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.

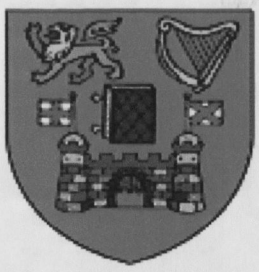
Access Agreement

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.

MOLECULAR GENETIC INVESTIGATIONS OF THE DIVERSITY AND ORIGINS OF OLD AND NEW WORLD CATTLE POPULATIONS

David Andrew Magee



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

Department of Genetics,

University of Dublin,

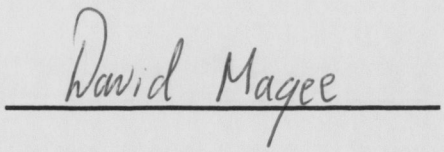
Trinity College

November 2002

TRINITY COLLEGE
18 MAR 2003
LIBRARY DUBLIN
THESIS
7190

DECLARATION

This thesis has not been previously submitted for examination to this or any other university. The work described herein has been carried out by the author alone, except where otherwise stated. The author agrees that the college library may lend or copy this thesis upon request.



David Magee

November 2002

1.5.4 Mitochondrial phylogenies in other domestic species 38

Chapter 2: Mitochondrial DNA phylogeography of domestic *Bos indicus* cattle 42

2.1 Introduction 43

2.1.1 Agricultural origins within South Asia 43

2.1.2 The origins of *Bos indicus* cattle 44

2.1.3 The dispersal of *Bos indicus* throughout Asia 46

2.1.4 A separate origin for South Indian *Bos indicus* 47

2.1.5 Modern Asian cattle 50

2.1.6 Research objectives 51

2.2 Materials and Methods 53

2.2.1 The domestic cattle populations involved in this study 53

2.2.2 *Bos indicus* DNA sample information 57

2.2.3 Mitochondrial DNA sequence data analysis 61

2.3 Results 66

2.3.1 Amplification of bovine mitochondrial DNA 66

2.3.2 Bovine mitochondrial DNA sequence variation 68

2.3.3 Genetic signatures of *Bos indicus* and *Bos taurus* introgression 69

2.3.4 Summary of *Bos indicus* mitochondrial DNA sequence variation 73

2.3.5 Intra-population genetic variation 76

2.3.6 Phylogenetic networks analysis 80

2.3.7 Regional *Bos indicus* mitochondrial DNA phylogenetic analysis 87

2.3.8 Mitochondrial DNA haplotypic distribution 91

2.3.9 Multi-dimensional scaling of inter-population genetic distances 96

2.3.10 Demographic history of *Bos indicus* cattle 98

2.3.11 Dating historic population expansions 105

2.4 Discussion and Conclusions 110

2.4.1 A South Asian origin for *Bos indicus* cattle 110

2.4.2 The genetic legacies of *Bos indicus* and *Bos taurus* migrations 111

2.4.3 The domestic and demographic history of *Bos indicus* cattle 113

Chapter 3: The genetic origins of Caribbean Creole cattle 123

3.1 Introduction 124

3.1.1 The historical origins of American Creole cattle 124

3.1.2 The genetic ancestry of *Bos taurus* cattle and research objectives 127

3.2 Materials and Methods 128

3.2.1 Cattle samples and DNA extraction 128

3.2.2 Mitochondrial DNA sequence analysis 130

3.2.3 Microsatellite allele scoring and phylogenetic analysis 132

3.3 Results 135

3.3.1 Amplification and sequencing of the most variable 240 bp DNA fragment of the mitochondrial genome 135

3.3.2 Typing and scoring of bovine microsatellite allelic variation 135

3.3.3 Summary of *Bos taurus* Creole mitochondrial DNA sequence variation 137

3.3.4 Phylogenetic network analysis 139

3.3.5 Mitochondrial DNA sequence pairwise mismatch distributions 142

3.3.6 Mitochondrial admixture proportions 148

3.3.7 Multi-dimensional scaling of mitochondrial DNA-derived inter-population genetic distances 148

3.3.8 Microsatellite allelic variation and admixture analysis 150

3.3.9 Multi-dimensional scaling of microsatellite-derived inter-population genetic distances 156

3.4 Discussion and Conclusions 158

3.4.1 The genetic legacies of historic introgressions to the Caribbean 158

3.4.2 Genetic founder effects 160

Chapter 4: Single nucleotide polymorphism variation in domestic cattle and a comparison with mitochondrial DNA and nuclear microsatellite variability 163

4.1 Introduction 164

4.1.1 Single nucleotide polymorphisms (SNPs) and their use in phylogenetic studies 164

4.1.2 Research objectives 166

4.2 Materials and Methods 167

4.2.1 Genotyping SNP alleles using the oligoligation assay (OLA) 167

4.2.2 SNP allele genotype data 168

4.2.3 Mitochondrial DNA sequence data and microsatellite genotypic data considered for comparative phylogenetic analysis 171
4.2.4 Phylogenetic analysis of molecular data 174
4.3 Results 178
4.3.1 Oligoligation assay (OLA) GENESCAN™ gel files 178
4.3.2 Intra-regional genetic diversity 178
4.3.3 The apportionment of genetic diversity between and within regional groups 189
4.3.4 Admixture analysis 195
4.3.5 Multi-dimensional scaling of inter-population genetic distances 198
4.4 Discussion and Conclusions 200
4.4.1 Intra-regional genetic diversity 200
4.4.2 Population admixture analysis 201
4.4.3 Inter-population genetic distances 204
4.4.4 The apportionment of bovine genetic diversity 205

Chapter 5: Laboratory methods 210
5.1 The sources of the molecular genetic data used for analysis in this thesis 211
5.1.1 The mitochondrial DNA sequence data used in this thesis 211
5.1.2 The microsatellite genotypic data used in this thesis 211
5.1.3 The SNP genotypic data used in this thesis 211
5.2 DNA extraction methods 225
5.2.1 DNA extraction from Caribbean Creole cattle serum samples 225
5.3 Mitochondrial DNA amplification and sequencing 227
5.3.1 Oligonucleotide primer information for mitochondrial DNA PCR amplification 227
5.3.2 PCR conditions for the amplification of bovine mitochondrial DNA sequences 227
5.3.3 Preparation of single stranded template DNA for sequencing 229
5.3.4 Sequencing mitochondrial DNA control region sequences 230
5.3.5 Polyacrylamide gel electrophoresis (PAGE) of sequenced cattle samples 232
5.3.6 Determination and recording of mitochondrial DNA sequence data 233
5.4 Microsatellite allele genotyping 234
5.4.1 The microsatellite loci genotyped in Chapter 3 of this thesis 234

5.4.2 PCR amplification of microsatellites 234
5.4.3 Polyacrylamide gel electrophoresis (PAGE) of PCR-amplified microsatellites 236
5.5 Single nucleotide polymorphism genotyping 238
5.5.1 Multiplex genotyping of SNP loci 238
5.5.2 Polyacrylamide gel electrophoresis (PAGE) of OLA products 244

Chapter 6: References 245

Appendix 1: Single nucleotide polymorphism discovery in the bovine genome 263

Appendix 2: Genetic distance matrices and multi-dimensional scaling co-ordinates 284

Appendix 2.A: *Bos taurus* mtDNA sequence diversity observed in 4 Morphologically zebu samples from the Indian subcontinent 285
Appendix 2.B: *Bos taurus* mtDNA sequence diversity observed in 13 cattle samples from the Near East 285
Appendix 2.C: Matrix of mtDNA pairwise F_{ST} genetic distances for the Asian zebu populations analysed in this study 286
Appendix 2.D: Two-dimension MDS plot co-ordinates calculated from mtDNA pairwise F_{ST} values in Appendix 2C 286
Appendix 2.E: Population pairwise F_{ST} genetic distances for European, African and Creole populations 287
Appendix 2.F: Two-dimensional MDS co-ordinates for European, African and Creole populations presented in Appendix 2.E 288
Appendix 2.G: Microsatellite D_A genetic distance estimates for European, African and Creole populations 289
Appendix 2.H: Two-dimensional MDS co-ordinates constructed from the data presented in Appendix 2.G 289
Appendix 2.I MtDNA pairwise F_{ST} genetic distances computed for European, African, Near Eastern and Indian populations 290
Appendix 2.J: MtDNA two-dimensional MDS plot co-ordinates for the data presented in Appendix 2.I 291

Appendix 2.K: Microsatellite-derived D_A genetic distances for European, African, Near Eastern and Indian populations 292

Appendix 2.L: Two-dimensional MDS plot co-ordinates for the data presented in Appendix 2.K 292

Appendix 2.M: SNP-derived DA genetic distance estimates for European, African, Near Eastern and Indian populations 293

Appendix 2.N Two-dimensional MDS plot co-ordinates for the data presented in Appendix 2.M 293

Appendix 3: Buffers and reagents **294**

FIGURES

Chapter 1

Figure 1.1: Systematic classification of wild and domestic species within the *Bovini* tribe 14

Figure 1.2: Postulated migratory routes of cattle throughout the Old World 20

Figure 1.3: The current distribution of domestic cattle throughout the Old World 21

Figure 1.4: A diagrammatic representation of the bovine mitochondrial genome 24

Figure 1.5: Bovine mtDNA- and microsatellite-derived phylogenies 37

Figure 1.6: Unrooted neighbor-joining mtDNA phylogenies for 6 Eurasian domestic ungulate species 39

Figure 1.7: The location of suggested livestock domestication centres throughout the Old World 40

Chapter 2

Figure 2.1: Ancient *Bos indicus* domestication and herding sites within South Asia 49

Figure 2.2: The approximate locations of the populations analysed for *Bos indicus* mtDNA variability in this study 56

Figure 2.3: PCR amplification of *Bos indicus* mtDNA 67

Figure 2.4: Partial DNA sequence of the most variable 240 bp fragment of the bovine mitochondrial control region 68

Figure 2.5a: The proportion of *Bos taurus* and *Bos indicus* mtDNA haplotypes in domestic populations from the Near East and the Indian subcontinent 71

Figure 2.5b: Patterns of *Bos taurus* introgression into the Indian subcontinent 72

Figure 2.6a: Intra-regional mtDNA sequence diversity expressed as the mean number of pairwise differences 78

Figure 2.6b: Intra-regional mtDNA sequence diversity expressed as Watterson's θ_S parameter 79

Figure 2.7a: A reduced median phylogenetic network featuring 149 *Bos indicus* mtDNA sequences 83

Figure 2.7b: Reduced median phylogenetic networks constructed for *Bos taurus* mtDNA sequences sampled in Europe, Africa, the Near East and Japan 84

Figure 2.7c: The phylogenetic relationship between the predominant *Bos taurus* and *Bos indicus* mtDNA haplotypes 85

Figure 2.8: Regional *Bos indicus* mtDNA reduced median phylogenetic networks 90

Figure 2.9: Z1 and Z2 haplogroup proportions for each assayed population 92

Figure 2.10: Regression plot of population Z1 haplogroup proportions with respect to degrees longitude 95

Figure 2.11: MDS plot and unrooted NJ phylogenetic tree summarising inter-population F_{ST} genetic distances among the zebu populations analysed in this study 97

Figure 2.12: *Bos indicus* and *Bos taurus* mtDNA sequence mismatch distributions 101

Figure 2.13: The distribution of mutational steps from the presumed ancestral haplotype within each major sampled *Bos indicus* and *Bos taurus* haplogroup 106

Chapter 3

Figure 3.1: Introductory routes of cattle to the Americas 126
Figure 3.2: The geographic location of the Creole samples used in this study 129
Figure 3.3: Microsatellite allelic variation at the *BM2113* locus 136
Figure 3.4a: A reduced median phylogenetic network constructed from 73 Creole *Bos taurus* mtDNA sequences 140
Figure 3.4b: Reduced median phylogenetic networks constructed for Western European, Iberian, Creole and African *Bos taurus* mtDNA sequences 141
Figure 3.5: MtDNA sequence mismatch distributions for the T1 and T3 haplogroups in European, Iberian, Creole and African cattle 145
Figure 3.6: A two-dimensional MDS plot featuring African, European, Iberian and Creole mtDNA sequences 149
Figure 3.7: Microsatellite allele frequencies for European, Iberian, Caribbean Creole, West African taurine (WAT), West African zebu (WAZ) and Indian zebu cattle 151
Figure 3.8: MtDNA- and microsatellite-derived estimates of African *Bos taurus* and *Bos indicus* admixture proportions 155
Figure 3.9: A two-dimensional MDS plot constructed for African, European, Iberian and Creole microsatellite derived genetic distances 157

Chapter 4

Figure 4.1: Genotyping SNP alleles using the oligoligation assay (OLA) 169
Figure 4.2: GENESCAN™ and GENOTYPER™ files showing bovine SNP genotypes assayed using OLA 179
Figure 4.3: SNP allele frequencies for all regional groups 188
Figure 4.4: Inter-population genetic differentiation for a number of large bodied species with large geographic ranges 193
Figure 4.5: *Bos taurus* and *Bos indicus* admixture proportions for all 3 genetic systems 197
Figure 4.6: Two-dimensional MDS plots constructed for 3 genetic systems 199

TABLES

Chapter 2

Table 2.1: The total *Bos indicus* mtDNA sequence data set considered for phylogenetic analysis in this survey 60

Table 2.2: Nucleotide variation observed in 149 *Bos indicus* mtDNA control region sequences 75

Table 2.3: *Bos indicus* mtDNA sequence diversity estimates for individual *Bos indicus* populations and regional groups assayed in this study 77

Table 2.4: Geographic divisions of *Bos indicus* mtDNA data for regional phylogenetic network analysis 87

Table 2.5: Hierarchical components of variation for *Bos taurus* and *Bos indicus* regional groups 88

Table 2.6: Chi-squared contingency table for geographic distribution of *Bos indicus* haplogroups 94

Table 2.7: Statistical parameters for the mtDNA sequence mismatch distributions presented in Figure 2.12 103

Table 2.8: Expansion dates for the predominant *Bos indicus* and *Bos taurus* mtDNA haplogroups 108

Chapter 3

Table 3.1: Caribbean Creole and Brazilian Nelore mtDNA sequence information 131

Table 3.2 Caribbean Creole DNA sample information for microsatellite analysis 132

Table 3.3: Nucleotide variation observed in 73 Creole *Bos taurus* mtDNA control region sequences 138

Table 3.4: Statistical parameters for T3 and T1 *Bos taurus* haplogroup mismatch distributions within each geographic region 147

Table 3.5: Proportion of African mtDNA haplotypes, West African taurine-specific and Indian zebu-specific microsatellite alleles for each Creole population analysed 153

Chapter 4

Table 4.1: SNP locus information and populations considered for phylogenetic analysis 170

Table 4.2: The mtDNA sequence data collection used for comparative analysis 172

Table 4.3: The microsatellite data used for comparative analysis 173

Table 4.4: Genetic diversity estimates for geographical regional groups 180

Table 4.5: SNP-based genetic diversity estimates for each individual cattle population 181

Table 4.6a: Intra-regional mtDNA genetic diversity as expressed in terms of the total number of haplotypes and number of unique haplotypes 184

Table 4.6b: Intra-regional microsatellite genetic diversity as expressed in terms of the total number of alleles and number of unique alleles 185

Table 4.6c: Intra-regional SNP genetic diversity as expressed in terms of the total number of alleles and number of unique alleles 185
Table 4.7a: The number of mtDNA haplotypes shared between regional groups 186
Table 4.7b: The number of nuclear microsatellite alleles shared between regional groups 187
Table 4.7c: The number of nuclear SNP alleles shared between regional groups 187
Table 4.8: The proportion of genetic variance attributable to population subdivision as represented by F_{ST} values 190
Table 4.9 A comparison of the total number of unique haplotypes/alleles, shared haplotypes/alleles and F_{ST} values for the total Old World data set analysed in this study 191
Table 4.10: Hierarchical AMOVA analysis showing the percentage of variation at each of the three levels of population hierarchy 195

Chapter 5

Table 5.1: Mitochondrial DNA data used for phylogenetic analysis in Chapters 2, 3 and 4 of this thesis 212
Table 5.2: The microsatellite allelic data used in Chapter 3 of this study 219
Table 5.3: The microsatellite allelic data used in Chapter 4 of this study 221
Table 5.4: The SNP data used in Chapter 4 of this thesis 223
Table 5.5: Oligonucleotide primers used for bovine mtDNA amplification 227
Table 5.6: Mitochondrial DNA PCR reagents 228
Table 5.7 Mitochondrial DNA PCR cycling parameters 228
Table 5.8 DNA sequencing annealing reaction components (per sample) 231
Table 5.9 DNA sequencing labelling reaction components (per sample) 231
Table 5.10 DNA sequencing termination reaction components (per sample) 232
Table 5.11 Microsatellite loci genotyped in Creole cattle populations 235
Table 5.12 Radiolabelled PCR amplification of microsatellite loci 236
Table 5.13 Radiolabelled PCR amplification cycling parameters 236
Table 5.14 Multiplex PCR conditions for SNP genotyping 239
Table 5.15 Multiplex OLA reaction conditions for SNP genotyping 239
Table 5.16 Primer and allele information for SNP genotyping 240

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my parents, Derek and Hilary, my brother Paul and my family who were the driving force behind this thesis - for their constant support, words of wisdom, encouragement and for always lending me their ears in times of need.

I would like to say a massive "thank you" to my supervisor, Dan Bradley, for giving me the opportunity to be part of the team and for his enthusiasm, scientific insight, endless patience and constant help at every level, right from day one. To paraphrase one famous Irish sports journalist – "he's not a good supervisor; he's a great supervisor".

I am deeply indebted to the other members of "Club Bovine", both past and present, who made it a great place to work. I would particularly like thank Ashie, Brian, Ceiridwen, Elaine, Emma, Laoise, Ruth, Stephen and Teresa – for their friendship, good humour and above all else their unending willingness to help. I wish them the very best of luck with everything in the future. A big thanks also to Chris, Ciaran, Dave, Emmy, Jill, John, Ronan, the Identi-girls and Paddy Cunningham for their help and advice over the years.

My sincere thanks to Dave, Paul, Brenda, Louis and Sue – unquestionably the heartbeat of the department - for their remarkable resourcefulness and for always being able solve the unsolvable. I am also grateful to Dr. Gregor Durstewitz, Dr. Alex Morrow, Dr. Pushpendra Kumar, Prof. Reudi Fries, Prof. Hide Mannen, Dr. Juha Kantenen, Ben Davies and Susan Harrison, whom I had the good fortune to meet and collaborate with during the course of the project.

This would not be complete without saying a huge, heartfelt thanks to my friends who were always there to provide laughs and pints, particularly Paulzer, Ric, Domhnall, Edders, the boys from footie (Mick McCarthy will be on the blower some day!!), the sirens of Firhouse/Ballyroan, Tom and Norrie for bed and lodgings (my cheque is in the post), Fred and Jacinda and my fellow postgrads who introduced me to that den of debauchery known as Renards!

Last, but by no means least, Kavo. Thanks for always understanding, for always being there and putting things into perspective - I really couldn't have done it without you.

ABBREVIATIONS

AD	Anno domini
APS	Ammonium persulphate
BC	Before Christ
bp	Base pairs
BTA	Bovine chromosome number
ca.	*circa*
Ci	Curie
D_A	Nei's standard genetic distance
dATP (A)	Deoxyadenosine triphosphate
dCTP (C)	Deoxycytosine triphosphate
ddATP	Di-deoxyadenosine triphosphate
ddCTP	Di-deoxycytosine triphosphate
ddGTP	Di-deoxyguanosine triphosphate
ddH_2O	Deionised, distilled water
ddTTP	Di-deoxythymidine triphosphate
dGTP (G)	Deoxyguanosine triphosphate
D-loop	Displacement loop
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucelic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
dTTP (T)	Deoxythymidine triphosphate
EDTA	Ethylenediamine tetra-acetic acid
F_S	Fu's statistic of selective neutrality
F_{ST}	Estimator of total genetic variance due to population subdivision
HEX	4, 7, 2', 4', 5', 7'-Hexachloro-6-carboxyfluorescein
HVR1	Hypervariable region 1
kb(s)	Kilobase pair(s)
M	Molar
mCi	Millicurie
MDS	Multi-dimensional scaling
$MgCl_2$	Magnesium chloride
ml(s)	Millilitre(s)
mM	Millimolar
$MnCl_2$	Manganese chloride
MNPD	Mean number of pairwise differences
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
MYA	Million years ago
n	number of samples/number of populations

n/a	not available
NaCl	Sodium chloride
NAD	Nicotinamide adenine dinucleotide
NaOH	Sodium Hydroxide
N_e	Effective population size
ngs	Nanograms
Obs.	Observed
°C	Degrees Celsius (temperature)
OLA	Oligoligation assay
OXPHOS	Oxidative phosphorylation
P	Probability
^{32}P-dCTP	^{32}Phosphorous-labelled dCTP
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEO	Polyethylene oxide
Pho	5'-phosphate residue
pers. comm.	personal communication
RFLP	Restriction fragment length polymorphism
ROX	6-Carboxy-X-rhodamine
rpm	Revolutions per minute
rRNA	Ribosomal RNA
^{35}S-dATP	^{35}Sulpher-labelled dATP
SDS	Sodium dodecyl sulphate
6-FAM	6-Carboxy fluorescein
SNP(s)	Single nucleotide polymorphism(s)
SSCP	Single-strand conformation polymorphism
SSD	Sum of square deviations
STR(s)	Simple tandem repeat(s)
TAMRA	*N, N, N', N'*-Tetramethyl-6-carboxyrhodamine
TBE	Tris-borate EDTA buffer
TCD	Trinity College Dublin
TEMED	*N,N,N',N'*-Tetramethylethylenediamine
3'-UTR	3' Untranslated region
tRNA	Transfer RNA
2*n*	Diploid chromosome number
μg	Microgram
μls	microlitres
μM	Micromolar
UV	Ultraviolet light
VNTR	Variable number of tandem repeats
W	Watts
YBP	Years before present

SUMMARY

(1) *Bos indicus* domestic origins and demographic history

One hundred and forty-nine *Bos indicus* mitochondrial DNA (mtDNA) control region sequences sampled from the Indian subcontinent, the Near East and East Asia were assessed in order to investigate the domestic origins and demographic history of *Bos indicus* cattle. Results were compared and contrasted to previously reported *Bos taurus* mtDNA sequences sampled in Europe, Africa, the Near East and Japan. Phylogenetic analysis reveals substantially elevated levels of *Bos indicus* mtDNA diversity in populations from the Indian subcontinent, which are comparable to those estimated for *Bos taurus* populations of the Near East - the purported cradle of *Bos taurus* domestication. Phylogenetic network analysis shows that *Bos indicus* mtDNA haplotypes cluster around one of two numerically and topologically predominant sequence variants producing star-like patterns of diversity, which are indicative of past population expansions. Time-depth estimates of these supposed episodes of demographic growth are consistent with the historic timeframe of *Bos indicus* animal herding, *circa* 9,000-8,000 years before present. Collectively, these data suggest a South Asian origin of domestic *Bos indicus* cattle and that modern matrilines are derived from at least two founder haplotypes that were incorporated into the domestic pool via the capture and taming of wild female progenitors during the domestication process. The geographic assortment of the two major *Bos indicus* mtDNA sequence clusters may also indicate that these putative domestication events were spatially separate. Furthermore, the detection of *Bos taurus* mtDNA sequence variants in *Bos indicus* samples from North India and a somewhat symmetrical detection of *Bos indicus* mtDNA haplotypes in Near Eastern *Bos taurus* populations may reflect ancient pastoralist movements and trade between the two regions.

(2) The genetic history of Caribbean Creole cattle

Seventy-eight cattle samples from Latin America (Creole cattle) were analysed for sequence variation in the hypervariable segment of the mitochondrial DNA control region. Seventy-three samples displayed *Bos taurus* mtDNA haplotypes; five displayed *Bos indicus* haplotypes. Phylogenetic analyses reveal that all sampled Creole *Bos taurus* sequences fall into two distinct clusters which have separate African and European origins. European sequences were encountered in each population. However, the distribution of African haplotypes was uneven, with the highest proportion of African influence found in the Guadeloupe Creole. The reduced levels of African haplotypic variation within the

Caribbean is consistent with founder effects during importation. Additionally, genetic variation at three microsatellite loci illustrate African influence uniquely in the Guadeloupe Creole. The data suggest that this African influence is, possibly, attributable to the historical importation of African cattle to the Americas. Furthermore, alleles of *Bos indicus* ancestry were detected at appreciable frequencies in all Caribbean Creole populations and are consistent with a history of predominantly male-mediated African and Indian *Bos indicus* introgression into the Americas.

(3) Single nucleotide polymorphisms (SNPs) variability in domestic cattle populations and a comparison with mitochondrial and microsatellite genetic systems

A panel of eleven nuclear SNP loci were genotyped in European, African and Near Eastern *Bos taurus* and Indian *Bos indicus* populations. Resulting patterns of SNP-derived variability were compared to previously published mtDNA sequence and nuclear microsatellite allelic data for domestic *Bos taurus* and *Bos indicus* populations of similar geography. Both mitochondrial and microsatellite systems reveal highest genetic diversity levels in Near Eastern populations, which is consistent with the view that this region represents a primary domestication centre for *Bos taurus* cattle. In contrast, SNP-derived diversity is higher in Europe but is probably explained by an ascertainment bias in which the assayed loci were initially identified as being polymorphic in European samples. Comparison of inter-regional population genetic differentiation estimates (F_{ST} values) shows that the majority of nuclear variability occurs between individuals within populations rather than between regional groups. Alternatively, mtDNA-based estimates of inter-regional differentiation are high. These results, (1) reflect the high disparity between male and female reproductive variance under managed breeding systems in which genetic exchange between locales is predominantly male-mediated, and (2) illustrate the differential evolutionary dynamics between nuclear and mtDNA genetic elements. The primary feature of the genetic distance estimates between domestic cattle populations using all three genetic markers is the marked divergence between *Bos taurus* and *Bos indicus* populations. The genetic distinction of African *Bos taurus* populations noted for all three data sets may also support archaeological assertions for the domestication of African *Bos taurus* cattle from an indigenous wild progenitor strain.

CHAPTER 1

GENERAL INTRODUCTION

1.1 THE ORIGINS OF AGRICULTURE

1.1.1 The advent of agriculture

The advent of agriculture is considered to be one of the most important innovations in human history. Prior to this anatomically modern humans had successfully managed to occupy most habitable and accessible regions of the globe, surviving as hunter-gatherers on a diverse range of wild plant and animal foods. The transition in human subsistence from nomadic forager to sedentary agriculturist had a number of dramatic cultural, economic and demographic consequences. Through the cultivation of crops and the herding of animals, humankind could for first time exploit natural surroundings, enabling increased food yields to be obtained and larger populations to be supported. This in turn, facilitated long-term sedentism and ultimately resulted in the development and progression of urban civilisation throughout the world (Cavalli-Sforza 1996; Harris 1996).

The conditions that prompted this transition in human subsistence most probably involved a series of coincidental demographic, geographic and climatic factors that were exclusive to the primary heartlands of agriculture, namely Southwest Asia (also referred to as the Near East), East Asia, Mesoamerica, North America and South America (Pringle 1998a; Diamond 2002), however some authors also cite the Indian subcontinent as another candidate centre for agricultural origins (Meadow 1996). According to the prevailing view, agricultural practices were first initiated in Southwest Asia some 10,000 years before present (YBP), within the region encompassed by the Fertile Crescent - a broad geographical arc stretching from the Jordan Valley in the southwest to the Zagros Mountains in the east (McCorriston & Hole 1991; Harris 1996; Zohary 1996). However, recent archaeobotanical evidence suggests that Near Eastern hunter-gatherer communities may have cultivated crops as early as 13,000 YBP. It is also possible that the peoples of East Asia, Mesoamerica and South America may have domesticated plants between

11,000–10,000 YBP – several millennia earlier than previously thought (Pringle 1998). However, despite initial advances towards an agrarian lifestyle, it seems that early plant cultivators continued to hunt and gather wild plants and animals, only becoming full-scale agriculturists *circa* 10,500 YBP in Southwest Asia, 9,000 YBP in Asia and 7,000 YBP in the Americas. Thus, the transition in human subsistence was a long, complex process rather than a dramatic cultural revolution (Wenke 1999; Lev-Yadun *et al.* 2000).

1.1.2 Climate change and agricultural origins

Numerous hypotheses have been proposed to explain the emergence of agriculture throughout the world. Some authors have attributed the domestication of plants and animals to humankind's ingenuity, stating that the process was executed in a co-ordinated and premeditated fashion (Isaac 1962). Other proposals, such as the "Edge-Zone Hypothesis" (Binford 1968 [cited by Wenke 1999]) contend that human population expansion in certain regions prompted a more efficient use of natural resources, which ultimately resulted in the domestication of plants and animals. However, the somewhat simultaneous emergence of agriculture throughout different regions of the world has prompted most authorities to invoke a climatic trigger.

The close of the Pleistocene (*circa* 13,000 YBP) within Southwest Asia, the oldest and most extensively investigated agricultural centre, was marked by a series of climatic improvements. During this period the cold, arid conditions established during the Late Glacial Maximum (*circa* 24,000-15,000 YBP) gradually gave way to increasingly warmer conditions characterised by seasonal precipitation regimes. However, the overall trend of climatic amelioration across the Pleistocene-Holocene boundary (*circa* 12,000 YBP) was not a stable process and was often punctuated with periods of increased warmth alternating with reversions to cooler, more arid conditions. One such episode of climatic cooling,

known as the Younger Dryas (*circa* 11,000-10,000 YBP), is believed to have caused a dramatic worldwide reduction in the diversity of plant and foods previously available to hunter-gatherers (Moore & Hillman 1992; Hole 1996). The corresponding decrease in resource availability would have induced stress on local hunter-gatherer communities and consequently, there was a need to develop methods which facilitated the progression and maintenance of a continued food supply (Diamond 2002). In turn, the dependency on plant foods that were easily stored and cultivated greatly increased (Hillman 1996).

Periods of prolonged drought and hyper-aridity, such as that experienced during the Younger Dryas, favour the growth and spread of geophytes (herbaceous perennials plants with subsurface storage organs) and large-seed annuals (plants such as the wild cereals that complete their lifecycle in a year or less) (Byrne 1987). As a result, plants of this type became the most abundant forms of vegetation throughout the early agriculture centres across the Pleistocene-Holocene boundary (Blumler 1996). Due to ease of storage and cultivation, large seed annuals ultimately became the first plant types to be domesticated within the primary agricultural centres (Byrne 1987; Hillman 1996; Zohary 1996; Lev-Yadun *et al.* 2001).

Once the trends of seasonality associated with the Holocene became more stable, *circa* 10,000 YBP, the transition in human subsistence from forager to farmer was well established. Humankind's greater understanding of the seasons and growing conditions greatly facilitated the spread of agriculture and peoples throughout the Old and New Worlds (Cavalli-Sforza 1993; Blumler 1996).

1.1.3 Geographical factors influencing the emergence of agriculture

In addition to climatic contributions, geographical factors are also believed to have played a major role in the development of agriculture in Southwest Asia. The modern landscapes of the three primary domestication centres were created during the Late Pleistocene as a result of plate tectonics, vulcanism, and cycles of erosion and deposition. As a result, numerous diverse topographical features, with a multitude of divergent, dynamic environments, such as coastal plains, mountains, well-irrigated river valleys, and desert habitats were combined within a restricted area (Moore & Hillman 1992; Sherratt 1996). It is argued that the intimate juxtaposition of these contrasting ecosystems provided new opportunities for early agrarians to transplant crops between habitats, allowing optimal environmental conditions necessary for successful crop cultivation to be established (Sherratt 1996; Diamond 2002).

1.2 ANIMAL DOMESTICATION

1.2.1 Why domesticate animals?

"Animal domestication, in the case of food species, can be defined as a selective diachronic process of change in human-animal relationships involving, at the very least, a shift of focus from the dead to the living animal and, more particularly, from obtaining and distributing the products of the dead animal to securing and selectively maintaining the most important product of the living animal - its progeny" (Meadow 1993).

The domestication of animals can be viewed as an elaboration of the predator-prey relationship, the primary aim of which was to secure and conserve animal food supplies ensuring human survival, especially against the threat of crop harvest failure or decline in the abundance wild food plants (Meadow 1989; Smith 1995). Although successful crop harvests would have increased the longevity and size of early semi-sedentary human groups, the hunting of wild animals remained a fundamental practice in order to provide the essential animal proteins necessary for survival (Bökönyi 1976; Hillman 1996). However, excessive predation would have eventually thinned out the wild fauna within the range of the settlement, greatly reducing humankind's animal protein resources (Bar-Yosef & Meadow 1995). Therefore, several authors argue that animal domestication was forced upon increasingly sedentary plant cultivators living in restricted areas (Bökönyi 1976; Bar-Yosef & Meadow 1995).

Whether the domestication of animals by humans occurred spontaneously or consciously is subject to continual debate. Uerpmann (1996) states that humans may have initially domesticated animals while unaware of their actions and consequences. Human nursing instincts may have resulted in the capture and taming of young, easily manageable herbivorous wild animals, such as wild goats (*Capra aegagrus*) and sheep (*Ovis*

orientalis). Although the rearing of young wild animals as pets may not have directly resulted in their domestication, successful maintenance of these animals within a few proto-Neolithic villages may have resulted in the awareness of their economic potential (Smith 1995). The coupling of this realisation with a few successive, good cereal harvests may have facilitated the development and progression of a number of small productive domestic flocks. Once the basic techniques of animal management were understood, humans could consciously focus their attention on the domestication of larger herbivorous animals such as wild cattle (*Bos primigenius*) and wild boars (*Sus scrofa*).

1.2.2 Archaeological evidence for animal domestication

There are two principal approaches by which scientists investigate the origins of animal (and plant) domestication, namely a biological and archaeological approach. Biological approaches, as first implemented by Nikolai Valivov in the 1930's, inspect the patterns and biological diversity of modern domesticates. Modern genetic diversity studies can be seen as an extension of this approach. Elevated biological and genetic diversity within a defined geographic area together with an overlapping distribution between domestic and pre-domestic forms of a species are collectively regarded as biological evidence for domestication. Alternatively, archaeological approaches document the progression of wild and domestic forms within the purported natural habitation zone of a species through the comprehensive analysis of the fossil record (Smith 1995). Inferring animal domestication and herding by means of archaeological data is the primary subject of this section.

When archaeologically evaluating evidence for animal domestication it is essential to distinguish between wild and domestic forms. Furthermore, it is important to differentiate between the presence of domestic animals and the process of animal domestication at a site. The former involves identifying the end products of the domestication process (*i.e.* the

domestic animal itself) and may indicate the practice of animal herding. Alternatively, a convincing candidate region for a domestication centre is one where a staged, temporal transition from characteristically wild to domestic faunal patterns in response to changes in human behaviour over time, may be demonstrated (Meadow 1989; Meadow 1993; Wenke 1999). The established criteria used to determine the status of faunal remains at archaeological sites are summarised below (Meadow 1989):

Evidence for animal herding at an archaeological site

(1) The age groups and sex ratios of a candidate domestic animal species, as inferred from faunal remains, are disproportionate to those encountered in wild populations. This suggests a planned culling rather than the more random distribution expected from a hunted sample;

(2) Domesticated species appear in an area in which their wild progenitor did not exist, at least since the Pleistocene;

(3) Artistic representations of domestic animals, in the form of rock paintings/carvings and sculptures and artefacts associated with animal herding exist at the site.

Evidence for animal domestication at an archaeological site

(1) The skeletal remains of the domestic animal, its purported wild progenitor and transitional forms exist at the site. One of the major effects of domestication on animal morphology is a gradual reduction in body size over time, due primarily to malnutrition and physiological demands on the animal (such as drawing of milk) within early human settlements (Bökönyi 1976; Smith 1995). Furthermore, distinctive changes in morphological proportions and the form and fine structure of bones, such as shortening of the skull, loss or changes in horn morphology, crowded or decreased teeth may also be apparent;

(2) There is a decrease in the number of wild species remains and a corresponding increase in domestic species remains in the bone assemblages recovered from successive strata at the site;

(3) The age and sex group proportions of the wild forms excavated from the site are not the same as those found normally in wild populations, possibly suggesting a pre-determined selection of the animals chosen for domestication. For example, young female animals may have been more easily captured and tamed than aggressive males;

(4) Possible artistic representations depicting wild animal captures are present at the site.

1.2.3 The first domestic animals of the Near East

To negotiate the transition from a wild to a domesticated state, an animal species must fulfil a defined set of behavioural and physiological attributes first outlined by Galton (1865 [cited by Clutton-Brock 1981]). Accordingly, an ideal candidate for domestication is a wild animal that:

(1) is already useful to humans either as a food source or provider of secondary products, such as hides and milk;

(2) is well-adapted to living and breeding in confined areas;

(3) is a placid, highly social animal with an inborn liking for man;

(4) does not require a specialised diet to survive.

The earliest indication of animal domestication in the Near East traces back to the Palaeolithic archaeological sites of the Nautifan culture of the Levant (*circa* 12,500-10,500 YBP). It is believed that during this period a symbiotic relationship developed between hunter-gatherers and the wolf (*Canis lupus*), which gave rise to the domesticated dog (*Canis familiaris*) (Reed 1977). Some 1,000 years later, evidence for ungulate domestication appears in the archaeological record of the Near East. These animals were

central to the Neolithic economies of Southwest Asia and although they were domesticated several centuries after the advent of crop cultivation, it is believed that both processes were inextricably linked (Garrard *et al.* 1996).

The first herbivores to be domesticated within Southwest Asia were the *ovicaprids* (sheep and goats) during the beginning of the Holocene period (11,000-9,000 YBP). These small ruminants were suitable for domestication due to their small size and ease of capture. Pronounced shifts in the faunal record and demographic evidence (such as sex and age ratios), suggest that domestic sheep (*Ovis aries*) were derived from the West Asiatic mouflon (*Ovis orientalis*) at several possible sites throughout the Near East, such as Cafer, Cayönü and Nevali Çori (all in Eastern Anatolia), Abu Hureyra and Tell Halula (both in Northern Syria), *circa* 10,500-9,000 YBP (Smith 1995; Legge 1996).

The Persian wild goat, or bezoar (*Capra aegagrus*) has been identified as the wild progenitor of the domestic goat (*Capra hircus*). Fossil and demographic evidence, namely an over-representation of adult females and the large percentage of sexually immature male animals when compared to wild populations, have been employed to infer possible goat domestication events at Cayönü, Tell Halula and at Ganj Dareh (Northwest Iran), *circa* 10,500-9,000 YBP (Bökönyi 1976; Smith 1995).

Cattle and pigs were the final two Near Eastern livestock species to be domesticated. Archaeological data from Nevali Çori and Cayönü provide the early evidence for the possible domestication of the pig (*Sus scrofa*) between 10,000-9,000 YBP (Bökönyi 1976; Smith 1995). During this time, the first steps towards cattle domestication took place in the Near East, the archaeological evidence for which is discussed in greater detail in **Section 1.3**.

1.3 CATTLE

1.3.1 Cattle taxonomy

The family Bovidae

Domesticated cattle belong to the mammalian *Bovidae* family, within the zoological order *Artiodactyla*, or even-toed ungulates (Lenstra & Bradley 1999). The *Bovidae* are characterised morphologically by the presence of non-deciduous, hollow horns, which distinguish them as a monophyletic clade (Gentry 1992). Monophyly of the *Bovidae* is also supported by several molecular investigations (Gatesy *et al.* 1997; Matthee & Davis 2001). Fossil records indicate that the *Bovidae* first appeared in the Miocene (*circa* 20 MYA) and it is believed that their diversification coincided with the widespread emergence of African savannah-based ecological niches 18-23 MYA (Allard *et al.* 1992). The evolution of the *Bovidae* is characterised by large-scale rapid adaptive radiations, global immigrations and mass extinctions, which in concert have given rise to more than 49 extant genera comprising over 140 extant species and some 300 extinct species (Nowak 1991). The majority of the extant species are endemic to Africa, while the present-day natural distribution of the family covers all of Africa, most of Eurasia and North America and some islands of the Arctic and East Indies (Nowak 1991; Gatesy *et al.* 1992).

Molecular and morphological investigations generally divide the extant members of the *Bovidae* family into at least seven subfamilies: the earliest diverging *Bovinae* (cattle, bison, yak, buffalo, nilgai and eland) and the more recent *Cephalophinae* (duikers), *Caprinae* (sheep, goats and related animals), *Hippotraginae* (roan antelope, sable antelope and gemsbok), *Antilopinae* (gazelles, springbok and related animals), *Reduncinae* (waterbuck) and *Alcelaphinae* (hartebeest and wildebeest) (Lenstra & Bradley 1999; Matthee & Davis 2001). Furthermore, phylogenetic inferences based on mitochondrial and nuclear genetic

elements assign the African impala to a separate *Aepycerotinae* subfamily (Matthee & Davis 2001).

The subfamily Bovinae and the tribe Bovini

There are three zoological tribes within the *Bovinae*, namely the *Boselaphini* (the nilgai and four horned antelope), the *Tragelaphini* (the kudu and eland) and the large-sized *Bovini*, which includes the wild and domesticated buffalo, yak and cattle species (Lenstra & Bradley 1999). Previous attempts to resolve the evolutionary relationships among the *Bovini* genera using molecular data have proved difficult. Evolutionary reconstructions using molecular data are believed to have been complicated by intra-specific introgression, the retention of ancestral polymorphisms and inter-specific hybridisation and often result in anomalous and incongruent phylogenies (Janecek *et al.* 1996; Lenstra & Bradley 1999).

Using nuclear microsatellite genetic markers, Ritz *et al.* (2000) argue for the division of the *Bovini* into four genera: (1) *Bos*, which consists of three subgenera, namely subgenus *Bos* (domesticated *Bos taurus* (taurine) and *Bos indicus* (zebu) cattle and their wild relatives), subgenus *Pöephagus* (domesticated and wild yak) and subgenus *Bibos* (domesticated mithan and Bali cattle and their respective wild relatives, the gaur and the banteng), (2) *Bison* (American bison and European bison or wisent), (3) *Bubalus* (water buffalo and their wild relatives, the arni, tamarao and anoa) and (4) *Syncerus* (African buffalo). This systematic classification of the *Bovini* tribe is depicted in **Figure 1.1**.

This phylogeny, however, does not concur with the results generated from principal component analysis of PCR-generated genetic fingerprints (Buntjer *et al.* 2002). This study, based on the variation detected at numerous genetic loci dispersed over the nuclear genome, reveals a clustering of the yak (*Pöephagus*) with the *Bison* species as opposed to

the *Bos* species. Other taxa groupings within the *Bovini* tribe supported by the data of Buntjer *et al.* (2002) include: (1) a clustering of the buffalo taxa (*Bubalus* and *Syncerus*) relative to the *Bos*, *Pöephagus* and *Bison* species, (2) a clustering of domesticated taurine and zebu cattle, and (3) an evolutionary proximity between the banteng and the gaur.

Figure 1.1: Systematic classification of wild and domestic species within the *Bovini* tribe

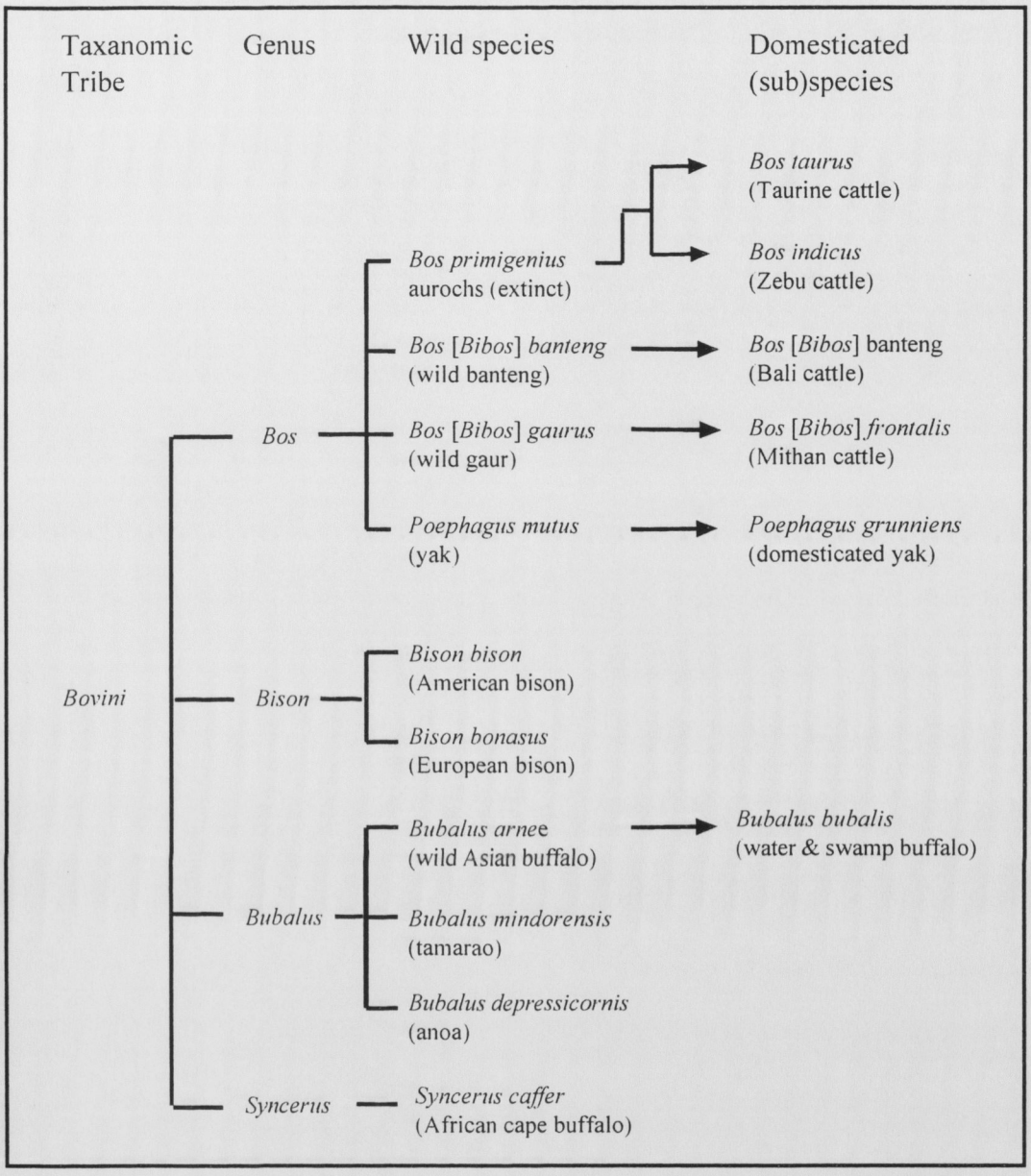


Figure 1.1. The taxanomic classification of wild and domesticated species within the *Bovini* tribe. Figure modified from Payne (1991) considering phylogenetic inferences of Ritz *et al.* (2000). Ritz *et al.* (2000) propose that the *Bovini* be divided into four genera, namely *Bos*, *Bison*, *Bubalus* and *Syncerus*. The clustering of both buffalo taxa (*Bubalus* and *Syncerus*) is supported by genetic fingerprint analysis, as is the clustering of, (1) the domestic cattle taxa (*Bos taurus* and *Bos indicus*), (2) the clustering of the banteng and the gaur, and (3) the clustering of both *Bison* taxa. However, this analysis groups the yak with the *Bison* species (Lenstra & Bradley; Buntjer *et al.* 2002).

