 to 62 accordingly. Amino acid abbreviations are as follows: alanine (A), aspartate (D), phenlyalanine (F), isoleucine (I), leucine (L), asparagine (N), proline (P), serine (S), threonine (T), and valine (V).

The picture is less clear from the neighbor-joining tree for the cytochrome *b* region (**Figure C.1b**), although *M. giganteus* forms an affinity to the group that includes red deer (*C. elaphus*), as appears to be the case with the D-loop tree and the tree constructed using amalgamated data from both regions (**Figure C.1c**). In fact, the sequence *C. elaphus* 1 shows the lowest genetic distance to the giant deer (**Table C.6**). Also, in all three neighbor-joining trees, *M. giganteus* is included within the group containing roe deer (*Capreolus capreolus*), with *C. dama* as an outgroup. The genetic distance comparisons for 186 bp of the cytochrome *b* region are greater than seen for the 80 bp of D-loop (**Table C.6**). This appears to be a feature of the cytochrome *b* data from all species comparisons, rather than an artefact of utilising sequence data from ancient remains.

a)	*M. giganteus*	*Cervus alfredi*	*C. elaphus*	*C. nippon*	*C. unicolor*	*C. eldi*	*C. timorensis*	*Capreolus capreolus*	*C. dama*
Cervus alfredi	**0.0388**								
C. elaphus	**0.0792**	0.0655							
C. nippon	**0.0939**	0.0799	0.0391						
C. unicolor	**0.0947**	0.0947	0.0261	0.0525					
C. eldi	**0.1236**	0.0806	0.0947	0.0525	0.1091				
C. timorensis	**0.1702**	0.1247	0.0813	0.0934	0.0955	0.0799			
Capreolus capreolus	**0.2222**	0.2057	0.1734	0.1912	0.2097	0.1734	0.2414		
C. dama	**0.2640**	0.2640	0.2464	0.2438	0.2244	0.2586	0.2613	0.1786	
Bos taurus	**0.5135**	0.4648	0.5135	0.4733	0.4758	0.4299	0.5209	0.4586	0.3970

b)	*M. giganteus*	*Cervus elaphus 2*	*C. nippon*	*C. albirostris*	*C. dama*	*C. elaphus 1*	*Capreolus capreolus*	*C. timorensis*	*C. unicolor*
Cervus elaphus 2	**0.1453**								
C. nippon	**0.1579**	0.0327							
C. albirostris	**0.1643**	0.0608	0.0494						
C. dama	**0.1695**	0.1629	0.1629	0.1895					
C. elaphus 1	**0.1778**	0.0496	0.0610	0.0725	0.1903				
Capreolus capreolus	**0.1858**	0.1727	0.1662	0.1597	0.1622	0.1865			
C. timorensis	**0.1909**	0.0496	0.0496	0.0901	0.1622	0.0904	0.1932		
C. unicolor	**0.1909**	0.0496	0.0383	0.0783	0.1622	0.0786	0.1799	0.0216	
Bos taurus	**0.2472**	0.2188	0.2258	0.2400	0.2170	0.2483	0.2230	0.2483	0.2777

c)	*M. giganteus*	*C. elaphus 2*	*C. nippon*	*C. elaphus 1*	*C. unicolor*	*C. timorensis*	*C. dama*	*Capreolus capreolus*
Cervus elaphus 2	**0.1251**							
C. nippon	**0.1385**	0.0345						
C. elaphus 1	**0.1473**	0.0345	0.0544					
C. unicolor	**0.1615**	0.0426	0.0425	0.0628				
C. timorensis	**0.1848**	0.0587	0.0624	0.0877	0.0426			
C. dama	**0.1960**	0.1864	0.1858	0.2063	0.1801	0.1901		
Capreolus capreolus	**0.1963**	0.1729	0.1734	0.1827	0.1884	0.2069	0.1669	
Bos taurus	**0.3132**	0.2911	0.2888	0.3142	0.3290	0.3173	0.2644	0.2820

Table C.6 Genetic distance estimates between the giant Irish deer ancient sequences using: **a)** 80 bp of the control region; **b)** 186 bp of the cytochrome *b* gene; and **c)** both regions together, including various sequences from GENBANK (see text). Genetic distances are ranked for comparisons between the giant Irish deer and the extant data (column 1).