1.3.2 The domestic taxa of cattle

Currently, there are approximately 1,300 million head of cattle (constituting some 800 extant cattle breeds) distributed across the globe, making them the most numerous and arguably, the most economically important of all the domestic livestock species (Bradley & Cunningham 1999). There are two recognised subspecies of domesticated cattle, namely the humpless taurine cattle of Europe, Northern Asia and West Africa (*Bos taurus*) and the humped zebu or indicine cattle of the Indian subcontinent and Africa (*Bos indicus*). *Bos taurus* cattle predominate in the temperate lands of Europe, West Africa and Northern Asia, whereas *Bos indicus* cattle generally inhabit the hot-arid or semi-arid regions of South Asia and Africa.

Zebu cattle differ from taurine cattle in a number of morphological and physiological aspects. Zebu animals are characterised externally by a prominent hump located either on the posterior part of the neck (cervico-thoracic) or on the shoulder (thoracic), a long face, usually steep upright horns and a dewlap (a fold of loose skin that hangs from the neck). The exact function of the hump remains unknown, however its biological composition (a mixture of muscle, fat and connective tissue) indicates that it is not a water storage organ (Payne 1991). Zebu cattle display a lower metabolic rate and nutrient requirement in response to drought and food shortages than their taurine counterparts in similar conditions. They are also more resistant to ticks, gastro-intestinal parasites and rinderpest than taurine cattle, although indigenous taurine animals exhibit comparative advantage in the tsetse-fly infected regions of Africa (Epstein & Mason 1984; Marshall 1989). Despite these differences, both domestic taxa possess the same number of chromosomes ($2n = 60$) and can freely interbreed, producing fertile offspring. Hence, some authors consider *Bos taurus* and *Bos indicus* cattle as subspecies within the *Bos* genus (Bökönyi 1997).

1.3.3 The origins of domestic cattle

The wild ancestor of domesticated cattle

It is widely accepted that the progenitor of all domesticated cattle was the morphologically similar aurochs, *Bos primigenius* (Zeuner 1963; Epstein & Mason 1984; Payne 1991). *Bos primigenius* is thought to have evolved from an Asiatic ancestor, *Bos acutifrons*, during the course of the Upper Pliocene, approximately 5-1.8 MYA (Zeuner 1963). The aurochs is believed to have ranged throughout much of Asia, Northern Africa and Europe reaching as far as the eastern coastlands of the Atlantic ocean, excluding Ireland (Epstein & Mason 1984). Although now extinct (the last surviving European aurochs died in a Polish animal reserve in 1627), the external appearance of the aurochs is well known from skeletal remains and contemporary descriptions. Aurochs bulls were pronouncedly larger than modern domesticates, with a height of up to six and a half feet at the shoulder and were often equipped with extremely long horns (Zeuner 1963). Furthermore, morphological differences in fossilised horn and body shapes have prompted some archaeozoologists to classify the aurochs into three major separate subspecies – an Eurasian subspecies (*Bos primigenius primigenius*), a South Asian subspecies (*Bos primigenius namadicus*) and a North African subspecies (*Bos primigenius opisthonomus*) (Clutton-Brock 1999).

The domestic origins of Bos taurus and Bos indicus cattle – archaeological evidence

The oldest recorded evidence for the domestication of cattle comes from several Neolithic sites situated across the Near East. At Çatal Hüyük (South-central Anatolia) progressive reductions in the sizes of fossil cattle bones through successive depositional layers have been documented (Perkins 1969). These faunal remains date to approximately 8,400 YBP and are most probably representative of early domestic humpless *Bos taurus* cattle (Perkins 1969; Smith 1995).

Cattle domestication within the Near East may not have been a localised phenomenon and evidence for the domestic manipulation of wild cattle has also been reported at sites along the western fringes of the Fertile Crescent at a similar time to those situated in Southern Anatolia and Northern Syria. Indeed, some archaeologists suggest that early, yet unsuccessful, attempts to domesticate the wild aurochs were made at Ain Ghazal in the Southern Levant, *circa* 9,400-7,600 YBP (Horwitz *et al.* 1999).

It has been postulated that both *Bos taurus* and *Bos indicus* cattle were domesticated from the western Eurasian aurochs (*B .p. primigenius*) within Southwest Asia during the early Neolithic phase of the pastoralism development. Within this scenario, *Bos indicus* cattle would have been derived from initial *Bos taurus* herds by means of artificial selection during early pastoral movements across the Iranian Plateau into the Indian subcontinent *circa* 6,000 YBP (Epstein 1971; Epstein & Mason 1984; Payne 1991; Payne 1999). However, a more widely held view is that *Bos indicus* cattle were domesticated independently or subsequently to *Bos taurus* cattle from a biologically distinct wild progenitor within the Indian subcontinent (Zenuer 1963). This assertion is supported by detailed craniometric and osteological measurements which reveal a strong similarity between *Bos indicus* and the South Asiatic aurochs, *Bos primigenius namadicus*, and *Bos taurus* and the Eurasian aurochs, *Bos primigenius primigenius* (Grigson 1978; Grigson 1980). These data point to *B. p. namadicus* being the ancestor of zebu cattle and *B. p. primigenius* being the ancestor of taurine cattle. Indeed, recent archaeological surveys located within Baluchistan (situated in present day Pakistan) have provided plausible evidence for the independent domestication of *Bos indicus* cattle most probably from local *Bos primigenius namadicus* populations, *circa* 7^{th}-6^{th} millennium BC (Meadow 1993; Meadow 1996; Bökönyi 1997). Divergent ancestries of taurine and zebu cattle are further

supported by biological and genetic data, which is discussed in greater detail in **Section 1.5**.

1.3.4 Dispersal of domestic cattle from the primary centres of domestication

Introductions of domestic cattle into Europe

Bos taurus cattle are thought to have been introduced into Europe as a result of early pastoralist migrations from the Near East, some 7,500 YBP (Epstein & Mason 1984; Medjugorac *et al.* 1994; Cavalli-Sforza 1996). Taurine animals, along with other Near Eastern domesticates, such as sheep, goats, barley and wheat, entered Europe via two routes; (1) a Danubian route, in which domesticates first entered Greece *circa* 6,000 BC, and later followed the inland course of the River Danube into Southeast Europe, finally reaching the Northwest coast mainland Europe, *circa* 4,000-3,000 BC, and (2) a more rapid maritime Mediterranean coastal route, via the Dalmatian coast, Sicily, Southern France, reaching Spain and Portugal, *circa* 5,400 BC (Waterbolk 1968; Diamond 2002).

Domestic cattle migrations in Africa and the origins of African cattle

Although pictorial representations and faunal remains indicate that the earliest cattle of the African continent were humpless *Bos taurus* (Smith 1995; Hanotte *et al.* 2002), their origins are controversial and two hypotheses are currently debated. Traditionally, it is held that that domesticated cattle were introduced into Africa, via the Isthmus of Suez (present day Egypt), from the Near East approximately 7,000-6,500 YBP (Epstein & Mason 1984). From the Nile Valley, domestic cattle dispersed southwards through Sudan and westwards across the North African coast and into West Africa (Epstein & Mason 1984).

Alternatively, it has been proposed by Wendorf and Schild (1994) that domestic African cattle were derived through the separate domestication of the North African aurochs, *Bos*

primigenius opisthonomus. In support they cite faunal remains of presumed domestic cattle bones from several archaeological sites located in the Eastern Sahara that date to *circa* 9,000-8,000 YBP, some 1,500 years older than the earliest domestic cattle remains from the Nile Valley. However, the faunal remains from this region are fragmentary and some archaeologists regard their specific identification and domestic status as insecure (Clutton-Brock 1989). This postulation is further weakened by the absence of any clear evidence of domestication-induced size diminution of the recovered cattle remains (Smith 1995).

Africa currently represents a mosaic of domestic cattle types, which are classified in several groups based on their morphology. The major groups are:

(1) Indigenous African *Bos taurus* cattle, which are often regarded as the modern descendants of the initial African *Bos taurus* domesticates. These cattle pre-dominate in the subtropical regions of Western Africa, where an inherited resistance to trypanosomiasis permits them to inhabit tsetse-fly infected areas.

(2) African *Bos indicus* cattle, which inhabit the more arid regions north of the trypanosomiasis belt, which stretches across much of sub-Saharan Africa. Zebu cattle are not native to Africa and archaeological evidence shows that they were first introduced to Egypt from South Asia, some 4,000 YBP. However, zebu introductions only became widespread *circa* 700 AD as a result of Arab migrations to the continent (Epstein & Mason 1984; Marshall 1989; Payne 1991).

(3) Indigenous *Bos taurus* cattle of Central and South Africa (also referred to as sanga cattle), some populations of which display varying degrees of zebu influence (Hanotte *et al.* 2002).

The postulated migratory routes of domestic cattle from the early domestication centres, incorporating the archaeological evidence discussed above, are depicted in **Figure 1.2**. The

present-day distributions of domestic *Bos taurus* and *Bos indicus* cattle are shown in **Figure 1.3**.

Figure 1.2: Postulated migratory routes of cattle throughout the Old World

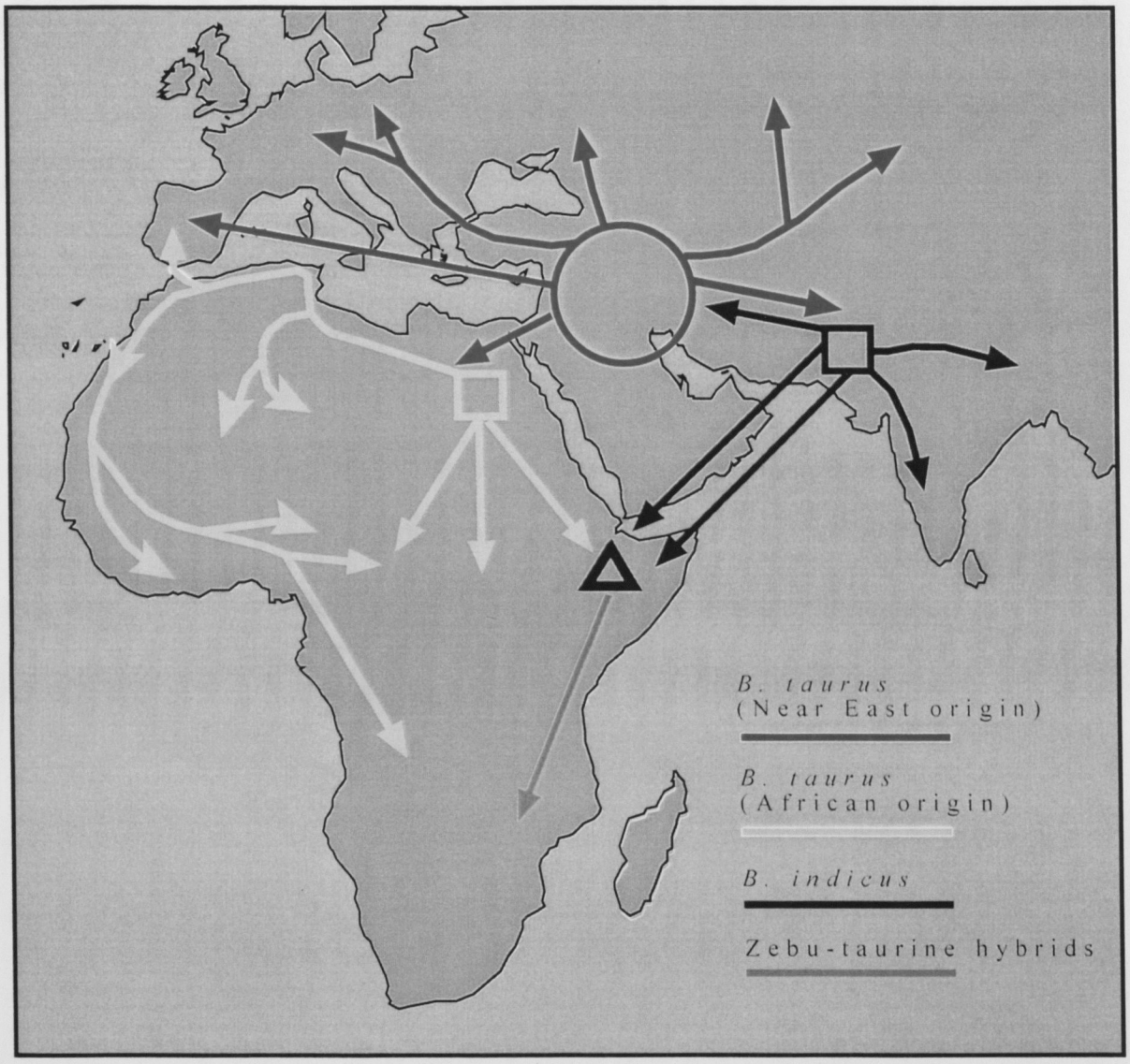
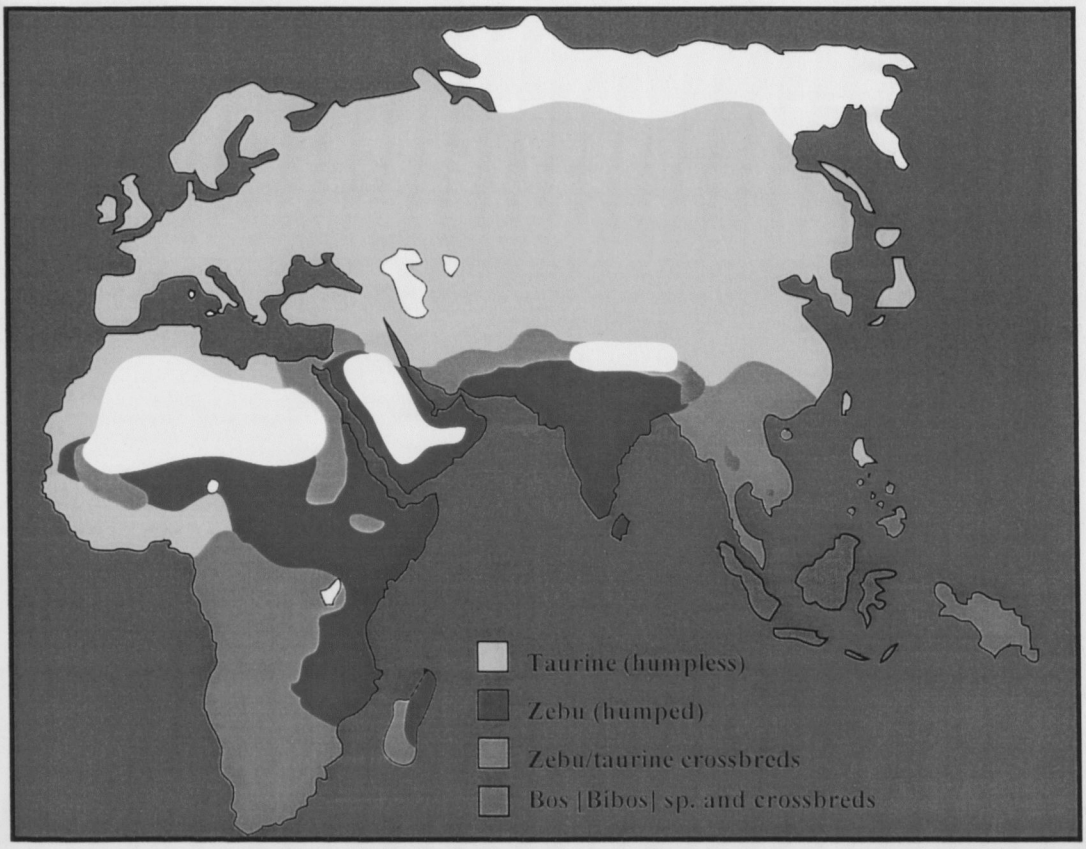


Figure 1.2. The postulated migratory routes of cattle throughout the Old World from purported centres of domestication. Archaeological evidence indicates that (1) Eurasian *Bos taurus* were domesticated within the Neolithic complexes of the Near East (red circle), (2) Indian *Bos indicus* cattle were domesticated in the early agricultural centres of Baluchistan (black square) and (3) African *Bos taurus* cattle were domesticated within sub-Saharan Africa (yellow square).

Figure 1.3: The current distribution of domestic cattle throughout the Old World



1.4 MOLECULAR PHYLOGENETICS

1.4.1 Molecular polymorphisms

Analyses of molecular variation have been illuminating evolutionary processes and phylogenetic relationships within and between species since the latter half of the 20th century. Patterns of genetic variation observed in modern populations are expected to display some continuity with past populations of similar geography. Historic demographic events such as population expansion, migrations and bottlenecks (*i.e.* major size reductions) leave characteristic imprints, in the form of altered gene frequencies, on the genome of a species which are transmitted to succeeding generations. Therefore, analysis of extant genetic variation allows aspects of the evolutionary past to be deciphered (Jorde *et al.* 1998).

There is considerable variety in the types of molecular polymorphisms currently employed in genetic surveys, each of which exhibits differential rates of evolution and modes of inheritance. Among the most extensively used markers are single nucleotide polymorphisms (SNPs), simple tandem repeat or microsatellite loci, mitochondrial DNA (mtDNA) sequence and Y-chromosomal DNA (reviewed by Avise 1994; Cavalli-Sforza 1998). In this thesis, mtDNA sequence variation, and allelic variability at several nuclear microsatellite and SNP loci, were employed to reconstruct the historical origins and demographic history of domestic cattle populations from the Old and New Worlds. Most cattle-based population genetic studies to date have involved mtDNA and nuclear microsatellite loci, the phylogenetic merits of which are discussed in **Sections 1.4.2** and **1.4.3**, respectively. The potential of SNPs in molecular investigations is detailed in **Section 4.1**.

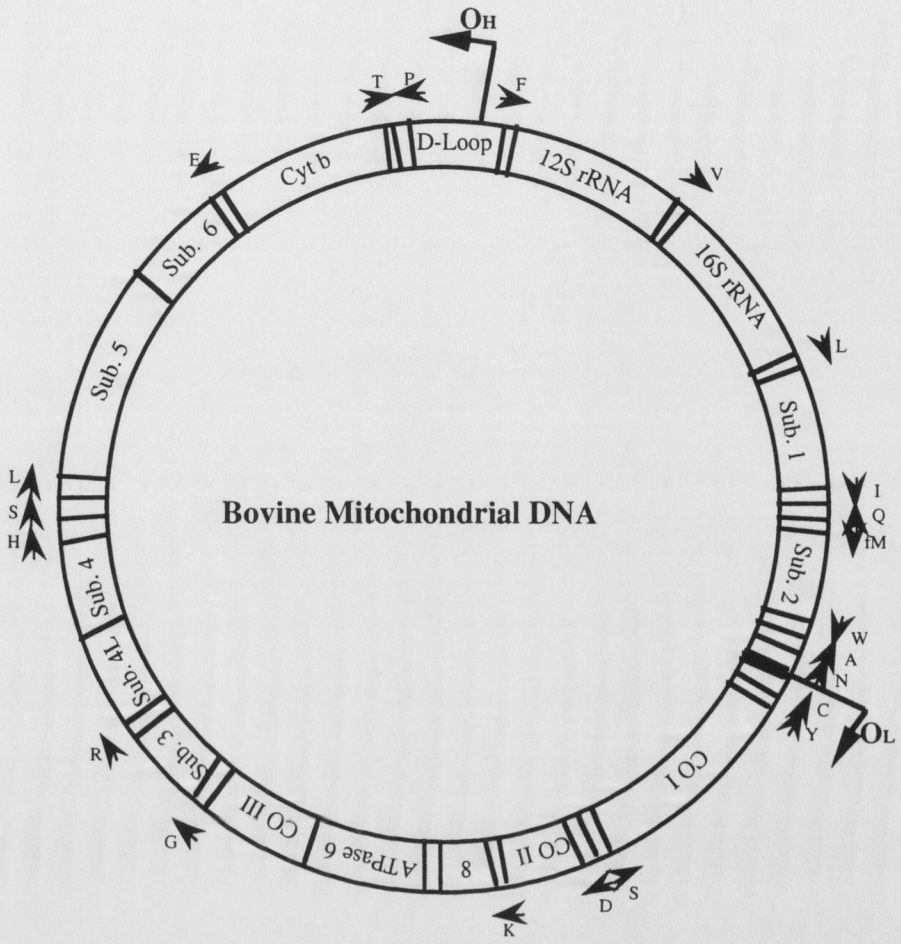
1.4.2 Mitochondrial DNA

The molecular structure and function of animal mitochondrial DNA

Mitochondria are intra-cellular organelles, which occur in most eukaryotic cells. They are primarily involved in the cellular energy-generating pathway, oxidative phosphorylation (OXPHOS), but they also contribute to the biosynthesis of numerous biomolecular substrates, such as amino acids and phospholipids (Attardi & Schatz 1988). Animal mitochondria contain multiple copies of a closed circular, duplex DNA molecule typically ranging between 15-20 kb in length. There are approximately 10^3-10^4 copies of the mitochondrial genome (mtDNA) in every animal cell, constituting less than 1% of the total cellular DNA content (Clayton 1991). Mitochondria, along with their genome, are semi-autonomous and have the capacity to replicate independently of the somatic nuclear division cycle (Ankel-Simons & Cummins 1996).

The animal mitochondrial genome is typically composed of 37 genes encoding 22 transfer RNA molecules (tRNA), 2 ribosomal RNA molecules (12S and 16S rRNAs) and 13 messenger RNA molecules (mRNA) specifying proteins involved in the cellular respiration (Wallace 1994). The remaining proteins necessary for the metabolic functions of the mitochondrion and for the replication, transcription and translation of the mitochondrial genome are encoded by the nuclear genome and are imported from the cytoplasm (Stoneking & Wilson 1989). Mitochondrial genes are tightly packed within the genome and non-coding DNA stretches such as introns, repetitive DNA sequences, pseudogenes and large spacer sequences are rare or absent. The largest non-coding region of the mitochondrial genome, the control region or D-loop, measures approximately 1 kb in length and contains the nucleotide sequences essential for the initiation of replication and transcription (**Figure 1.4**).

Figure 1.4: A diagrammatic representation of the bovine mitochondrial genome



Mitochondrial DNA as a phylogenetic marker

Surveys of mtDNA variation continue to propel molecular phylogeographic surveys in a wide range of species (Avise 1998; Avise 2000). This is due to several distinguishing, advantageous properties, which have resulted in the extensive use of the molecule for inferring evolutionary processes and phylogeny. These are discussed in detail below:

(1) Ease of accessibility

The cytoplasmic location and high copy number of mitochondrial chromosomes within each eukaryotic cell (10^3-10^4 molecules per cell) facilitates the retrieval of mtDNA from difficult substrates. This is one reason, for example, why most ancient DNA investigations have focused primarily on mtDNA.

(2) Rapid rate of evolution

Mammalian mtDNA is highly polymorphic due to a rapid rate of evolution, which is approximately 5-10 times that of nuclear DNA (Brown *et al.* 1979). This represents an average evolution rate across the whole genome of 2-4% per MYA (or 2-4 substitutions per 100 nucleotides per MYA). Consequently, mtDNA is extremely useful for phylogenetic comparisons involving closely related species or populations of the same species, as a sufficient number of substitutions will have accrued over short periods of evolutionary time, such as that associated with the time frame of animal domestication (Stoneking 1996).

Evolutionary rates, however, are not uniform across the molecule with some regions more susceptible to nucleotide substitutions than others. This is particularly evident within the control region, which according to human data mutates some 5-10 times faster than coding mtDNA sequences (Vigilant *et al.* 1991; Richards & Macaulay 2001). Furthermore, human data have revealed considerable variation in the rate of evolution between nucleotide sites situated within the control-region. For example, hypervariable sites, which are hot spots for mutation, tend to mutate faster than other sites within the control region (Stoneking 2000).

Each of the 13 mtDNA-encoded polypeptides exhibits their own characteristic evolutionary rates, which vary across different vertebrate species. Both tRNA and rRNA genes display the lowest evolutionary rates of all mitochondrial gene sequences, yet they evolve approximately 100 times faster than their nuclear counterparts (Brown 1985). Salient features of the accelerated rate of mtDNA sequence evolution include an exceptionally high proportion of silent (> 90%) versus replacement substitutions (<10%) and a higher transition (purine-purine or pyrimidine-pyrimidine) to transversion (purine-pyrimidine) ratio (of the order of 10:1) (Gray 1989).

The evolutionary rate of DNA sequences is a function of the rate of nucleotide substitution and the rate at which new nucleotide variants become fixed (Wilson *et al.* 1985). Both processes occur at elevated rates for mtDNA and two general explanations have been proposed to account for this. Firstly, it has been suggested that mtDNA can tolerate a greater number of mutations in tRNA and rRNA genes due to relaxed functional constraints on the mitochondrial translational apparatus (Wilson *et al.* 1985). Secondly, an enhanced mtDNA evolution rate may be attributable to an inefficient DNA repair mechanism that is coupled with a high incidence of oxidative damage due to the generation of oxygen free radicals during cellular respiration (Avise 1991).

(3) Intra-individual homoplasmy

All mtDNA molecules within an individual are thought to be identical, a status known as homoplasmy. However, homoplasmy is not absolute as a low frequency of mtDNA variants continually arise within somatic and germline cells. For germline cells, intra-individual mtDNA sequence homogeneity is believed to be maintained by a bottlenecking process at the pre-implantation stage of embryogenesis. This process greatly reduces the number of founding mtDNA molecules and ensures mtDNA homoplasmy within the developing embryo (Chinnery *et al.* 2000). For somatic mtDNA mutations, the number of mutant mitochondrial molecules that are transmitted to daughter cells are subject to random genetic drift which eventually leads to fixation of the mutant or normal molecule after many rounds of cell replication, hence re-establishing homoplasmy (Wallace 1994; Bertranpetit *et al.* 1996). In conclusion, Avise (2000) states that for phylogenetic purposes there is a close approximation to homoplasmy for somatic cell mtDNA molecules within most individuals.

(4) Maternal inheritance and lack of recombination

It is widely accepted that mtDNA is inherited exclusively along the maternal lineage with inconsequential paternal contribution. Thus, mtDNA lineages are regarded as clonal, information-rich, singly inherited units or haplotypes, with which matriarchal evolutionary histories can be reconstructed without the complication of bi-parental recombination (Avise 1994). However, within the past decade several authors have presented molecular evidence suggesting that the inheritance of mtDNA is not clonal, thus, challenging previous evolutionary inferences based on mtDNA sequence data.

There are at least three possible pathways by which the clonality of mtDNA can be disrupted: (1) recombination between normal and mutant mtDNA molecules in a heteroplasmic individual, (2) recombination between mtDNA molecules and mtDNA pseudogene sequences encoded by the nuclear genome and (3) recombination between maternally- and paternally- derived chromosomes as a result of paternal leakage (Eyre-Walker & Awadalla 2001). The latter provides the most likely route for mtDNA recombination as it is now known that paternally-derived mitochondria do enter the oocyte during fertilisation in humans (Ankel-Simons & Cummins 1996) and in other mammalian species (Gyllensten *et al.* 1991). Although paternal mitochondria are eliminated shortly after zygote formation, presumably by an intra-specific recognition and destruction mechanism (Stoneking & Soodyall 1996), it is probable that these mechanisms are not infallible (Eyre-Walker & Awadalla 2001). Furthermore, it has been demonstrated that mitochondria possess the enzymes necessary for homologous intra-molecular recombination (Thyagarajan *et al.* 1996).

Initial studies suggesting mtDNA recombination argued that the high proportion of sequence homoplasies (*i.e.* the occurrence of the same mutation at the same nucleotide site

in different mtDNA lineages which have been derived independently of each other) within human mtDNA could not be accounted for solely by mutation and that the only remaining possible explanation is recombination (Eyre-Walker *et al.* 1999). However, Macaulay *et al.* (1999) have shown that some of the data on which this conclusion was based are insecure and if excluded from analysis the findings of Eyre-Walker *et al.* (1999) are no longer valid.

Secondly, Awadalla *et al.* (1999) illustrated that linkage disequilibrium (LD, the non-random association of allelic variants on chromosomal regions within a population) decreases as a function of physical distance between nucleotide sites in human mtDNA data sets, a phenomenon typically observed in the recombining nuclear genome. This analysis has been criticised by several authors on a number of grounds, including poor data quality, the accuracy of the choice of LD measure and level of consistency of results, who deduce that the reconsideration of previous mtDNA phylogenies is not required within the human data (Parsons & Irwin 2000; Kivisild & Villems 2000; Jorde & Bamshad 2000; Kumar *et al.* 2000). Indeed, this finding is now accepted by some of the original proponents of mitochondrial recombination (Eyre-Walker & Awadalla 2001) and Stoneking (2000) states that the high number of mtDNA control region homoplasies are most likely the result of hypervariable mutational hot spots.

The apparent lack of recombination and maternal inheritance of mtDNA ensures that mutation is the only source of mtDNA sequence variation, which is the most clock-like evolutionary process (Stoneking 1996). Hence, divergence times between mtDNA sequences can be calibrated by means of a molecular clock (Avise 2000).

(5) MtDNA as a sensitive indicator of demographic events

As mtDNA is haploid and inherited in a uni-parental fashion, it has an effective population size [*i.e.* the theoretical number of breeding individuals in a population (N_e)] one-quarter that of nuclear DNA, which is diploid and inherited bi-parentally. Consequently, population differentiation due to genetic drift tends to proceed more rapidly for mtDNA than for nuclear genetic elements (Jorde *et al.* 2000). Furthermore, the lower N_e of mtDNA implies that it is a more sensitive indicator of population dynamics than nuclear DNA. For example, population bottlenecks can dramatically reduce mtDNA diversity within a population, while having little effect on nuclear variability (Wilson *et al.* 1985).

Theoretical considerations for mtDNA phylogenetic studies

Despite its advantageous characteristics, phylogenetic reconstruction using mtDNA is not without complication. Firstly, most mtDNA-based intra-specific phylogenetic studies involve the analysis of the sequence variation contained within the non-coding control region. However, hypervariability within the control region suggests that the rate of evolution is not constant at every nucleotide site, thus confounding estimates of mtDNA lineage divergence. Conventional approaches calibrate molecular clocks by reference to the quantitative divergence between the species under consideration and a closely related outgroup species, for example the divergence between humans and chimpanzees (Vigilant *et al.* 1991) and cattle and bison (Bradley *et al.* 1996). However, over such long evolutionary periods, hypervariable sites within the mtDNA control region can revert back to their original unmutated state, resulting in an underestimation of the long-term mtDNA mutation rate, also referred to as the phylogenetic rate (Gibbons 1998; Jorde *et al.* 1998). Alternatively, control region mutation rates estimated over short evolutionary periods, such as the pedigree rates of evolution obtained from the analysis of mutation patterns in human familial pedigree studies (Howell *et al.* 1996), tend to be based primarily on hypervariable

sites and can often lead to an overestimation of mtDNA control region mutation rates (Gibbons 1998). Although estimates from human data suggest that the phylogenetic rate (which approximates to 1 substitution every 10,000 years) is markedly slower than the pedigree rate (approximately 1 substitution every 800 years), Macaulay *et al.* (1997) suggest that the phylogenetic rate is more consistent with evolutionary events dating to 1,000 years or greater.

Secondly, hypervariability also generates a large number of homoplasies within the control region and as a result many phylogenetic topologies are equally plausible, rendering the selection of the most parsimonous phylogenetic tree difficult (Disotell 1999; Richards & Macaulay 2001). To counter such dilemmas, alternate phylogenetic approaches have been developed, including statistic evaluations of phylogenetic tree landscape (Penny *et al.* 1995) and the construction of phylogenetic networks (Bandelt *et al.* 1995), which resolve homoplasies as separate mutational events.

Thirdly, it must be emphasised that phylogenetic inferences drawn from mtDNA data are based on one type of genetic system. Alternative evolutionary scenarios may be deduced by analysing variation at different genetic loci, such as nuclear diversity and Y-chromosomal and X-chromosomal polymorphisms (Jorde *et al.* 2001). Additionally, genetic reviews continually highlight the potential of mtDNA to invoke past demographic events, particularly episodes of population growth, which are often characterised star-type phylogenies in which derivative haplotypes radiate from a centrally-positioned ancestral variant (Jorde *et al.* 1998; Jorde *et al.* 2001). However, similar patterns of diversity can be produced by natural selection, in which the frequency of a favourable variant increases until it is carried by all members of an ancestral population and consequently, all modern mtDNA chromosomes are descendants of the selected variant (Harpending & Rogers

2000). Caution must therefore be exercised when interpreting putative demographic events from mtDNA data.

1.4.3 Microsatellite genetic markers

Microsatellite loci: a definition

Microsatellite or simple tandem repeat (STR) markers consist of tandemly arranged reiterated units of non-coding DNA sequence, typically ranging between 2-5 bp in length, which are ubiquitously distributed throughout eukaryotic genomes. These genetic elements are often highly polymorphic due to variation in the number of repeat units, which account for a multitude of distinguishable alleles that are inherited in a Mendelian fashion (Bruford & Wayne 1993). Microsatellite marker variation is presumed to be neutral, *i.e.* without the complication of natural selection unless closely linked to a selected coding region. However, as microsatellite repeats are rarely longer than 30 units, it appears that they are under selective pressure with regards to allele length (Jarne & Lagoda 1996).

Mutational dynamics of microsatellites

Microsatellite loci are fast-evolving genetic markers, with estimated evolutionary rates range between 10^{-5} and 10^{-2} mutations per generation (Rosenbaum & Deinard 1998). It is generally agreed that novel microsatellite polymorphisms are generated primarily by intra-allelic DNA strand slippage during DNA replication. Under this model, repeat units on the nascent and template DNA strands anneal out of register, resulting in either the expansion or contraction of the locus on the newly derived daughter strand. Normally, this length difference is corrected by the DNA mismatch repair system, however, it seems probable that naturally occurring microsatellite mutations escape this correction process (Tautz & Schlötterer 1994).

Application of microsatellite markers to molecular phylogenetics

The high level of heterozygosity, Mendelian inheritance and ease of genotyping, via polymerase chain reaction (PCR) amplification, has made microsatellites an attractive marker for molecular phylogenetic studies (Rosenbaum & Deinard 1998). However, technical and theoretical considerations must be applied when applying microsatellite data to molecular population genetic studies. For example, null alleles (*i.e.* non-amplifying alleles due to incorrect annealing of primers in flanking DNA regions) can result in an apparent deficiency of heterozygotes, leading to an underestimation of intra-population variability (Schlötterer & Pemberton 1998). Also, the evolutionary dynamics of microsatellite loci can result in allelic homoplasy due to convergent evolution such that alleles of identical size may be identical only in state rather than by descent. Microsatellite allelic homoplasy can underestimate the actual divergence between populations, hence complicating phylogenetic reconstructions (Chambers & MacAvoy 2000; Ritz *et al.* 2000).

1.5 GENETIC VARIATION AND DOMESTIC CATTLE ORIGINS

1.5.1 Mitochondrial DNA variation and domestic cattle origins

The independent domestications of Bos taurus and Bos indicus

The origins of domestic cattle have been a long-standing subject of study, which has invoked contributions from several disciplines. However, some of the major events surrounding bovine prehistory have only recently been elucidated with the maturation of molecular phylogenetics within the past decade. In this regard, mtDNA has represented the most extensively employed genetic element for addressing domestic livestock origins.

As mentioned previously, mtDNA phylogenetic reconstructions are based solely on a single segregating locus and are therefore not necessarily representative of the ancestry of the whole genome. However, in the context of animal domestication this prerequisite is actually advantageous. The matriarchal transmission of mtDNA implies that all domestic mtDNA lineages must at some time in their evolutionary history have been adopted into the domestic pool through the capture and taming of wild females. Alternatively, nuclear gene histories may have been complicated by matings (ancient or recent) between wild males and domesticated female animals (MacHugh & Bradley 2001). Furthermore, under managed domesticate breeding systems there is a high disparity between male and female reproductive variance, in which genetic exchange between local domestic livestock populations is predominantly male-mediated. Thus, maternal lineages are likely to show some geographical inertia, which can facilitate detailed phylogenetic analysis and comparison of information-rich mitochondrial haplotypes between geographical regions (Bradley 2000).

Early phylogenetic investigations involving domestic cattle (Loftus *et al.* 1994a; Loftus *et al.* 1994b) demonstrated a duality in mtDNA control region sequences in which haplotypes

invariably clustered into one of two groups (bovine mtDNA sequence duality, together with mtDNA phylogenies constructed from other domestic livestock species, is illustrated in **Figure 1.6**). A dominant internal branch separates the two distinct mtDNA sequence groups producing a double-headed broomstick-like phylogeny. Furthermore, these clades are geographically distributed along an east-west axis, with the division clearly following previous morphologically-established taxanomic classifications, namely that between *Bos taurus* and *Bos indicus*. A molecular clock-calibrated time depth of the divergence between the two major clades yielded a minimum estimate of 275,000 YBP for the common ancestry of the two domestic *Bos* lineages (Bradley *et al.* 1996).

These data are clearly inconsistent with a common domestic origin for all cattle *circa* 10,000 YBP. To invoke a single common origin for domestic cattle requires either an assumption of a highly erratic molecular clock or the incorporation of two highly divergent lineages into the domestic pool that have been maintained at appreciable frequencies throughout the 10,000 year history of domestication. Although the former explanation is a remote possibility due to the high evolutionary rate of the mitochondrial genome, the apparent lack of intermediate branching patterns and tight sequence clustering at the extremities of the topologies implies that fast divergence of mtDNA does not satisfactorily account for the observed lineage duality. Equally, it is plausible that a single domestication event resulted in the adoption of diverse matrilines into the domestic mitochondrial pool. However, it is a laboured interpretation to assert that the mtDNA sequence clades became geographically partitioned and fixed separately into western and eastern populations over a 10,000-year period.

Instead the patterns of bovine mtDNA diversity are suggestive of the domestication of both *Bos taurus* and *Bos indicus* cattle from two genetically distinct, geographically disparate

wild progenitor populations. Notably, the amount of diversity within each clade is of a similar magnitude and is probably derived from the attenuated sampling of wild variation associated with localised domestications. Furthermore, the extant diversity also constitutes the limited variability accrued through mutation within each clade throughout domestic history (MacHugh & Bradley 2001).

1.5.2 Mitochondrial DNA variation and *Bos taurus* ancestry

Analysis of *Bos taurus* mtDNA control region sequence data have revealed a strong phylogeographical structuring of mtDNA haplotypes throughout Eurasia and Africa. Troy *et al.* (2001) demonstrated that *Bos taurus* mtDNA haplotypes in Europe are composed primarily of a single haplogroup (designated T3), which is one of three encountered at appreciable frequencies within the Near East (namely T, T2 and T3). This finding was interpreted as evidence for the adoption of at least three distinct, yet closely related, wild matrilines into the domestic pool within the Near Eastern complex during the phase of aurochs capture and taming. Furthermore, the predominance of the T3 haplogroup in Europe, which is a subset of the variation encountered in the Near East, is suggestive of a derived Near Eastern origin for the cattle of Europe.

Despite substantial historic introgression of zebu cattle from South Asia, all African *Bos taurus* and African *Bos indicus* cattle possess similar *Bos taurus* mtDNA sequences, which are believed to be descendent from the initial founding taurine cattle of the continent (Troy 1998; MacHugh *et al.* 1997; Troy *et al.* 2001). African mtDNA sequences form a distinct *Bos taurus* haplogroup (T1) rarely encountered elsewhere (Troy 1998; Troy *et al.* 2001). The near absence of this haplogroup from the Near East provides some support for archaeological suppositions of a separate indigenous domestication of African cattle from a local wild progenitor that was biologically similar to the ancestor of European cattle

(Wendorf & Schild 1994). Mannen *et al.* (1998) have additionally identified a *Bos taurus* mtDNA haplotypic lineage exclusive to samples of North East Asian origin, which may represent a local adoption of wild matriarchal lineages into the domestic pool.

1.5.3 Microsatellite variation and the origins of African cattle

Consistent with mtDNA-based phylogenies, nuclear microsatellite diversity reveals the deepest phylogenetic division between *Bos taurus* and *Bos indicus*, with calibrated estimates of divergence time between 610,000 and 850,000 YBP. This figure is approximately three times greater than that estimated from mtDNA diversity and reflects the inherent errors and difficulty associated with estimating divergence time from biological data (MacHugh *et al.* 1997). Microsatellite data also support a phylogenetic distinction between African and European *Bos taurus* cattle, with an estimated divergence time of 150,000-250,000 YBP (MacHugh *et al.* 1997).

Comparison of mtDNA and nuclear microsatellite phylogenies illustrates an ambiguity in the positioning of African zebu populations (**Figure 1.5**). As mentioned previously, all surveyed African *Bos indicus* cattle exclusively possess *Bos taurus* T1 haplogroup sequences and hence cluster with African *Bos taurus* cattle in the mtDNA-derived phylogeny. In contrast, microsatellite-based analyses show that African *Bos indicus* group with Indian *Bos indicus*, suggesting a hybrid origin for African *Bos indicus* cattle. Genetic analysis indicate that African *Bos indicus* populations have pre-dominantly zebu autosomal genomes which have been derived from introgressing *Bos indicus* males from the Indian subcontinent, primarily over the past 1,500 years (MacHugh *et al.* 1997; Bradley *et al.* 1998; Loftus *et al.* 1999; Hanotte *et al.* 2002). A recent survey of autosomal variation in sub-Saharan populations points towards the Horn of Africa as the focal point for zebu introductions into Africa, which arrived via the Arabian Sea (Hanotte *et al.* 2002). From

the Horn, zebu cattle dispersed and introgressed with native taurine herds, resulting in an east-west cline of zebu admixture across Africa (MacHugh *et al.* 1997; Hanotte *et al.* 2002).

Similarly, studies of paternally-inherited Y-chromosome polymorphisms have mirrored the findings of autosomal data, revealing the highest proportion of zebu-specific alleles in East African breeds, particularly those from the Lake Victoria region (Hanotte *et al.* 2000). Collectively, nuclear and Y-chromosomal data, together with apparent absence of zebu mtDNA haplotypes in Africa *Bos indicus* cattle suggest that zebu introgression in Africa was pre-dominantly male-mediated (MacHugh *et al.* 1997; Bradley *et al.* 1998).

Figure 1.5: Domestic cattle mtDNA- and microsatellite-derived phylogenies

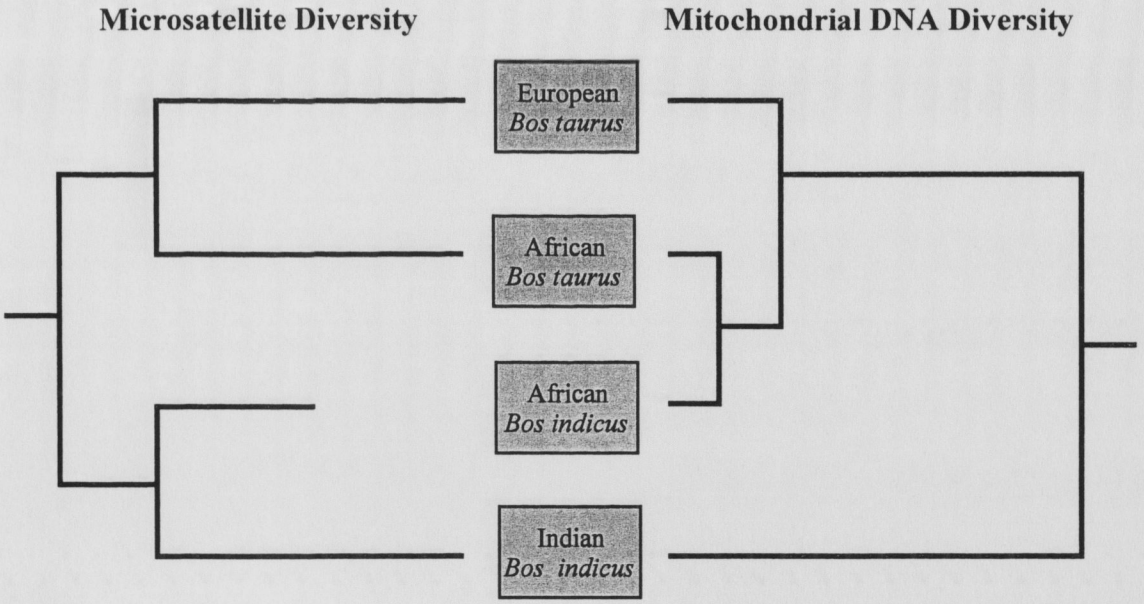


Figure 1.5. Summary of phylogenetic tree topologies derived from microsatellite and mitochondrial DNA diversity (reprinted from Bradley *et al.* 1998). Both data collections reveal the deepest phylogenetic division between *Bos taurus* and *Bos indicus* cattle. However, the positioning of the African *Bos indicus* is ambiguous. For microsatellites, they cluster with Indian *Bos indicus* mtDNA and for mtDNA they cluster with African *Bos taurus*. This suggests that African *Bos indicus* are hybrid animals with the majority of their nuclear genome derived from introgressing *Bos indicus* cattle, while their mitochondrial genomes are representative of the original *Bos taurus* domesticates of the continent. Furthermore, both phylogenies show divergence between European and African *Bos taurus*, suggesting that they also have divergent ancestries.