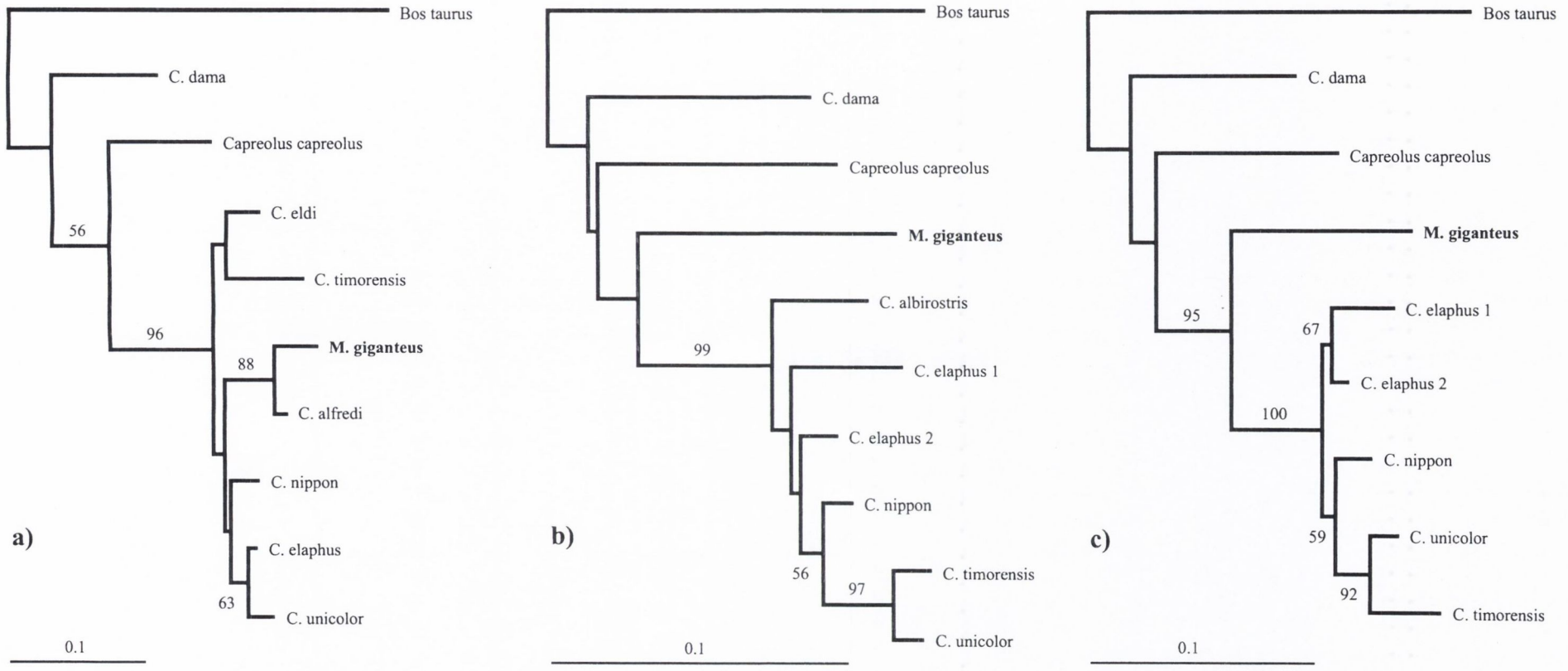


Figure C.1 Neighbor-joining phylogenies constructed using: **a)** 80 bp of cervid control region sequence; **b)** 186 bp of the cytochrome *b* region; and **c)** 266 bp comprising both D-loop and cytochrome *b* data. In all cases, cattle has been included as an outgroup. Bootstrap values (threshold of 50% after 1000 replicates) are reported on the nodes. The position of the ancient giant deer sequence is shown in bold. Interestingly (although these data are preliminary) *Megaloceros* clusters with red deer (*C. elaphus*) rather than fallow deer (*C. dama*). The scales represent units of genetic distance and are derived from the original comparison matrices (Table C.6).

To further investigate the authenticity of the ancient sequence, the neighbor-joining tree and alignment data from the cytochrome *b* data were analysed using the MacClade program (Version 3.05, Maddison & Maddison 1992). This was used to visualise all unambiguous changes in the cytochrome *b* neighbor-joining tree, and this is shown pictorially in **Figure C.2**. From this tree, the average transversion-transition ratio calculated over all branches leading to the modern samples, minus the branch leading to *M. giganteus*, was 0.14. The ratio on the *M. giganteus* branch is similar (0.11), which lends weight to the veracity of the giant deer cytochrome *b* sequence.