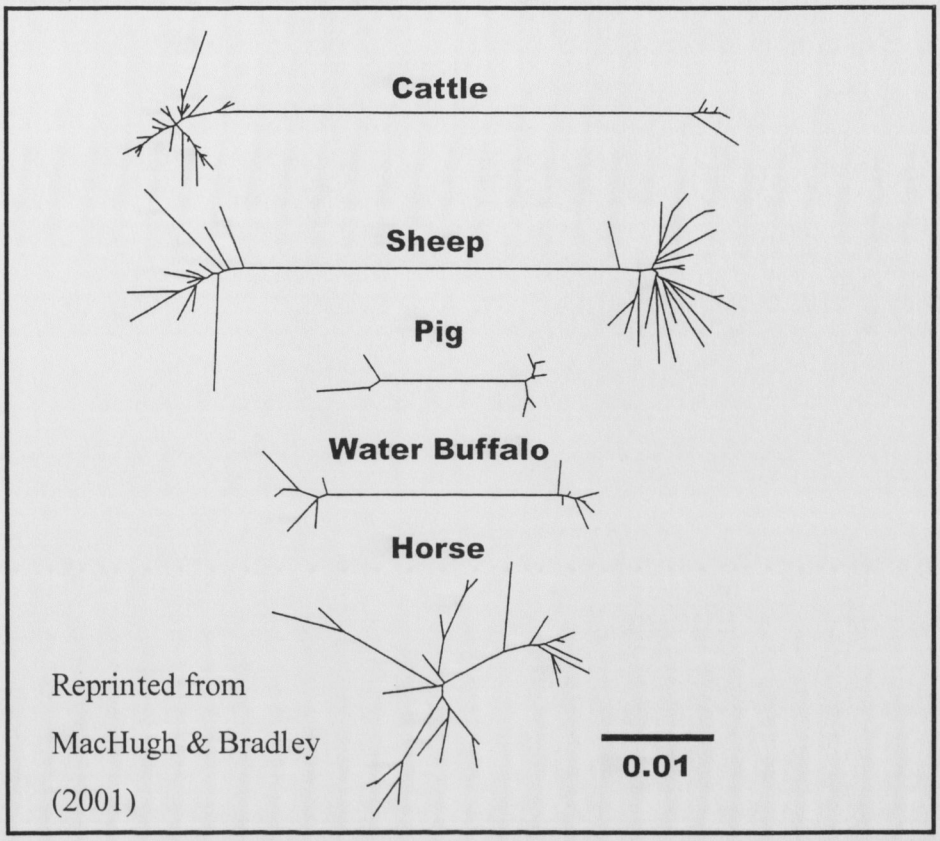
1.5.4 Mitochondrial phylogenies in other domestic species

As with bovine mtDNA phylogenetic patterns, duality has been identified in domestic sheep (Hiendleder *et al.* 1998; Hiendleder *et al.* 2002), pig (Giuffra *et al.* 2000) and water buffalo (Tanaka *et al.* 1996; Lau *et al.* 1998) mtDNA lineages (**Figure 1.6**). Notably, the sequence clades identified in these species are spatially-assorted along an east-west geographic division. For water buffalo, this divide follows the morphological distinction between river buffalo and swamp buffalo and in pigs this divide occurs between European and Asian samples. Although the phylogeographic pattern in sheep is not as clear as those for water buffalo, pigs and cattle, one clade was found in European samples and the other was detected in samples of European and Asian provenance. Furthermore, calibrations of time depths corresponding to the divergences between these domestic mtDNA sequence clades have been estimated at several hundred thousand YBP. The quantitative divergence together with the geographic partitioning of mtDNA sequence clades provide strong support for at least two domestications of sheep, pig and water buffalo from biologically distinct wild progenitor populations (MacHugh & Bradley 2001).

In contrast, the emerging mtDNA phylogeographic distributions presented for domestic goats suggest at least three geographically and temporally disparate captures of wild female progenitors during the formation of early domestic herds (Luikart *et al.* 2001). Also, the more complex patterns of mtDNA diversity within domestic horse populations are indicative of numerous widespread integrations of wild matrilines associated extensive domestication events along the Eurasian steppes (Vilà *et al.* 2001).

The approximate locations of suggested domestication centres throughout the Old World for sheep, pig, cattle, goats, horses and water buffalo are shown in **Figure 1.7**.

Figure 1.6: Unrooted neighbor-joining mtDNA phylogenies for six Eurasian domestic ungulate species



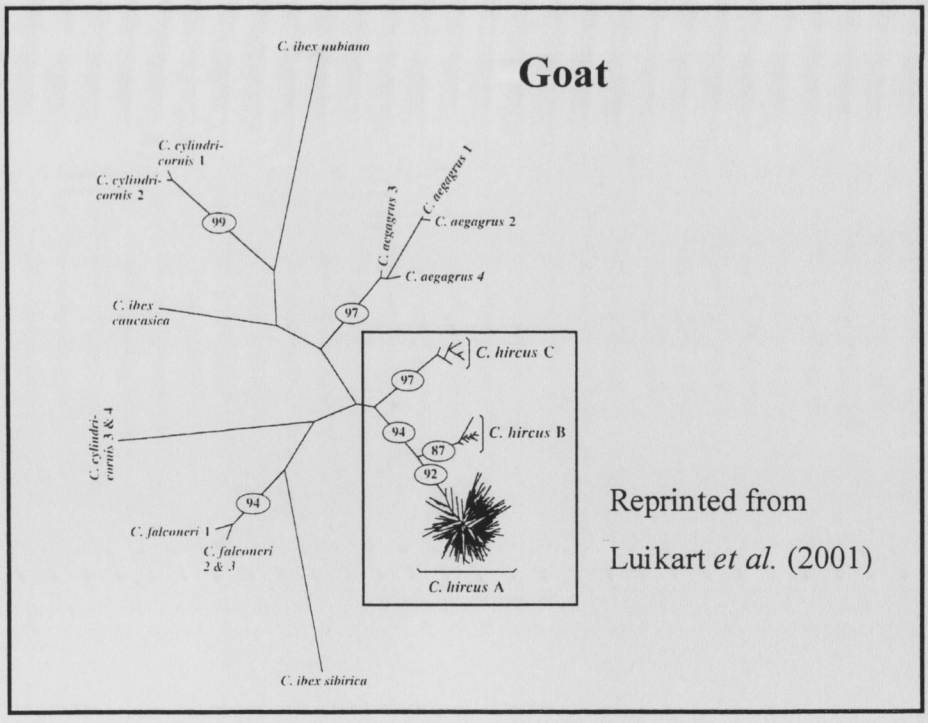


Figure 1.6. MtDNA sequence neighbor-joining phylogenies for six domestic ungulate species. A duality in mtDNA diversity is evident in cattle, sheep, pig and water buffalo sequences. Three domestic mtDNA lineages are observed in goat, denoted A, B and C (inset). Horse samples display multi-divergent matrilines.

Figure 1.7: The location of suggested livestock domestication centres throughout the Old World

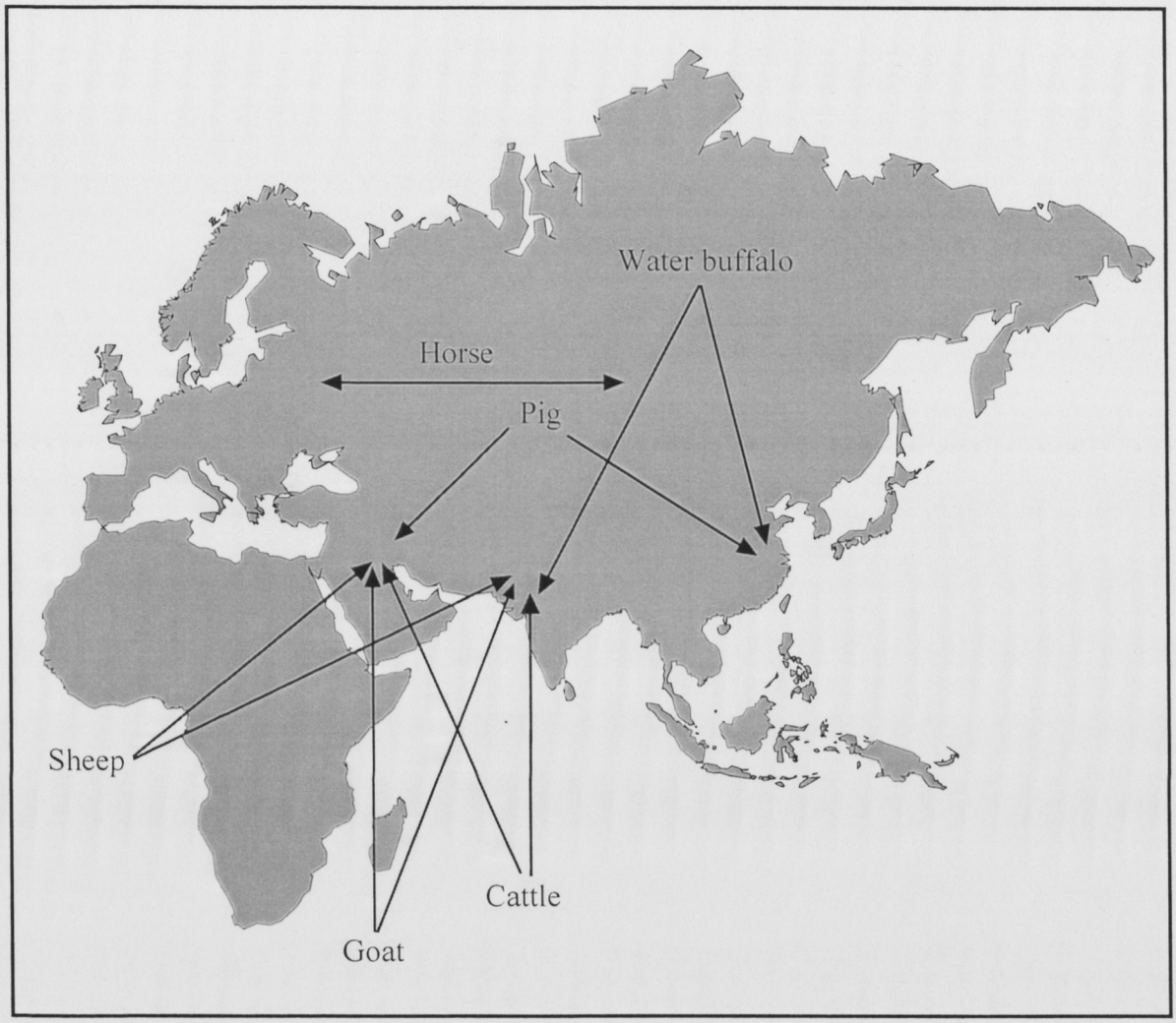


Figure 1.7: Possible domestication centres of the major Eurasian domestic ungulate species. Possible domestication centres for sheep, cattle and goats are the Near East and the Indian subcontinent; evidence for pig domestication comes from the Near East and they may have also been domesticated in South China; water buffalo were probably domesticated in the Indian subcontinent and the Yangtse Valley in China; horses are believed to have been domesticated along the Eurasian steppes.

Based on an accumulation of genetic and archaeological data, it appears that the theme of multiple domestications is common in the origins of domestic livestock. These data also suggest that domestication was not restricted to peoples inhabiting the Neolithic complexes of the Near East. It is well-documented the wild ancestors of most of the major Old World domestic animals had large geographic ranges, covering vast tracts of Eurasia, Eastern Asia and in the case of wild cattle, possibly North Africa (Meadow 1993; Bellwood 1996; Clutton-Brock 1999). Thus, these animals were widely available for domestication in different regions, *circa* 10,000 YBP.

It is now increasingly accepted that there were five primary centres of domestication across the globe at which crop cultivation occurred independently - the Near East, Eastern Asia, Mesoamerica, South America and North America (Diamond 2002). It possible that these primary heartlands covered large geographical regions and contained multiple secondary domestication centres. Although geographically separated, domestication at these secondary sites may have been triggered by the arrival of farmers from elsewhere, which suggests that they were culturally linked. This hypothesis seems particularly appropriate for the purported domestication centres situated within the Indian subcontinent, which Meadow (1993) believes were on the receiving end of the cultural and technological innovations developed within the Near East. Thus, livestock domestication within the Indian subcontinent may actually be viewed as part of a larger pattern of agro-pastoralist innovations beginning in the Near East and which later diffused eastwards prompting similar, somewhat later developments further east (Meadow 1996; Diamond 2002).

CHAPTER 2

MITOCHONDRIAL DNA PHYLOGEOGRAPHY OF DOMESTIC *BOS INDICUS* CATTLE

2.1 INTRODUCTION

2.1.1 Agricultural origins within South Asia

The emergence of agro-pastoralism in South Asia is believed to have occurred some 9,000-8,000 YBP, with the earliest documented evidence originating from Baluchistan, a region situated in present day Pakistan, to the west of the Indus River. Lying on the easternmost periphery of the Middle East, this region has often been regarded as being on the receiving end of the technological and cultural innovations developed in Neolithic Southwest Asia, particularly animal rearing and crop cultivation (Meadow 1993). Unlike Europe, where most domestic species were introduced to the continent from the Near East as a result of pastoralist migrations (Blumler 1996), Baluchistan occurs within the natural range of the wild ancestors of many of the domestic species that dominated most of the South Asian Neolithic economies. Consequently, this region may have served as a centre of agricultural innovation that was substantially influenced by cultural developments from more advanced areas to the west (Meadow 1993).

The archaeological site of Mehrgarh, situated within the North Kachi Plain of Baluchistan, currently provides the earliest evidence for the development of agriculture in South Asia (Jarrige & Meadow 1980). The North Kachi Plain lies within the vast drainage system of the Indus River and is characterised agriculturally by well-irrigated, mineral-rich alluvial soil deposits (Allchin & Allchin 1982; Meadow 1993). Palaeobotanical evidence from Mehrgarh suggests that the North Kachi Plain was potentially a centre of crop cultivation, particularly barley (*Hordeum vulgare* sp.) (Jarrige & Meadow 1980). Ancient remains of other floral domesticates such as einkorn and emmer wheat, have also been uncovered at Mehrgarh. However, as no wild forms of wheat have been uncovered within Baluchistan, it seems plausible that these domestic forms were introduced from the early cultivation centres of the Near East. The appearance of Near Eastern-derived subsistence species in

the archaeobotanical record of the Indian subcontinent, *circa* 9,000-8,000 YBP strengthens claims that the Indian subcontinent was a secondary centre of agricultural development in which local domestication was triggered by the arrival of agro-pastoralists from the Near East.

2.1.2 The origins of *Bos indicus* cattle

In addition to the emergence of crop cultivation, excavations at Mehrgarh (**Figure 2.1**) also provide the earliest known evidence for the use of domestic animals in South Asia (Jarrige & Meadow 1980; Allchin & Allchin 1982; Meadow 1993). The earliest bone assemblages of this site (*circa* 6,300 – 5,300 BC) display the remains of numerous wild animals including gazelles (*Gazella bennetti*), wild sheep (*Ovis orientalis*), wild goats (*Capra aegagrus*), wild deer (*Cervus duvauceli*), wild boar (*Sus scrofa*), wild water buffalo (*Bubalus arnee*) and wild cattle (most probably *Bos primigenius namadicus*), all of which were presumably hunted for food, skins and fibre (Meadow 1996).

The presence of such wild fauna at Mehrgarh, particularly wild sheep, goat, cattle and buffalo, suggests that they comprised the potential ancestral stock for South Asian domestic livestock (Meadow 1993; Misra 2001). In support of this view, marked shifts from faunal assemblages dominated by the bones of wild animals to assemblages characterised almost exclusively by bones of domesticated ungulates have been reported within the early Neolithic depositional layers of Mehrgarh (Meadow 1993). The earliest of these domestic faunal remains date to the 7^{th}-6^{th} millennium BC, indicating that the development of animal husbandry in the North Kachi Plain was somewhat contemporaneous with similar developments in Southwest Asia (Jarrige & Meadow 1980).

Wild animal remains constitute more than 50% of the total bone assemblages excavated from the earliest levels of Mehrgarh and this number greatly declines in the successive strata (Meadow 1984). In particular, the representation of cattle remains rises dramatically from approximately 4% in the earliest levels of Mehrgarh to almost 65% in assemblages dating to *circa* 6,000-5,000 BC (Meadow 1993). Moreover, these later skeletal remains display a marked decrease in size when compared to their wild counterparts. The trend towards size diminution coincident with increasing representation within the fossil record of Mehrgarh provides a strong basis for the local domestication of cattle, with several authors agreeing that the resulting domestic form was most probably *Bos indicus*, the humped taxa of South Asia (Jarrige & Meadow 1980; Meadow 1993; Meadow 1996; Bökönyi 1997). Additional support also comes from artistic evidence, which shows that the earliest terracotta figurines from Baluchistan depict humped cattle, which are physically similar to modern zebu breeds (Allchin & Allchin 1968).

The excavations from Mehrgarh strongly suggest that *Bos indicus* cattle were domesticated from a wild progenitor that was morphologically and geographically distinct from the purported wild ancestor of *Bos taurus* cattle, *B. p. primigenius*. This assertion is supported by the osteological analyses performed by Grigson (1980), which have revealed anatomical similarities between the skeletal remains of *Bos indicus* and those of the South Asian aurochs, *B. p. namadicus*. Furthermore, archaeological data indicate that wild *B. p. namadicus* occupied the dry lands of the Indian subcontinent during the late Pleistocene-early Holocene boundary and was therefore available for domestication by early farming communities (Grigson 1985; Bökönyi 1997). Thus, it seems likely that *Bos indicus* cattle are the modern descendants of locally domesticated *B. p. namadicus* animals (Zeuner 1963; Allchin & Allchin 1974; Naik 1978; Meadow 1984).

Archaeological data also indicate that the geographic distribution of *B. p. namadicus* was not restricted to the Indian subcontinent and may have extended westwards to the Great Salt and Luc Deserts of Eastern Iran (Epstein 1971; Bökönyi 1997). Indeed, horn cores believed to be those of *B. p. namadicus* together with *Bos indicus* figurines, dating to the 4th and 3rd millennia BC, have been recovered from Shahr-I-Sokhta (Sistan, Iran), providing some evidence for the local domestication of zebu along the westernmost borders of Baluchistan (Epstein 1971; Bökönyi 1997).

2.1.3 The dispersal of *Bos indicus* throughout Asia

Successful plant cultivation and controlled animal management within Baluchistan resulted in a marked increase in human population size and promoted the development of settled village life (Meadow 1993). These demographic changes ultimately precipitated the diffusion of farmers and domestic livestock from the North Kachi Plain eastward into neighbouring areas of the Indus Valley. By the beginning of the 4th millennium BC, more advanced human societies had begun to develop the more complex civilisations of the Indus Valley, which were responsible for the creation of the large urban states of Harappa and Mohenjo-daro (Misra 2001; **Figure 2.1**).

By the 3rd millennium BC zebu cattle were prominent components of Indus Valley economies (Zeuner 1963; Allchin & Allchin 1968; Meadow 1996). It is interesting to note that excavations at Harappa and Mohenjo-daro have recovered artefacts, such as circular stamp seals and terracotta figurines (*circa* 3,000 BC), depicting both humped zebu and humpless taurine animals (Zeuner 1963; Allchin & Allchin 1968). In addition, *Bos indicus* fossil remains have been reported in the faunal assemblages in certain Neolithic settlements of Iran and Mesopotamia (Zeuner 1963; Clason 1978; Bökönyi 1997). These data have been interpreted as evidence for the westward dispersal of domestic zebu cattle

into the Near East and a counter flow of domestic taurine cattle eastward into the Indus Valley. Most authors believe that the appearance of both domestic taxa in the Near East and the Indus Valley is representative of ancient trade routes between the civilisations, which may date back to the 4^{th}-3^{rd} millennium BC (Zeuner 1963; Misra 2001).

2.1.4 A separate origin for South Indian *Bos indicus*

Several authors argue, somewhat controversially, that *Bos indicus* may have also been domesticated within peninsular India. This hypothesis is based primarily on the observation that humped cattle breeds from peninsular India display different morphological and behavioural characteristics to those originating from the Indo-Pakistan border.

Ancient artistic impressions from the Indus Valley and peninsular India suggest that geographic variants of *Bos indicus* may have existed as far back as the 3^{rd} millennium BC as a result of independent domestic origins (Allchin 1963; Allchin & Allchin 1974). The zebu bulls depicted in Harappan artefacts display morphological characters reminiscent of modern Northwest Indian breeds, such as heavily built bodies, pronounced dewlaps and massive horns. In contrast, rock art and terracotta figurines from peninsular India represent thinner, long-horned cattle with pronounced humps and less distinct dewlaps. Similar physical features are observed in several modern Desi populations of the region (Naik 1978).

Differences between Northwest and Southern Indian *Bos indicus* have also been observed at the molecular level. Blood group studies have demonstrated the existence of a haemoglobin variant exclusive to the Khillari breed of the Deccan region which was absent

in all assayed North Indian breeds (Naik 1978). These observations have been interpreted as possible separate origins for Northern and Southern Indian *Bos indicus*.

The earliest evidence for animal herding in peninsular India comes from the archaeological sites of the Deccan Plateau, which date to *circa* 3,000 BC (Allchin & Allchin 1968; Allchin & Allchin 1974; Allchin & Allchin 1982; Misra 2001). Excavations at two principal sites of the region, Kupgal and Utnur, have revealed the existence of distinctive ash-mounds, which are believed to represent heaps of ash produced by the burning of cattle dung, which provided fuel and a building material. It has been deduced that the Deccan ashmounds are indicative of Neolithic cattle-pens, which may have served as enclosures for the capture and taming of local wild aurochs animals (Allchin & Allchin 1968). In addition, metric analysis of the faunal assemblages from prehistoric sites in peninsular India suggest that *B. p. namadicus* survived into the Holocene and was available for domestication, possibly as late as 2,000 BC (Grigson 1985; Joglekar & Thomas 1992).

Figure 2.1: Ancient *Bos indicus* domestication and herding sites within South Asia

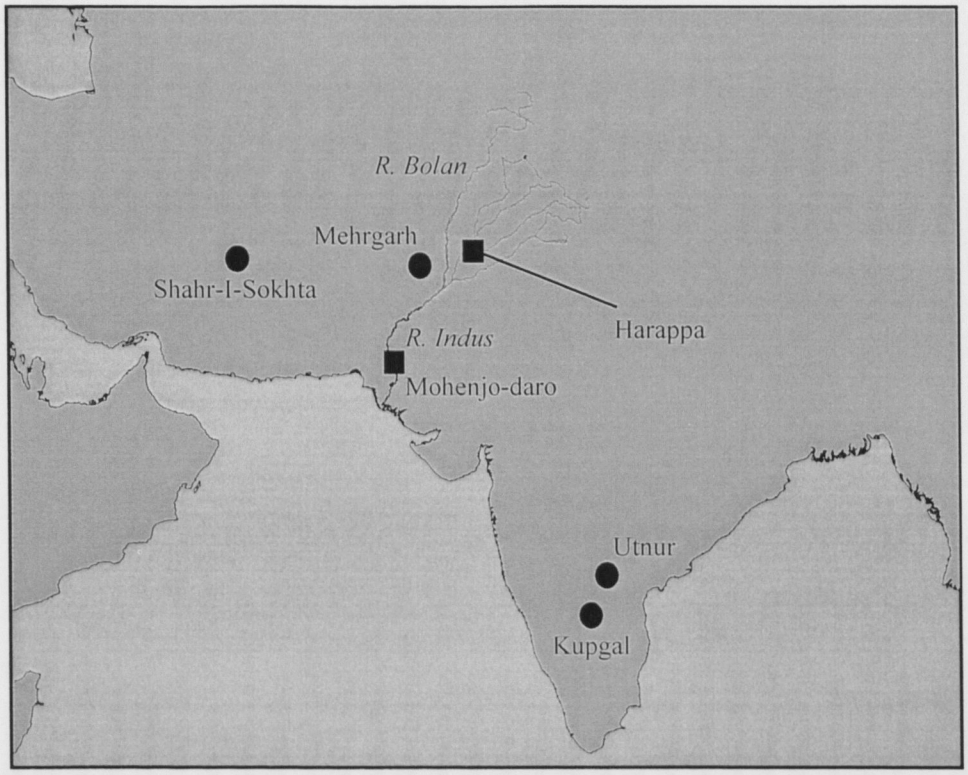


Figure 2.1. A map illustrating the approximate locations of ancient archaeological sites in South Asia displaying evidence for the possible domestication and herding of *Bos indicus* cattle. Shaded squares denote putative herding sites; shaded circles denote putative domestication centres. The approximate time depths of these sites are as follows: Mehrgarh (9,000 YBP), Harappa and Mohenjo-daro (6,500-6,000 YBP), Shahr-I-Sokhta (6,000-5,000 YBP), Kupgal and Utnur (5,000 YBP).

2.1.5 Modern Asian cattle

It is estimated that over 400 million cattle exist on the continent of Asia, accounting for approximately 30% of the world's domestic cattle resources (Payne 1999). Similar to African cattle types, the domestic cattle of Asia can be subdivided into *Bos taurus* and *Bos indicus*. In Southeast Asia approximately 1.5 million Bali cattle (domesticated banteng, *Bos* [*Bibos*] *banteng*) are found on the island of Bali and other Indonesian islands. Additionally, several thousand domesticated gaur (*Bos* [*Bibos*] *gaurus*) inhabit the tropical forests and woodlands of India, Indochina and the Malay peninsula (Lenstra & Bradley 1999).

Humpless *Bos taurus* cattle predominate in Northern Asia. Those occupying Western Russia and countries of the former Soviet Union are believed to share a common Near Eastern origin with those of the rest of Europe (Porter 1991). Alternatively, the indigenous *Bos taurus* cattle of Northeast Asia, often referred to as Turano-Mongolian cattle, are distinguished from other cattle types by their distinctive skull formation and upturned horns (Felius 1995). It is unclear whether or not Turano-Mongolian cattle were domesticated from local wild aurochs populations, however phylogenetic analysis of mtDNA variation in Japanese Black *Bos taurus* cattle suggests that modern Japanese sequences are descended from a mixture of early Near Eastern derived *Bos taurus* migrants and a locally domesticated strain of wild aurochs (Mannen *et al.* 1998).

Most South Asian cattle breeds are *Bos indicus* or zebu types. *Bos indicus* cattle predominate in India and Pakistan, however they are also the major domestic cattle variety found in Afghanistan and Iran (Bradley & Cunningham 1999; http://dad.fao.org). Zebu cattle are also found in Laos, Vietnam, South China, Taiwan and Indonesia, although in

most cases this has probably involved the introduction of domestic zebu breeds from India (**Figure 1.3**).

There are approximately 30 zebu cattle breeds that are native to the Indian subcontinent (National Bureau of Animal Genetic Resources (ICAR), http://karnal.nic.in/res_nbagr.asp). The majority of these humped Indian cattle exist as the non-descript 'village zebu' animal often referred to as *Desi* or *Nadudana* cattle (Epstein & Mason 1984). Desi cattle prevail in the mountain and plateau regions of India where they are well adapted to harsh environments and poor water supplies. As they are permitted to breed freely, the general quality of local Desi herds is poor.

As part of intensive foreign aid programmes initiated during the 20th century, crossbreeding indigenous zebu breeds with imported *Bos taurus* cattle, primarily from Europe, became widespread throughout India (Epstein & Mason 1984). This however, resulted in the loss of valuable pure zebu breeds, especially those populating the northern part of the subcontinent. In more recent times, carefully formulated schemes have been implemented to establish more efficient and structured breeding programmes, in order to improve the genetic potential of indigenous breeds (Porter 1991).

2.1.6 Research objectives

Molecular analyses involving cattle samples of wide provenance have provided key insights into the origins and phylogeography of domestic populations. However, while the majority of these investigations have focused on the divergence between *Bos taurus* and *Bos indicus* cattle (Loftus *et al.* 1994a; Loftus *et al.* 1994b; MacHugh *et al.* 1997) and the origins of *Bos taurus* cattle (Bradley *et al.* 1996; Troy *et al.* 2001), little is known about the

genetic ancestry of *Bos indicus* cattle. Therefore, the primary research objectives of this chapter were to employ mtDNA sequence data analysis to:

(1) Increase the size and range of the *Bos indicus* breeds used in previous genetic surveys and to obtain samples from previously unexamined populations;

(2) Assess mtDNA sequence diversity within and among the zebu breeds of Asia;

(3) Inter-relate the genetic data with previous archaeological investigations in order to elucidate the domestic origins and demographic history of *Bos indicus* cattle.

(4) To compare and contrast the patterns of *Bos indicus* mtDNA variation to those established for previously published *Bos taurus* mtDNA sequences.

2.2 MATERIALS AND METHODS

2.2.1 The domestic cattle populations involved in this study

General information regarding the Asian cattle breeds involved in this study is described in summary below. A map illustrating the approximate origins of these breeds is shown in **Figure 2.2**.

Red Sindhi

Red Sindhi zebu cattle are believed to be the descendants of local small red hill cattle from Baluchistan (present-day Pakistan), purportedly the cradle of zebu domestication. The breed is similar in many respects to the Sahiwal breed of the Punjab region of India and may also be related to the red type of cattle of Afghanistan (Joshi & Phillips 1953). Today, they dominate in the Southern Pakistani state of Sind, and the home base of the breed is now situated near Karachi. Due to their durability, heat-resistance and superior milk yields several of them have been transported to many parts of India and numerous countries throughout Austral-Asia, Africa and the Americas (Payne 1999).

Sahiwal

Sahiwal zebu cattle are characterised by their substantial build and distinctive convex shaped forehead. The Sahiwal breed is believed to have originated in the Punjab region situated on the present Indian-Pakistan border (Joshi & Phillips 1953). Sahiwal are related to Red Sindhi and Red Afghan cattle, suggesting a northern origin for the ancestors of this breed. They are noted for their excellent milk production and heat resistance, which has resulted in their exportation to other tropical regions such as Africa, Australia and the Caribbean (Payne 1999)..

Hariana

Hariana zebu cattle originated in the Haryana state situated to the west of Delhi in Northern India (Joshi & Phillips 1953; Payne 1999). Today, they exist as the most widespread and numerous zebu cattle breed in North India. These compact animals are used primarily as draught animals but are also used as milk producers.

Ongole and Nelore

Originating in the South Indian state of Andhra Pradesh, the Ongole breed exists as the southernmost shorthorned grey-white zebu breed of the Indian subcontinent. Physically, Ongole animals are similar to Hariana cattle and are characterised by their extreme docility. Ongole bulls are used as draught animals and cows are generally used for milk production. The Nelore zebu breed of Brazil are the modern descendents of Ongole cattle transported to Brazil between 1895 and 1964 in order to improve the adaptability of local cattle herds to tropical conditions (http://dad.fao.org).

Tharparker

This zebu breed originated in the Thar Desert situated in the Sind district of Pakistan. Tharparker animals exhibit tremendous endurance and resistance to the poor feeding and drought conditions that exist within this region (Payne 1999). Tharparker animals are strongly built and are considered to be the best dual-purpose milk-work breeds existing within the Indian subcontinent.

Iraqi cattle

Collectively, Iraq and Mesopotamia form an area considered to be one of the primary cradles of *Bos taurus* domestication and native taurine populations are regarded as the descendants of the first domestic tauirne animals (Bradley & Cunningham 1999; Troy *et*

al. 2001). Historic agro-pastoralist migrations, resulting in the transfer of South Asian zebu cattle into Iraq, have greatly modified the genetic composition of native Iraqi taurine populations (Loftus *et al.* 1999).

Figure 2.2: The approximate locations of the populations analysed for *Bos indicus* mtDNA variability in this study

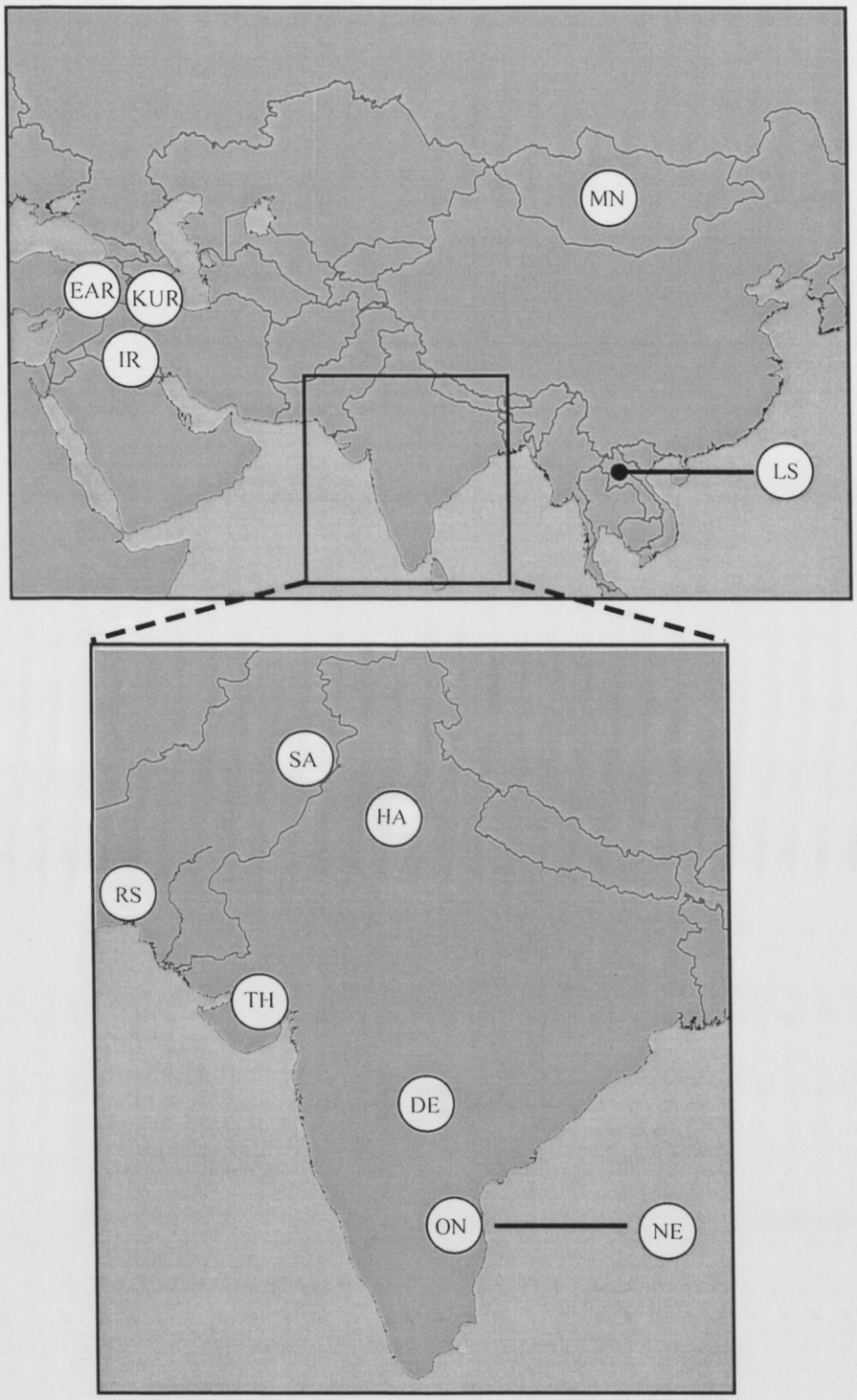


Figure 2.2. Approximate locations of the Asian populations analysed in this study. Population codes: EAR, East Anatolian Red; IR, Iraqi; KUR, Kurdi; RS, Red Sindhi; SA, Sahiwal; TH, Tharparker; HA, Hariana; DE, Desi; ON, Ongole; LS, Laotian; MN, Mongolian; NE, Nelore. The Nelore population of Brazil are believed to be the descendants of Ongole cattle exported to the Americas during the 19th and 20th centuries. *Bos indicus* mtDNA sequences sampled in East Anatolian Red and Kurdi populations and some Iraqi samples were previously published by Troy *et al.* (2001).

2.2.2 *Bos indicus* DNA sample information

The DNA samples sequenced and reported for the first time in this study

A total of 91 samples were assayed for the first time as part of this study. These comprised 6 domestic Asian cattle populations (18 Iraqi samples, 22 Red Sindhi samples, 3 Hariana samples, 5 Sahiwal samples, 15 Tharparker samples, 19 Desi samples), plus 9 Brazilian Nelore samples. All 91 samples were surveyed for mtDNA sequence variation within the most variable 240 bp fragment of the bovine mitochondrial control region (nucleotide positions 16,023-16,262 according to nomenclature of Anderson *et al.* 1982). This 240 bp mtDNA fragment has been sequenced and analysed in previous surveys of *Bos taurus* phylogeography (Cymbron *et al.* 1999; Troy *et al.* 2001).

MtDNA sequencing was performed using the procedures outlined in **Section 5.3**. All cattle samples were procured as DNA that was isolated by various members of the Molecular Population Genetics laboratory (TCD) prior to the initiation of this study. *Bos taurus* and *Bos indicus* mtDNA sequences were identified by alignment with previously published *Bos taurus* and *Bos indicus* mtDNA sequences (Troy 1998), using the muliple sequence alignment option incorporated in the GENEJOCKEY II software package (Taylor 1993).

Of the 91 samples that were assayed for the first time in this molecular survey, 25 samples displayed *Bos taurus* mtDNA sequences, while the remaining 66 samples were of *Bos indicus* mtDNA type. Thirteen of the 25 *Bos taurus* mtDNA sequences reported here were encountered in the morphologically taurine Iraqi samples, 8 were detected in the Brazilian Nelore samples and were assumed to be of South American Creole origin (discussed and analysed in Chapter 3 of this thesis) and 4 were encountered as minority types in three morphologically zebu populations of the Indian subcontinent (1 Hariana sample, 2 Red Sindhi samples and 1 Tharparker sample).

Of the 66 samples displaying *Bos indicus* mtDNA sequences reported for the first time in this study, 5 were encountered in the morphologically taurine Iraqi samples analysed here and were classified as introgressors from the Indian subcontinent, 14 were detected within the Tharparker samples assayed, 2 were detected in the Hariana samples assayed, 20 were detected in the Red Sindhi samples assayed, 19 were detected in the Desi samples assayed, 1 was detected in the Nelore samples assayed and 5 were detected in the Sahiwal samples assayed.

Additional Bos indicus mtDNA sequence data used in this study

The 5 *Bos indicus* introgressor mtDNA sequences encountered in the morphologically taurine Iraqi samples surveyed here for the first time were pooled with the 9 *Bos indicus* mtDNA introgressor sequences detected in morphologically taurine Near Eastern samples, as reported by Troy *et al.* (2001). The combined 14 *Bos indicus* sequences from the Near East were treated as a discrete population for phylogenetic analysis and are herein referred to collectively as 'Near East zebu'.

The *Bos indicus* mtDNA sequence data presented for the first time here were also supplemented with an additional 35 previously published Indian *Bos indicus* mtDNA sequences (Troy 1998). These comprised 7 Hariana samples, 7 Sahiwal samples, 6 Tharparker samples, 4 Nelore samples and 11 Ongole samples. Also, 39 unpublished *Bos indicus* mtDNA sequences sampled in Laos and Mongolia, were contributed by Prof. H. Mannen, (University of Kobe, Japan), and were also considered for phylogenetic analysis here.

The details regarding the total *Bos indicus* mtDNA sequence data collection used for phylogenetic analysis in this survey is outlined in **Table 2.1**. Overall, the number of *Bos indicus* mtDNA sequences representing each population ranged between 5 individuals (Nelore) and 30 individuals (Laotian). Previous bovine mtDNA phylogeography studies have analysed between 3-12 samples per population (Troy 1998; Cymbron *et al.* 1999; Troy *et al.* 2001).

Table 2.1: The total *Bos indicus* mtDNA sequence data set considered for phylogenetic analysis in this survey

Population name (Population code)	**Total number of *B. indicus* mtDNA seqs.** (seqs reported for 1st time here /previously reported seqs)	**Geographic origin**	**Geographic co-ordinates Latitude; Longitude**	**Sample type** (seqs reported for 1st time here /previously reported seqs)
Near East zebu (NEZ)	14 (5/9)	Iraq, Kurdistan and Anatolia	33.2N; 44.26E	Hair[1]/mtDNA seqs[4]
Red Sindhi (RS)	20 (20/0)	Baluchistan, Pakistan	24.5N; 67.02E	Hair[2]
Sahiwal (SA)	12 (5/7)	Punjab region, Indo-Pakistan border	30.41N; 73.11E	Blood[1]/mtDNA seqs[4]
Hariana (HA)	9 (2/7)	Haryana district, Northern India	28.4N, 77.14E	Blood[1]/mtDNA seqs[4]
Tharparker (TH)	20 (14/6)	Thar region, Indo-Pakistan border	24.26N; 71.44E	Hair[2]/mtDNA seqs[4]
Desi (DE)	19 (19/0)	Central India	18.4N; 78.05E	Hair[2]
Ongole (ON)	11 (0/11)	Andhra-Pradesh district, India	16.2N; 80.27E	mtDNA seqs[4]
Nelore (NE)	5 (1/4)	Brazil	16.2N; 80.27E	extracted DNA[3]/mtDNA seqs[4]
Laotian (LS)	30 (all unpublished)	Laos	19.53N; 102.1E	mtDNA seqs[5]
Mongolian (MN)	9 (all unpublished)	Mongolia	47.54N; 106.52E	mtDNA seqs[5]
Total	**149 (66/83)**			

Table 2.1. The total combined *Bos indicus* mtDNA sequence data collection considered for phylogenetic analysis in this survey (also listed in **Table 5.1**). The approximate geographic origins for each population were determined by plotting the region of origin of each population, as recorded during sampling (where possible) on a detailed map of the Asian continent. The co-ordinates of the Brazilian Nelore breed are considered to be identical to that for the Ongole breed as they are the direct descendants of Ongole cattle exported to the Americas during the 19th and 20th centuries (Mason 1988).

Sampling information for the DNA samples analysed for the first time in this study: [1] Blood samples collected by Dr. Ronan Loftus and Dr. David MacHugh, Trinity College Dublin, Ireland; [2] Blood samples collected by Dr. Pushpendra Kumar, University of Bareilly, India; [3] DNA samples contributed by Dr. Claude Gaillard, University of Geneva, Switzerland.

Source of additional data included in the data set for phylogenetic analysis: [4] mtDNA sequences previously published by Troy (1998); [5] previously unpublished mtDNA sequences contributed by Prof. Hide Mannen, Kobe University, Japan.

2.2.3 Mitochondrial DNA sequence data analysis

DNA sequence alignment

Alignment of all mtDNA sequences was performed using the CLUSTALX computer package (Thompson *et al.* 1997).

Intra-population and intra-regional genetic diversity

Intra-regional and intra-population genetic diversity estimates for the 240 bp mtDNA sequences are expressed here as, (1) the mean number of pairwise differences (MNPD) (Tajima 1983), and (2) the θ_S parameter (Watterson 1975), a diversity measure based on the number of variable sites in non-recombining DNA which is corrected for the number of samples within a population.

The mean number of pairwise differences computes the proportion of nucleotide differences between all pairs of sequences in a population and is given by (Tajima 1983):

$$\text{Mean number of pairwise differences (MNPD)} = \sum_{i=1}^{k} \sum_{j<i} p_i p_j d_{ij}$$

where d_{ij} is an estimate of the number of mutations having occurred since the divergence of haplotypes i and j, k is the number of haplotypes and p_i is the frequency of haplotype i.

The θ_S parameter (Watterson 1975) is a diversity measure estimated from the infinite-site equilibrium relationship between the number of variable sites (S), the sample size (n) and the population parameter θ [where θ is the product of twice the mutation rate (θ) and the effective female population size (N_{ef}) for mtDNA (*i.e.* $2\theta N_{ef}$)], for a sample of non-recombining DNA. The θ_S parameter is given by:

$$\theta = \frac{S}{a_1}$$

where $a1$ equals:

$$a1 = \sum_{i=1}^{n-1}$$

with respect to haplotype i.

Analysis of demographic events from mtDNA sequence data

Several authors have argued that historical population dynamics can be inferred by analysing mtDNA sequence pairwise mismatch distributions - histograms showing the distribution of the number of nucleotide differences between pairs of mtDNA sequences (Rogers and Harpending 1992, Harpending *et al.* 1993). Smooth, uni-modal distributions are indicative of rapid, past population expansions, during which the expectancy of mtDNA lineage survivorship is high. Ragged, multi-modal distributions are consistent with demographic stability. Fu's F_S statistic of selective neutrality (Fu 1997; Excoffier & Schneider 1999; Excoffier & Schneider 2000), which is sensitive to demographic growth was also employed to investigate putative episodes of historic population growth. All

calculations were computed using the ARLEQUIN VERSION 2.0 software package (Schneider *et al.* 1999).

Inter-population genetic distances

Pairwise F_{ST} values (Wright 1965; Slatkin 1991; Slatkin 1995), calculated for every pairwise population combination using the ARLEQUIN VERSION 2.0 package, were employed as estimates of genetic distance among Asian zebu populations. Pairwise F_{ST} values, which are generally used as estimates of genetic differentiation due to population subdivision (Wright 1965; **Section 4.2.4**), are considered as an appropriate genetic distance estimate when surveying inter-population genetic variation over shallow time-depths (Slatkin 1991; Slatkin 1995). An unrooted neighbor-joining (NJ) tree was constructed from the resulting matrix of F_{ST} values using the NEIGHBOR algorithm (Saitou & Nei 1987) incorporated in the PHYLIP VERSION 3.572 package (Felsenstein 1993). Inter-population F_{ST} genetic distances were also represented in two-dimensions following multi-dimensional scaling (MDS, Kruskal 1964) using the SPSS VERSION 8.0 software package. MDS can be used as a statistical technique to illustrate population divergences based on a matrix of inter-population genetic distances. Unlike phylogenetic trees, MDS does not impose an ancestral status upon the populations under consideration and instead demonstrates the affinities among populations, relative to each other, based on the genetic distances calculated between them. Moreover, each dimension is orthogonal and may display separate genetic trends within the data.

Hierarchical analysis of molecular variation (AMOVA)

The hierarchical structure of molecular variation was studied using the analysis of molecular variance (AMOVA) approach incorporated within the ARLEQUIN VERSION 2.0 package (Schneider *et al.* 1999). AMOVA estimates the proportion of the total molecular

variance in a given data set which attributable to the three major levels of population subdivision, namely (1) genetic differences between regional groups; (2) differences between populations within regional groups; and (3) differences between individuals within populations within regional groups.

Unlike conventional ANOVA-based analysis of variance, which compares average haplotype frequencies between populations, AMOVA incorporates information regarding haplotype frequencies plus the amount of genetic differences (*i.e.* nucleotide differences) between haplotypes (Excoffier 2001). For haplotypic data, such as mtDNA sequences, the AMOVA model assumes that the *i*-th haplotype frequency vector from the *j*-th population in the *k*-th group is a linear equation given by:

$$\mathbf{x}_{ijk} = \mathbf{x} + \mathbf{a}_k + \mathbf{b}_{jk} + \mathbf{c}_{ijk}$$

where x is the vector of the expected haplotypic states. The effects are **a** for the group, **b** for population and **c** for the individual and are assumed to be random, additive and independent with corresponding covariance components σ^2_a, σ^2_b and σ^2_c (Excoffier 2001).

The ARLEQUIN VERSION 2.0 package calculates the sum of squared Elucidean distances (*i.e.* mutational differences) between all pairs of haplotypes within populations, between populations within groups and between groups and corresponding haplotype frequencies for all data entered into the package. From the resulting matrix, the major variance components are calculated. The total molecular variance (σ^2) is sum of the covariance components due to differences among haplotypes within populations (σ^2_c), differences among haplotypes between populations within groups (σ^2_b), and the variance components due to differences among haplotypes among groups (σ^2_a) (Excoffier 2001).

Statistical analysis of haplogroup distributions

Statistical testing of *Bos indicus* mtDNA haplogroup distributions with respect to geography were performed using, (1) the data analysis options incorporated within the MICROSOFT EXCEL© software package (Chi-squared tests of association, regression analysis), and (2) the SPSS VERSION 8.0 software package (Spearman rank tests). Descriptions of these statistics are given in Mead *et al.* (1993) and Owen & Jones (1982).

Phylogenetic networks

Reduced median phylogenetic networks, illustrating the primary relationships between the mtDNA haplotypes presented in the data, were also employed as an alternative visualisation of the mtDNA sequence data. These were constructed manually according to the principles outlined by Bandelt *et al.* (1995).

2.3 RESULTS

2.3.1 Amplification of bovine mitochondrial DNA

PCR amplifications and electrophoresis of DNA samples were performed according to the procedures outlined in **Section 5.3**. **Figure 2.3** shows the PCR amplification of a 375 bp fragment of the bovine mtDNA control region (nucleotide positions 15,960-16,335, as per Anderson *et al.* 1982) for 13 Red Sindhi samples together with negative controls to ensure that no DNA contamination was present in the PCR reaction. From these amplified fragments, the most variable 240 bp portion of the bovine mtDNA control region was sequenced.

Figure 2.3: PCR amplification of *Bos indicus* mtDNA

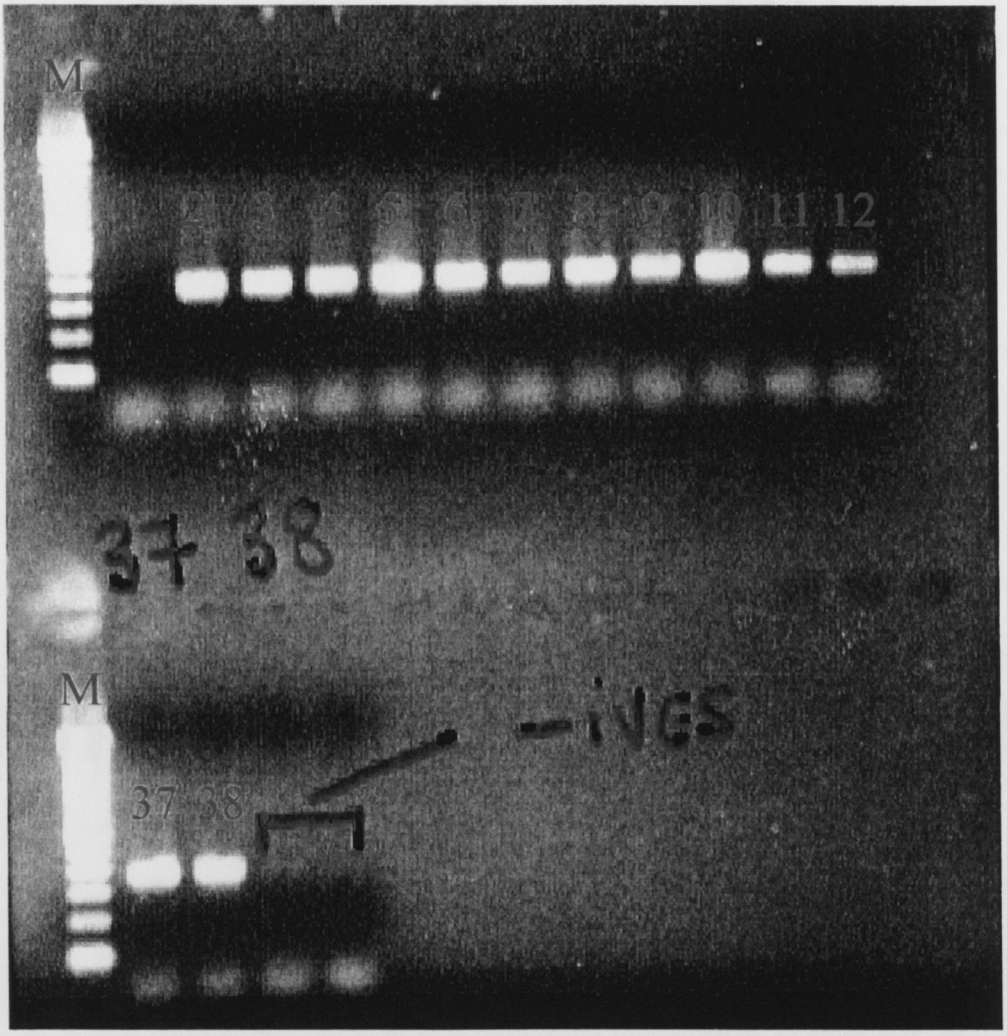


Figure 2.3. PCR amplification of Red Sindhi samples 2-12, 37 and 38. Negatives controls are also shown. 'M' denotes 100 bp ladder DNA.

2.3.2 Bovine mitochondrial DNA sequence variation

Figure 2.4 shows a typical sequencing autoradiograph produced following the procedures outlined in **Section 5.3**. In the majority of cases the entire 240 bp region of interest was determined using a single PAGE run of 165 minutes.

Figure 2.4: Partial DNA sequence of the most variable 240 bp fragment of bovine mitochondrial control region

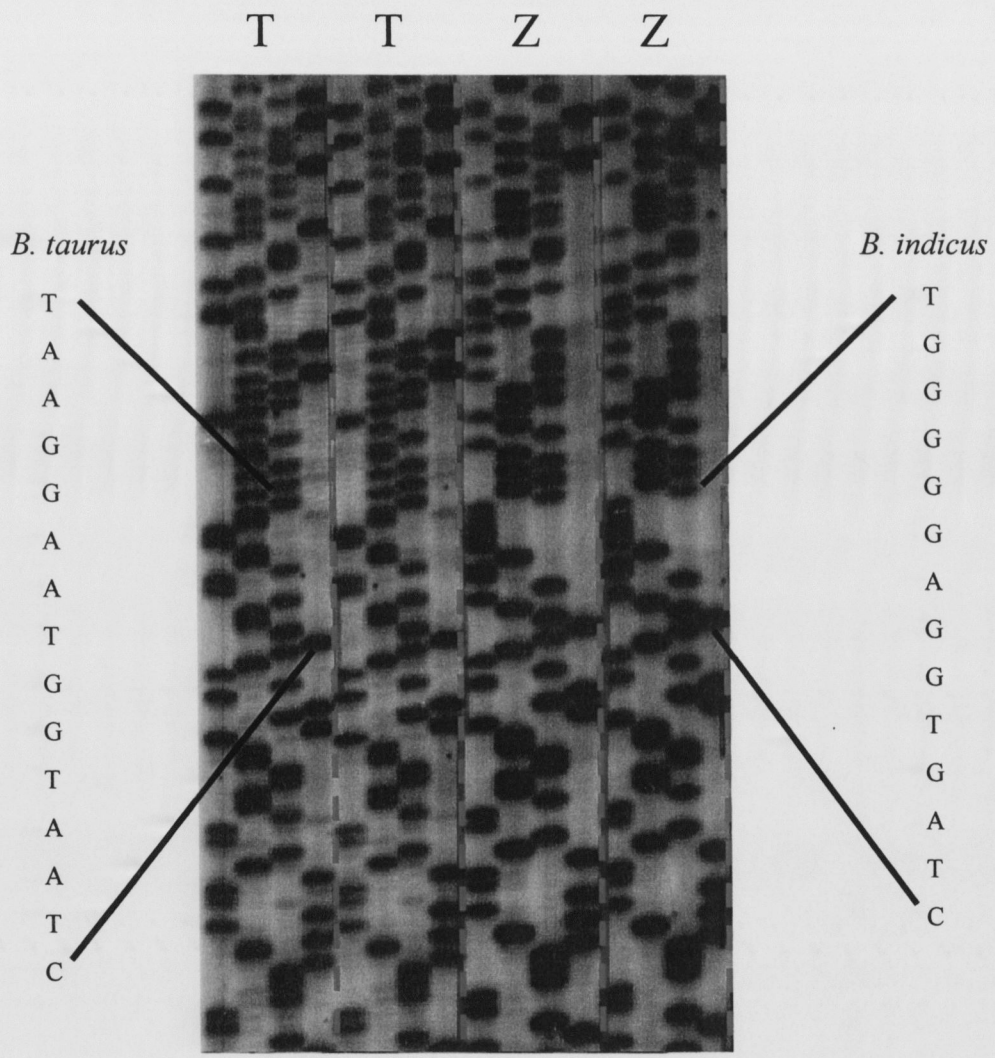


Figure 2.4: MtDNA sequences are shown for Iraqi samples 30, 37, 40 and 54 (left to right). Samples displaying *Bos taurus* mtDNA sequences are denoted T; those displaying *Bos indicus* sequences are denoted Z. Partial *Bos taurus* and *Bos indicus* sequences are given. Individual lanes from left to right represent the bases G, A, T and C, respectively.

2.3.3 Genetic signatures of *Bos indicus* and *Bos taurus* introgression

Bos indicus introgression into the Near East

Recent molecular surveys have demonstrated an east-west geographical cline in the distribution of zebu admixture proportions in morphologically *Bos taurus* breeds of the Near East. These findings have been interpreted as introgression of *Bos indicus* cattle from the Indian subcontinent into the taurine breeds of Southwest Asia, believed to have that occurred over the past 6,000 years (Loftus *et al.* 1999). Genetic signatures of these ancient zebu migrations are also evident from the mtDNA sequence data analysis presented here. When the Near Eastern mtDNA sequence data in this study was combined with that reported by Troy *et al.* (2001), a total of 14 *Bos indicus* introgressors were detected as minority types in morphologically taurine populations from the Near East. As with microsatellite data (Loftus *et al.* 1999) the pattern of zebu introgression is distributed along an east-west axis, with populations in closer proximity to the Indian subcontinent displaying a higher proportion of *Bos indicus* mtDNA variants (**Figure 2.5a**)

Bos taurus introgression into the Indian subcontinent

Phylogenetic analyses have demonstrated that all surveyed *Bos taurus* mtDNA control region sequences fall into 1 of 5 geographically distributed haplotypic clusters (or haplogroups) (Mannen *et al.* 1998; Troy *et al.* 2001). One of these (T3 haplogroup) predominates in Europe and is one of three (T, T2 and T3 haplogroups) encountered at considerable frequencies in domestic taurine populations of the Near East, another (T1 haplogroup) is almost exclusive to Africa, while one (J5 haplogroup) occurs only in Japan.

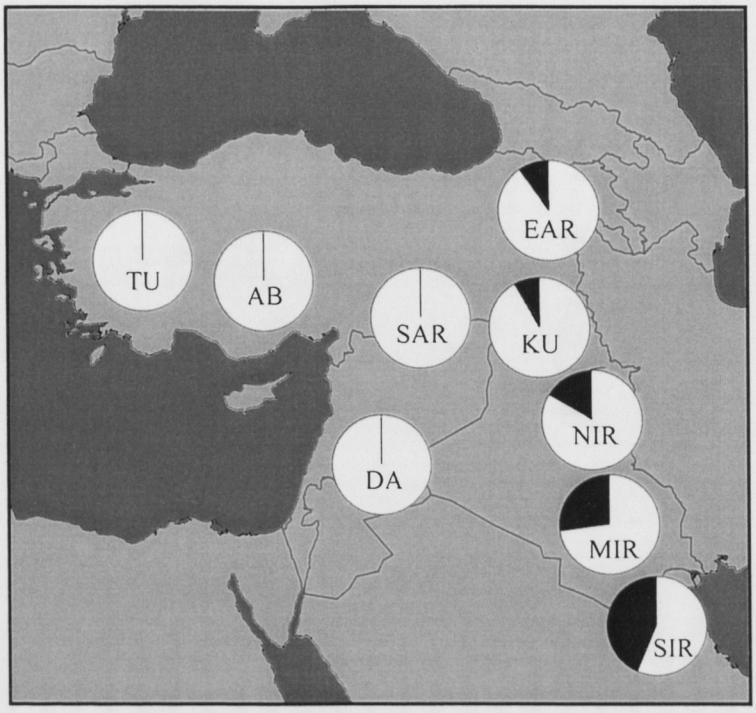
In the data presented in this study, a total of 4 South Asia *Bos indicus* samples displayed mtDNA sequences that were of *Bos taurus* origin. These sequences, comprising 3 *Bos taurus* haplotypes, were detected within 3 zebu populations originating from the Northwest

of the Indian subcontinent (Red Sindhi [n = 2], Tharparker [n = 1] and Hariana [n = 1]). The sequence variations defining the three *Bos taurus* mtDNA introgressor haplotypes are shown in **Appendix 2.A**.

A reduced median network, illustrating the primary relationship between sequences (discussed in **Section 2.3.6**), was constructed for all 4 *Bos taurus* introgressor mtDNA sequences (**Figure 2.5b**). The resulting patterns of diversity resemble those produced for sequences of Near Eastern ancestry, with 3 of the *Bos taurus* introgressor mtDNA sequences clustering with the T haplogroup, while the remaining taurine introgressor sequence group with the T2 haplogroup. As the *Bos taurus* T3 mtDNA haplogroup, which predominates in Europe, is not encountered among these introgressor sequences, this finding may be suggestive of ancient introgression from the Near East rather than recent crossbreeding with imported European *Bos taurus* animals. This assertion is supported by archaeological evidence suggesting that *Bos taurus* cattle were present in the early agricultural societies of the Indus Valley, *circa* 4^{th}-3^{rd} millennia BC and which may have presented an opportunity for gene-flow between the two domestic taxa (Zeuner 1963; Kumar *et al.* in press).

Bos taurus haplogroup geographical structuring across the Near East was assessed in order to evaluate any putative genetic affinities between the 4 *Bos taurus* introgressor sequences and Near Eastern taurine populations located in closer proximity to the Indian subcontinent (**Figure 2.5b**). This analysis, however, did not reveal any significant *Bos taurus* haplogroup structuring across the Near East.

Figure 2.5a: The proportion of *Bos taurus* and *Bos indicus* mtDNA haplotypes in domestic populations from the Near East and the Indian subcontinent



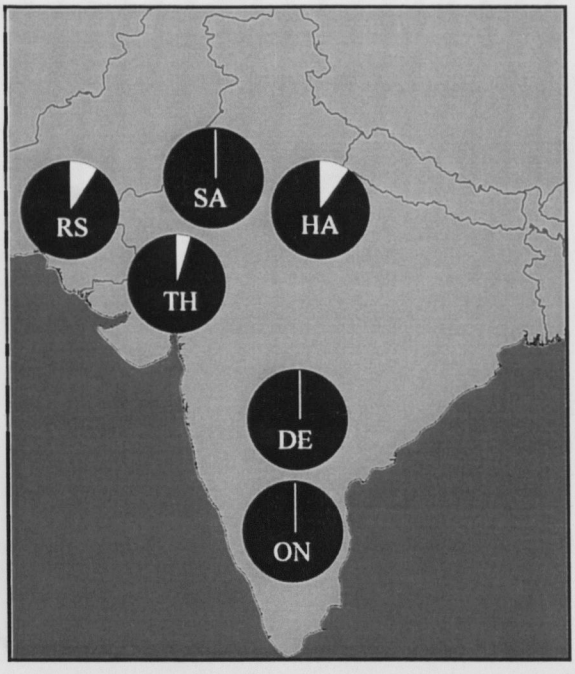


Figure 2.5b: Patterns of *Bos taurus* introgression into the Indian subcontinent



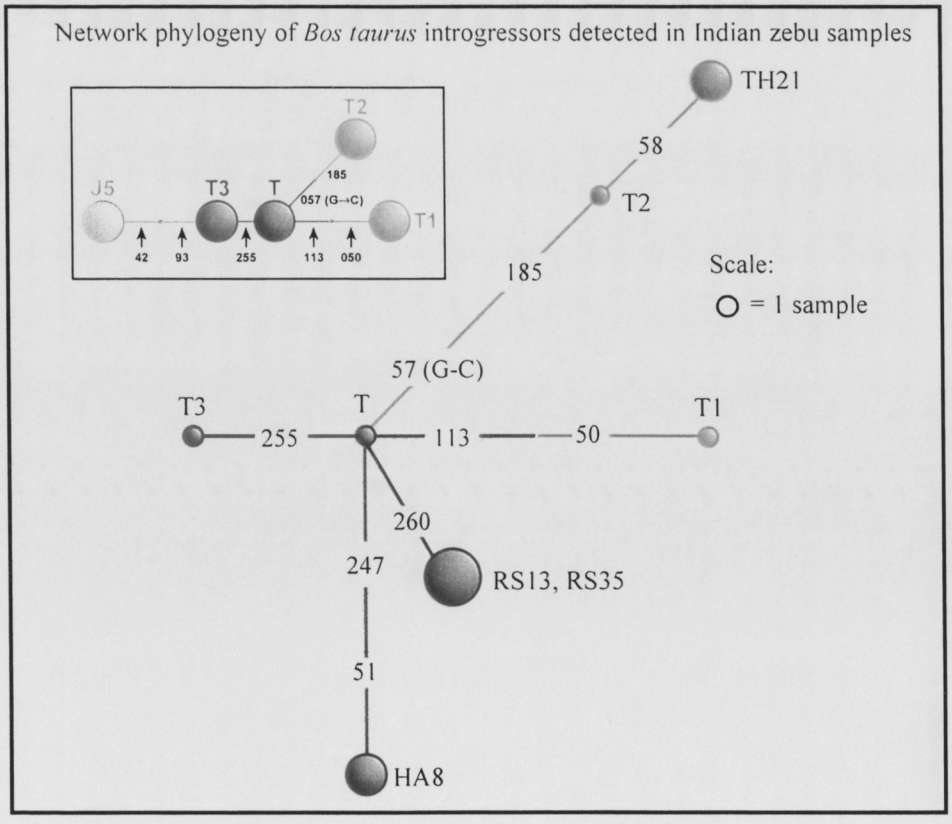


Figure 2.5a. The proportion of *Bos indicus* mtDNA lineages within each Near Eastern population is shaded black. Pie chart positioning represents the approximate geographic location of each population. A cline of *Bos indicus* maternally-mediated introgression is evident with populations in closer proximity to the Indian subcontinent showing a higher proportion of *Bos indicus* mtDNA haplotypes. Similar patterns of introgression have been observed in Near Eastern populations using nuclear microsatellite data (Loftus *et al.* 1999). Population codes: AB (Anatolian Black), DA (Damascus), EAR (East Anatolian Red), KU (Kurdi), MIR (Middle Iraqi), NIR (North Iraqi), SAR (South Anatolian Red), SIR (South Iraqi) and TU (Turkish Grey). Sample sizes range between 9-14 individuals per population.

Figure 2.5b. Patterns of *Bos taurus* introgression into the morphologically zebu populations of the Indian subcontinent. The pie charts (top) represent the proportions of the three major Eurasian *Bos taurus* haplogroups [T (blue), T2 (green) and T3 (red)] and the major African *Bos taurus* haplogroup [T1 (yellow)] as defined by Troy *et al.* 2001, within each morphologically taurine Near Eastern population. Population codes are identical to those in **Figure 2.5a**. The reduced median phylogenetic network (bottom), constructed according to the methods of Bandelt *et al.* (1995), illustrates the primary relationship between the *Bos taurus* introgressor sequences detected within the breeds of the Indo-Pakistan border (phylogenetic networks are discussed in more detail in **Section 2.3.6**). Circles denote sequence haplotypes and circle area is proportional to the frequency of haplotype representation. Lines connecting circles represent nucleotide substitutions; numbers correspond to the site of substitution within the mtDNA control region, minus 16,000, as per Anderson *et al.* (1982). Introgressor sequences are indicted by sample number and their relationship to the four primary *Bos taurus* sequence haplotypes reported by Troy *et al.* (2001) (T, T1, T2, T3; depicted in the inset) is denoted by circle colouration. The pattern of sequence variation observed for the *Bos taurus* introgressor mtDNA sequences reflect the patterns of mtDNA sequence diversity within the Near East and are probably reflective of ancient maternal introgression from the Near East. The source of the mtDNA sequence data used in both in **Figure 2.5a** and **2.5b** are listed in **Table 5.1**.

2.3.4 Summary of *Bos indicus* mitochondrial DNA sequence variation

Alignment and comparison of all 149 *Bos indicus* mtDNA sequences presented in **Table 2.1** reveals a total of 41 haplotypes, differentiated at a total of 34 polymorphic sites comprising 31 transitions, 3 transversions and 1 single base-pair insertion. No deletions were detected in the data. The ratio of transitions to transversions as estimated from the

reduced median phylogenetic network illustrated in **Figure 2.7a** is estimated here as 16:1, respectively. This pattern of variability is comparable to those previously described for domesticated cattle and confirms the strong transitional bias that exists within the bovine mitochondrial control region (Loftus *et al.* 1994a; Bradley *et al.* 1996; Mannen *et al.* 1998; Cymbron *et al.* 1999).

The sequence variations defining each *Bos indicus* haplotype are presented in **Table 2.2**. The two most frequently represented haplotypes, denoted Z1 and Z2, were detected 31 times and 39 times, respectively. The Z1 and Z2 haplotypes are differentiated by four substitutions situated at positions 16,084, 16,085, 16,141 and 16,232. Derivative haplotypes are labelled by suffixes, which indicate the position of nucleotide substitution within the control region, less 16,000 (numbering according to Anderson *et al.* 1982). Of the remaining haplotypes, one was represented 12 times ($Z2_{74b}$), one occurred 6 times, one was shared by 5 samples, six occurred in triplicate, eight were duplicated and twenty-two were unique. Haplotypic diversity was found within each population examined with the exception of the five Nelore *Bos indicus* sequences, all of which were identical to the Z2 haplotype.

From **Table 2.2**, sequences clustering with the Z1 central haplotype are referred to as the Z1 haplogroup, while sequences displaying consistent substitutions with the Z2 central haplotype, are referred to as the Z2 haplogroup. Also, some sequences group with the $Z2_{74b}$ haplotype forming a minor haplotypic cluster within the Z2 haplogroup. The $Z2_{74b}$ haplotype is differentiated from the predominant Z2 central variant by a single transition at nucleotide position 16,074. The $Z2_{74b}$ haplotype and the sequence cluster centred round the $Z2_{74b}$ haplotype are herein referred to in the text as the Z3 haplotype and Z3 haplogroup, respectively.

Table 2.2: Nucleotide variation observed in 149 *Bos indicus* mtDNA control region sequences

| Haplotype name | Samples within each haplotype | Nucleotide sequence positions as per Anderson *et al.* (1982), minus 16,000. | | |

Table 2.2. Sequence variation observed among 149 *Bos indicus* 240 bp mtDNA control region sequences sampled from 9 Asian populations and 1 Brazilian population of Indian descent. Haplotype codes are given in the first column. Only variable sites are shown. Differences are given, full stops (.) denote identity. The two predominant haplotypes observed in the data are denoted, Z1 and Z2. Derivative haplotypes from either predominant haplotype are labelled by suffixes, which indicate the position of nucleotide substitution within the control region less 16,000 (numbering according to Anderson *et al.* 1982). Samples shown in bold type are those which were assayed for the first time in this study. Additional data are referenced in the text and also in **Table 5.1**.

2.3.5 Intra-population genetic variation

Genetic diversity estimates and pattern of nucleotide variability for each *Bos indicus* population and 4 geographically defined regional samples [West Asia (n = 34), Northwest India (n = 41), South India (n = 30, excluding the Brazilian Nelore sequences) and East Asia (n = 39)] are shown in **Table 2.3**. Previous studies have shown that domestic cattle populations from a purported centre of origin exhibit greater allelic and haplotypic diversity due to the retention of ancestral variation (Troy 1998; Loftus *et al.* 1999; Troy *et al.* 2001). Successive outward migrations from centres of origin typically result in reduced genetic variation in derived populations. Accordingly, *Bos indicus* mtDNA genetic variation is visibly higher within populations from the Indian subcontinent than populations from the Near East and Eastern Asia. No mitochondrial sequence diversity was observed within the five *Bos indicus* sequences encountered in the Brazilian Nelore population, each of which is identical to the Z2 haplotype.

Table 2.3: *Bos indicus* mtDNA sequence diversity estimates for individual *Bos indicus* populations and regional groups assayed in this study

Population	*n*	MNPD	θ_S	Total no. pol. sites	Total no. transts.	Total no. transvs.	Total no. inserts
NEZ	14	2.165 ± 1.276	1.887 ± 0.996	6	5	1	0
RS	20	3.247 ± 1.747	3.100 ± 1.362	11	10	1	0
TH	20	3.432 ± 1.830	2.819 ± 1.264	10	10	0	0
SA	12	2.409 ± 1.405	1.987 ± 1.067	6	6	0	0
HA	9	3.111 ± 1.779	4.047 ± 2.010	11	11	0	0
DE	20	4.152 ± 2.160	5.436 ± 2.187	19	18	1	0
ON	11	4.036 ± 2.182	3.073 ± 1.527	10	9	0	1
NE	6	0.000 ± 0.000	0.000 ± 0.000	0	0	0	0
LS	30	0.975 ± 0.681	1.767 ± 0.834	7	7	0	0
MN	9	0.556 ± 0.499	0.368 ± 0.368	1	1	0	0
West Asia	34	2.777 ± 1.506	2.935 ± 1.189	12	11	2	0
Nwest India	41	3.378 ± 1.766	3.973 ± 1.462	17	17	0	0
South India	30	4.166 ± 2.130	5.806 ± 2.113	24	22	1	1
East Asia	39	0.955 ± 0.667	1.892 ± 0.841	8	8	0	0

Standard deviations are given for each intra-population diversity estimate. **Abbreviations**: *n* (number of samples); pol (polymorphic); transts (transitions); transvs (transversions); inserts (insertions). *Bos indicus* regional groups consisted of the following populations: West Asia (NEZ, RS); Northwest India (HA, SA, TH); South India (DE, ON); East Asia (LS, MN). To ensure that the data for the South India group was not skewed by Nelore data, all 5 *Bos indicus* sequences which are identical to the Z2 central variant sequence were removed.

Figure 2.6a: Intra-regional mtDNA sequence diversity expressed as the mean number of pairwise differences

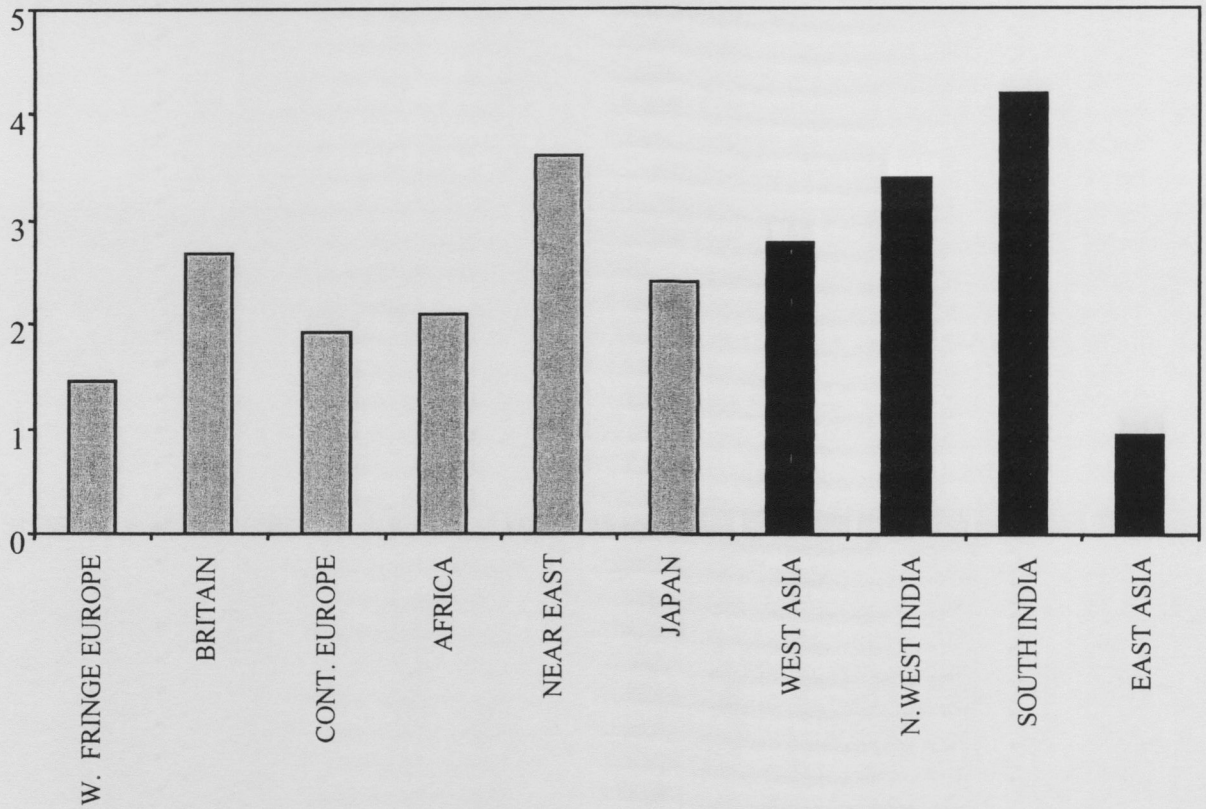


Figure 2.6a. Intra-regional number of pairwise differences. Sample sizes are as follows: Western Fringe Europe (n = 51), Britain (n = 65), Continental Europe (n = 91), African (n = 95), Near East (comprising Anatolia and the Middle East; n = 93), Japan (n = 32), West Asia (n = 34), Northwest India (n = 41), South India (n = 30) and East Asia (n = 39). The Nelore samples were not included with the South Indian group. *Bos taurus* mtDNA sequences are highlighted in grey and *Bos indicus* mtDNA sequence data are shaded black. The source of the *Bos taurus* mtDNA sequence data used is detailed in **Table 5.1**.

Figure 2.6b: Intra-regional mtDNA sequence diversity expressed as Watterson's θ_S parameter

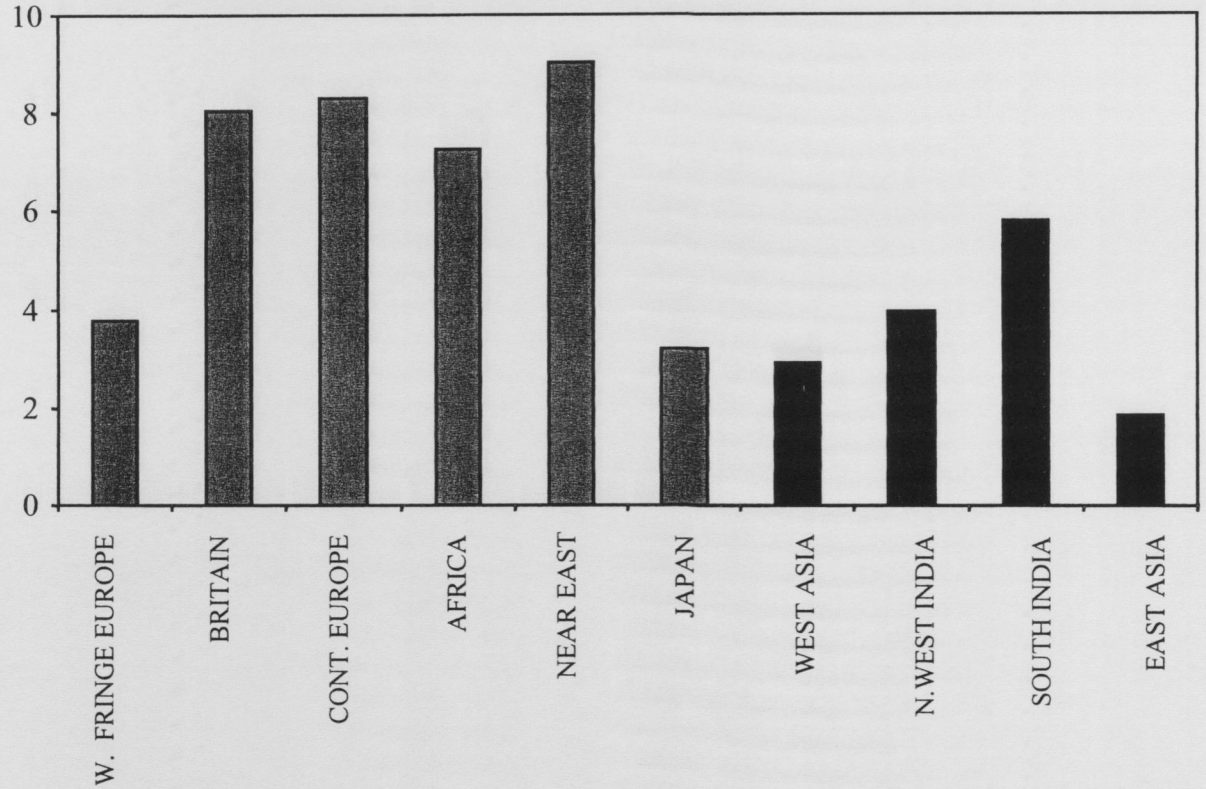


Figure 2.6b. Intra-regional θ_S values. Sample sizes are as follows: Western Fringe Europe (n = 51), Britain (n = 65), Continental Europe (n = 91), African (n = 95), the Near East (comprising Anatolia and the Middle East; n = 93) Japan (n = 32), West Asia (n = 34), Northwest India (n = 41), South India (n = 30) and East Asia (n = 39). The Nelore samples were not included with the South Indian group. *Bos taurus* mtDNA sequence data are highlighted in grey and *Bos indicus* mtDNA sequence data are shaded black. The source of the *Bos taurus* mtDNA sequence data used is detailed in **Table 5.1**.

The Desi from central India display the highest diversity values (MNPD = 4.125; θ_S = 5.436). All other populations from the Indian subcontinent exhibit moderately high levels of genetic diversity, with MNPD values ranging between 4.036 (Ongole) and 2.409 (Sahiwal) and θ_S values ranging between 4.407 (Hariana) and 1.987 (Sahiwal). When treated as a discrete population, the 14 NEZ sequences display appreciable levels of genetic variation (MNPD = 2.165; θ_S = 1.887). Conversely, the zebu sequences encountered in the two East Asian populations reveal the lowest diversity estimates of all Asian populations examined. Only two *Bos indicus* haplotypes were observed in the sample of nine Mongolian cattle (MNPD = 0.556; θ_S = 0.368). These were differentiated by a single transition at position 16,074. The thirty Laotian cattle samples exhibit seven haplotypes but genetic diversity for this population is also low (MNPD = 0.975; θ_S = 1.767).

A comparison of intra-regional MNPD values (**Figure 2.6a**) shows that the diversity contained within the *Bos indicus* mtDNA sequences sampled in West Asia, Northwest Indian and South India groups is comparable to that estimated for *Bos taurus* mtDNA sequences sampled in the Near East - the major purported centre of Eurasian *Bos taurus* domestication. However, a comparison of θ_S values (**Figure 2.6b**) shows that most *Bos taurus* intra-regional mtDNA variation exceeds that found in regional *Bos indicus* groups.

2.3.6 Phylogenetic network analysis

A median phylogenetic network is a diagrammatic presentation that includes all most parsimonious trees supported by the data. Reduced median networks simplify median networks by detecting mutations that are likely to have occurred more than once at the same site (parallel mutations or homoplasies). This allows recurrent mutations to be resolved as separate mutational events. Phylogenetic networks are considered to be

particularly useful for shallow genetic differentiation encountered in intra-specific studies of molecular diversity (Bailey *et al.* 1996; MacHugh *et al.* 1999). **Figure 2.7a** places each of the 149 *Bos indicus* sequences in a reduced median network phylogeny, constructed manually according to the methods outlined by Bandelt *et al.* (1995).

For comparative purposes, network phylogenies are shown for the *Bos taurus* mtDNA sequences sampled in Europe, the Near East, Africa and Japan (**Figure 2.7b**). These data comprise the mtDNA sequences previously reported by Mannen *et al.* (1998) and Troy *et al.* (2001), plus the 13 Near Eastern *Bos taurus* sequences reported for the first time in this study. The sequence variants defining each *Bos taurus* mtDNA haplotype encountered in the 13 Near Eastern reported here are presented in **Appendix 2.B**.

Phylogenetic network analysis demonstrates that Eurasian and African *Bos taurus* mtDNA sequences form star-like phylogenies positioned around one of four frequently represented ancestral haplotypes, namely T, T1, T2 and T3 (Troy *et al.* 2001). The 13 Near Eastern *Bos taurus* mtDNA sequence presented for the first time in this study also fall into one of these 3 predominant haplogroups (**Figure 2.7b**). Phylogenetic analysis of *Bos taurus* mtDNA sequences sampled in Japanese Black taurine cattle indicates that 53% of sequences cluster with the haplogroups of Near Eastern origin and are perhaps a legacy of the first introductions of Near Eastern domesticates to East Asia, *circa* 5,000 YBP (Mannen *et al.* 1998). The remaining Japanese *Bos taurus* sequences cluster with the J5 haplotype, believed to derived from a locally domesticated strain of aurochs (Mannen *et al.* 1998; **Figure 2.7b**).

The reduced median network in **Figure 2.7a** reveals that all 149 *Bos indicus* sequences fall into two distinct haplotypic clusters. Each *Bos indicus* sequence cluster consists of a highly

represented, centrally placed haplotype, namely Z1 or Z2, through which all other sequences root to the rest of the phylogeny. In addition, a minor haplotypic clade centred round the $Z2_{74b}$ haplotype (the Z3 haplogroup) was encountered as a subset of the mtDNA variability within the Z2 haplogroup. The Z3 haplogroup consists of a total of 20 mtDNA sequences, comprising 6 mtDNA haplotypes. The phylogenetic relationship between the four predominant *Bos taurus* haplotypes (T, T1, T2, T3 and J5) described by Troy *et al.* (2001) and Mannen *et al.* (1998) and the two predominant *Bos indicus* mtDNA haplotypes (Z1 and Z2) as described here is illustrated in **Figure 2.7c**.

Notably, most *Bos indicus* sequences fall within one or two base-pair substitutions of the presumed ancestral haplotypes, resulting in star-like phylogenies. These patterns of mtDNA diversity may be consistent with the derivation of almost all sequence variants from one of the two ancestral *Bos indicus* haplotypes within domestic history (Bradley *et al.* 1996; Cymbron *et al.* 1999; Troy *et al.* 2001). Star-like phylogenies are often indicative of small populations that have undergone demographic expansions at some point in the recent evolutionary past (Jorde *et al.* 1998). Thus, as with Eurasian and African *Bos taurus* mtDNA variability, the resulting star-like phylogeny surrounding each ancestral *Bos indicus* haplotype is suggestive of past population expansion events.