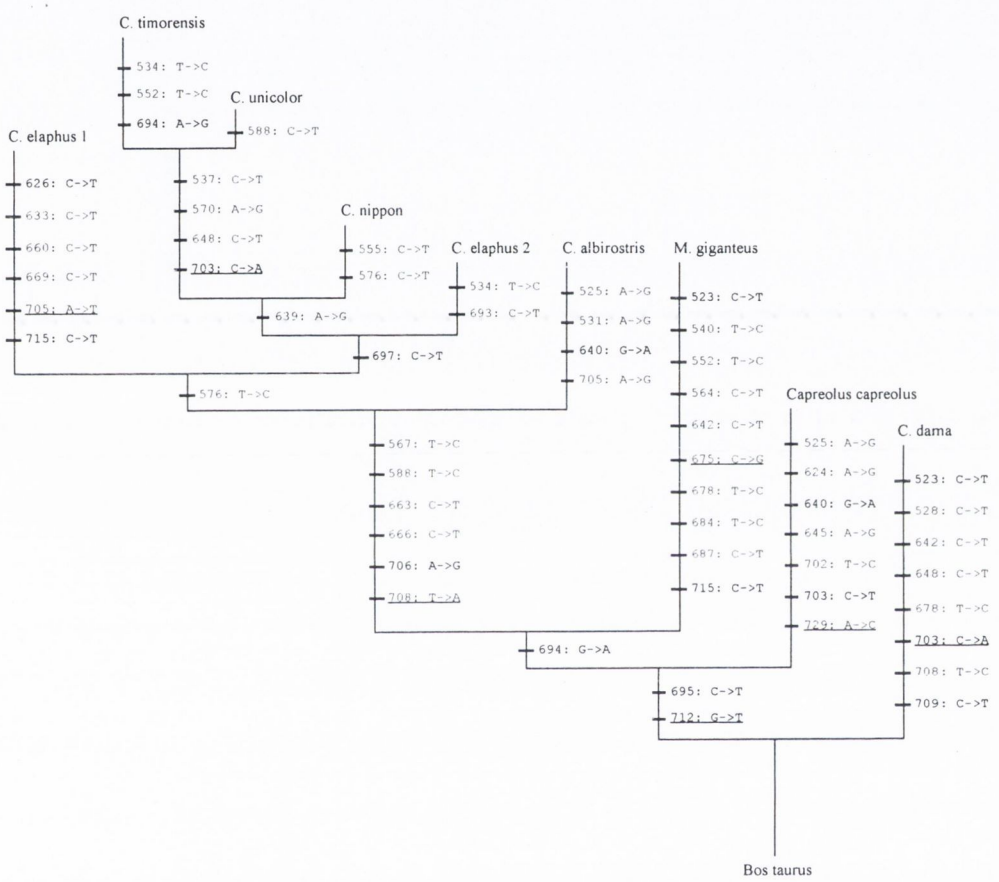


Figure C.2 Manipulated cytochrome *b* tree using MacClade. Unambiguous changes are shown and the branch lengths are proportional. Numbers relate to the base position of the mutation in the 1140 bp cytochrome *b* gene (see text). Sites in red are third positions changes, and those underlined represent transversions.

C.5 Discussion

Two experts believe the fallow deer (*Cervus dama*) to be the closest living relative of the Irish elk (Lister 1981, 1994b; Monaghan 1995), the relationship being largely based upon the similarity of the palmation of their antlers, along with a correlation between the structure of bones and teeth. However, the preliminary data presented here show a possible closer association between *Megaloceros* and red deer (*C. elaphus*) (**Table C.6**). Indeed, the high-crowned dentition of the elk is similar to that of red deer (Stuart 1991), and the upright orientation of some of the giant deer antlers would have caused the head to be lowered into a fighting position similar to that of the red deer (Lister 1994b).

However, this supposed close affinity between *M. giganteus* and red deer has to be considered in the context of the relationships shown between the extant species included in **Figure C.1**. In all three trees, the position of the *C. dama* sequence appears to be anomalous. In the phylogeny generated from the whole 1140 bp of the cytochrome *b* gene by Randi *et al.* (1998), the fallow deer clusters with fellow Old World deer (Cervinae and Muntiacinae), as opposed to the New World deer (Odocoileinae and Hydropotinae). This structuring is in agreement with the morphological data, whereby Old World deer belong to a plesiometacarpal group and New World deer belong to a telemetacarpal group. To test the validity of the fallow deer's position in the trees shown in this study, other trees were constructed including sequence data from a further three New World species: *Alces alces* (moose), *Odocoileus hemionus* (mule deer) and *Rangifer tarandus* (reindeer). Comparisons were also made between tree topologies obtained when the cattle outgroup was replaced with each of five other outgroup species (horse, giraffe, goat, pig and sheep). In all additional trees examined, the positioning of the fallow deer in relation to its Old World family members remained similar, and therefore these trees are not included here. The position of *C. dama* in the topologies shown in **Figure C.1** therefore indicate that the mitochondrial regions targeted are not informative for the phylogeny of the giant Irish deer.

This is the first phylogenetic study of *Megaloceros giganteus*, and the giant deer sequences obtained group with the majority of the extant cervids. However, before any concrete conclusions can be made with regards to the relative positioning of the giant deer in the cervid phylogeny, more specimens from both Ireland and the Isle of Man need to be analysed, and larger fragments of the mitochondrial genome, for example the complete cytochrome *b* gene, need to be targeted.