Figure 2.7a: A reduced median phylogenetic network featuring 149 *Bos indicus* mtDNA sequences

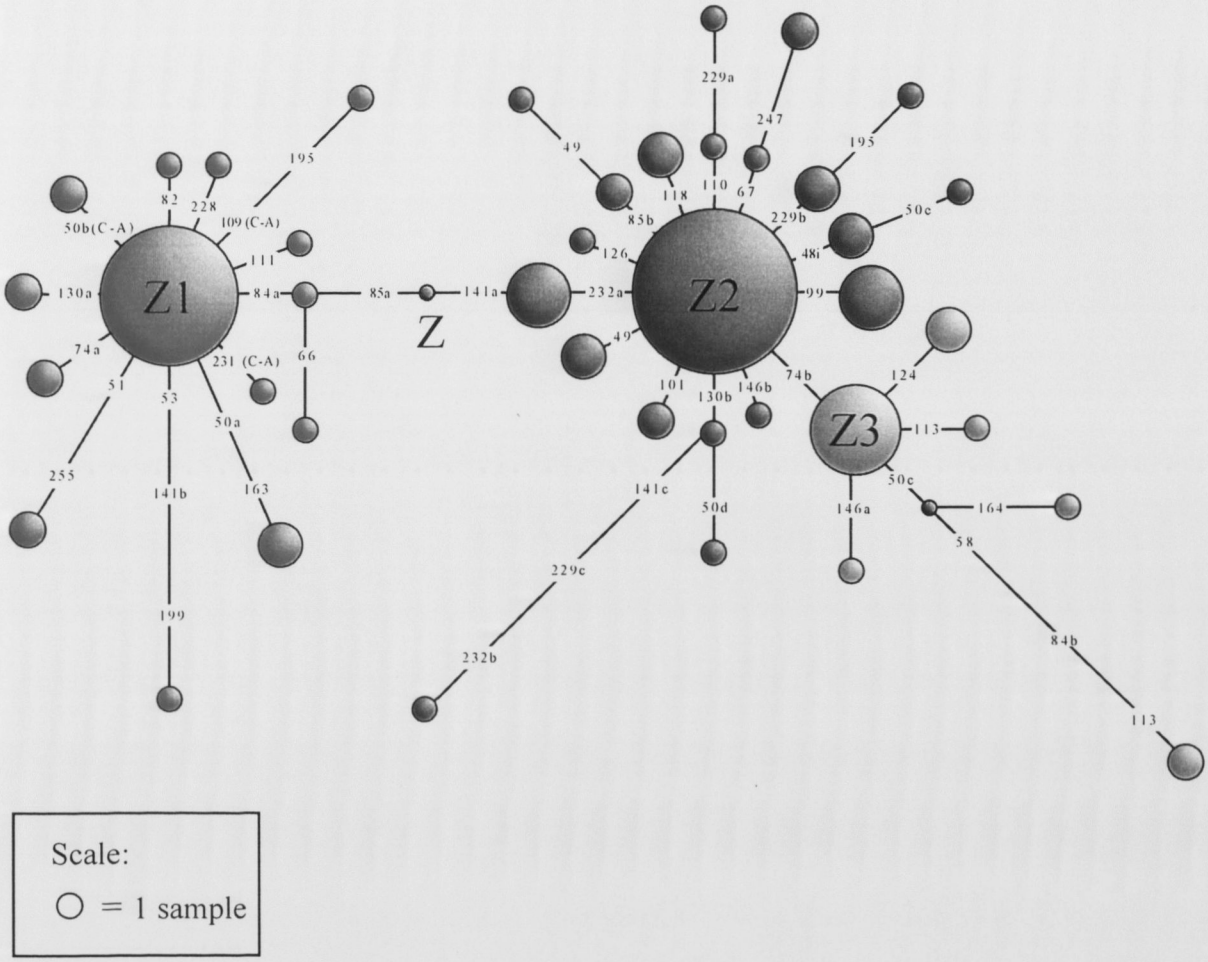
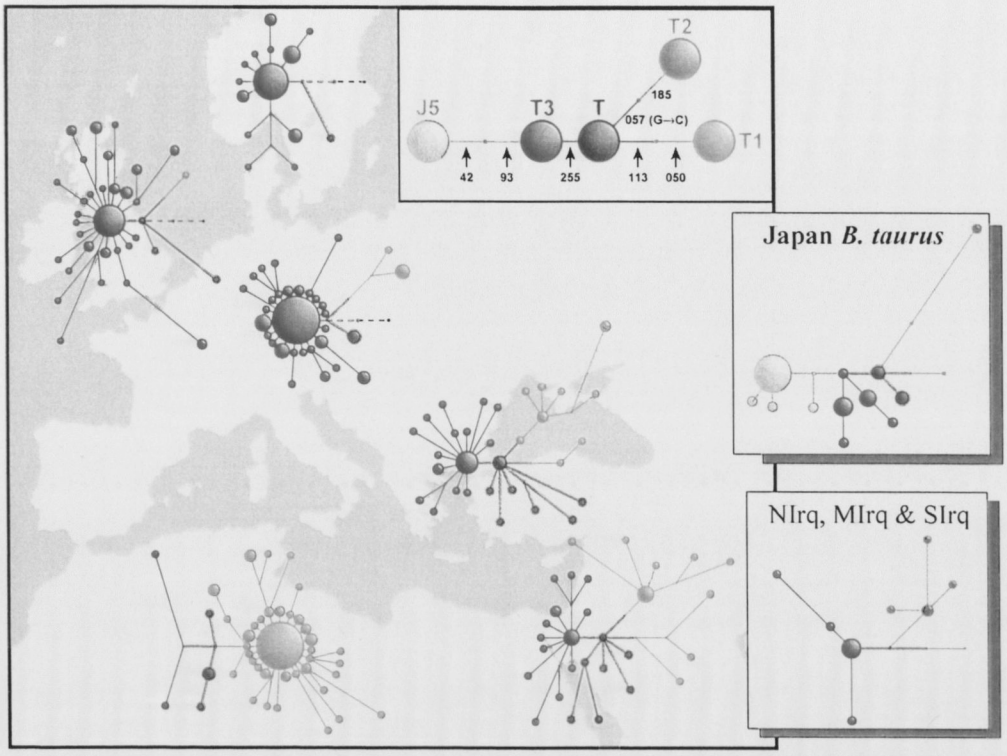


Figure 2.7b: Reduced median phylogenetic networks constructed for *Bos taurus* mtDNA sequences sampled in Europe, Africa, the Near East and Japan



Phylogenetic network constructed for North, Middle and South Iraqi *Bos taurus* mtDNA sequence reported for the first time in this thesis

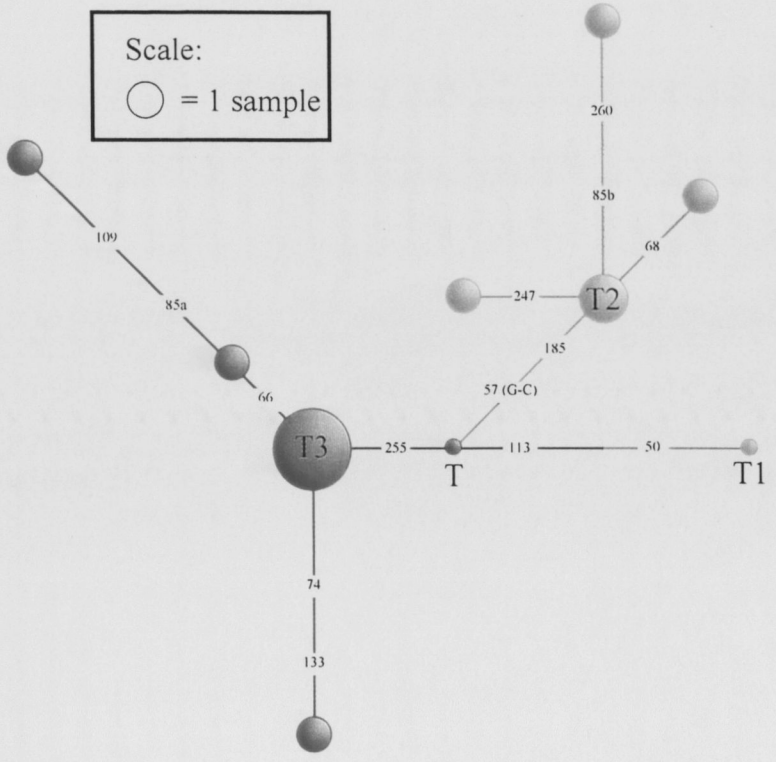


Figure 2.7c: The phylogenetic relationship between the predominant *Bos taurus* and *Bos indicus* mtDNA haplotypes

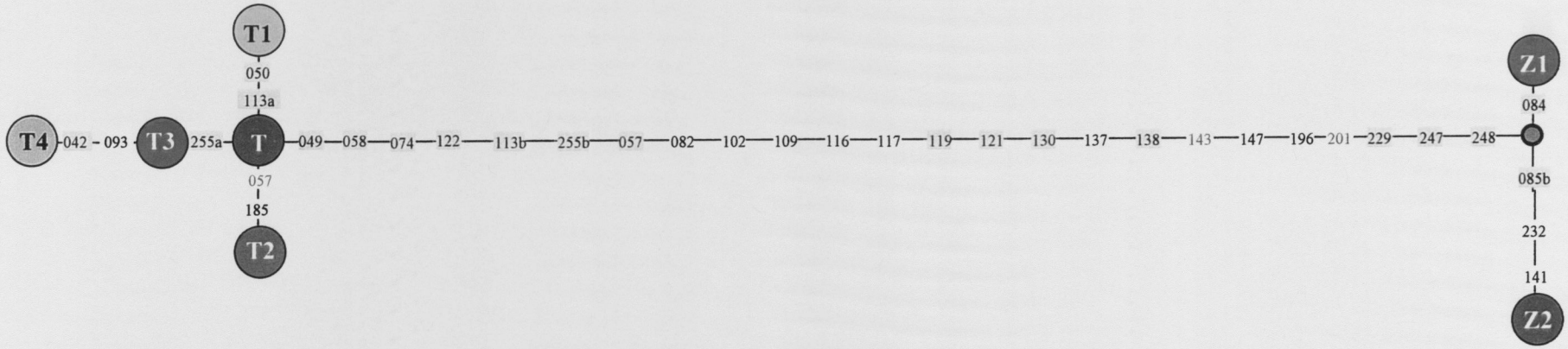


Figure 2.7a shows a reduced median phylogenetic network constructed from 149 *Bos indicus* mtDNA sequences reported in 9 Asian populations and 1 Brazilian cattle breed of Indian descent. Circles represent sequence haplotypes; the area is proportional to the frequency of the haplotype. The two most frequent haplotypes are labelled Z1 (n = 31) and Z2 (n = 39), respectively, and the moderately represented $Z2_{74b}$ (or Z3) haplotype (n = 12) is also shown. Red circle colouration denotes coalescence to the Z1 haplotype; blue indicates coalescence to the Z2 haplotype; green represents clustering with the Z3 haplotype. The Z3 sequence cluster (or Z3 haplogroup) is a subset of the variation within the Z2 haplogroup. Unsampled intermediate haplotypes are denoted by black dots. The unsampled putative ancestral haplotype of all sampled *Bos indicus* mtDNA sequences (the Z haplotype) is shown. Lines connecting sequence haplotypes represent nucleotide substitutions and are labelled as per Anderson *et al.* (1982), minus 16,000. All substitutions are transitions unless otherwise stated. Recurrent substitutions (homoplasies) are indicated by **a**, **b**, **c**, **d** and **e**. This network is reduced at 9 nucleotide positions - 16,050, 16,074, 16,084, 16,085, 16,130, 16,141, 16,146, 16,229 and 16,232. Recurrent mutations have been identified at 3 of these sites (16,050, 16,074 and 16,085) in a previous study of *Bos taurus* mtDNA sequence diversity (Troy *et al.* 2001). A total of 23 homoplasies were identified and resolved in this data set. The star-like phylogeny of derived mtDNA variants surrounding each of the two predominant *Bos indicus* haplotypes are consistent with past population expansions.

Figure 2.7b. The reduced median phylogenetic network constructed for *Bos taurus* sequences sampled in Europe (n = 207), Anatolia and the Middle East (collectively referred to here as the Near East; n = 80), Africa (n = 95) and Japan (n = 32; shown in inset) are shown in the top panel. The source of these data are referenced in **Table 5.1**. A network illustrating all 13 Near Eastern *Bos taurus* sequences presented for the first time in this study is presented in the bottom panel and in the inset in the top panel. Troy *et al.* (2001) demonstrated that almost all European mtDNA sequences belong to the T3 haplogroup, which is one of three haplogroups (T, T2 and T3) detected at appreciable frequencies in the Near East. Alternatively, African diversity consists primarily of members of the T1 haplogroup, which is encountered rarely elsewhere. The variation centred round the T, T1, T2 and T3 haplogroups produce star-type genealogies and are consistent with demographic growth (Troy *et al.* 2001). Mannen *et al.* (1998) demonstrated that some Japanese mtDNA sequences cluster with the *Bos taurus* haplogroups of Near Eastern origin (T, T2 and T3), while other mtDNA sequences group with the J5 haplotype, which is exclusive to cattle of North East Asia.

Figure 2.7c. A reduced median phylogenetic network illustrating the phylogenetic relationship between the four predominant *Bos taurus* mtDNA haplotypes (T, T1, T2, T3 and J5) as defined by Troy *et al.* (2001) and Mannen *et al.* (1998) and the two predominant *Bos indicus* mtDNA haplotypes (Z1 and Z2) described for the first time in this study. Numbers represent nucleotide positions within the control region less 16,000 as per Anderson *et al.* (1982). Numbers in red indicate transversions; numbers in blue indicate indels and numbers in grey indicate fast mutating sites as identified by Dr. Jill Bailey (Dept. of Genetics, TCD, personal communication). This figure has been modified from an original phylogenetic constructed by Dr. J. Bailey.

2.3.7 Regional *Bos indicus* mitochondrial DNA phylogenetic analysis

In order to assess the distribution of haplotypic variation at a regional level, populations were subdivided into four geographical groupings, based on their place of origin (**Table 2.4**). Phylogenetic networks constructed for each geographical region were constructed as per Bandelt *et al.* (1995) and are presented in **Figure 2.8**.

Table 2.4: Geographic divisions of *Bos indicus* mtDNA data for regional phylogenetic network analysis

Regional Group	Populations in each regional division	Number of mtDNA sequences within each geogroup
West Asia	NEZ, RS	34
Northwest India	HA, SA, TH	41
South India	DE, NE, ON	35
East Asia	LS,MN	39
Total		**149**

The 5 *Bos indicus* Nelore sequences were assumed to be of South Indian origin (Mason 1988) and were included with the South Asian regional group for phylogenetic network construction.

The genetic structure of *Bos indicus* sequence variation within and between the four designated regional groups was investigated using an AMOVA approach (Excoffier *et al.* 1992). The hierarchical structuring of genetic variation was compared and contrasted for

Bos indicus and *Bos taurus* regional mtDNA data collections and the results are presented in **Table 2.5**.

Table 2.5: Hierarchical components of variation for *Bos taurus* and *Bos indicus* regional groups

	mtDNA data	
Source of variation	***Bos taurus*[1]**	***Bos indicus*[2]**
Among groups	35.41% ($P \leq 0.001$)	13.12% ($P = 0.050$)
Among populations within	6.36% ($P \leq 0.001$)	7.17% ($P \leq 0.010$)
Within populations within groups	58.24% ($P \leq 0.001$)	79.72% ($P \leq 0.001$)

[1] ***Bos taurus* regional groups for AMOVA analysis:**

Europe (n = 207) - Aberdeen Angus, Berrenda, Charolais, Friesian, Galloway, German Black, Hereford, Highland, Icelandic, Jersey, Kerry, Limousin, Norwegian Red, Romagnola, Simmental, Sykia, Telemark, Westland Fjord; **the Near East** (n = 93) - Anatolian Black, Damascus, East Anatolian Red, Kurdish, Middle Iraq, North Iraq, South Anatolian Red, South Iraq, Turkish Grey; **Africa** (n = 95) - Butana, Egyptian, Kenana, Kapsiki, Kuri, Namchi, N'Dama, Somba, White Fulani; **Japan** (n = 32) - Japanese mtDNA sequences of Near Eastern origin; Japanese mtDNA sequences of North East Asian origin.

Note: All *Bos taurus* mtDNA sequence data used for AMOVA analysis comprised the data of Mannen *et al.* (1998), Troy *et al.* (2001) and the Near Eastern *Bos taurus* mtDNA sequence data presented in this study (**Table 5.1**).

[2] **Asian *Bos indicus* regional groups for AMOVA analysis:**

West Asia (n = 34) - Near East Zebu, Red Sindhi; **Northwest India** (n = 41) - Hariana, Sahiwal, Tharparker; **South India** (n = 30) - Desi, Ongole; **East Asia** (n = 39) - Laotian, Mongolian.

Note: For *Bos indicus* AMOVA analysis, the 5 Nelore *Bos indicus* sequences were omitted.

A substantial proportion of the total *Bos taurus* variance (35.41%) is attributable to differences between regional groups, namely Europe, the Near East and Africa. This is explained by the spatial assortment of the four major *Bos taurus* haplotypic clusters as reported by Troy *et al.* (2001), in which European mtDNA diversity is a subset of that

encountered in the Near East and African mtDNA variability is composed of members of a separate haplogroup rarely detected elsewhere (**Figure 2.7b**). With some symmetry, a considerable proportion of the total *Bos indicus* variance (13.12%) is accounted for by regional differences. This is most probably explained by the geographic distribution of the predominant *Bos indicus* haplogroups across Asia. The Z1 haplogroup predominates in Western Asia (68%), while only the Z2 haplogroup is encountered in the East Asian (100%). The proportion of Z2 haplogroup was also high in the Northwest Indian (Z2 = 63%) and South India (60%, Nelore data excluded).

Figure 2.8: Regional *Bos indicus* mtDNA reduced median phylogenetic networks

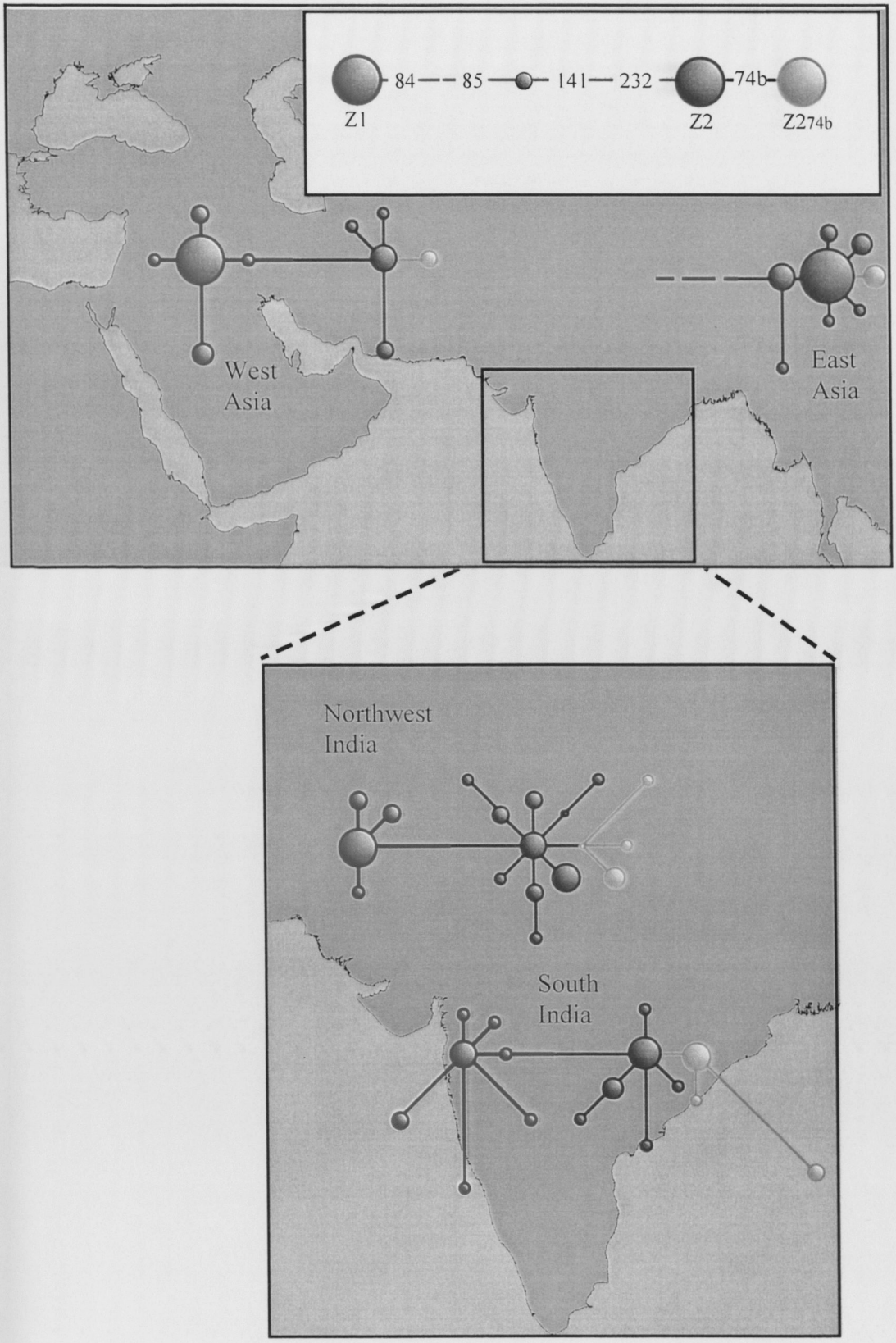


Figure 2.8. Reduced median networks constructed for *Bos indicus* mtDNA sequences sampled from four geographical regions – West Asia, Northwest India, South India and East Asia. The relationship between Z1 and Z2 haplotypes plus the $Z2_{74b}$ haplotype (denoted Z3) are shown in the inset. The spatial arrangement and colour coding of the skeletal network is preserved and presented for full data networks, which are placed on a background map according to region of origin. Haplotypes encountered in each region and unsampled intermediate nodes, are represented by coloured circles and small points, respectively. Haplotype frequency is proportional to the area of each circle. Colour coding indicates which of the two haplogroups sequences belong to. A cline in haplogroup distribution is evident from the regional networks, whereby the West Asian group displays the highest proportion of the Z1 haplogroup (Z1 = 68%, Z2 = 32%) and the East Asian group displays the highest proportion of the Z2 haplogroup (Z1 = 0%, Z2 = 100%).

2.3.8 Mitochondrial DNA haplotypic distribution

The geographic distribution of Z1 and Z2 haplogroup frequencies along an east-west axis, as described in **Section 2.3.6**, is also evident at the individual population level. Z1 haplogroup frequencies are highest in the two populations belonging to the West Asian group (NEZ = 71%; RS = 65%) and in the Sahiwal breed (66%) of the Northwest Indian group. Z1 haplogroup frequencies gradually decrease in an eastward direction with a complete absence of Z1 variants in both East Asian populations. The Z2 haplogroup is detected at high frequencies in the remaining Indian cattle populations (ON = 72%; DE = 51%; TH = 70%; HA = 88%). All 5 *Bos indicus* sequences detected in the Brazilian Nelore samples were identical to the Z2 haplotype. The proportion of Z1 and Z2 haplogroups for each cattle population is depicted graphically in **Figure 2.9**.

When the 20 Z3-type sequences (Z3 represents a subset of the mtDNA variation detected in the Z2 major haplogroup) are treated as a discrete haplotypic cluster, 45% of the Z3 mtDNA variants (9 sequences out of 20) are found within the populations of South India, 25% occur in Northwest Indian samples, 20% are encountered in East Asian populations and 10% are detected in West Asian samples.

Figure 2.9: Z1 and Z2 haplogroup proportions for each assayed population

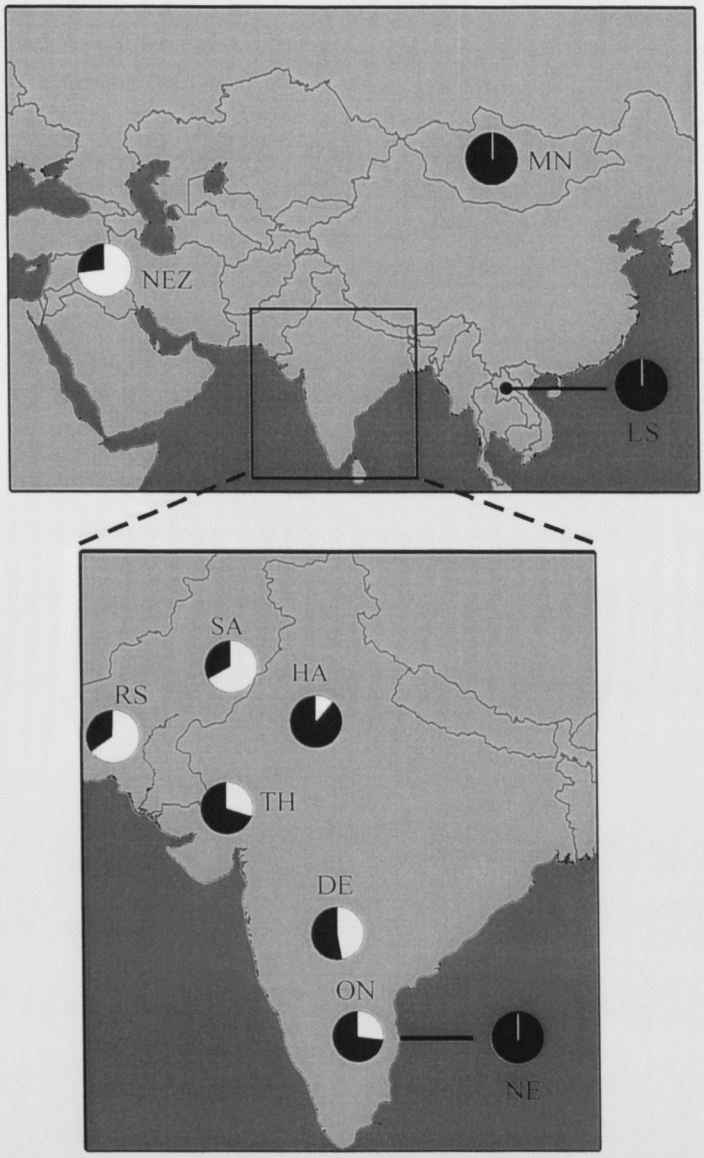


Figure 2.9. Z1 and Z2 *Bos indicus* mtDNA haplogroup proportions within all assayed Asian population Z1 and Z2 haplogroup proportions are shaded white and black, respectively. Pie char positioning represent approximate geographic locations of populations. Population codes are as p **Table 2.1**. An east-west geographic cline in haplogroup distribution is evident, with members of the Z haplogroup predominanting in West Asia and Z2-type sequence predominating in East Asi Z1 haplogroup proportions of *Bos indicus* mitochondrial chromosomes sampled in each population are as follow NEZ (71%); RS (65%); SA (66%); TH (30%); HA (11%); DE (47%); ON (27%); NE (0%), LS (0%); MN (0%

Statistical testing of haplogroup distributions

Three alternative statistical tests were performed to test the significance of the east-west distribution of Z1 and Z2 haplogroups. In order to concentrate on the distribution of the two predominant *Bos indicus* haplogroups (Z1 and Z2) across Asia, the 5 zebu mtDNA types encountered in the Nelore samples, all of which are identical, were omitted from all three statistical analyses.

(a) Chi-squared test of significance

For chi-squared tests of association between Z1 and Z2 haplogroup proportions and geography, populations were assigned in roughly equal numbers to a West Asian or East Asian category based on the longitude of population origin (given in **Table 2.1**). The chi-squared contingency table is shown below (**Table 2.6**). This statistical approach reveals a significantly higher incidence of the Z1 haplogroup in more westerly located populations and a higher incidence of the Z2 haplogroup in more easterly located populations across Asia ($P \leq 0.001$). This result was maintained when, (1) the Hariana population was assigned to the East Asian category (*i.e.* West Asia sample size = 66; East Asia sample size = 78) ($P \leq 0.001$) and, (2) the Desi population was assigned to the West Asian category (*i.e.* West Asia sample size = 94; East Asia sample size = 50) ($P \leq 0.001$).

(b) Spearman rank test

For Spearman rank-correlations, the proportion of Z1 and Z2 haplogroups within each population (given in **Figure 2.9**) was ranked with the longitudinal co-ordinates for each population as given in **Table 2.1**. Significant associations between the higher incidence of Z1 in more westerly originating populations ($P \leq 0.05$) were revealed.

Table 2.6: Chi-squared contingency table for geographic distribution of *Bos indicus* haplogroups

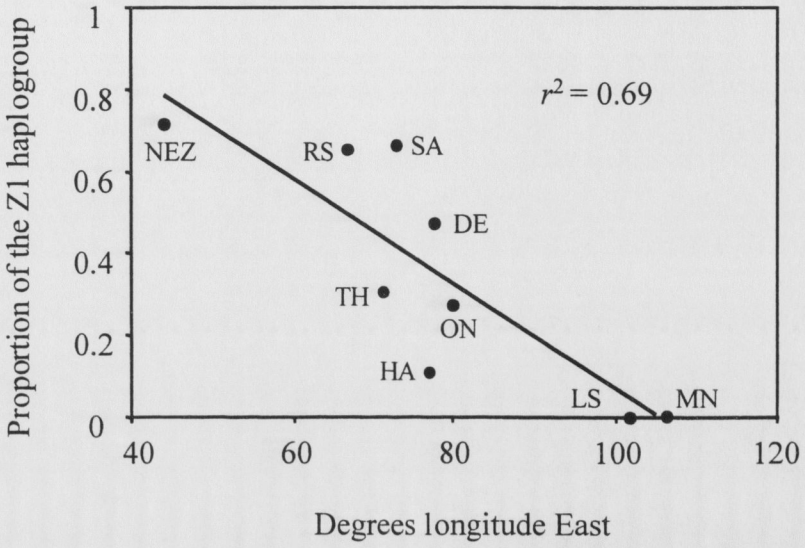
Western populations	No. samples with Z1/Z2 haplotypes	Eastern populations	No. samples with Z1/Z2 haplotypes
Near East Zebu	14 (10/4)	Desi	19 (9/10)
Red Sindhi	20 (13/7)	Ongole	11 (3/8)
Sahiwal	12 (8/4)	Laotian	30 (0/30)
Tharparker	20 (6/14)	Mongolian	9 (0/9)
Hariana	9 (1/8)		
Total	75	Total	69

Geographical division	Number of Z1 haplotypes	Number of Z2 haplotypes	Total
West	38	37	75
East	12	57	69
Total	50	94	144

(c) Regression analysis

The proportions of the Z1 and Z2 haplogroup were calculated for each population and plotted against their longitudinal co-ordinates (**Figure 2.10**). A significantly negative correlation was observed between the incidence of the Z1 haplogroup and longitude (r^2 = 0.69, $P \leq 0.05$).

Figure 2.10: Regression plot of population Z1 haplogroup proportions with respect to degrees longitude



2.3.9 Multi-dimensional scaling of inter-population genetic distances

To elucidate the genetic relationships among the Asian *Bos indicus* populations analysed here, pairwise F_{ST} values (**Appendix 2.C**) were plotted in two-dimensions following multi-dimensional scaling (**Figure 2.11**). Two-dimensional MDS co-ordinates are presented in **Appendix 2.D**. For comparative purposes, an unrooted neighbor-joining (NJ) phylogenetic tree, also constructed from these pairwise F_{ST} values, is shown in **Figure 2.11**. Comparison of the spatial arrangement of the populations in the MDS plot and NJ tree reveal some similarities. Firstly, the Near Eastern zebu (NEZ), Sahiwal (SA) and Red Sindhi (RS) populations and the Laotian (LS) and Nelore (NE) populations form two population clusters, which occur at extremities of both the MDS plot and NJ tree. It is therefore plausible that dimension 1 of the MDS plot and the major internal branch of the NJ tree accounts for the differentiation of *Bos indicus* populations displaying the highest proportion of the Z1 and Z2 haplogroups, respectively. Secondly, the Mongolian (MN) population is encountered as an outlier in the MDS plot, which is differentiated from the LS-NE population group along dimension 2. Similarly for the NJ tree, the MN population is separated from this central population cluster by a long external branch. Thirdly, the Ongole (ON), Tharparker (TH) and Hariana (HA) populations form a population cluster which occupies a central position along dimension 1 in the MDS plot and the major internal branch of the NJ tree, with the Desi (DE) population displaying some divergence from this cluster in both.

Figure 2.11: MDS plot and unrooted NJ phylogenetic tree summarising inter-population F_{ST} genetic distances among the zebu populations analysed in this study

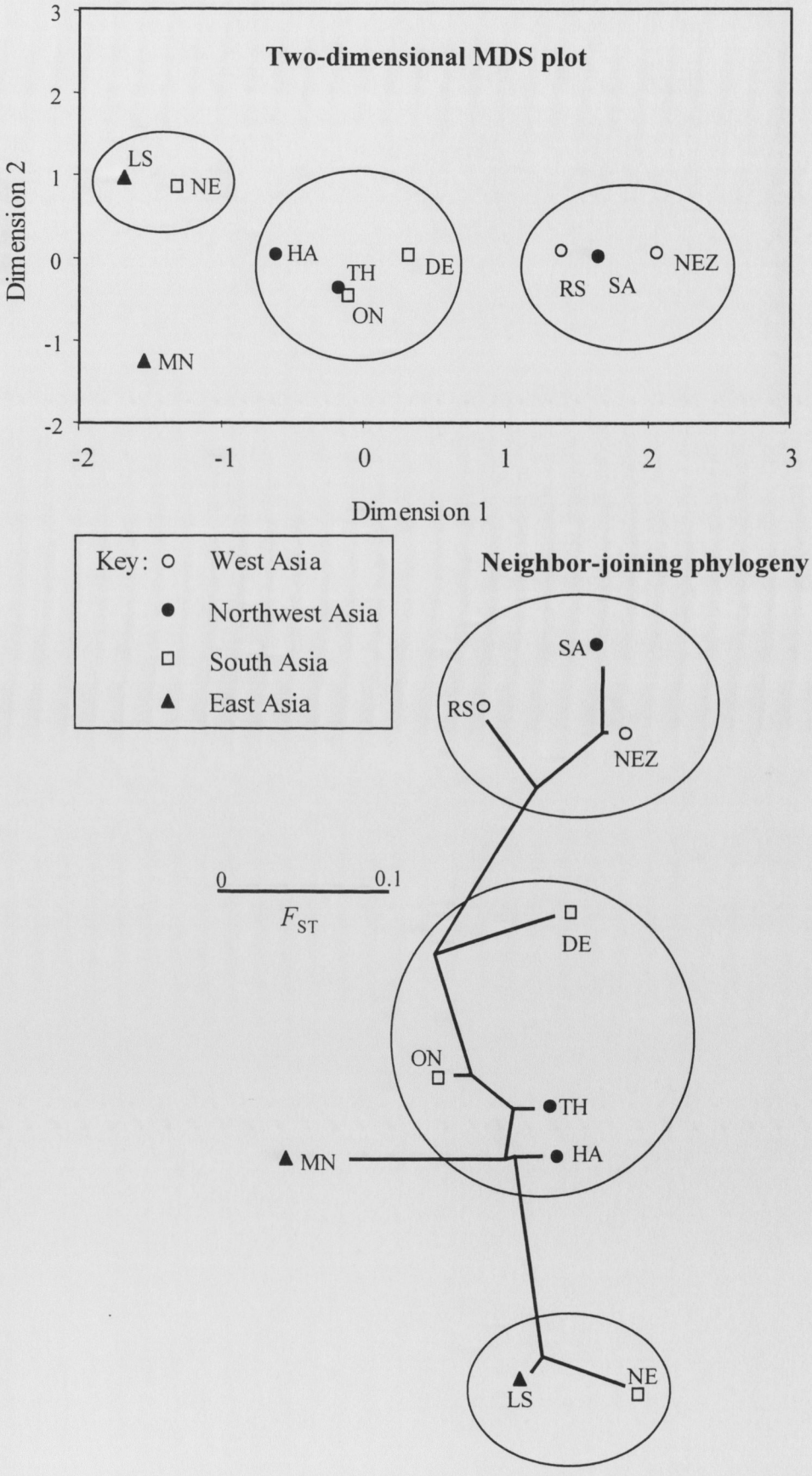


Figure 2.11. Two-dimensional MDS plot and NJ phylogenetic tree summarising pairwise F_{ST} genetic distances (Slatkin 1991; Slatkin 1995) between the zebu populations analysed in this study. The two-dimensional MDS plot was constructed following multi-dimensional scaling of population pairwise F_{ST} values. The MDS dimensions summarise 97.2% of the total *Bos indicus* mtDNA variation. Population codes: DE (Desi), HA (Hariana), LS (Laotian), MN (Mongolian), NE (Nelore), NEZ (Near East zebu), ON (Ongole), RS (Red Sindhi), SA (Sahiwal), TH (Tharparker).

2.3.10 Demographic history of *Bos indicus* cattle

Rogers & Harpending (1992) have argued that historical population dynamics can be inferred by analysing the properties of mtDNA mismatch distributions. Ragged, multi-modal distributions are consistent with long-term demographic stability. Alternatively, past population expansions, during which mtDNA lineage survivorship is high, produce smooth, uni-modal (single peaked), bell-shaped mismatch distributions (Rogers & Harpending 1992; Sherry *et al.* 1994). Episodes of population growth generate waves in mismatch distributions that move to the right along the horizontal axis (which represents the number of nucleotide differences between pairs of mtDNA sequences) as mutational time accrues. Thus, the more ancient an expansion event is, the greater the distance the wave travels along this axis and consequently, the modal value of the mismatch distribution is directly related to the time depth of expansion events (Excoffier & Schneider 2000).

Pairwise mismatch distributions were constructed for the each of the four regional *Bos indicus* mtDNA groups (presented in **Table 2.4**) and for all sampled *Bos indicus* mtDNA sequences. For comparative purposes, mismatch distribution curves were constructed for *Bos taurus* mtDNA sequences sampled in the Near East, Europe, Africa and Japan. A mismatch distribution was also constructed for a combined *Bos taurus* mtDNA sequences data set incorporating all mtDNA sequences sampled in Europe, the Near East, Africa and

Japan (Mannen *et al.* 1998; Troy *et al.* 2001 supplemented with 13 Near Eastern *Bos taurus* sequences reported for the first time in this thesis). All pairwise mtDNA sequence mismatch distribution curves are presented in **Figure 2.12**.

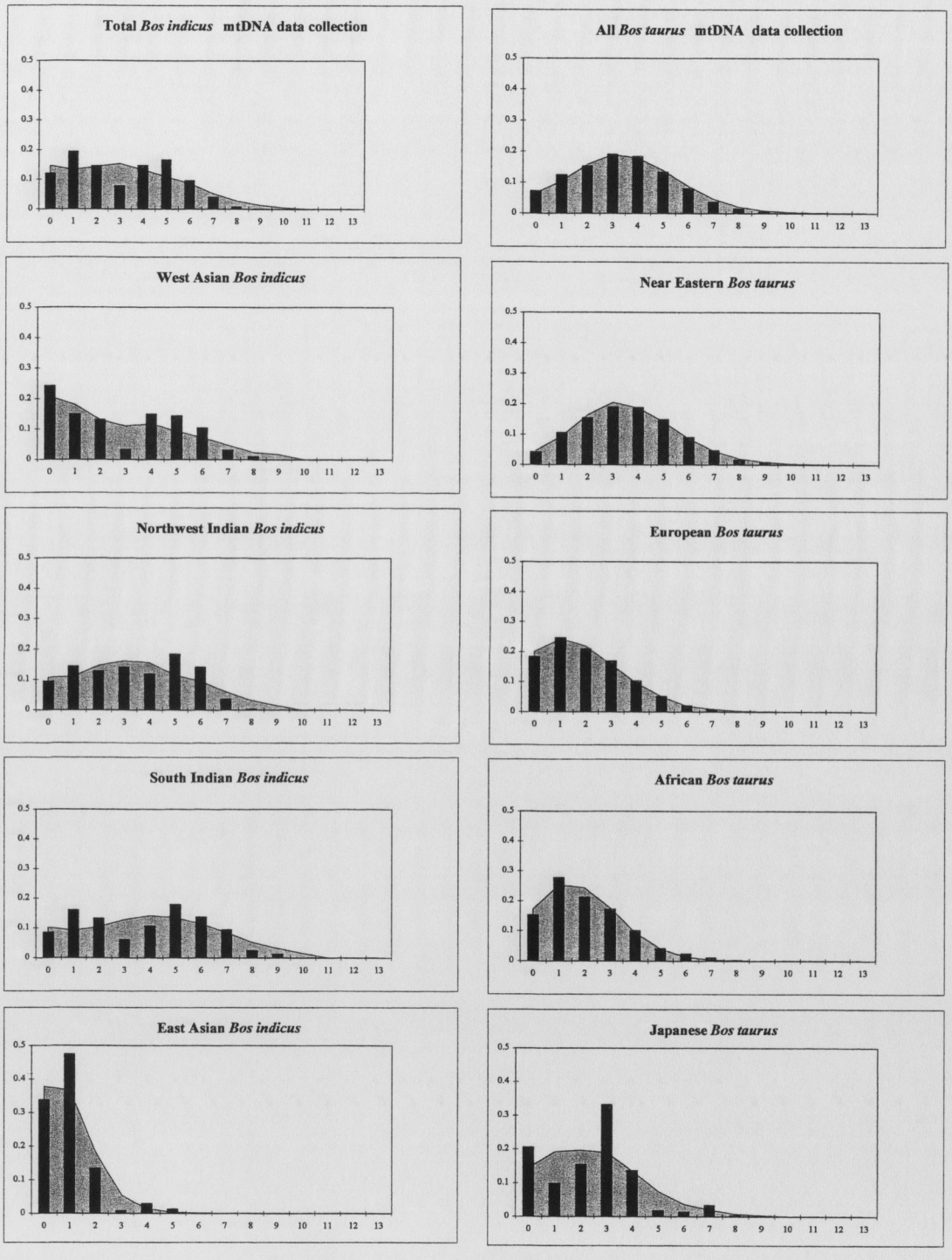
Two approaches were used to assess the validity of each of the observed mismatch distributions under a model of demographic expansion. The first method uses a parametric bootstrap approach implemented by Excoffier & Schneider (1999), which is incorporated in the ARLEQUIN VERSION 2.0 software package. This method compares the sum of square deviations (SSD) between the observed mismatch distributions and simulated distributions expected in expanding populations. Simulated mismatches are generated under a model of population growth using 3 demographic parameters estimated from the observed mismatch curve (Rogers & Harpending 1992): θ_1, θ_2 [θ before and after an expansion, respectively, where θ equals the product of the female effective population size (N_{ef}) and twice the mutation rate (2μ) (*i.e.* $2\mu N_{ef}$)] and τ (the date of the expansion in units of mutational time, given by $\tau = 2\mu t$, where t is time in generations). The sum of square deviations (SSD) generated for the simulated distributions is statistically compared to the SSD generated for the observed mismatch and a hypothesis of expansion is accepted if the proportion of simulated mismatches displaying a smaller SSD than the observed mismatch is less than 5% (Excoffier & Schneider 2000).

The second approach employs Fu's F_S statistic of selective neutrality (Fu 1997), which is particularly sensitive to episodes of population growth. The purpose of this test is to detect significant deviations from neutrality characterised by an excess of derivative haplotypes compared to a demographically stable population, which tend to yield significant negative F_S values (Excoffier & Schneider 2000). Excoffier & Schneider (1999) state that an F_S

value should only be considered as significant if its corresponding *P*-value is less than or equal to 0.02.

In addition, the "smoothness" of each mismatch distribution was evaluated using the raggedness statistic (Harpending *et al.* 1993). Low raggedness values of this statistic are indicative of smooth, uni-modal mismatch distribution curves and are suggestive of demographic expansion. Low observed raggedness values are considered significant if the proportion of simulated values less than the observed values is smaller than 5%. The statistical parameters for each of the observed sequence mismatch distribution illustrated in **Figure 2.12** are presented in **Table 2.7**.

Figure 2.12: *Bos indicus* and *Bos taurus* mtDNA sequence mismatch distributions



Black shading indicates the observed mismatch distribution; grey shading indicates the theoretically expected mismatch for population expansion.

Figure 2.12. MtDNA control region mismatch distribution curves for *Bos taurus* and *Bos indicus* mtDNA regional groups and total *Bos taurus* and *Bos indicus* data collections. The number of sequence mismatches is denoted on the horizontal axis; relative frequencies are represented on the vertical axis.

***Bos indicus* mtDNA regional group sample sizes**: West Asia (n = 34); Northwest India (n = 41); South India (n = 35) and East Asia (n = 39); all *Bos indicus* data (n = 149).

***Bos taurus* mtDNA regional group sample sizes**: the Near East (n = 93; Troy *et al.* 2001 supplemented with the 13 Near Eastern *Bos taurus* sequences reported in this thesis); Europe (n = 207; Troy *et al.* 2001), Africa (n = 95; Troy *et al.* 2001); Japan (n = 32; Mannen *et al.* 1998). The mismatch distribution for the total *Bos taurus* data collection (n = 427) comprised all European, Near Eastern, African and Japanese *Bos taurus* mtDNA sequences.

Table 2.7: Statistical parameters for the mtDNA sequence mismatch distributions presented in Figure 2.12

	Bos indicus mtDNA sequence data					*Bos taurus* mtDNA sequence data				
Statistical analysis	**All *B. indicus*** ($n = 149$)	**WA** ($n = 34$)	**NWI** ($n = 41$)	**SI** ($n = 34$)	**EA** ($n = 39$)	**All *B. taurus*** ($n = 427$)	**NE** ($n = 93$)	**EUR** ($n = 207$)	**AFR** ($n = 95$)	**JAP** ($n = 32$)
Sum of square deviations (SSD)	0.0122	0.0167	0.0092	0.0139	0.0178	<0.001	0.0003	0.0004	0.0023	0.0395
P (sim. SSD $\geq$ obs. SSD)	0.5400	0.7400	0.4300	0.4900	0.0100	0.9400	0.9100	0.9200	0.6900	0.0400
Raggedness (rag) statistic	0.0265	0.0394	0.0220	0.0275	0.1512	0.0100	0.0156	0.0150	0.0304	0.0998
P (sim. rag $\geq$ obs. rag)	0.7400	0.8100	0.7700	0.6500	0.0400	0.8700	0.8000	0.9700	0.7200	0.1100
Observed F_S values	-10.2344	-1.4429	-5.1772	-4.6123	-3.6316	<-100.00	-13.7307	-10.5294	-11.3318	-4.1453
P value of obs F_S value	($P \leq 0.003$)	($P \leq 0.256$)	($P \leq 0.021$)	($P \leq 0.030$)	($P \leq 0.015$)	($P \leq 0.008$)	($P \leq 0.002$)	($P \leq 0.005$)	($P \leq 0.001$)	($P \leq 0.020$)

Table 2.7. Statistical parameters for each mtDNA control region sequence mismatch distributions presented in **Figure 2.12**. Two approaches are used to infer an episode of demographic expansion. Firstly, a hypothesis of demographic expansion is accepted if the proportion of simulated (sim) sum of square deviations (SSD) less than the observed (obs) SSD is smaller than 5% (Excoffier & Schneider 2000). Secondly, departures from Fu's F_S statistic of selective neutrality (denoted by significantly negative F_S values) are interpreted as evidence for population growth (Fu 1997; Excoffier & Schneider 2000). Significant low raggedness statistic values (Harpending *et al.* 1993) are indicative of smooth, uni-modal mismatch distributions, which are consistent with a model of demographic growth as proposed by Rogers & Harpending (1992). Abbreviations for the mtDNA control region sequence data collections are as follows: WA (West Asian *Bos indicus*); NWI (Northwest Indian *Bos indicus*); SI (South Indian *Bos indicus*), EA (East Asian *Bos indicus*); NE (Near Eastern *Bos taurus*); EUR (European *Bos taurus*); AFR (African *Bos taurus*); JAP (Japanese *Bos taurus*). Sample sizes for each mtDNA data collection are shown in brackets.

Smooth, uni-modal sequence mismatch distributions are evident for the total *Bos taurus* mtDNA sequence data collection and all individual *Bos taurus* regional groups, with the exception of the *Bos taurus* mtDNA sequences sampled in Japanese Black cattle. The mismatch curves generated for the Near Eastern, European *Bos taurus* and total combined *Bos taurus* mtDNA sequence data collections are close to the theoretically expected uni-modal distributions for population expansions. Also, all 4 regional *Bos taurus* data sets and the total combined *Bos taurus* mtDNA data collection display significant negative F_S values, which are indicative of demographic expansion.

Notably, the mismatch distribution curve generated for all sampled *Bos indicus* mtDNA sequences is bi-modal. Presumably, the leftmost portion of the curve (modal value of 1) is representative of the number of mismatches within each of the two predominant *Bos indicus* haplogroups, whereas the rightmost curve (modal value of 5) describes the pattern of variation between the predominant *Bos indicus* haplogroups. None of the observed *Bos indicus* mismatch distributions conform to the theoretical curves generated under a model

of demographic growth as proposed by Excoffier & Schneider (1999). However, the data set comprising all sampled *Bos indicus* mtDNA variants and 3 of the regional *Bos indicus* mtDNA collections reveal significantly negative F_S values. The West Asian *Bos indicus* group does not display significantly negative F_S values and may be suggestive of prolonged demographic stability or a population bottleneck.

2.3.11 Dating historic population expansions

Troy *et al.* (2001) have previously shown that time depths of putative population expansions may be estimated from mtDNA-derived star-like phylogenetic topologies as the product of the average mutational distance of derivative haplotypes from a central variant within haplotypic clusters [the parameter 'rho'] and the estimated rate of mutation for the DNA sequence under consideration. The distribution of mutational steps from the presumed ancestral haplotype within each *Bos indicus* (Z1 and Z2) and *Bos taurus* (T, T1, T2, T3 and J5) haplogroup are shown in **Figure 2.13**. The distribution of the number of mutational steps for all sampled *Bos indicus* and *Bos taurus* mtDNA sequences around the putative ancestral *Bos indicus* Z and *Bos taurus* T haplotypes are also shown in **Figure 2.13**. To test if the number of mutations within the control region are occurring at a constant evolutionary rate (*i.e.* in a molecular clock-like manner), Poisson distributions, which can be used to predict a number of events (*i.e.* the number of mutations) over a given time period (*i.e.* within the history of domestication), were calculated for all *Bos* haplogroups using rho values. Notably, the distribution of mutations from the founder mtDNA variant within each *Bos* haplogroup follows a Poisson distribution, suggesting that mutational events within the bovine control region are occurring at a constant evolutionary rate. Furthermore, this finding supports the use of rho values and the mutation rate to estimate the time depth of putative demographic expansions from mtDNA sequence data.

Figure 2.13: The distribution of mutational steps from the presumed ancestral haplotype within each major sampled *Bos indicus* and *Bos taurus* haplogroup

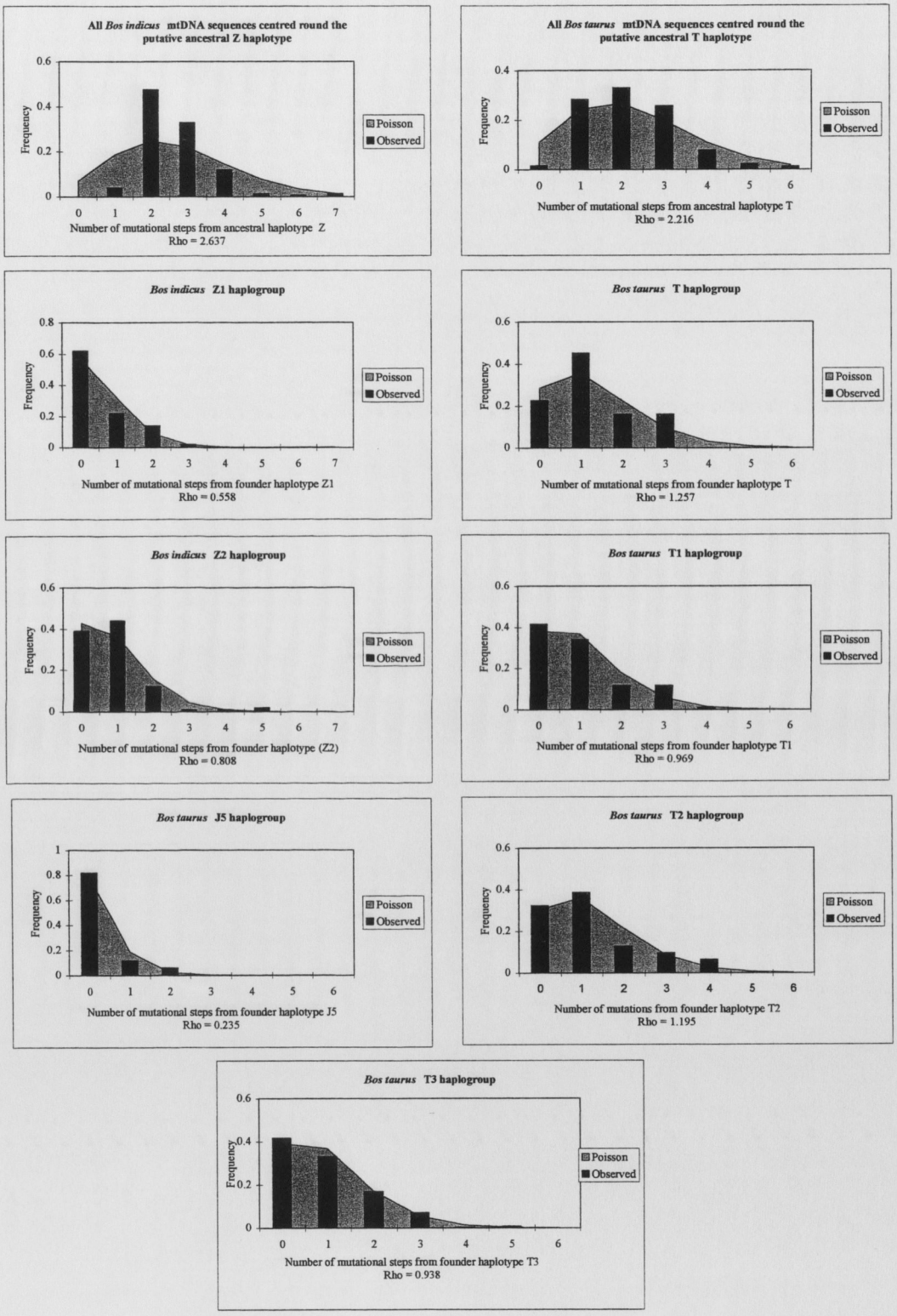


Figure 2.13. The distribution of mutational steps from each of the predominant founding *Bos taurus* (T, T1, T2, T3 and J5) and *Bos indicus* mtDNA lineages (Z1 and Z2). In addition, the mutational distributions are shown for all sampled *Bos indicus* and *Bos taurus* samples centred round the putative ancestral Z (**Figure 2.7a**) and T, mtDNA variants (**Figure 2.7b**). The number of mutational steps (ignoring insertions and deletions) from the ancestral mtDNA sequence and the corresponding number of sequence are shown along the horizontal and vertical axes, respectively. Rho values (*i.e.* the average number of mutations from the ancestral haplotype) are given for each *Bos* haplogroup and the total sampled *Bos indicus* and *Bos taurus* mtDNA sequences around the Z and T haplotypes, respectively. Notably, the distribution of mutations within each *Bos* haplogroup follows a Poisson distribution suggesting that mutation events within the bovine control region are occurring a constant rate of evolution.

Troy *et al.* (2001) have previously estimated the one-lineage substitution rate for the 240 bp region of the mitochondrial control region analysed in this study as 38% per million years, which equates to 1 substitution every 10,928 years. Using this mutation rate and corresponding rho values, Troy *et al.* (2001) estimated the time depths of the expansion events surrounding the four predominant Eurasian (T, T2 and T3) and African (T1) *Bos taurus* haplogroups. This method was used to estimate the 95% confidence intervals (CI) for the time depths of the putative expansions within each of the two predominant *Bos indicus* haplogroups using the CRED computer package (Macaulay 1998). This approach was also used to estimate the overall coalescence of all sampled *Bos indicus* mtDNA sequences around the presumed ancestral Z sequences. Information regarding the time depth estimates for the expansion surrounding each *Bos* haplogroup is presented in **Table 2.8**.

Table 2.8: Expansion dates for the predominant *Bos indicus* and *Bos taurus* mtDNA haplogroups

mtDNA haplogroup	No. of mutational steps from central haplotype	Total number of samples in haplogroup	Rho (average mutational distance from central haplotype)	Haplogroup expansion time (YBP)	95% confidence interval range for haplogroup expansion (YBP)
All *Bos taurus* (T)	945	427	2.216	24,200	not calculated
T	39	31	1.258	13,750	9,800-18,400
T1	91	94	0.968	10,580	8,500-12,900
T2	37	31	1.195	13,000	9,200-17,600
T3	238	254	0.937	10,230	9,000-11,600
J5	4	17	0.235	2,570	700-5,600
All *Bos indicus* (Z)	393	149	2.637	28,800	26,000 - 31,700
Z1	28	50	0.558	6,100	4,100 - 8,600
Z2	80	99	0.808	8,800	7,000 – 10,800

Table 2.8. Estimated time depths for the demographic expansions around the Z1 and Z2 *Bos indicus* haplotypes and 5 major *Bos taurus* (T, T1, T2, T3, J5) mtDNA haplogroups. *Bos taurus* mtDNA sequences sampled in Europe, the Near East and Africa were taken from Troy *et al.* (2001) and supplemented with the additional 13 Near Eastern *Bos taurus* sequences reported for the first time in this study, while those sampled in Japan were reported by Mannen *et al.* (1998). The overall coalescence times of all sampled Asian *Bos indicus* mtDNA sequences (n = 149) around the central Z sequence and all sampled Eurasian, African and Japanese *Bos taurus* mtDNA sequences (n = 427) around the central T sequence are also presented. Expansion dates were estimated by multiplying the average mutational distance from the presumed ancestral haplotype (ignoring insertions and deletions) within each *Bos* haplogroup (denoted rho) by the average mutation rate for the 240 bp fragment of the bovine mtDNA control region under consideration, previously estimated by Troy *et al.* (2001) as 38% per MYA (1 million years).

In accordance with the observations of Troy *et al.* (2001), the 95% CI range estimates for the putative expansions around the T, T1, T2 and T3 haplotypes, may be viewed as somewhat consistent with the historic time frame for *Bos taurus* domestication as suggested by archaeological data. The 95% credible regions for the major Z1 (4,100 – 8,600 YBP) and Z2 (7,000 – 10,800 YBP) haplogroup expansion times also concur with the 8,000-year history of *Bos indicus* domestication (Meadow 1993; Meadow 1996). The mean expansion times around the Z2 haplogroup (8,800 YBP) pre-dates the expansion around the Z1 haplogroup (6,100 YBP). Furthermore, the mean expansion times estimates for the four predominant Eurasian and African *Bos taurus* haplogroups pre-date those estimated for the primary *Bos indicus* haplogroups. The overall coalescence for all sampled domestic *Bos indicus* samples around the hypothetical ancestral sequence, Z, is estimated here as 26,000-31,700 YBP.

2.4 DISCUSSION AND CONCLUSIONS

2.4.1 A South Asian origin for *Bos indicus* cattle

Genetic and archaeological data have provided evidence for the derivation of both domestic cattle taxa, *Bos taurus* and *Bos indicus*, from at least two geographically disparate, genetically distinct aurochsen subspecies (Zeuner 1963; Grigson 1980; Meadow 1993; Loftus *et al.* 1994a; Loftus *et al.* 1994b; Bradley *et al.* 1996; Meadow 1996; MacHugh *et al.* 1997). It seems likely that *Bos taurus* cattle were first domesticated from the Eurasian aurochs, *Bos primigenius primigenius*, within the Neolithic complexes of the Near East *circa* 8,400 YBP (Perkins 1969). There is also a body of molecular and fossil evidence which support, albeit controversially, the separate domestication of African *Bos taurus* cattle from a localised strain of aurochs (Wendorf & Schild 1994; Smith 1995; Bradley *et al.* 1996; Troy *et al.* 2001; Hanotte *et al.* 2002). Alternatively, fossil evidence points towards the derivation of *Bos indicus* cattle from the South Asian aurochs, *B. p. namadicus*, within the emerging civilisations of Baluchistan and the Indus Valley, during the 7th millennium BC (Grigson 1980; Meadow 1984; Meadow 1993; Meadow 1996).

Samples from a domestic centre of origin are expected to retain more ancestral variation and exhibit greater genetic variability. Conversely, one expects an accompanying reduction in derived populations. This has been successfully illustrated for *Bos taurus* mtDNA sequences, where diversity estimates in populations from the cradle of *Bos taurus* domestication, the Near East, greatly exceed those obtained for derivative European breeds (Loftus *et al.* 1999; Troy *et al.* 2001). Here, within *Bos indicus* samples, mtDNA sequence variability is clearly highest within domestic populations from the Indian subcontinent, with pronounced reductions in haplotypic and nucleotide variation apparent in *Bos indicus* mitochondrial chromosomes sampled from the Near East and East Asia.

Notably, the mean number of pairwise differences estimated for the zebu populations of the Indian subcontinent are comparable to those computed for Near Eastern *Bos taurus* haplotypes, thus suggesting a domestic origin of *Bos indicus* cattle within the Indian subcontinent. However, an alternative diversity estimate, θ_S, shows that in the majority of cases intra-regional diversity is higher in regional samples of *Bos taurus* mtDNA chromosomes, compared with regional *Bos indicus* mtDNA samples, suggesting that the most regional *Bos taurus* populations have a higher effective population size than regional *Bos indicus* groups.

2.4.2 The genetic legacies of *Bos indicus* and *Bos taurus* migrations

The detection of *Bos indicus* mtDNA lineages within morphologically taurine breeds of the Near East are consistent with the genetic legacies of historic secondary introgressions of *Bos indicus* to the region. Notably, as with autosomal admixture proportions (Loftus *et al.* 1999), there is a geographic cline in the distribution of *Bos indicus* mtDNA types across the Near East. Near Eastern breeds occurring in closer proximity to the Indian subcontinent exhibit higher proportions of *Bos indicus* mtDNA variants.

The data presented here also suggest symmetrical eastward movements and secondary introductions of *Bos taurus* into South Asia. Four *Bos taurus* introgressors were identified at low frequencies in 3 morphologically zebu populations situated to the Northwest of the Indian subcontinent. Analysis of these taurine introgressors shows that they cluster with two of the three predominant *Bos taurus* haplogroups that occur at appreciable frequencies within the Near East, namely T and T2 (Troy *et al.* 2001). The absence of *Bos taurus* T3 type sequences suggests that the observed *Bos taurus* introgressor mtDNA sequences are the result of historic admixture with cattle from the Near East and not the result of recent introductions of European cattle to the region as part intensive breeding schemes (Epstein

& Mason 1984). Historic eastward *Bos taurus* introgression into the Indian subcontinent is further supported by a recent autosomal microsatellite allelic variation analysis in which *Bos taurus* ancestry has been detected within South Asian zebu cattle populations. Notably, *Bos taurus* introgression appears to be stronger in North Indian populations than those from peninsular India (Kumar *et al.* in press).

The distribution of *Bos indicus*-specific genetic markers across the Near East, together with the apparent partial *Bos taurus* ancestry of domestic South Asian zebu is suggestive of ancient introgression events between the two domestic cattle taxa. It is plausible that these genetic exchanges occurred *circa* the 4th millennium BC as a result of extensive trade routes between civilisations of the Near East and the Indus Valley, thus providing opportunities for genetic exchange (Zeuner 1963; Allchin & Allchin 1968; Naik 1978). Faunal and artistic evidence indicate that domestic zebu cattle were present in Sistan (present day Iran) and other more westerly situated sites in Mesopotamia and the Lebanon *circa* 3,000-1,000 BC (Zeuner 1963; Clason 1978). In addition, evidence from stamp seals and terracotta figurines indicate that *Bos taurus* domesticates were present in the Neolithic complexes of Harappa and Mojenho-daro of the Indus Valley, during the 3rd millennium BC (Zeuner 1963; Allchin & Allchin 1968; Naik 1978).

Analysis of the archaeobotanical record also provides evidence for contact between the civilisations of the Indus Valley and Near East during this period. Domesticated pulses of Southwest Asian antiquity, such as lentils, flax (linseed), pea, and chickpea are believed to have dispersed eastwards into South Asia between the 3rd and 2nd millennia BC (Meadow 1996). Similarly, some domesticated species, most notably the water buffalo (Zeuner 1963) and sesame (Charles 1993 [cited by Kumar *et al.* in press]), spread westwards from South Asia. Thus, it appears that the influx of zebu cattle into the Near East and the

counter flow of taurine cattle into South Asia formed part of the continuous transfer of subsistence species between the two regions (Meadow 1996).

2.4.3 The domestic and demographic history of *Bos indicus* cattle

Founding Bos taurus and Bos indicus mtDNA lineages

Domestication entails the recruitment of a limited sampling of a subset of the genetic variation that exists within wild populations into the domestic pool. As mtDNA is maternally inherited *sans* recombination, mutation is the only source of mtDNA sequence variation and consequently one expects that all extant domestic mtDNA lineages will, at some point in their evolutionary history, coalesce to a founder haplotype, which represents a wild mitochondrial variant adopted into the domestic pool via the capture and taming of wild aurochsen females (MacHugh & Bradley 2001).

Troy *et al.* (2001) and Mannen *et al.* (1998) have demonstrated that all mtDNA sequences sampled in Eurasian, African and Japanese *Bos taurus* cattle coalesce to 1 of 5 founder haplotypes, designated T, T1, T2, T3 and J5. With some similarity, the reduced median network phylogenetic analysis presented in this study reveals that all sampled *Bos indicus* mtDNA sequences coalesce to one of two ancestral haplotypes (denoted Z1 and Z2) that are probably representative of moderately divergent wild matrilines incorporated into the domestic *Bos indicus* pool during the early phases of zebu domestication. The data also reveal the existence of a minor *Bos indicus* mtDNA haplogroup (Z3 haplogroup) that constitutes a subset of the variation within the Z2 lineage.

The paucity of domestic *Bos taurus* and *Bos indicus* founder lineages is in itself an interesting observation. Presumably, the process of cattle domestication would have involved the recruitment of a greater number of wild matrilines into the domestic pool than

the six encountered in extant domestic *Bos* populations. Thus, the *Bos taurus* and *Bos indicus* mtDNA data suggests that that a substantial number of wild matrilines have become extinct during the past 9,000 years of cattle domestic history. This finding highlights the 'self-pruning' nature of mtDNA evolutionary trees in which lineages are stochastically lost within the population if they are not transmitted to succeeding generations along the female line (Avise *et al.* 1987). Conversely, the mtDNA sequence members of the T, T1, T2, T3, J5, Z1 and Z2 domestic *Bos* haplogroups represent the extant mtDNA descendants, modified by mutation, of those wild matrilines that have survived and been maintained throughout the course of domestic history into modern extant domestic populations.

Genetic evidence for demographic expansion

Troy *et al.* (2001) have demonstrated that all *Bos taurus* mtDNA sequences sampled in Eurasia and Africa form star-like phylogenies centred around one of the four founder haplotypes, namely T, T1, T2 and T3. The star-like patterns of derived mtDNA variants surrounding each ancestral haplotypes comprises the sequence variation that has accrued via mutation during domestic history and are indicative of past population expansion events, during which mtDNA lineage survivorship is high. Conversely, phylogenetic analysis of Japanese *Bos taurus* mtDNA does not reveal a star-like genealogy centred round the J5 haplotype.

As with *Bos taurus* variability, *Bos indicus* mtDNA haplotypes derived from either of the two founding haplotypes (Z1 and Z2) produce star-like phylogenetic topologies, which are consistent with population growth. Two alternative treatments were used to infer a model of demographic expansion for *Bos indicus* cattle. Firstly, star-like phylogenetic topologies generate smooth, uni-modal curves in pairwise sequence mismatch distributions (Rogers &

Harpending 1992), such as those generated for 3 *Bos taurus* geographical groupings (namely the Near East, Europe and Africa) and the total combined *Bos taurus* mtDNA data set. Here, the variation contained within all sampled *Bos indicus* mtDNA sequences generates a smooth bi-modal mismatch distribution curve. Bi-modal curves are also observed for most regional *Bos indicus* groupings with the exception of the East Asian *Bos indicus* mismatch distribution. Consequently, none of the *Bos indicus* mtDNA mismatch distributions conform to the theoretically expected uni-modal curve for demographic expansions. It is probable that the leftmost curve within these bi-modal distributions corresponds to nucleotide differences within *Bos indicus* haplogroups, while the rightmost curve represents nucleotide differences between *Bos indicus* haplogroups. Notably, the inter-haplogroup curve markedly leads the inter-haplogroup curve by approximately four substitutions suggesting that the two putative ancestral *Bos indicus* haplotypes were genetically differentiated prior to domestication.

Secondly, significant negative F_S values (Fu 1997), which are suggestive of population growth (Excoffier & Schneider 2000), are observed for almost all regional *Bos taurus* and *Bos indicus* mtDNA samples. Only the West Asian *Bos indicus* grouping exhibits non-significant departures from neutrality, which may be indicative of demographic equilibrium or alternatively, a bottleneck-induced reduction in haplotypic diversity.

Troy *et al.* (2001) have concluded that the star-shaped phylogenetic topologies, bell shaped mismatch distributions curves and significant deviations from Fu's F_S statistic noted for *Bos taurus* mtDNA geographical groupings are consist with a model of demographic expansions from a limited number of founder individuals adopted into the domestic pool during domestication. Accordingly, the significantly negative F_S values generated for almost all regional *Bos indicus* mtDNA samples and star-shaped phylogenetic topologies

are interpreted here as evidence for demographic growth. Under the assumption that the advent of agro-pastoralism would have resulted in a profound and continued increase in domestic livestock numbers from an initially shallow genetic and numeric foundation, the patterns of *Bos indicus* variation are most probably the genetic signatures of domestication-induced population expansions.

Demographic expansion times

Troy *et al.* (2001) have argued that the 95% confidence interval range of estimates presented for the 4 predominant *Bos taurus* haplotypic clusters sampled in Eurasia (T = 9,800-18,400 YBP; T2 = 9,200-17,600 YBP; T3 = 9,000-11,600 YBP) and Africa (T1 = 8,500-12,900 YBP), although subject to large margins of error, are roughly consistent with the historic 10,000-year timeframe for *Bos taurus* domestication (Smith 1995; Bradley *et al.* 1996; Troy *et al.* 2001). Similarly, the 95% credible region estimates for the expansion times around the Z1 (4,100-8,600 YBP) and Z2 (7,000-10,800 YBP) haplogroups concur with the archaeologically established 8,000-year history of *Bos indicus* domestication within South Asia (Meadow 1993; Meadow 1996). The overall coalescence time of all sampled domestic *Bos indicus* mtDNA sequences around the hypothetical ancestral node Z (26,000-31,700 YBP). This shallow divergence of all domestic *Bos indicus* sequences indicates that modern zebu domesticates were derived from an aurochs population that may have displayed limited mtDNA sequence diversity.

Of note, the mean expansion times estimated for the Eurasian and African *Bos taurus* haplogroups are appreciably older than those computed for the primary *Bos indicus* haplogroups. This finding tentatively suggests that the domestication and subsequent expansion of *Bos taurus* cattle within the Neolithic Near Eastern complexes pre-dated the emergence of domestic *Bos indicus* cattle within South Asia. This observation is supported

by archaeological data, which indicate that the initial domestication of *Bos taurus* within the Near East (*circa* 8,400 YBP; Perkins 1969), may have occurred some 500-1000 years prior to the initial domestication of *Bos indicus* (*circa* 8,000-7,000 YBP) (Meadow 1993; Meadow 1996).

The Eastern margin of the Middle East, which encompasses parts of Afghanistan, Eastern Iran and Baluchistan (present-day Pakistan, west of the River Indus), is often considered to have been on the receiving end of the numerous technological and cultural developments initiated in more advance areas situated further west (Vishnu-Mittre 1978; Meadow 1984; Meadow 1993). Consequently, Baluchistan and surrounding regions may actually be secondary centres of domestication in which the transition to agro-pastoralism was triggered by the arrival of early farmers, who introduced the ideas and techniques associated with agriculture and animal husbandry, together with some floral domesticates, to South Asia from the Near East. Palaeontological analysis suggests that the emergence of agro-pastoralism within South Asia, *circa* 9,000-8,000 YBP, involved imported domestic taxa of Near Eastern origin, such as barley, einkorn wheat and emmer wheat, and locally domesticated taxa, including cotton, dates, gram bean, sheep and goats (Meadow 1993; Meadow 1996; Bradley 2000; Luikart *et al.* 2001). Thus, the domestication of zebu cattle, along with that of other domestic fauna and floral species, within South Asia may be viewed as part of a larger pattern of agro-pastoralist innovations which began in the Near East *circa*, 8,500 BC and which diffused eastwards prompting similar, somewhat later developments further east, *circa* 7,000 BC (Vishnu-Mittre 1978; Meadow 1996; Diamond 2002).

It is important to emphasise, however, that while these mtDNA-derived time depth estimates are consistent with a domestication-induced expansion of *Bos taurus* and *Bos*

indicus cattle within the historic time frame of animal herding, they are subject to margins of error. This is primarily attributable to the presence of hypervariable sites, which mutate faster than other nucleotide sites within the control region, resulting in mutation rate heterogeneity and contributing to the high number of recurrent mutations (homoplasies) observed within the control region (Stoneking 2000). However, the distribution of mutations within each of the predominant *Bos taurus* and *Bos indicus* haplogroups follow a Poisson distribution, suggesting that mutation events within the bovine mtDNA control region occur at a relatively constant rate of evolution, thus strengthening claims for animal domestication/herding-induced population expansions.

Phylogeographic structuring of Bos indicus haplotypes

The four major *Bos taurus* haplotypic clusters described by Troy *et al.* (1998) display marked geographic distributions. One of these clusters (T3) predominates in Europe and is one of three (T, T2 and T3) that are encountered at appreciable frequencies within sequences of Near Eastern origin – supporting a Near East origin for European cattle. In contrast, African mtDNA sequence diversity is almost exclusively comprised of members of a separate haplogroup (T1), which is rarely encountered in other regions. Additionally, Mannen *et al.* (1998) have identified a mitochondrial haplogroup that seems to be particular to *Bos taurus* populations of North East Asian provenance.

A similar geographic distribution of moderately divergent *Bos indicus* haplotypes is also evident across Asia. Populations from West Asia display higher proportions of the Z1 haplogroup while the Z2 haplogroup predominates in domestic populations from East Asia. This finding is also reflected in the appreciable genetic differentiation observed between regional *Bos indicus* groupings (13.12%; $P = 0.05$) as indicated by hierarchical AMOVA analysis. Considerable frequencies of both haplogroups are detected within

Indian populations. Phylogeographic structuring of the primary *Bos indicus* haplogroups is further supported by three alternative statistical tests, namely chi-squared tests of association, Spearman rank correlations and regression analysis, each of which reveal a significant correlation between the higher frequencies of mtDNA sequence members of the Z1 haplogroup in Western South Asia and of the Z2 haplogroup in Eastern Asia. Phylogeographic structuring of *Bos indicus* haplogroups is also evident from the spatial arrangement of zebu populations in both the MDS plot and NJ phylogenetic tree. In these populations exhibiting the highest proportion of Z1 and Z2 haplotypes form distinct clusters, which occur at the extremities.

A model of domestication for Bos indicus cattle

Based on the levels and patterns of mitochondrial sequence variation presented in this study two scenarios may be postulated regarding the domestication and demographic history of *Bos indicus* cattle within South Asia:

Scenario 1

The data may point to a single domestic origin for *Bos indicus* cattle within South Asia, in which at least two moderately distinct wild progenitor mitochondrial lineages were incorporated into the domestic pool during the phase of aurochs capture. Available archaeological evidence suggests that *Bos indicus* domestication may have occurred in the early Neolithic complexes of Baluchistan, some 9000-8,000 years ago, which were most probably influenced by the earlier agrarian innovations of the Near East (Jarrige & Meadow 1980; Meadow 1993; Meadow 1996). The phylogenetic signatures of expansion associated with the predominant *Bos indicus* Z1 and Z2 haplotypes are consistent with the dramatic, sustained increases of flocks due to successful animal management. Undoubtedly, increased stock yields prompted the diffusion of zebu animals southwards

into peninsular India and eastwards into Southern and Eastern India and further into East Asia. Although a single domestic origin for *Bos indicus* cattle in which both founder haplotypes were incorporated into the domestic pool is plausible, it is difficult to envisage how the two predominant haplogroups became spatially assorted across South and Central Asia within the given 8,000-year timeframe of zebu domestication.

Scenario 2

It may be postulated that the domestication of *Bos indicus* cattle involved the recruitment of at least two genetically distinct maternal founders at more than one location. These two events may have been spatially and temporally independent of each other or may have consisted of one primary and a subsequent secondary event, suggesting that the domestication processes were somehow culturally intertwined.

The geographic division of the Z1 and Z2 *Bos indicus* haplogroups across Asia tentatively suggests that these putative domestication centres may have been separated along an east-west axis, such that the Z1 and Z2 haplogroups originated in more westerly and easterly locales, respectively. This supports previous, albeit controversial, archaeological assertions suggesting that there were at least two separate domestications of zebu cattle. Some authors contend that the distinctive ashmounds of the Deccan Plateau, which date to the 3rd millennium BC, represent ancient cattle pens which may have served as enclosures for the capture and taming of local wild aurochs animals (Allchin & Allchin 1968; Allchin & Allchin 1974; Allchin & Allchin 1982). Furthermore, the detection of a haemoglobin variant, exclusive to South Indian zebu populations, has lead Naik (1978) to deduce a separate origin of Southern and Northern Indian *Bos indicus*. It is interesting to note, however, that the mean time depth for the expansion centred round the more easterly located Z2 haplogroup, pre-dates that for the more westerly Z1 haplogroup. If the Z2

haplogroup originated within peninsular India, these data suggest that domestication-induced population growth first occurred in peninsular India, rather than in Baluchistan. On the other hand, it is possible that the members of Z2 haplogroup represents the modern descendants of the first wild founder haplotype to enter the domestic pool at the site of Mehrgarh, *circa* 8,000-7,000 YBP and that the earlier expansion times and increased Western Asian incidence associated with Z1 haplogroup are the genetic signatures of a zebu domestication that occurred later than, and to the west of Mehrgarh. Indeed, faunal remains and hand-made figurines provide evidence for the co-existence of *B. p. namadicus* and domestic *Bos indicus* at the site of Shahr-I-Sokhta in Eastern Iran, *circa* 4,000-3,000 BC, which may indicate a local Iranian domestication of zebu some 2,000 years after their purported initial domestication at Mehrgarh (Bökönyi 1997). Thus, the extant mtDNA sequence variation within South Asia may be representative of a primary domestication in Baluchistan which involved the incorporation of the Z2 haplotype into the domestic pool and a more recent, westerly situated, secondary domestication event involving the recruitment of the Z1 haplotype.

While separate domestic origins of South Asian *Bos indicus* cattle is an attractive hypothesis, the data presented in this study are not totally conclusive regarding this issue. Intensified sampling of domestic *Bos indicus* populations from other regions, particularly the Ganges Valley in Northeast India, peninsular India, the Iranian Plateau and Bangladesh would undoubtedly aid the further elucidation of the domestic history of *Bos indicus* cattle and may reinforce or refute multiple domestications of *Bos indicus* cattle. Ancient DNA analysis mtDNA sequence data isolated from faunal remains of the presumed progenitor of *Bos indicus*, *Bos namadicus*, may also add novel layers of inference pertaining to the domestic origins of *Bos indicus* cattle. However, as the arid climatic conditions that exist

within South Asia do not encourage the preservation of ancient genetic material, this would undoubtedly be problematic.

CHAPTER 3

THE GENETIC ORIGINS OF CARIBBEAN CREOLE CATTLE

3.1 INTRODUCTION

3.1.1 The historical origins of American Creole cattle

Creole or Criollo cattle, is the term given to American domestic cattle populations of Iberian descent. Domestic cattle did not exist on the American continent until 1493, when Christopher Columbus landed a small number of Spanish cattle on the Caribbean Island of Hispanola (Wilkins 1978). During the early phases of European colonisation, cattle were continually introduced to the Caribbean archipelago as attempts to initiate and develop agro-pastoralism on the islands intensified. It is understood that the majority of the cattle transported to the Americas during this time were of Andalucian (Southwest Spain) origin. However, Rouse (1973) states that the initial domesticates introduced by Columbus were loaded on the Canary Islands, which had been populated with breeds from Northwest Spain since becoming a Spanish colony in 1479 (Primo 1992). Indeed, several authors have described numerous physical similarities between modern Creole animals and the modern breeds of Northwest Spain and Andalucia, such as the Retinta Andaluza and the Berrenda.

The difficulties associated with long-distance sea fearing ensured that the numbers of Spanish animals imported to the Americas during this period were minimal. Despite this, multiplication of the founding population was extremely rapid. By 1525 domestic cattle populated most Caribbean islands and had dispersed throughout much of Central and South America, with the exception of Brazil (de Alba 1978). Alternatively, the first domestic cattle of Brazil were introduced by Portuguese colonisers in 1534. Some of these animals were imported directly from Portugal however, indirect shipments of Portuguese cattle from the Cape Verde Islands (a Portuguese colonial islands situated off the west coast of Senegal) have also been documented (Primo 1992).

During the 16th and 18th centuries African *Bos taurus* cattle were introduced to the Americas, possibly as a consequence of slave trade routes. It is believed that these cattle were shipped directly from the French, Portuguese and Spanish colonies located along the Guinean Gulf coast (Maillard *et al.* 1993; Maillard & Maillard 1998). Purchases of West African cattle by the later-established French settlers of Martinique and Guadeloupe during the 17th and 18th century have also been reported (Maillard & Maillard 1998).

Towards the beginning of the 19th century, importations of humped *Bos indicus* cattle became widespread throughout the Americas. Zebu animals were initially imported from Senegal and were employed to supply the traction necessary for crop cultivation (Rouse 1973). During the 20th century, African zebu animals were gradually replaced by Indian zebu domesticates, which exhibit increased durability in tropical environments. Crossbreeding of zebu cattle with native Creole animals led to marked rises in cattle productivity in tropical regions. Consequently, large-scale importations of Indian zebu cattle became widespread throughout the Americas during this time (Wilkins 1978). **Figure 3.1** illustrates the reported introductory routes of domestic cattle to the Americas.

Figure 3.1: Introductory routes of cattle to the Americas

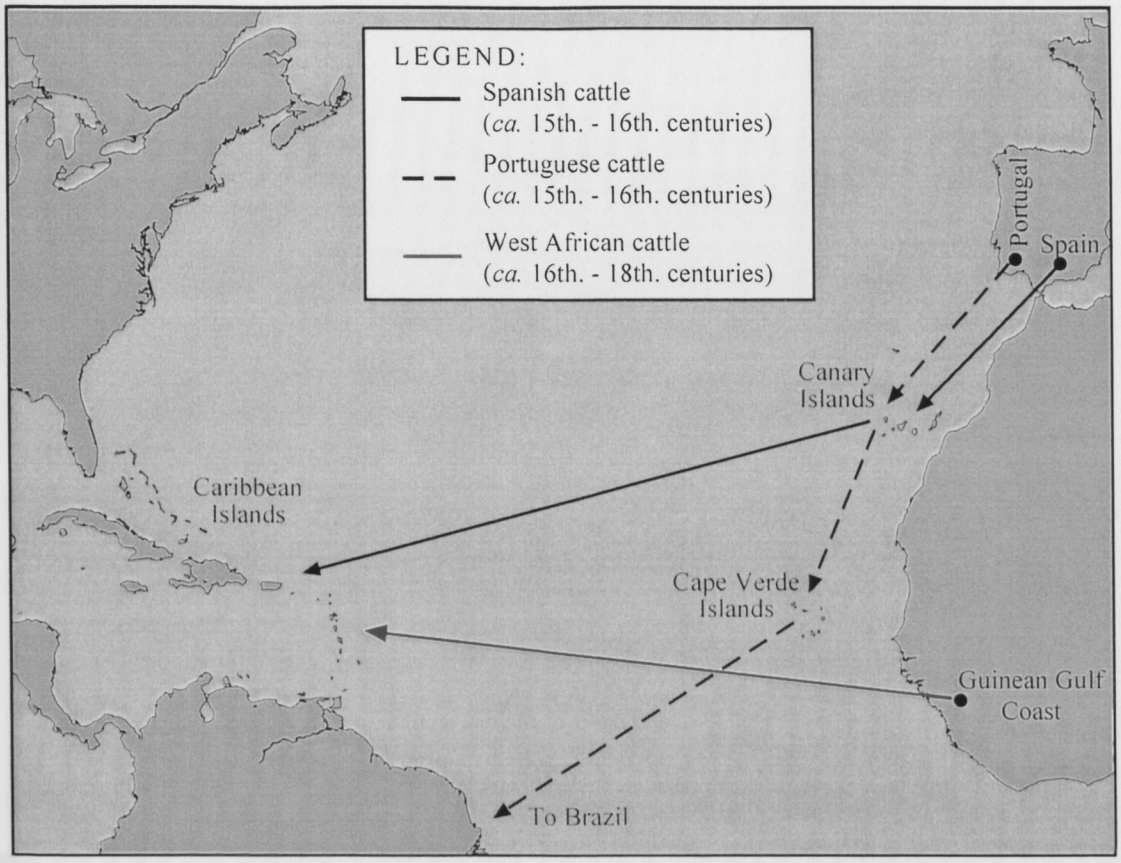


Figure 3.1. Migratory routes of domestic cattle to the Americas. Although some Iberian cattle were brought to the Americas indirectly via the Canary Islands and the Cape Verde Islands, direct shipments from Spain and Portugal have been documented (de Alba 1978). Large shipmenats of *Bos indicus* animals from India were also brought to the Americas during the 19th and 20th centuries (not shown on map).

3.1.2 The genetic ancestry of *Bos taurus* cattle and research objectives

Examination of mtDNA control sequence variation in domestic cattle population samples of wide geographical range, have been pivotal in the elucidation of bovine phylogeography. Initial investigations identified a deep, pre-domestic divergence between *Bos taurus* and *Bos indicus* cattle (Loftus *et al.* 1994a). Subsequent phylogenetic analyses demonstrate that all *Bos taurus* mtDNA sequences fall into one of five ancestral star-like haplotypic clusters, which are geographically distributed across the Old World (Mannen *et al.* 1998; Troy *et al.* 2001; depicted in **Figure 2.7b**). The observed phylogeographic partitioning of *Bos taurus* haplogroups allows geographical exceptions to be securely identified as secondary introductions, such as the detection of African haplotypes in southern Portuguese breeds, where they are probably a legacy of historical North African influence (Cymbron *et al.* 1999).

Rouse (1973) states that the waves of cattle importation to the Americas since the 16th century have greatly altered the Iberian-derived genetic foundations of Creole cattle. In this study, mtDNA sequence variation was employed to reconstruct the genetic ancestry of Caribbean Creole cattle. Comparisons between the resulting patterns of variation and those previously established for European, Iberian and African breeds allowed the effects of secondary introgressions and phylogenetic relationships to be evaluated. A survey of allelic variation at three microsatellite loci was also used to investigate the consequences of secondary influxes on the autosomal genome of extant Creole populations. MacHugh *et al.* (1997) have shown that the selected loci display a series of alleles, which are exclusive to West African *Bos taurus* populations, while other alleles are specific to populations of *Bos indicus* ancestry. Analysis of the frequencies of these diagnostic alleles permitted autosomal admixture proportions to be quantified.

3.2 MATERIALS AND METHODS

3.2.1 Cattle samples and DNA extraction

Serum samples from Antiguan and St. Lucian Creole cattle populations were collected by Dr. Alex Morrow (University of Edinburgh) during several field excursions between 1987 and 1988. DNA was extracted from these samples in January 2001 using the techniques outlined in **Section 5.2**. DNA extracted from Guadeloupe Creole and Brazilian Nelore cattle samples was donated to the Molecular Population Genetics laboratory (TCD) by various collaborators prior to the initiation of this study. A map illustrating the approximate locations of each assayed population is presented in **Figure 3.2**.

Figure 3.2: The geographic locations of the Creole populations analysed in this study

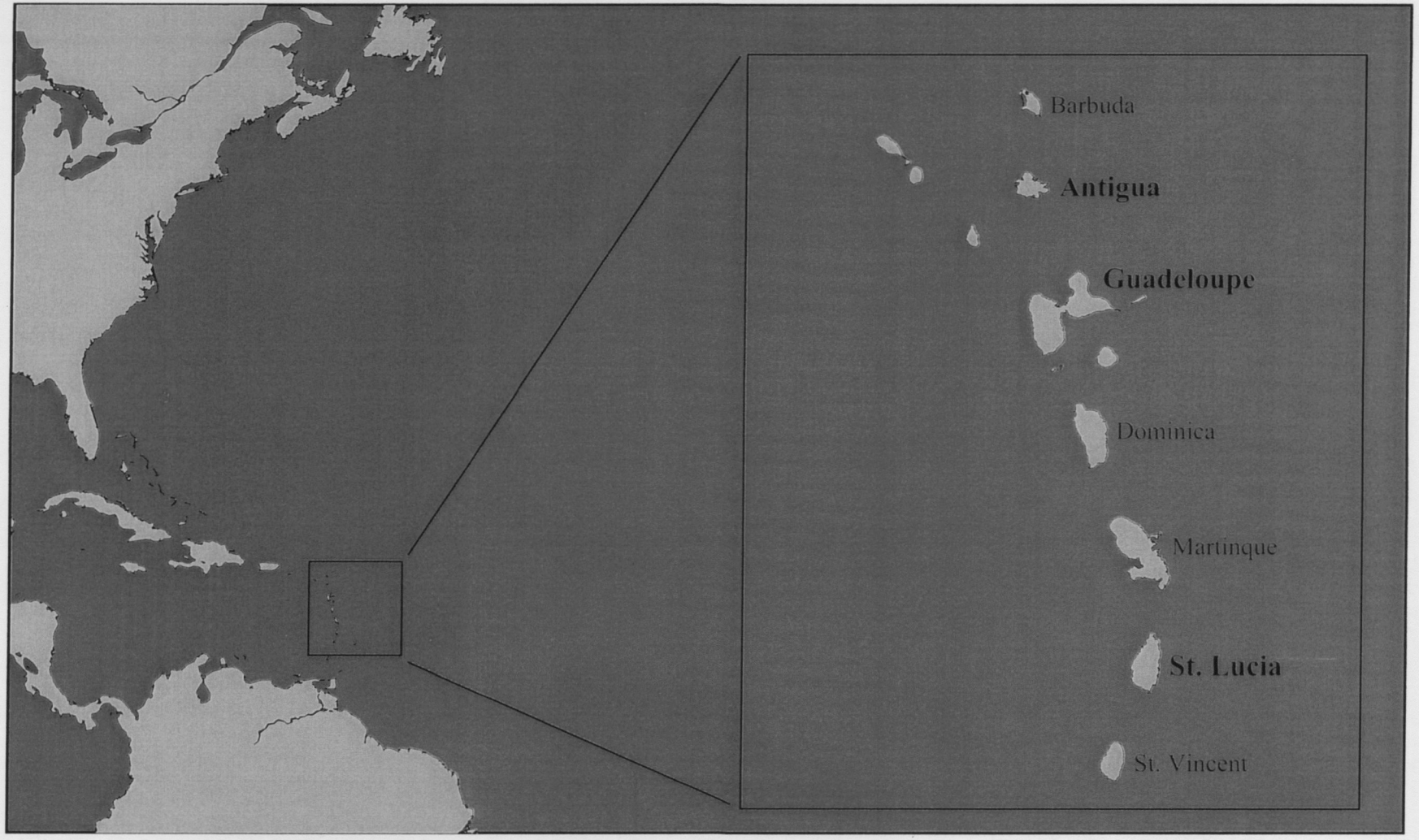


Figure 3.2. The geographic locations of the American Creole populations assayed in this study. Antigua, Guadeloupe and St. Lucia are highlighted in bold. Neighbouring islands are also shown. Samples from the Brazilian Nelore population were also used in this study but are not shown on this map.

3.2.2 Mitochondrial DNA sequence analysis

MtDNA sequence data reported for the first time in this study

In this study, the most variable 240 bp fragment of the mitochondrial control region (nucleotide positions 16023-16262, as per Anderson *et al.* 1982) was sequenced and analysed in a total of 59 samples comprising three Caribbean Creole populations [Antigua (n = 21), St. Lucia (n = 13) and Guadeloupe (n = 25)], plus 9 Brazilian Nelore samples. This mtDNA fragment has been widely examined in previous surveys of bovine phylogeny (Cymbron *et al.* 1999; Troy *et al.* 2001; Chapter 2 of this thesis). Partial mitochondrial control regions were amplified, purified and sequenced using the primers and conditions described in **Section 5.3**.

Table 3.1 details the mtDNA sequences used for phylogenetic analysis in this survey. The Caribbean Creole and Nelore mtDNA sequence data presented for the first time in this study were supplemented with an additional 10 Nelore mtDNA sequences, previously published by Troy (1998). *Bos taurus* and *Bos indicus* mtDNA sequences were identified by alignment with previously published *Bos taurus* and *Bos indicus* mtDNA sequences (Troy 1998), using the multiple sequence alignment option incorporated in the GENEJOCKEY II software package (Taylor 1993). All 59 Caribbean Creole sequences displayed *Bos taurus* mtDNA sequences and 14 Nelore samples displayed *Bos taurus* mtDNA sequences and were assumed to be of Creole origin. The remaining 5 Nelore samples, which exhibited *Bos indicus* mtDNA lineages, have been analysed and discussed in Chapter 2 of this thesis.

Table 3.1: Caribbean Creole and Brazilian Nelore samples used for mtDNA sequencing

Population code	Geographic origin	*Bos taurus* mtDNA sequences	*Bos indicus* mtDNA sequences	Total number of mtDNA sequences
Antiguan Creole (AC)	Antigua	21	0	21
Guadeloupe Creole (GC)	Guadeloupe	25	0	25
St. Lucian Creole (SC)	St. Lucia	13	0	13
Nelore (NE)	Brazil	14	5	19
Total		**73**	**5**	**78**

All Caribbean Creole mtDNA sequences plus 9 Nelore mtDNA sequences are reported for the first time in this study. An additional 10 Nelore mtDNA sequences were previously published by Troy (1998) (also listed in **Table 5.1**).

Phylogenetic reconstruction and mtDNA-derived inter-population genetic distance

All 73 Creole *Bos taurus* mtDNA sequences were aligned using the CLUSTALX software package (Thompson *et al.* 1997). Sequence variation in control region sequences were defined by direct comparison with the *Bos taurus* T3 central haplotype as defined by Troy *et al.* (2001). A reduced median phylogenetic network featuring all Creole *Bos taurus* mtDNA sequence was constructed manually according to the methods outlined by Bandelt *et al.* (1995). The resulting phylogeny was compared directly to networks constructed for continental European, Iberian and African populations using 188 previously published *Bos taurus* mtDNA sequences (Cymbron *et al.* 1999; Troy *et al.* 2001; **Table 5.1**). Sequence pairwise mismatch pairwise distributions for geographical haplotypic clusters were computed using the ARLEQUIN VERSION 2.0 software package.

To further quantify the distinctiveness of the Caribbean and Brazilian Creole mtDNA sequences, inter-population pairwise F_{ST} distances, using the Creole *Bos taurus* mtDNA sequences, together with all European and African *Bos taurus* mtDNA sequence data

published by Cymbron *et al.* (1999) and Troy *et al.* (2001) (**Table 5.1**), were generated according to Slatkin (1991; 1995) using the ARLEQUIN VERSION 2.0 package (**Section 2.3.9**). These were represented in two dimensions following multi-dimensional scaling (SPSS VERSION 8.0 software package).

3.2.3 Microsatellite allele scoring and phylogenetic analysis

Microsatellite allele scoring

Each Caribbean Creole population (**Table 3.2**) was assayed for allelic length variation at three autosomal microsatellite loci: *BM2113* (Sunden *et al.* 1993), *HEL1* (Kaukinen & Varvio 1993) and *ILSTS001* (Brezinsky *et al.* 1992). Microsatellite allele typing and scoring was performed using the techniques outlined in **Section 5.4**.

Table 3.2: Caribbean Creole DNA sample information for microsatellite analysis

Population (code)	Origin	Number of samples
Antiguan Creole (AC)	Antigua	70
Guadeloupe Creole (GC)	Guadeloupe	52[b]
St. Lucian Creole (SC)	St. Lucia	39
Total		161

Microsatellite genotype data for the Antiguan and St. Lucian Creole samples are reported in this study for the first time. Genotypic data for the Guadeloupe Creole samples were contributed by Ciaran Meghan (Department of Genetics, TCD).

Admixture analysis

Allele frequencies for each microsatellite locus were determined for each Caribbean Creole population using the MICROSOFT EXCEL©-based MICROSATELLITE TOOLKIT (Park 2002). The *BM2113* and the *ILSTS001* loci have been described previously as displaying a total of 4 alleles which are specific to West African *Bos taurus* breeds, while 3 alleles are found

across the *HEL1* and *ILSTS001* loci which are exclusive to populations exhibiting *Bos indicus* ancestry (MacHugh *et al.* 1997). The percentages of the West African *Bos taurus*-specific and *Bos indicus*-specific alleles were averaged to give an estimate of the proportion of West African *Bos taurus*-specific and *Bos indicus* autosomal admixture for each of the Caribbean Creole populations, as per MacHugh *et al.* (1997) and Loftus *et al.* (1999). The proportion of West African *Bos taurus*-specific and *Bos indicus*-specific microsatellite alleles were also tabulated for sample European, West African taurine and Indian zebu populations using data previously published by MacHugh *et al.* (1997) and a sample Iberian population using data generated by T. Cymbron (Department of Genetics, TCD, personal communication).

Microsatellite-derived inter-population genetic distance

Inter-population genetic distances (D_A) were estimated according to Nei *et al.* (1983) using the DISPAN computer package (Ota 1993). This genetic distance measure is particularly useful for recovering the phylogenetic relationships based on allele frequency data and is given by:

$$\text{Nei's genetic distance } (D_A) = \sum_{k=1}^{r} \left(1 - \sum_{i=1}^{m_k} \sqrt{x_{ik} y_{ik}}\right)/r$$

in which x_{ik} and y_{ik} are the frequencies of the i-th allele at the k-th locus in populations X and Y and where m_k and r are the number of alleles at the k-th locus and the number of the loci examined, respectively. This measure gives values ranging between 0 and 1, the latter value being obtained when the two populations under investigation share no common alleles. Small values of D_A are expected to increase in a linear fashion with respect to evolutionary time (Nei 1987).

D_A inter-population genetic distances were calculated using allelic data from all 3 Caribbean Creole populations together with allelic data previously published by MacHugh *et al.* 1997 for three European breeds (Charolais, Friesian and Simmental), six West African taurine breeds (N'Dama breeds from Guinea, Guinea Bissau, Gambia, Mali, Nigeria and Senegal), two West African zebu breeds (Gobra and Maure) and three Indian zebu breeds (Hariana, Sahiwal and Tharparker) (the source of these data are listed in **Table 5.2**). Allelic data from three Portuguese breeds were also included in the analysis (Alentejana, Aroquesa and Mertolenga). Inter-population D_A distances were represented in two dimensions following multidimensional scaling (MDS) using the SPSS VERSION 8.0 software package.

3.3 RESULTS

3.3.1 Amplification and sequencing of the most variable 240 bp DNA fragment of the mitochondrial genome

A PCR amplification of a 375 bp fragment of the bovine mtDNA control region [nucleotide positions 15,960-16,335, as per Anderson *et al.* (1982)] and typical sequencing autoradiograph for the most variable 240 bp fragment of the mitochondrial genome [nucleotide positions 16,023-16,262, as per Anderson *et al.* (1982)] have been presented previously in **Figure 2.3** and **Figure 2.4**, respectively.

3.3.2 Typing and scoring of bovine microsatellite allelic variation

Figure 3.3 shows an autoradiograph showing allelic length variation at the *BM2113* locus. The procedures employed for PCR amplifications and electrophoresis of microsatellite loci are outlined in **Section 5.4**.

Figure 3.3: Microsatellite allelic variation at the *BM2113* locus

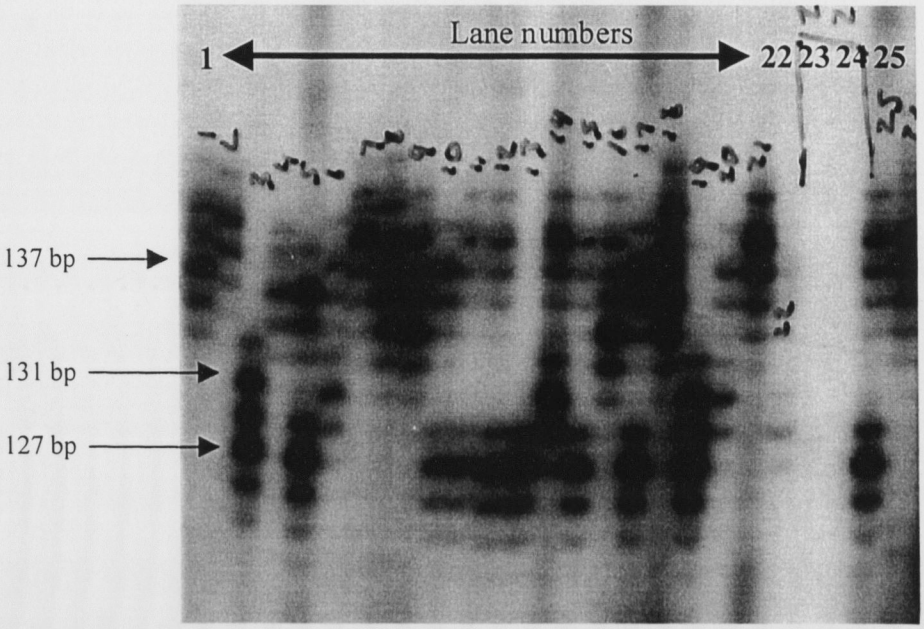


Figure 3.3. Microsatellite allelic variation at the *BM2113* locus in Antiguan Creole samples 1-23 (lanes 1-22 and 25). Negative PCR controls are shown in lanes 23 and 24.

3.3.3 Summary of Creole *Bos taurus* mitochondrial DNA sequence variation

In this study, the most variable 240 bp region of the mitochondrial control region was sequenced in a total of 59 cattle samples constituting three Caribbean Creole populations. Also, 14 *Bos taurus* sequences encountered in the Brazilian Nelore breed were assumed to have a South American Creole origin and were included in the analysis. The remaining 5 Brazilian sequences were typical *Bos indicus* lineages and have been presented in Chapter 2 of this thesis. Comparison of all Creole *Bos taurus* mtDNA sequences revealed a total of 17 haplotypes, differentiated at 20 polymorphic sites, all of which displayed transitions. This conforms to the strong transitional bias described in previous surveys of bovine mtDNA variation (Loftus *et al.* 1994a; Bradley *et al.* 1996; Cymbron *et al.* 1999; Troy *et al.* 2001).

The sequence variants defining each Creole *Bos taurus* haplotype are shown in **Table 3.3**. Both the European and African central haplotypes were detected and are denoted T3 and T1, respectively, following the nomenclature of Troy *et al.* (2001). Derivative haplotypes are labelled by suffixes, which indicate the transition position within the control region less 16,000 (numbering according to Anderson *et al.* 1982). In total, the European central haplotype was encountered 16 times while only one animal possessed the central African haplotype. Three other haplotypes ($T1_{195}$ [3 samples]; $T1_{53,122,139,196}$; [18 samples] and $T1_{53,122,139,196,231}$ [1 sample] shared all three African-defining substitutions. The remaining haplotypes were typically European and of these, one ($T3_{58,130}$) was encountered 12 times, one was shared by five individuals, another was found in 4 specimens, one was encountered three times, 2 occurred in duplicate and 6 were detected only once. In addition to the T3 and T1 central sequences, only 5 of the 18 haplotypes encountered ($T3_{42}$, $T3_{51}$, $T3_{139}$, $T3_{141}$, $T3_{247}$) have been previously described (Troy 1998; Cymbron *et al.* 1999; Troy *et al.* 2001).

Table 3.3: Mitochondrial control region sequence variations observed in 73 Creole *Bos taurus* mtDNA sequences

Haplotype codes are given in the first column and the European and African consensus sequences, denoted T3 and T1, respectively, as per Troy *et al.* (2001). Only variable sites are shown. The nucleotide positions from the T3 ancestral *Bos taurus* mtDNA sequence are given above each column (Anderson *et al.* 1982). Differences are given, full stops (.) denote identity. Suffixes indicate sites of nucleotide change from the consensus T3 sequences. Samples in bold indicate those seqeunced for the first time as part of this study. Population codes as per **Table 3.1**.

Haplotype	Samples within each haplotype	Nucleotide poisiton in bovine mitochondrial genome minus 16,000, as per Anderson *et al.* (1982)																							Times represented in each American population				
		42	50	51	53	58	61	113	119	122	125	130	139	141	143	183	191	195	196	231	247	250	255		Antigua	Guadeloupe	Nelore	St. Lucia	Total
T3	**AC15, AC10, AC33, AC2, GC50, GC49, GC48, GC15, SC149, SC166, SC142, NE19, NE22, NE21, NE27,** NE12	T	C	T	T	C	T	T	T	T	A	T	C	T	A	C	C	A	G	C	C	A	T		4	4	5	3	16
$T3_{58,130}$	**GC44, AC3, AC28, AC23, AC21, AC12, AC29, SC173, SC169, SC164, SC163, SC146**	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.		6	1	-	5	12
$T3_{250}$	**SC171, AC14, AC32, SC162, SC152**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.		2	-	-	3	5
$T3_{141}$	**AC9, AC25, GC36, GC34,**	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.		2	2	-	-	4
$T3_{58,119}$	**AC35, AC36**	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.		2	-	-	-	2
$T3_{139}$	**GC19, AC5**	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.		1	1	-	-	2
$T3_{125}$	**AC27**	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.		1	-	-	-	1
$T3_{51}$	**AC16**	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		1	-	-	-	1
$T3_{51,61}$	**AC22**	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		1	-	-	-	1
$T3_{58,130,231}$	**GC43**	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	T	.	.	.		-	1	-	-	1
$T3_{42}$	**GC30**	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		-	1	-	-	1
$T3_{122,143}$	**NE18,** NE10, NE8	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.	.	.	.	.	.		-	-	3	-	3
$T3_{247}$	**SC144**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.		-	-	-	1	1
T1	**GC27**	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C		-	1	-	-	1
$T1_{195}$	NE5, NE7, **NE24**	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C		-	-	3	-	3
$T1_{53,122,139,196}$	**AC19, GC52, GC51, GC5, GC45, GC42, GC41, GC39, GC38, GC37, GC35, GC33, GC31, GC26, GC13, NE26, NE20,** NE2	.	T	.	C	.	.	C	.	C	.	.	T	.	.	.	.	.	A	.	.	.	C		1	14	3	-	18
$T1_{53,122,139,196,231}$	**SC170**	.	T	.	C	.	.	C	.	C	.	.	T	.	.	.	.	.	A	T	.	.	C		-	-	-	1	1
Total																									21	25	14	13	73

3.3.4 Phylogenetic network analysis

Figure 3.4a places each of the 73 *Bos taurus* Creole sequences in a reduced median network. For comparative purposes, networks constructed from previously published Western European, Iberian and African *Bos taurus* mtDNA sequences (Cymbron *et al.* 1999; Troy *et al.* 2001) are shown with the Creole *Bos taurus* network in **Figure 3.4b**. Almost all European mtDNA variation is clustered around the predominant centrally positioned haplotype, T3, through which derivative haplotypes root. With some symmetry, African mtDNA sequences coalesce to an alternative, predominant variant, T1. The star-like phylogeny of both clusters is consistent with past population expansions associated with the domestication process (Bradley *et al.* 1996). Within Iberia, sequences fall into either the European (T3) or African (T1) haplotypic clusters, indicating that these Iberian cattle samples share a partial African ancestry, which is most likely a legacy of historical North African influence in the peninsula (Cymbron *et al.* 1999).

All Creole *Bos taurus* sequences fall into one of the two distinct European and African predominant continental clusters. Sequences of European origin produce a roughly, star-like phylogeny centred round the predominant, ancestral T3 haplotype. Most Creole T3 sequences are identical to, or fall within one or two base pair substitutions of the ancestral node, while only one haplotype ($T3_{58,130,231}$) deviates by three substitutions. The Creole T1 cluster is atypical of the pattern of diversity found within the parental African continental population. Notably, T1 is represented only by a single sample while the most frequently represented African sequence ($T1_{53,122,139,196}$) is separated from the ancestral haplotype by a total of four nucleotide substitutions.

Figure 3.4a: A reduced median phylogenetic network constructed from 73 Creole *Bos taurus* mtDNA sequences

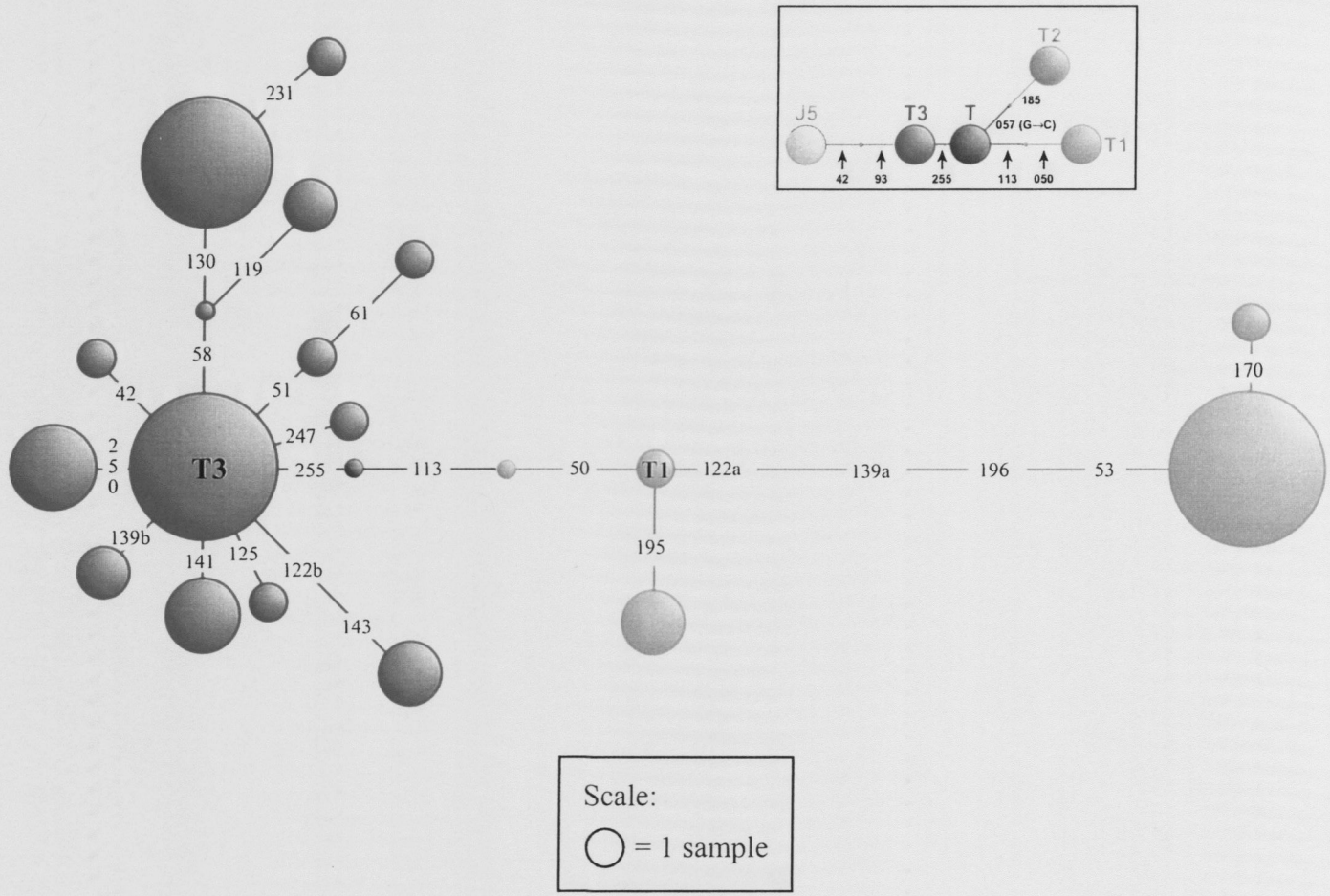


Figure 3.4b: Reduced median phylogenetic networks constructed for Western Europea Iberian, Creole and African *Bos taurus* mtDNA sequences

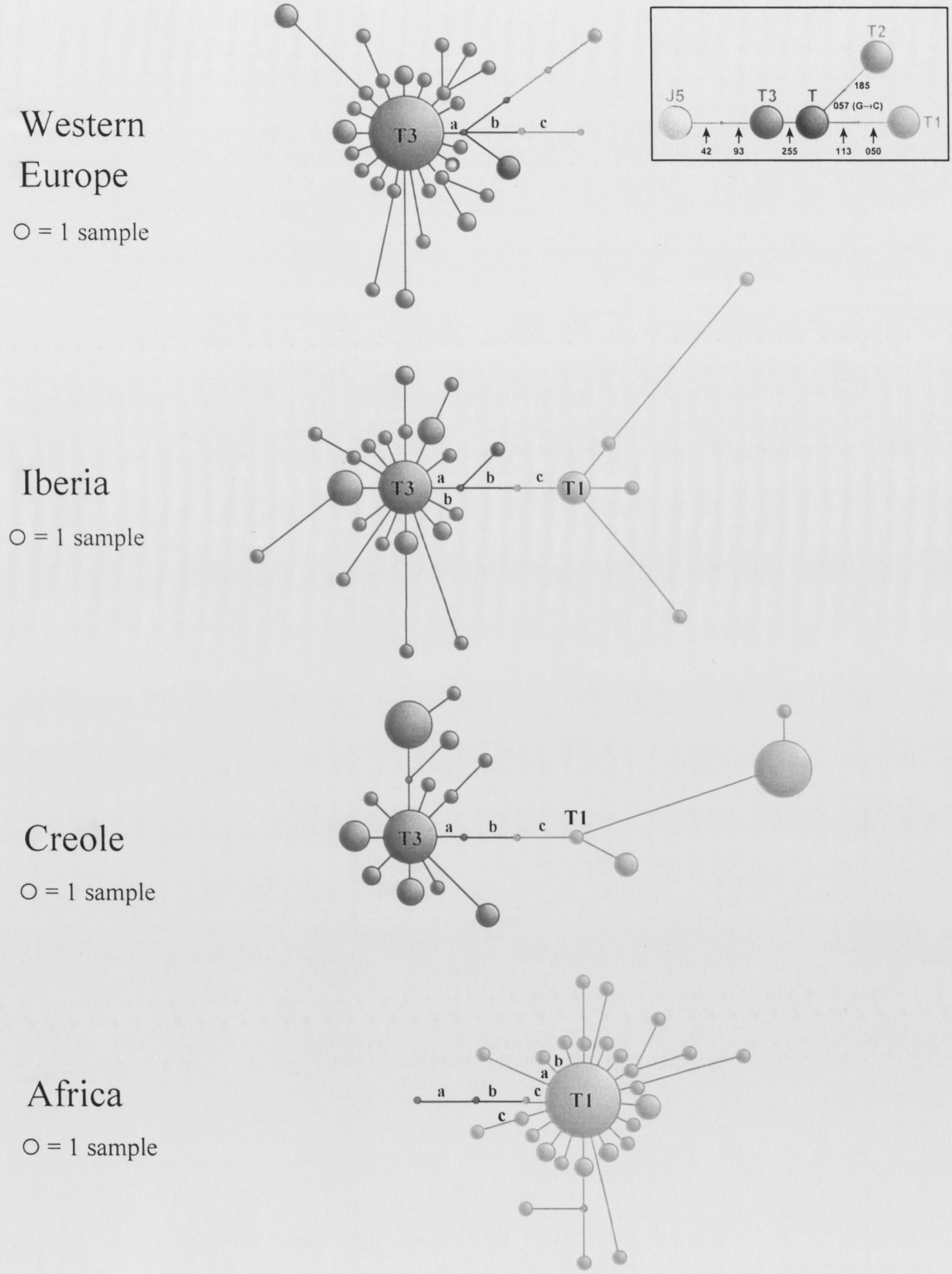


Figure 3.4a. A reduced median network (Bandelt *et al.* 1995) featuring *Bos taurus* mtDNA lineages sampled in 59 Caribbean Creole cattle from three Caribbean islands and 14 Nelore samples from Brazil. Haplotypes encountered in each region (shaded circles) and unsampled intermediate nodes (small points) are shown. Numbers indicate nucleotide positions of mutations, minus 16,000 as per Anderson *et al.* 1982. Both the central European haplotype (T3, shown in red) and African haplotype (T1, shown in yellow) are presented. The relationship between European (T3) and African (T1) central haplotypes are defined by transitions at nucleotide positions 16255, 16113 and 16050, as defined by Troy *et al.* (2001). This network was reduced at two nucleotide positions, 16,139 and 16,122 and recurrent mutations at these positions are denoted **a** and **b**, respectively. Circle areas are proportional to the frequency of each haplotype and shading indicates which of the two major haplotypes they root to. The Creole *Bos taurus* mtDNA sequences show an overall reduction in haplotypic diversity when compared with those for other geographical groups.

Figure 3.4b. For comparative reasons, reduced median networks were constructed for three regional groupings, namely, Western Europe (67 samples comprising Charolais, Friesian, German Black, Limousin, Romagnola, Simmental), Iberia (60 samples comprising Alentejana, Aroquesa, Barrosa, Barrenda, Maronesa, Mertolenga, Preta) and Africa (61 samples comprising Kapsiki, Kuri, Namchi, N'Dama, Somba, White Fulani) using previously published data (Cymbron *et al.* 1999; Troy *et al.* 2001; **Table 5.1**) and are shown with the Creole *Bos taurus* network. The relationship between European (T3) and African (T1) central haplotypes are defined by transitions at nucleotide positions 16255, 16113 and 16050 (denoted **a**, **b** and **c**, respectively). Haplotypes encountered in each region (shaded circles) and unsampled intermediate nodes (small points) are shown. Circle areas are proportional to the frequency of each haplotype and shading indicates which of the two skeleton network haplotypes they root to. The Creole *Bos taurus* mtDNA show an overall reduction in haplotypic diversity when compared with those for other geographical groups. The topology of the Creole *Bos taurus* network is indicative of contributions from both European and African cattle to the mtDNA gene pool of modern Creole populations.

3.3.5 Mitochondrial DNA sequence pairwise mismatch distributions

Several authors have argued that mismatch distributions preserve a record of past population histories (Rogers and Harpending 1992; Harpending *et al.* 1993). Smooth, uni-modal distributions are indicative of rapid, past population expansions, during which the expectancy of mtDNA lineage survivorship is high. Ragged, multi-modal distributions are

consistent with demographic stability. Ragged distributions may also result from a population bottleneck or founder event, in which a small number of individuals survive from a parental population.

Mismatch curves were constructed for both the T3 and T1 *Bos taurus* haplogroups within each major geographical region (Europe, Iberia, Creole and Africa) and are presented in **Figure 3.5**. The validity of all observed mismatch distributions was tested under the population expansion model (Rogers & Harpending 1992) by computing the probability of the sum of square deviations (SSD) between the observed and the expected mismatch curves, as described in **Section 2.3.10**. The smoothness of each mismatch distribution was evaluated by computing statistical conformity with the raggedness statistic (Harpending *et al.* 1993; **Section 2.3.10**). Fu's F_S statistic of selective neutrality was also used to describe the diversity within each regional haplogroup (Fu 1997; Excoffier & Schneider 1999; **Section 2.3.10**). The statistical parameters generated for each mismatch curve are presented in **Table 3.4**.

Smooth, uni-modal distributions, consistent with population expansions, are evident for all T3 regional haplogroups and the African T1 haplogroup mismatch curves. Each of these regional haplogroups displays significant departures from selective neutrality (denoted by significantly negative F_S values), signifying demographic growth. In contrast, ragged, multi-modal distributions are observed for the African-derived T1 sequences encountered within the Iberian and Creole populations. Notably, non-significant departures from neutrality are recorded for both the Iberian and Creole T1 haplogroups and may be representative of a population bottleneck or genetic founder effect, which can erase signals of past demographic expansion (Excoffier & Schneider 2000).

To ensure that the ragged multi-modal distributions and non-significant departures from neutrality noted for the Iberian and Creole T1 haplogroups are consistent with bottleneck or founder effect and not due to a sample size effect, F_S statistics were computed for two random panels of 50 sample files, each containing 11 and 23 African T1 sequences, respectively. Random sampling was performed with replacement using the DNA SEQUENCE ANALYSIS TOOLKIT software package devised by Dr. Stephen Park (Park 2002). The resulting data were compared to those obtained for the Iberian and Creole T1 haplogroup sequences. A total of 74% (37 out of 50) of all ARLEQUIN VERSION 2.0 sample files containing 11 randomised African T1 sequences displayed significant negative F_S values, while 100% of all sample files containing 23 randomised African T1 sequences displayed significant negative F_S values. Based on these findings, it seems likely that the patterns of variation noted within the Creole and Iberian T1 haplogroup are consistent with a bottleneck or genetic founder effect and not an artefact of sample size.

Figure 3.5: MtDNA sequence mismatch distributions for the T1 and T3 haplogroups in European, Iberian, Creole and African cattle

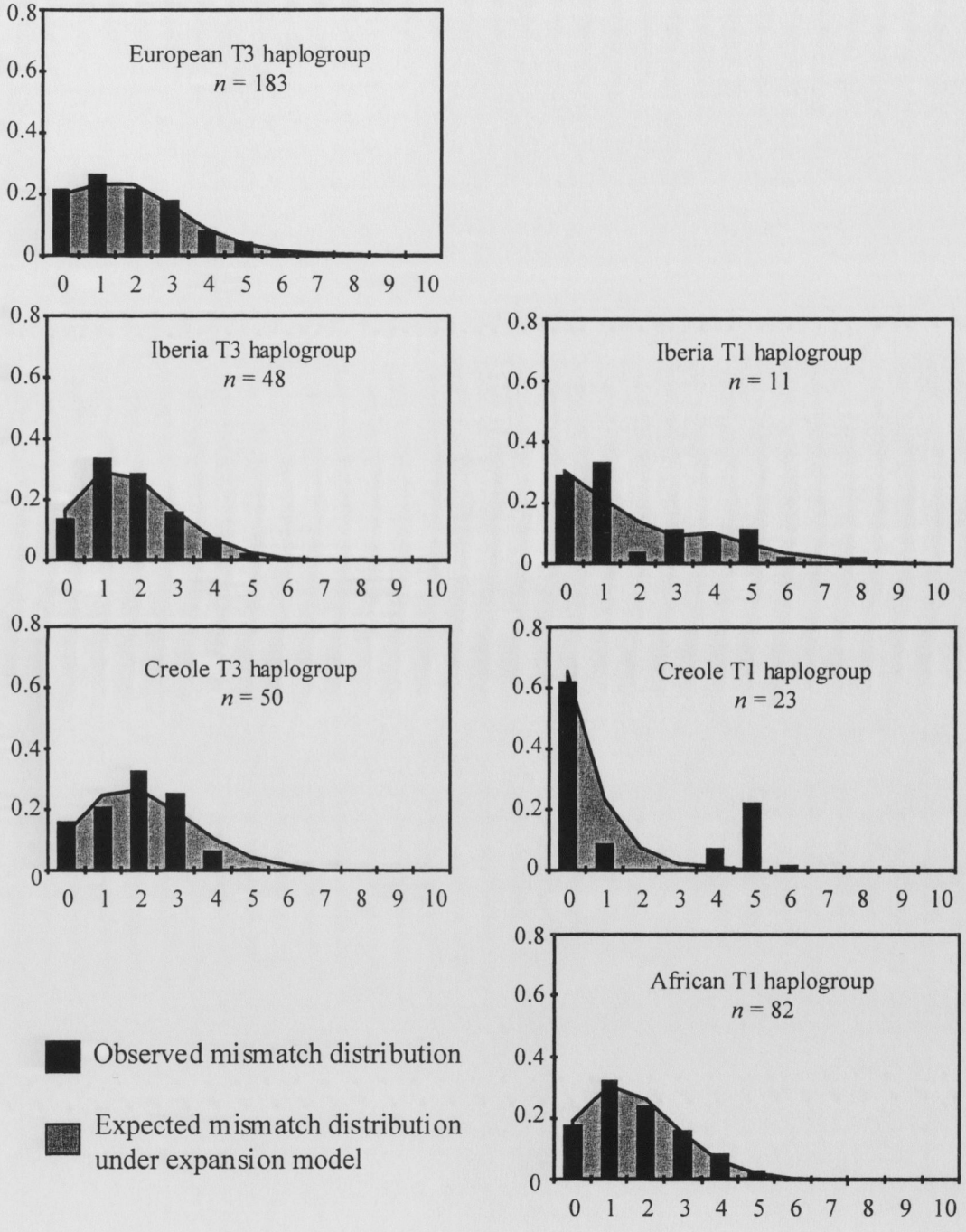


Figure 3.5. DNA sequence mismatch distributions for the T3 and T1 haplogroups found within *Bos taurus* mtDNA sequences from European, Iberian, Creole and African samples. The number of nucleotide differences and corresponding frequencies are represented on the horizontal and vertical axes, respectively. Black shading denotes the mismatch distribution constructed from the observed data; grey shading represents the theoretically expected mismatch distribution curve generated by demographic expansion (Excoffier & Schneider 2000). Ragged, multi-modal curves are observed for the Creole and Iberian T1 mtDNA haplogroups and may be suggestive of a recent bottleneck or genetic founder effect, resulting in the removal of past demographic signals.

Table 3.4: Statistical parameters for T3 and T1 *Bos taurus* haplogroup mismatch distributions within each geographic region

	Regional *Bos taurus* mtDNA haplogroups					
Statistical analysis	Europe T3 ($n = 183$)	Iberia T3 ($n = 48$)	Iberia T1 ($n = 11$)	Africa T1 ($n = 82$)	Creole T3 ($n = 50$)	Creole T1 ($n = 23$)
Sum of square deviations (SSD)	<0.001	0.003	0.030	0.001	0.013	0.078
P (sim. SSD ≥ obs. SSD)	0.940	0.340	0.790	0.370	0.060	0.100
Raggedness (rag) statistic	0.018	0.069	0.101	0.045	0.060	0.360
P (sim. rag ≥ obs. rag)	0.949	0.210	0.759	0.550	0.180	0.5199
Observed F_S value	-10.359	<-100.000	0.411	-11.089	-5.364	-0.188
P value of obs F_S value	0.003	0.001	0.608	0.001	0.008	0.746

Significant departures from neutrality, consistent with population growth are evident for all T3 haplogroups and the African T1 haplogroup. Deviations from neutrality were not recorded for the T1 sequences encountered in the Iberian and Creole samples, and are most probably indicative of a bottleneck or genetic founder effect.

3.3.6 Mitochondrial admixture proportions

For each Caribbean island population, the proportion of sequences possessing all three substitutions defining T1 (nucleotide positions 16050, 16113 and 16255; Bradley *et al.* 1996) are given in **Table 3.5**. The proportion of African haplotypes in the Guadeloupe samples was 60%, while 42.8% of all *Bos taurus* Nelore sequences were of African origin. Conversely, the incidence of African haplotypes in the St. Lucian and Antiguan populations was low, 7.7% and 4.7%, respectively.

3.3.7 Multi-dimensional scaling of mitochondrial DNA-derived inter-population genetic distances

Pairwise F_{ST} values may be used as an estimate of genetic distances between populations over shallow time depths (Slatkin 1991; Slatkin 1995). These were calculated using the Creole mtDNA sequences together with an additional 340 published mtDNA sequences, comprising 32 breeds of European and African origin (Cymbron *et al.* 1999; Troy *et al.* 2001). Pairwise F_{ST} values (**Appendix 2.E**) were represented in two dimensions following multi-dimensional scaling (**Figure 3.6**). Two-dimensional mtDNA-derived MDS plot co-ordinates are shown in **Appendix 2.F**.

The principal feature of this two-dimensional plot is the major divergence between European and African *Bos taurus* lineages, as clearly distinguished by the first dimension. The second dimension also differentiates European and African *Bos taurus* groups, albeit to a lesser extent, but also separates some members within the European grouping. Both the Guadeloupe and Nelore *Bos taurus* mtDNA samples occupy intermediate positions between the European and African groups and provide a strong indication of major contributions from both continental lineages. The Antiguan and St. Lucian Creole populations are positioned within the European grouping.

Figure 3.6: A two-dimensional MDS plot featuring African, European, Iberian and Creole mtDNA sequences

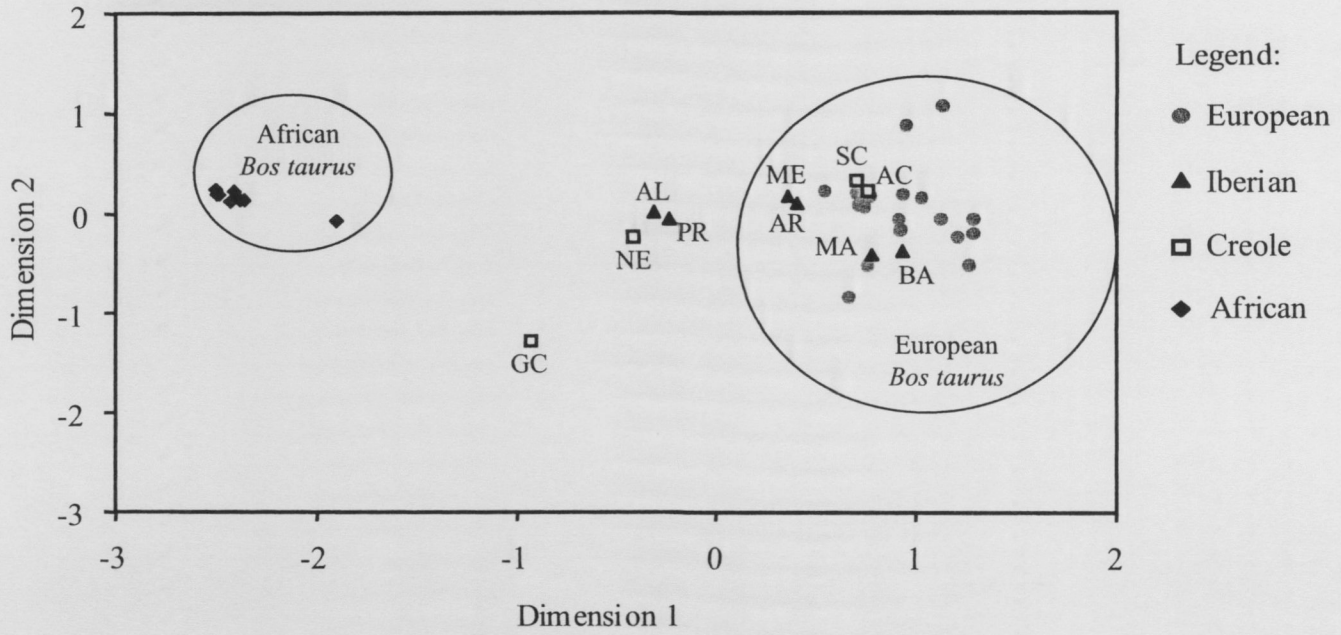


Figure 3.6. Two-dimensional scaling plot of genetic F_{ST} distances based on mtDNA data, based on the data presented in this study together with an additional previously published 340 *Bos taurus* sequences (Cymbron *et al.* 1999, Troy *et al.* 2001; **Table 5.1**). Groups were comprised of following populations: **Africa** (Butana, Kapsiki, Kenana Kuri, Namchi, N'Dama, Somba, White Fulani), **American Creole** (Antigua, Guadeloupe, Nelore, St. Lucia) and Europe (Alentejana, Angus, Aroquesa Barrosa, Berrenda, Charolais, Friesian, Galloway, German Black, Hereford, Highland, Iceland, Jersey, Kerry, Limousin, Mertolenga, Marronesa, Preta, Romangola), Sykia, Telemark, Westland Fjord). The breeds indicated above are as follows: AL (Alentejana), AC (Antiguan Creole), GC (Guadeloupe Creole), NE (Nelore), PR (Preta) and SC (St. Lucian Creole).

3.3.8 Microsatellite allelic variation and admixture analysis

In total, 29 alleles were detected within the Caribbean Creole samples assayed across all three autosomal microsatellite loci. Allele frequencies for each locus are presented in **Figure 3.7**. Notably, the *BM2113* and the *ILSTS001* loci display alleles which are specific only to West African *Bos taurus* populations, while the *ILSTS001* and *HEL1* loci exhibit alleles which are of *Bos indicus* origin. Based on the analysis performed by MacHugh *et al.* (1997), the frequencies of these groups of alleles were averaged to give an estimate of the frequency of West African taurine-specific and zebu-specific alleles in each of the individual Creole populations (**Table 3.5**).

The mean percentage of West African *Bos taurus*-specific alleles in a West African *Bos taurus* sample population was estimated as 39.4%. This figure reduces considerably to 1.5% within Iberian populations, while no West African *Bos taurus*-specific alleles were detected within continental European populations. Cattle from Guadeloupe were the only Creole cattle population to exhibit West African *Bos taurus*-specific alleles, with a mean percentage of 6.6% and no such alleles were encountered in the Antiguan and St. Lucian Creole breeds. Contributions from introgressing zebu populations were also observed in each of the Creole populations. An average frequency of 67.4% *Bos indicus*-specific alleles was obtained for an Indian zebu sample population, while mean *Bos indicus*-specific allele frequencies of 22.5% (Guadeloupe), 21.0% (Antigua) and 11.9% (St. Lucia) were calculated for each Creole population. A comparison of mtDNA- and microsatellite-derived estimates of African *Bos taurus* introgression are depicted graphically in **Figure 3.8**.

Figure 3.7: Microsatellite allele frequencies for European, Iberian, Caribbean Creole, West African taurine (WAT), West African zebu (WAZ) and Indian zebu cattle

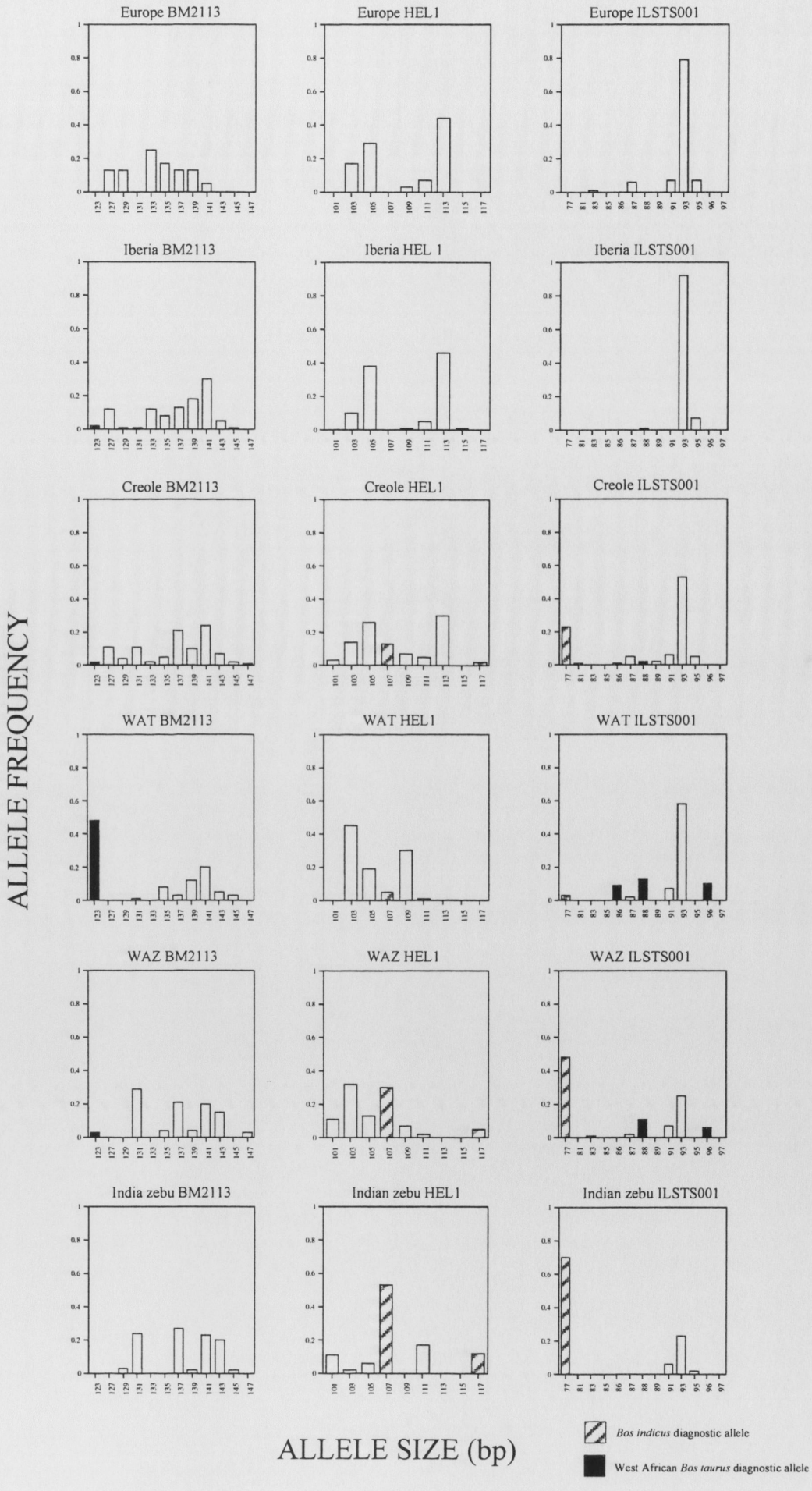


Figure 3.7. Microsatellite allele frequencies for Caribbean Creole cattle for all three assayed loci. The *BM2113* and the *ILSTS001* loci display alleles which are specific to West African *Bos taurus* populations (shaded in black), while the *ILSTS001* and *HEL1* loci exhibit alleles which are of *Bos indicus* ancestry (shaded with grey diagonal lines) (MacHugh *et al.* 1997).

For comparative purposes allele frequencies are also shown for the following regional sample populations:

(1) Europe [112 samples comprising the French Charolais (n = 36), and Friesian populations (n = 40) and the Swiss Simmental population (n = 36) (MacHugh *et al.* 1997)].

(2) Iberia [101 samples comprising 3 Portuguese populations namely, Aletejana (n = 30), Aroquesa (n = 34) and Mertolenga (n = 37) (T. Cymbron, TCD, personal communication)].

(3) West African taurine [(161 samples comprising three N'Dama populations from Guinea (n = 63), Guinea Bissau (n = 54) and Mali (n = 44) (MacHugh *et al.* 1997)].

(4) West African zebu [(114 samples comprising Gobra samples from Senegal (n = 59) and Maure samples from Mauritania (n = 55) (MacHugh *et al.* 1997)].

(5) South Asia zebu [(33 samples from the 3 populations from the Indian subcontinent, namely Hariana (n = 10), Sahiwal (n = 13) and Tharparker (n = 10) (MacHugh *et al.* 1997)].

The sources of these data are also listed in **Table 5.2**.

Table 3.5: The proportion of African *Bos taurus* mtDNA haplotypes, West African *Bos taurus*-specific and *Bos indicus*-specific microsatellite alleles estimated for the Caribbean Creole and Brazilian Nelore populations analysed in this study

	mtDNA	Microsatellite data	
Population	**African T1 haplotypes (%)**	**West African taurine-specific alleles (%)**	***Bos indicus*-specific alleles (%)**
Continental Europe[a]	0.0	0.0	0.0
Iberia[b]	18.3	1.5	0.0
Antiguan Creole (AC)	4.7	0.0	21.0
Guadeloupe Creole (GC)	60.0	6.6	22.5
Nellore (NE)	42.8 (31.5)	not available	not available
St. Lucian Creole (SC)	7.7	0.0	11.9
West African taurine[c]	100.0	39.4	2.6
Indian zebu[d]	0.0	0.0	67.4

Table 3.5. The proportion of African *Bos taurus* mtDNA haplotypes and West African *Bos taurus*-specific and *Bos indicus*-specific microsatellite alleles estimated for the Caribbean Creole and Nelore populations analysed in this study.

MtDNA sequence data (also listed in Table 5.1):

(1) Western Europe *Bos taurus* mtDNA data (67 samples comprising Charolais, Friesian, German Black, Limousin, Romagnola, Simmental). Data published by Troy *et al.* (2001);

(2) Iberia *Bos taurus* mtDNA data (60 samples comprising Alentejana, Aroquesa, Barrosa, Barrenda, Maronesa, Mertolenga, Preta). Data published by Cymbron *et al.* (1999) and Troy *et al.* (2001);

(3) West Africa *Bos taurus* mtDNA data (61 samples comprising Kapsiki, Kuri, Namchi, N'Dama, Somba, White Fulani). Data published data by Troy *et al.* (2001).

The 42.8% value given for the Nelore mtDNA sequences refers to the percentage of African *Bos taurus* haplotypes encountered in the 14 Nelore *Bos taurus* sequences (6 sequences out of 14). The 31.5% value (in brackets) refers to the percentage of African *Bos taurus* haplotypes encountered in all 19 Nelore mtDNA sequences, which includes the 5 Nelore *Bos indicus* sequences (6 sequences out of 19).

Microsatellite allelic data (also listed in Table 5.2):

[a] European sample population consisting of 112 samples [Charolais (n = 36), Friesian (n = 40) and Simmental (n = 36)] (MacHugh *et al.* 1997).

[b] Iberian sample population consisting of 101 samples [Alentejana (n = 30), Aroquesa (n = 34) and Mertolenga (n = 37)] (T. Cymbron, TCD, personal communication).

[c] West African *Bos taurus* sample population consisting of 161 samples [N'Dama populations sampled from Guinea (n = 63), Guinea Bissau (n = 54) and Mali (n = 44)] (MacHugh *et al.* 1997).

[d] Indian zebu sample population consisting of 33 samples [Hariana (n = 10), Sahiwal (n = 13) and Tharparker (n = 10)] (MacHugh *et al.* 1997).

Figure 3.8: MtDNA- and microsatellite-derived estimates of African *Bos taurus* and *Bos indicus* admixture proportions

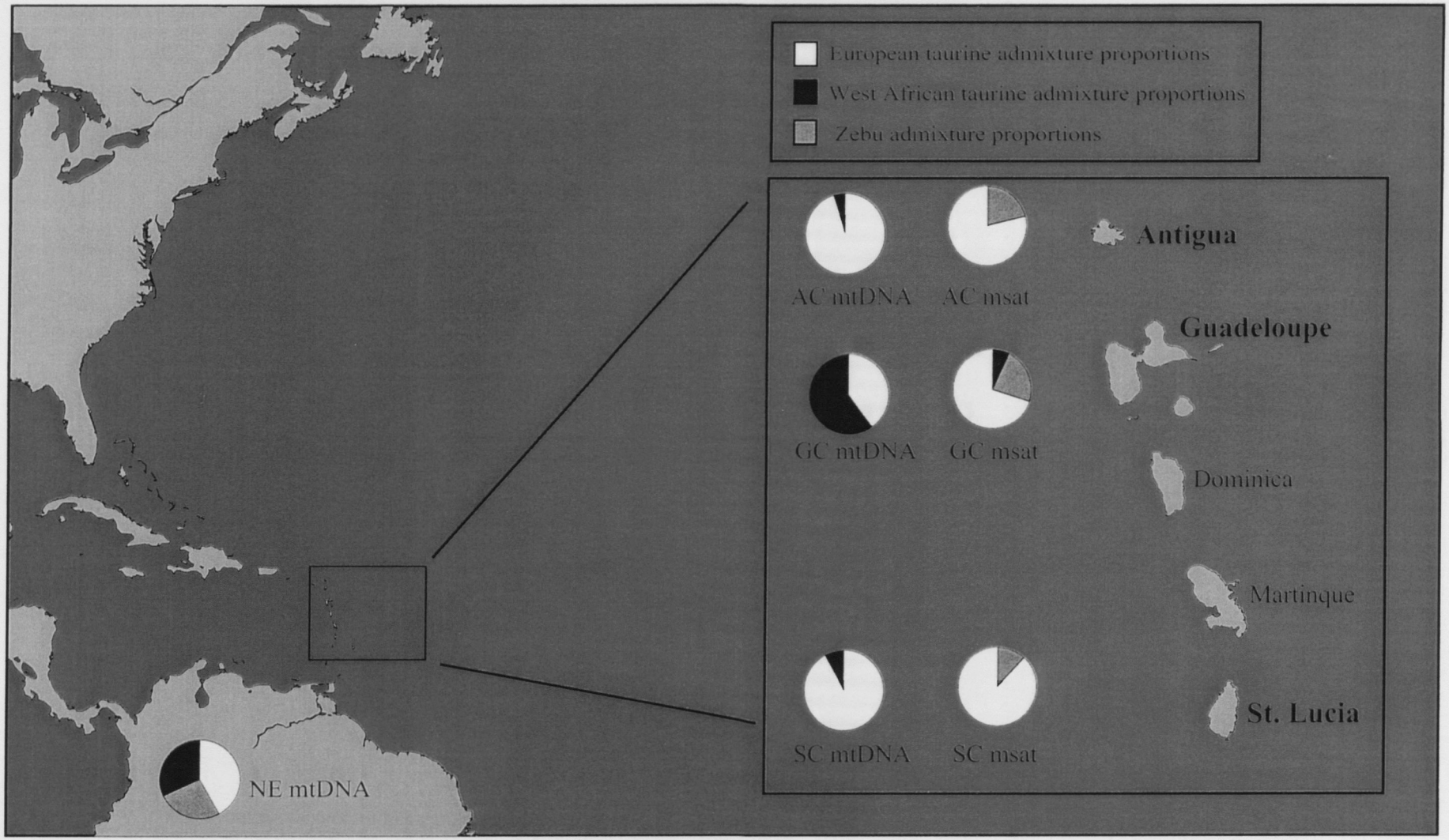


Figure 3.8. West African *Bos taurus* and *Bos indicus* admixture proportions for the Caribbean Creole and Nelore samples assayed for both mtDNA and microsatellite systems. MtDNA admixture estimates were calculated as the proportion of African taurine and zebu haplotypes occurring within each population. Microsatellite (msat) admixture estimates were calculated as the proportion of West African taurine- and zebu-specific alleles within [illegible] (1997) and Loftus *et al.* (1999). Population codes as per **Table 3.1**.

3.3.9 Multi-dimensional scaling of microsatellite-derived inter-population genetic distances

A two-dimensional MDS plot was constructed from a matrix of D_A genetic distances (**Appendix 2.G**) calculated for the 3 Caribbean Creole populations and previously published data for breeds from Europe, Iberia, West Africa and India, listed in **Section 3.2.3**. The two-dimensional MDS co-ordinates for the microsatellite-derived plot are shown in **Appendix 2.H**. The microsatellite-derived MDS plot is shown in **Figure 3.9**.

Dimension 1 of this plot illustrates the large divergence between cattle of *Bos taurus* and *Bos indicus* origin. Alternatively, dimension 2 accounts for the divergence between European and African *Bos taurus*. Notably, the Caribbean Creole populations occupy an intermediate position between the European *Bos taurus* and Indian *Bos indicus*. However, the Guadeloupe Creole population displays some genetic affinity with African *Bos taurus* cattle along dimension 2.

Figure: 3.9: A two-dimensional MDS plot constructed for African, European, Iberian and Creole microsatellite-derived genetic distances

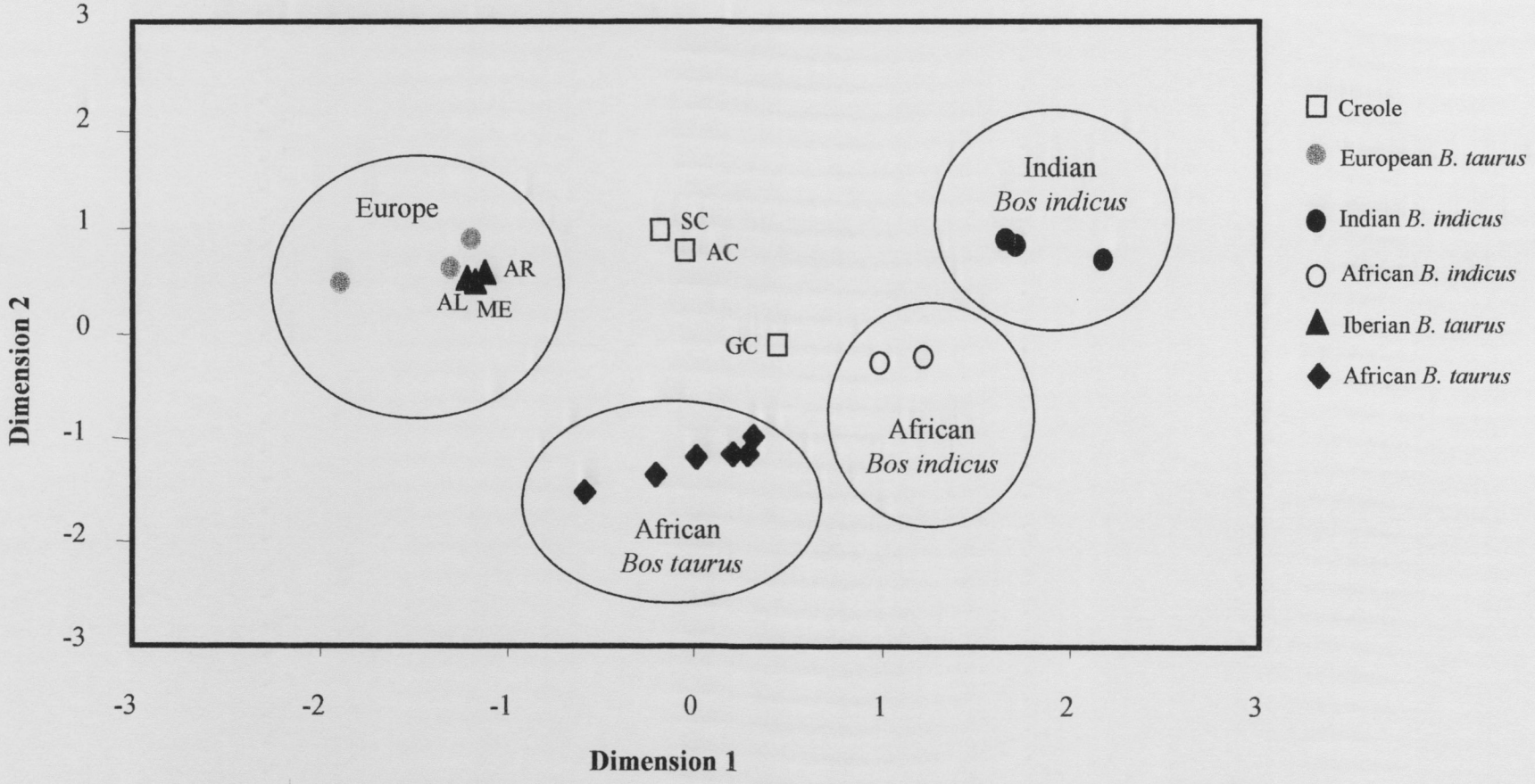


Figure 3.9. Two-dimensional MDS plot of D_A genetic distances based on allelic variation at three nuclear microsatellite loci. Regional groups consisted of data from the following populations: **Europe** [Charolais, Friesian, Simmental], **Iberia** [Alentejana (AL), Aroquesa (AR), Mertolenga (ME)], **African *Bos taurus*** [N'Dama populations sampled from Gambia, Guinea, Guinea Bissau, Mali, Nigeria and Senegal], **African *Bos indicus*** [Gobra, Maure] and **Indian *Bos indicus*** [Hariana, Sahiwal and Tharparker]. The source of these data are listed in **Table 5.2.** Dimensions 1 and 2 account for 98.5% of the data.

3.4 DISCUSSION AND CONCLUSIONS

3.4.1 The genetic legacies of historic introgression to the Caribbean

MtDNA phylogenetic analysis demonstrates that all sampled Creole *Bos taurus* sequences fall into two separate haplotypic clusters, which are representative of European and African continental lineages, as previously described by Troy *et al.* (2001). The contribution of African mitochondrial sequences to the total *Bos taurus* Creole mtDNA pool is 32%, peaking in the Guadeloupe Creole (60%) and Brazilian Nelore (42.8%) populations. The extent of this African influence in each of these populations is also reflected in their intermediate positioning between the European and African continental groups in the two-dimensional scaling plot (**Figure 3.6**). Conversely, the mtDNA pools of the St. Lucian and Antiguan Creole populations are predominantly European, with African mtDNA admixture proportions of 7.7% and 4.7%, respectively. West African *Bos taurus* influence is also evident on the nuclear genomes of the Guadeloupe Creole samples assayed in this study, which display an average of 6.6% of alleles that are specific to *Bos taurus* populations of West African origin. However, no such alleles were detected in the Creole populations sampled in St. Lucia and Antigua.

A recent survey of *Bos taurus* mtDNA phylogeography has identified African mtDNA haplotypes in Iberian cattle, which are undoubtedly a legacy of African migration, (Cymbron *et al.* 1999). Here, an estimated 18.3% of the total analysed Iberian mtDNA pool were of African origin, while an average of 1.5% of Iberian microsatellite alleles were diagnostic of West African taurine ancestry. As the founding cattle of the Americas were of Iberian descent, the results presented in this study may suggest that the African influence observed in modern Creole cattle may be attributable to three possible sources; (1) the partial African ancestry associated with the Iberian-derived founding population of

the Americas, (2) the direct importation of African cattle to the Americas, presumably during the 16th-18th centuries AD, or (3) a combination of both.

It is conceivable that a portion of this African influence may have originated from an unsampled Iberian population displaying increased African ancestry compared to the Iberian populations considered in this survey. However, the greater proportion of African ancestry in Guadeloupe relative to the Iberian populations analysed, would suggest that this is at least partially due to the direct importation of cattle from the African continent. This interpretation of the data is further supported by immunological studies, which suggest some affinity between Guadeloupe Creole and African N'Dama cattle sampled in Guinea (Maillard *et al.* 1993). In contrast, it appears that African genetic contributions to St. Lucian and Antiguan Creole populations have been substantially reduced compared to Guadeloupe Creole.

The direct introduction of African cattle to the Caribbean is further evidenced by the considerable proportion of *Bos indicus*-specific alleles on the autosomal genome of all 3 Caribbean Creole populations assayed. Here, the mean frequencies of the *Bos indicus*-specific alleles were estimated at 22.5% in Guadeloupe, 21.0% in Antigua and 11.8% in St. Lucia. Conversely, no *Bos indicus* mtDNA lineages were detected in any of the assayed Caribbean Creole populations; a feature also associated with the zebu populations of the African continent despite substantial zebu introgression from the Indian subcontinent over several millennia (MacHugh *et al.* 1997; Troy *et al.* 2001; Hanotte *et al.* 2002). Several authors contend that West African zebu cattle, possibly from Senegal, were imported to the Americas during the 18th and 19th centuries (Rouse 1973; Maillard *et al.* 1993). Furthermore, MacHugh *et al.* (1997) have shown that the *Bos indicus*-specific alleles assayed in this study are detected in some West African taurine and zebu populations at

moderate to low frequencies as a result of predominantly male-mediated zebu introgressive admixture. Thus, it is plausible that a portion of the zebu ancestry noted in Caribbean Creole cattle is of West African origin, as African zebu cattle introduced to the Americas would have carried both African *Bos taurus* mitochondrial chromosomes and zebu-specific alleles, the genetic signatures of which are observed in modern Creole populations.

It may also be speculated that some of the *Bos indicus* ancestry observed in the Caribbean Creole population surveyed here is the result of male-mediated introgression from the Indian subcontinent. Historical accounts state that Indian *Bos indicus* cattle, particularly males, were introduced to in large numbers Brazil during the 19th and 20th, which were later exported into neighbouring South American countries and to the Caribbean in order to improve local Creole stocks to tropical environments (Rouse 1973; Mason 1988; Bradley & Cunningham 1999). Indeed, the detection of a zebu-specific Y-chromosomal variant at high frequencies in Bolivian and Brazilian Creole cattle populations is congruent with this view (Giovambattista *et al.* 2000). However, it has also been suggested that zebu-specific Y-chromosomal variants could have reached South America *circa* 16th century AD, as a result of direct zebu introductions by Portuguese settlers from their African colonies (Giovambattista *et al.* 2000). Furthermore, the introduction of zebu dams from the Indian subcontinent to Brazil is also supported by the occurrence of 5 *Bos indicus* mtDNA sequences in the Nelore samples analysed in this study.

3.4.2 Genetic founder effects

Troy (1998) has demonstrated that approximately 5 haplotypes are encountered in a typical European breed sample size of 12 individuals. A similar value was obtained for the Creole T3 cluster, in which 13 haplotypes were encountered in 50 sequences assayed in a total of 4 breeds from the Caribbean and Brazil. The topology and uni-modal mismatch

distribution of the Creole T3 haplogroup is roughly consistent with those observed for European populations, suggesting that this diversity is a relatively even sample of the diversity encountered within continental Europe. In contrast, the diversity within the Creole T1 haplogroup is markedly different from that found within continental Africa. Troy *et al.* (2001) have demonstrated that almost all mtDNA diversity sampled from continental Africa coalesces to a predominant haplotype (T1), which is encountered rarely elsewhere. Also, it has been demonstrated that 12 samples from a continental African breed typically display an average of 5 haplotypes (Troy 1998). Conversely, only 3 haplotypes were encountered in a total of 23 Caribbean Creole and Brazilian samples displaying T1 haplotypes, with the $T1_{53,122,139,196}$ haplotype comprising 78% of all the African haplotypic variation. Furthermore, the African ancestral variant (T1) was only represented once in the Creole *Bos taurus* mtDNA sequence data.

The overall reduction in haplotypic diversity, the numerical predominance of the topologically peripheral $T1_{53,122,139,196}$ mtDNA haplotype and ragged mismatch distribution of the T1 haplotypes sampled in the Creole populations analysed here are collectively suggestive of a bottleneck or founder effect, in which a restricted subset of the ancestral T1 haplotypic variation survived and is displayed in extant populations. This sampling bottleneck of African mtDNA diversity may have occurred; (1) in Iberia, as a result of cattle migrations from Africa (Cymbron *et al.* 1999), (2) during the purported direct introduction of West African cattle in the Americas, *circa* 16^{th}-18^{th} centuries AD, or (3) a combination of both demographic processes. The former is supported by the pattern of diversity produced by the T1 haplotypes reported in Iberia, which is atypical to that encountered in continental Africa; thus indicating that there has been an bottlenecking of African mtDNA diversity within Iberia. The latter scenario (*i.e.* a bottleneck *en route* to the Caribbean) is suggested by the considerable proportion zebu-specific and taurine-specifc

alleles within modern Creole samples, which are presumably of West African origin, indicating a direct importation of cattle from the African continent.

The key answer to this question will undoubtedly be locating the geographic origin of the $T1_{53,122,139,196}$ haplotype, which thus far has not been documented elsewhere in previous studies of *Bos taurus* mtDNA phylogeography. It is possible that this T1-derivative may be detected in a wider sampling of Iberian mtDNA diversity, which would strengthen claims that Iberia was a source for the observed African ancestry in Creole cattle. Alternatively, the absence of this haplotype in an intensified Iberian sample would increase the likelihood that this African ancestry was indeed the result of cattle introductions directly from the African continent.

CHAPTER 4

SINGLE NUCLEOTIDE POLYMORPHISM VARIATION IN DOMESTIC CATTLE AND A COMPARISON WITH MITOCHONDRIAL DNA AND NUCLEAR MICROSATELLITE VARIABILITY

4.1 INTRODUCTION

4.1.1 Single nucleotide polymorphisms (SNPs) and their use in phylogenetic studies

The human genome-sequencing project has highlighted the wealth of DNA sequence variants that occur within mammalian genomes. The majority of these variants exist as single nucleotide polymorphisms or SNPs, single base-pair variations in genomic DNA, defined where the least common allele has a frequency of 1% or greater within a population (Brookes 1999; Kruglyak & Nickerson 2001). The surge of SNP data has simulated large-scale interest in applying genetic analysis to characterise several aspects of molecular biology, especially in molecular medicine and forensics (Schafer & Hawkins 1998; Gray *et al.* 2000; Jorde *et al.* 2001; Jobling 2001). SNPs also have the potential to assess the distribution of intra-specific genetic diversity, having important implications for evolutionary and phylogenetic studies (Chakravarti 1999; Jorde *et al.* 2000; Cann 2001; Przeworski *et al.* 2000; Yu *et al.* 2002).

Preliminary estimates based on human data suggest SNP loci occur approximately once for every kilobase (kb) of DNA assayed (Wang *et al.* 1998; Collins *et al.* 1997; Miller & Kwok 2001). This estimate has now been refined to 1 SNP every 1,331 bp of DNA with the advent of the human genome sequence (The International SNP Map Working Group 2001, Marth *et al.* 2001). To date, no such extensive studies have been performed in other mammalian species, however the limited data currently available from cattle indicate a high density of SNPs in *Bos taurus* cytokine genes, where a comparison of any two chromosomes have yielded an average of 1 SNP per 443 bp of DNA sequence (Heaton *et al.* 2001).

SNPs have low mutation rates and are considered to be extremely stable polymorphic markers. The mutation rate for SNPs, based on human data, has been estimated at 10^{-8}

changes per generation (Jorde *et al.* 1998; Ellegren 2000) and is markedly lower than that estimated for microsatellite DNA markers (10^{-3}-10^{-4} mutations per generation) (Ellegren 2002). The slow mutation rate of SNPs provides a means of investigating the more ancient history of a species, however, they are likely to be relatively uninformative regarding recent history (Jorde *et al.* 1998).

SNPs are digital in nature as allele scoring involves querying each polymorphic site for the absence or presence of a particular nucleotide base (Kruglyak 1997). As SNP genotyping is based directly on DNA sequence information, allelic determination does not require the inclusion of standard samples (*i.e.* samples of known genotype) (Fries & Durstewitz 2001). The digital nature of SNPs implies that allelic variants can be genotyped in a rapid, highly parallel fashion and numerous genotyping techniques are available (Lander 1999; Kwok 2001; Vignal *et al.* 2002).

Theoretically, SNPs can be tri- or tetra-allelic due to the four possible nucleotide bases that occur within mammalian DNA sequences. In practice, however, the vast majority of SNPs are bi-allelic. This is due to the low mutation rate associated with SNPs and a mutational bias towards nucleotide transitions at SNP loci (Vignal *et al.* 2002). Consequently, the information content of SNPs is severely reduced when compared to microsatellites, which are often multi-allelic (Mir & Southern 2000). Several SNP loci are therefore required appropriately evaluate the levels of genetic variability within populations.

Finally, the interpretation of SNP data may be complicated by ascertainment bias, such that an SNP may be more polymorphic in the population used in its discovery relative to other populations, resulting in an overestimation of diversity (Stoneking 2001).

4.1.2 Research objectives

To date, most phylogenetic analyses involving domesticated cattle have focused primarily on mtDNA sequence or microsatellite allelic variation. SNPs offer a vast, yet little exploited source of genetic variation that may provide a novel layer of information pertaining to the evolutionary relationship of modern domestic populations. The primary objective of the research presented in this chapter was to assess the distribution of nuclear SNP diversity within and between modern domestic cattle populations from four major geographical regions; Europe, Africa, the Near East and the Indian subcontinent. The resulting SNP-derived patterns of diversity were compared to those generated for alternate genetic systems, namely mtDNA and nuclear microsatellites.

4.2 MATERIALS AND METHODS

4.2.1 Genotyping SNP alleles using the oligoligation assay (OLA)

In this study, SNP alleles were genotyped using a modification of the OLA protocol described by Nickerson *et al.* (1990) and Tobe *et al.* (1996) devised by Dr. Gregor Durstewitz and Prof. Ruedi Fries (Department of Animal Science, Technical University of Munich, Germany). This methodology involves the use of two oligonucleotide probes that hybridise to complimentary target DNA sequences, such that the 3'end of one oligonucleotide probe is immediately adjacent to the 5' terminal of the second oligonucleotide probe, with the join situated at the position of a known SNP locus. If both probes hybridise perfectly to the target DNA sequence thermostable *DNA ligase* can covalently bind the two probes together under optimal conditions producing an allele-specific OLA product.

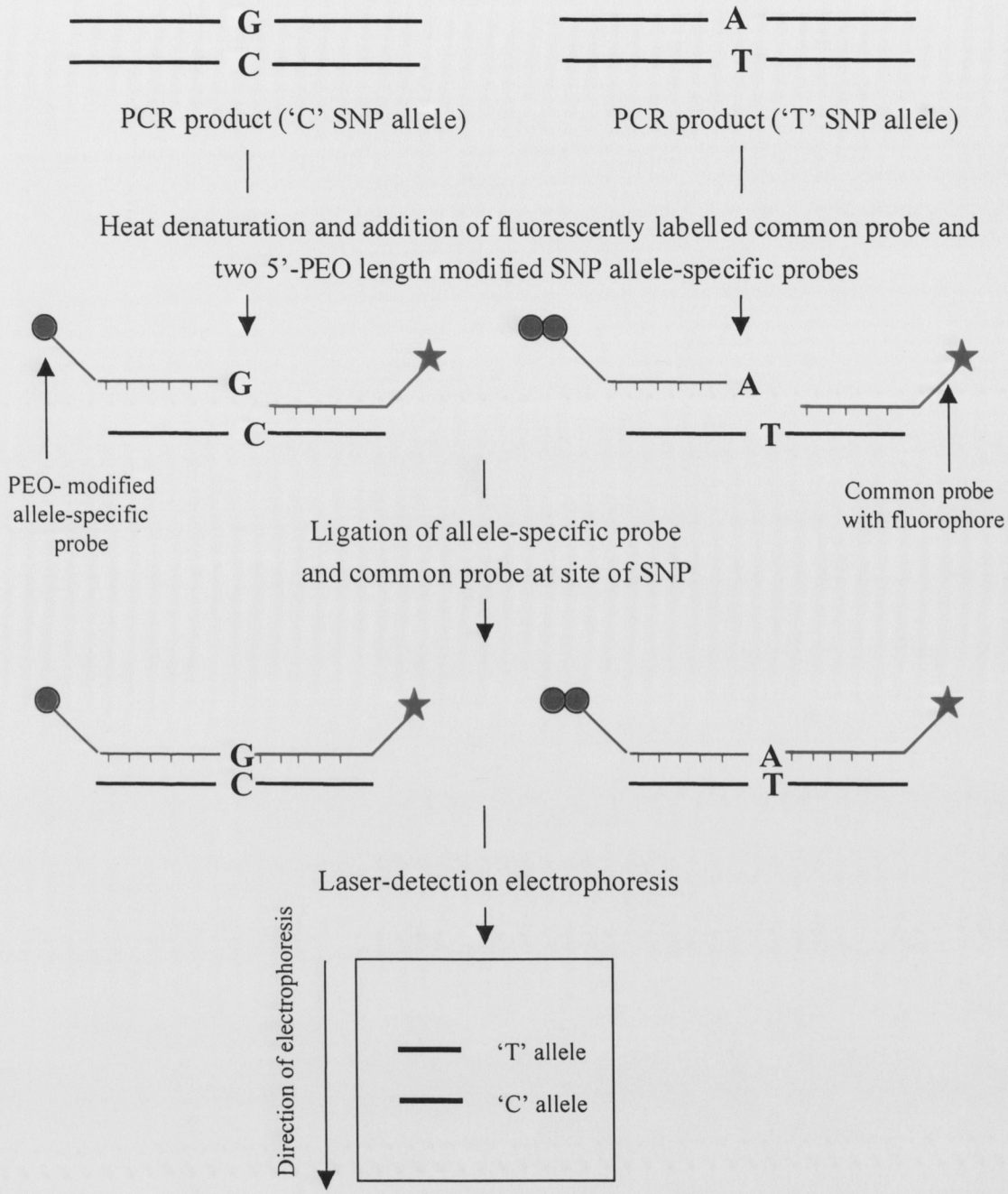
Separate ligation reactions are performed for each SNP allele. This is achieved by including two SNP-allelic specific probes and one common probe, together with PCR-amplified target DNA possessing a known SNP locus, in the OLA assay. PCR and OLA experimental conditions are described in **Section 5.5**. Both allelic-specific probes differ in sequence composition only at their 3' ends, which depends on the allelic form present. In addition, both of these probes posses varying numbers of flexible, neutral polyethylene oxide (PEO) synthetic tethers at their 5'ends, which confer the equivalent of an additional ~2.4 nucleotides to each allelic-specific probe (Cload & Schepartz 1991; Dr. G. Durstewitz, personal communication). Alternatively, a fluorescent dye molecule is attached to the 3' end of the common probe. This molecular system permits allelic forms to be determined based on the size produced by each of the OLA reactions which are detected by means of fluorescence electrophoresis using an ABI373 automated sequencer (APPLIED BIOSCIENCES). This procedure is described diagrammatically in **Figure 4.1**.

Raw SNP genotypic data were collected by use of the GENESCAN™ software and gel files were analysed using the GENOTYPER™ software package (APPLIED BIOSCIENCES). The OLA genotyping system yields a simple numerical output, whereby SNP alleles are denoted "Allele 1" or "Allele 2", based on the resulting OLA profiles, that can be imported directly into a Microsoft Excel © worksheet. This allows SNP genotypic data to be assessed using various sophisticated phylogenetic packages.

4.2.2 SNP allele genotype data

Eleven (11) SNP loci were assayed for allelic variation in a total of 390 samples constituting 14 extant cattle populations from Europe, the Near East, Africa and the Indian subcontinent. Genotypic data for an additional 94 individuals, comprising 3 breeds of German origin, were contributed by Dr. Gregor Durstewitz. All 11 SNP loci were developed by members of the Department of Animal Science, Technical University of Munich, Germany and have subsequently been deposited in the livestock SNP database (www.SNPzoo.de). Sample information for each cattle breed surveyed for SNP allelic variation together with details of the SNP loci assayed are presented in **Table 4.1**.

Figure 4.1: Genotyping SNP allels using the oligoligation assay (OLA)



Alleles are detected by differences in OLA product size as determined by the number PEO molecules attached to each allele-specific probe. Each PEO unit confers an equivalent 2.4 nucleotides to each oligoligation product.

Table 4.1: SNP locus information and populations considered for phylogenetic analysis

SNP locus name	Chromosome	SNP alleles	Breed assayed	No. samples	Geographic origin	Geographic group
TCRA	BTA 10q15	C / T	Kerry (KE)	30	Ireland	
CAC	BTA 4q18	C / G	Jersey (JE)	26	UK	
421-10	BTA 1q43-44	C / G	Aberdeen Angus (AA)	29	UK	
417-16	BTA 4q22-24	G / A	Hereford (HE)	30	UK	
463-67	BTA 17q12-14	C / T	Highland (HL)	25	UK	
454-G11	BTA 17q24-26	C / G	East Finn (EF)	30	Finland	
436-C10	BTA 4q13-14	C / T	Vosges (VO)	27	France	European *B. taurus*
CNTFR	BTA 8q21	C / G	Pinzgauer (PN)	27	Germany	
437	BTA 14q14-16	C / A	Braunvieh (BV)	30	Germany	
486-67	BTA 3q31-32	C / T	Fleckvieh (FV)	29	Germany	
444-66	BTA 8q18-22	C / T	Holstein (SB)	35	Holland	
			Hungarian Grey (HG)	30	Hungary	
			Total	**348**		
			Anatolian Black (AB)	25	Central	
			South Anatolian Red (SAR)	30	South Turkey	Near Eastern *B. taurus*
			Total	**55**		
			Guinean N'Dama (ND)	27	Guinea	African *B. taurus*
			Total	**27**		
			North Indian zebu (NIZ)	25	Northwest	
			Ongole (ON)	29	South India	Indian *B. indicus*
			Total	**54**		
			Overall total	**484**		

Table 4.1. The SNP loci and the populations analysed for SNP allelic variation in this study. Chromosomal positions of SNP loci were provided by Dr. Gregor Durstewitz. SNP alleles shown in red refer to 'Allele 1' and alleles shown in blue refer to 'Allele 2' according to the genotyping assay numerical output, discussed in **Section 4.2.2** and **Section 5.5**. All populations were genotyped for the first time in this study with the exception of the Braunvieh (BV), Fleckvieh (FV) and Holstein (SB) populations, which were genotyped by Dr. Gregor Durstewitz (personal communication). All assayed cattle samples were procured as DNA that was extracted by various members of the Molecular Population Genetics laboratory (TCD) prior to the initiation of this project with the exception of the East Finn samples, which were collected by Dr. Juha Kantanen (Agricultural Research Centre, Jokioinen, Finland). Population codes are shown in brackets. The *Bos indicus* population is a composite of SNP data from 3 populations originating from the Indo-Pakistan border (Sahiwal, $n = 6$; Hariana, $n = 10$; Tharparker, $n = 9$).

4.2.3 Mitochondrial DNA sequence data and microsatellite genotypic data considered for comparative phylogenetic analysis

Mitochondrial DNA sequence data

The mtDNA sequence data employed in this study comprised a total of 504 mtDNA sequences detected in 42 extant domestic cattle populations from Europe, Africa, the Near East and the Indian subcontinent. The mtDNA data collection used for comparative phylogenetic analysis is shown in **Table 4.2** and the sources of all these data are listed in **Table 5.1**.

Autosomal microsatellite data

The microsatellite data collection (**Table 4.3**) comprises 806 samples constituting a total of 21 modern cattle populations from Europe ($n = 10$), African ($n = 2$), the Near East ($n = 7$) and India ($n = 2$) using data previously reported by Schmid *et al.* (1999), Loftus *et al.* (1999), Edwards *et al.* (2000) and Teresa Cymbron (Department of Genetics, TCD, personal communication). The sources of these data are also listed in **Table 5.3**.

Table 4.2: The mtDNA sequence data collection used for comparative analysis

Regional groups									
Europe				Africa		Near East		Indian subcontinent	
Population name	No. seqs	Population name	No. seqs	Population name	No. seqs	Population name	No. seqs	Population name	No. seqs
Icelandic (IC)	12	Jersey (JE)	11	Butana (BU)	12	Turkish Grey (TU)	11	Red Sindhi (RS)	2 / 20
Norwegian red (NO)	12	Berrenda (BE)	11	Egyptian (EG)	11	Anatolian Black (AB)	12	Tharparker (TH)	1 / 20
Telemark (TE)	5	Charolais (CH)	12	Guinean N'Dama (ND)	11	East Anatolian Red (EAR)	9 / 1	Sahiwal (SA)	12
Westland Fjord (WF)	12	Friesian (FR)	12	Kapsiki (KA)	12	South Anatolian Red (SAR)	11	Hariana (HA)	1 / 9
Kerry (KE)	10	German Black (GE)	10	Kenana (KN)	7	Damascus (DA)	12	Desi (DE)	19
Aberdeen Angus (AA)	12	Limousin (LI)	12	Kuri (KU)	11	Kurdish (KUR)	11 / 1	Ongole (ON)	11
Galloway (GA)	18	Romagnola (RO)	11	Namchi (NA)	12	North Iraq (NIR)	10 / 2		
Highland (HI)	12	Simmental (SI)	11	Somba (SO)	12	Middle Iraq (MIR)	8 / 3		
Hereford (HE)	12	Sykia (SY)	12	White Fulani (FU)	7	South Iraq (SIR)	9 / 7		
Total number of seqs. 207				**Total number of seqs. 95**		**Total number of seqs. 107**		**Total number of seqs. 95**	

Table 4.2. The mtDNA data set employed for comparative analysis comprised a total of 504 sequences detected in 42 populations from Europe, Africa, the Near East and the Indian subcontinent. The source of the data used in this study is outlined in detail in **Table 5.1**. Red and blue colouration indicates the number of *Bos taurus* and *Bos indicus* sequences in each population, respectively. Population codes are given in brackets.

Table 4.3: The microsatellite data used for comparative analysis

Population	No. samples	Population origin	Regional group	Morphology
Jersey (JE)[1]	44	UK	Europe	*Bos taurus*
Hereford (HE)[2]	34	UK		
Highland (HI)[2]	37	UK		
Simmental (SI)[3]	50	Switzerland		
Holstein (HO)[3]	50	Holland		
Charolais (CH)[1]	36	France		
Vosges (VO)[4]	27	France		
Pustertaler (PS)[4]	31	Germany-Italy		
Pinzgauer (PN)[4]	30	Germany-Austria		
Hungarian Grey (HG)[1]	44	Hungary		
Total	**383**			
Turkish Grey (TU)[1]	43	Turkey	Near East	*Bos taurus*
Anatolian Black (AB)[1]	41	Central Turkey		
South Anatolian Red (SAR)[1]	43	South Turkey		
East Anatolian Red (EAR)[1]	44	East Turkey		
Damascus (DA)[1]	41	Syria		
Iraqi (IR)[1]	42	Iraq		
Kurdish (KUR)[1]	36	North Iraq		
Total	**290**			
Guinean N'Dama (ND)[1]	44	Guinea	Africa	*Bos taurus*
Egyptian (EG)[1]	32	Egypt		
Total	**76**			
Omgole (ON)[1]	30	South India	India	*Bos indicus*
Nelore (NE)[1]	27	Brazil		
Total	**57**			
Overall total	**806**			

Table 4.3. Microsatellite data used for comparative genetic analysis presented in this study.

Microsatellite loci: The microsatellite allelic data was based on 18 nuclear loci: *BM1818, BM1824, BM2113, CSRM60, CSSM66, ETH10, ETH225, Hel1, Hel5, Hel9, Hel13, ILSTS005, ILSTS006, INRA005, INRA023, INRA032, INRA035, INRA063.*

Sources of data (also listed in Table 5.3):

[1] Data published by Loftus *et al.* (1999).

[2] Data contributed by Teresa Cymbron (Department of Genetics, TCD, personal communication).

[3] Data published by Schmid *et al.* (1999).

[4] Data published by Edwards *et al.* (2000).

4.2.4 Phylogenetic analysis of molecular data

Genetic diversity estimates

(a) MtDNA genetic diversity estimates

For mtDNA sequence data, intra-regional group diversities are given as the mean number of sequence pairwise differences (MNPD) (**Section 2.2.3**) and were calculated using the ARLEQUIN VERSION 2.0 software programe (Schneider *et al.* 1999).

(b) Nuclear microsatellite and SNP genetic diversity estimates

For nuclear genetic data, estimates of intra-regional group genetic diversity were calculated using Nei's unbiased gene diversity (Nei 1987), which is an indicator of genetic diversity that is less sensitive to sample size than the number of distinct alleles detected at a given locus. For diploid organisms this diversity estimate is equivalent to expected heterozygosity (*i.e.* the proportion of heterozygotes expected in a population under random mating). For a random sample of n individuals in which the frequency of the genotype A_iA_j is denoted $\hat{X}_{ij}$, Nei's unbiased gene diversity is given by the formula (Nei 1987):

$$\hat{h} = 2n\left(1 - \sum \hat{x}_i^2\right)/(2n-1)$$

where

$$\hat{x}_i = \hat{X}_{ii} + \sum_{i?j} \hat{X}_{ij} / 2$$

Average gene diversity (across r loci) is given by:

$$\hat{H} = \sum_{j=1}^{r} \hat{h}_j / r$$

Nei's unbiased genetic diversity for all nuclear genetic systems was calculated as using the MICROSATELLITE TOOLKIT (Park 2002).

For microsatellite data, average sample-size adjusted mean number of alleles (MNA) were calculated for each regional population using a resampling with replaced with 1000 replicates. This was performed using the MICROSATELLITE TOOLKIT program written by Dr. Stephen Park (Park 2002).

Estimates of the proportion of genetic variance attributable to population subdivision and hierarchical analysis of molecular variation (AMOVA)

The proportion of genetic variance attributable to population subdivision (*i.e.* differences between subpopulations) was estimated using the fixation index, F_{ST} (Wright 1965). F_{ST} values, as defined by Wright (1965), is the correlation between two alleles chosen at random within subpopulations relative to alleles sampled at random in the total population (Balloux & Lugon-Moulin 2002). In terms of inbreeding coefficients (Weir & Cockerham 1984) and coalescence times (Slatkin 1991) this F_{ST} statistic can be expressed as:

$$F_{ST} = \frac{f_0 - f_1}{1 - f_1} \quad \text{or} \quad F_{ST} = \frac{t_1 - t_0}{t}$$

where f_0 is the probability of identity by descent of two alleles drawn from the same subpopulation (*i.e.* within-subpopulation genetic diversity), f_1 is the probability of identity

by descent of two alleles drawn from different subpopulations (*i.e.* between-subpopulation diversity), t_1 is the mean coalescence time of two alleles drawn from two different populations and t_0 is the mean coalesce time of two alleles drawn from the same population (Balloux & Lugon-Moulin 2002). F_{ST} values range from 0 (all genetic diversity within a species is shared equally by all populations with no genetic differences among populations) to 1 (all intra-specific genetic diversity is found as fixed differences among populations with no genetic diversity within populations) (Templeton 1998).

Hierarchical analysis of molecular variation (AMOVA; **Section 2.2.3**) analysis and corresponding tests of significance were also computed using the ARLEQUIN VERSION 2.0 software package

Admixture analysis

SNP allele frequencies for each locus were calculated using the MICROSATELLITE TOOLKIT (Park 2002). SNP allele frequencies were used to estimate the proportion of Indian zebu introgression into Near Eastern and African *Bos taurus* cattle populations using the method developed by Bertorelle & Excoffier (1998). This procedure is based on Bernstein's (1931) model of admixture in which allele frequencies in the admixed population are linear combinations of the allele frequencies from the two source populations and the effects of drift and mutation are presumed to be negligible. In this model, allele frequencies in the admixed population are given by:

$$p_h = p_1 m + p_2 (1 - m)$$

where m is the admixture proportion of the first source population and where p_h, p_1 and p_2 are the allele frequencies in the admixed and the two source populations, respectively.

The ADMIX VERSION 1.0 software (Bertorelle 1998) was used to obtain the standard least-squares estimate of this model (*m*R) which is based on fitting a regression through the points defined by $p_1 - p_2$ and $p_h - p_2$. For this, the Kerry, Jersey, Aberdeen Angus, Hereford, Highland, East Finn, Vosges, Pinzgauer, Braunvieh, Fleckvieh and Holstein cattle were pooled and used to represent a putative European parental population, while the Ongole population was employed as a putative *Bos indicus* source population.

Inter-population genetic distance estimates

For nuclear polymorphisms, the genetic relationships among populations were estimated using Nei's standard distance, D_A (Nei 1987; **Section 3.2.3**). D_A estimates were generated using the DISPAN software package (Ota 1993). For mtDNA data, population pairwise F_{ST} values were employed as genetic distance estimates among populations (Slatkin 1991; Slatkin 1995; **Section 2.3.9**). These were computed using the ARLEQUIN VERSION 2.0 software package (Schneider *et al.* 1999). The resulting genetic distances generated for each genetic marker (**Appendices 2.I-2.N**) were plotted in two-dimensions following multi-dimensional scaling (MDS) (Kruskal 1964, see **Section 2.3.9**) using the SPSS (VERSION 8.0) software package.

4.3 RESULTS

4.3.1 Oligoligation assay (OLA) GENESCAN™ gel files

Figure 4.2 shows a typical OLA gel file produced after 4 hours of gel electrophoresis using an ABI371 automated sequencer (APPLIED BIOSCIENCES). Gel electrophoresis was performed as described in **Section 5.5**.

4.3.2 Intra-regional genetic diversity

Genetic diversity estimates

Intra-regional genetic diversity estimates calculated for the three genetic marker systems analysed are shown in **Table 4.4**. As demonstrated previously, mtDNA and microsatellite-based estimates of diversity are visibly highest in Near Eastern domestic populations (Loftus *et al.* 1999; Troy *et al.* 2001). MtDNA-based intra-regional diversity is also high within the domestic zebu populations of the Indian subcontinent and are comparable to those yielded to for the Near East (Chapter 2 of this thesis).

In contrast, nuclear SNP genetic diversity levels are pronouncedly higher in the European regional group (0.377) with reductions in the Near East (0.318), Africa (0.218) and India (0.118). This pattern is also evident when each surveyed cattle population is considered individually, whereby genetic diversity estimates range between 0.109 for the Indian zebu Ongole population and 0.393 for the British taurine Highland population (**Table 4.5**).

Figure 4.2. GENESCAN™ and GENOTYPER™ files showing bovine SNP genotypes assayed using OLA

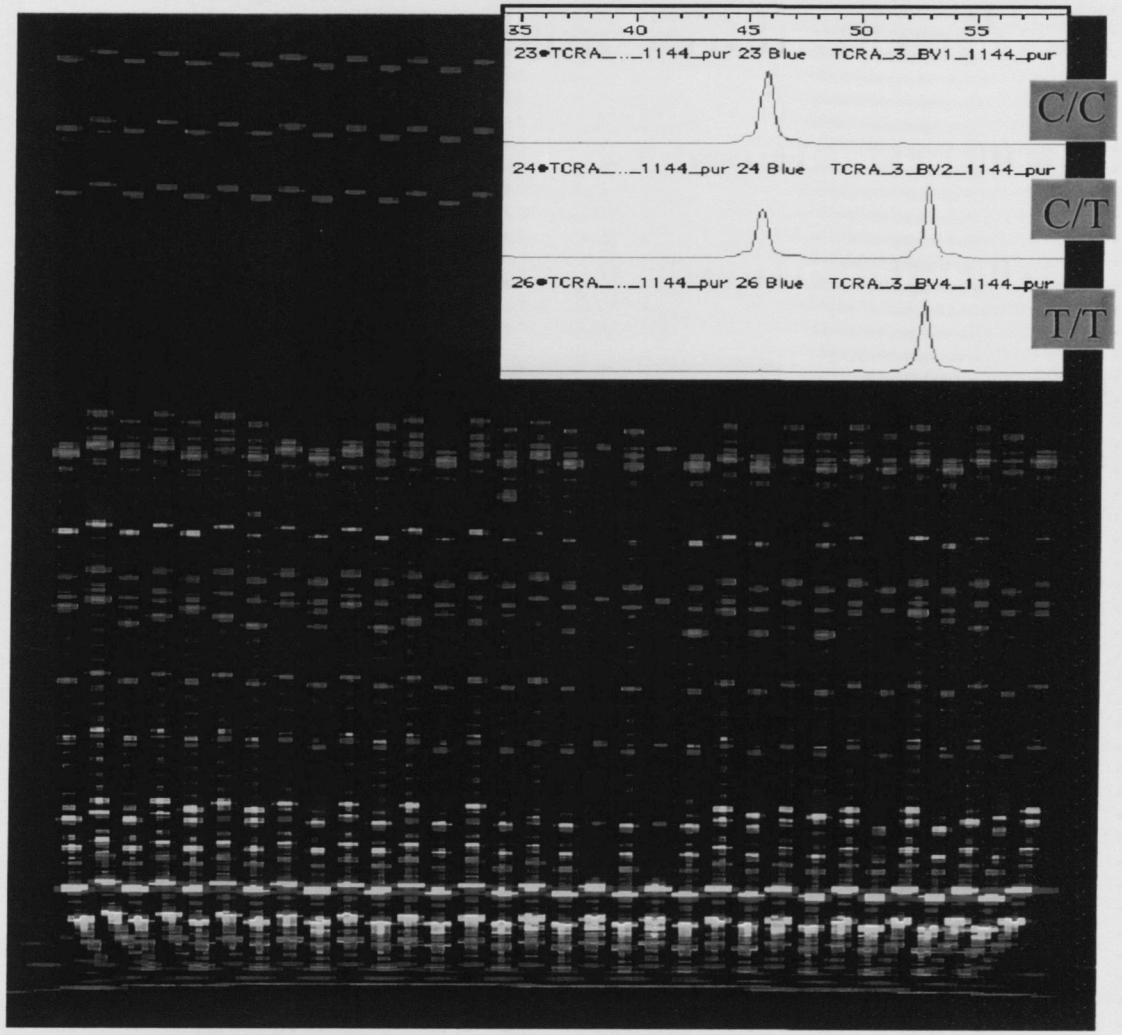


Figure 4.2. SNP OLA GENESCAN™ gel file. Inset depicts a typical GENOTYPER™ results file showing SNP "allele 1" and "allele 2" peaks for the TCRA locus. Other loci typed as part of this multiplex (coloured yellow, blue and green) are visible; red indicates size standards. This diagram is reprinted by kind permission of Dr. Gregor Durstewitz.

Table 4.4: Genetic diversity estimates for geographical regional groups

	Genetic diversity		
Regional group	**SNP[1]**	**Microsatellite[1]**	**MtDNA[2]**
Europe	0.377 (0.042)	0.709 (0.030)	2.06 (1.16)
Near East	0.318 (0.055)	0.787 (0.019)	8.80 (4.09) / 3.61 (1.85)
Africa	0.218 (0.066)	0.684 (0.032)	2.09 (1.18)
India	0.118 (0.038)	0.674 (0.027)	5.54 (2.69) / 3.69 (1.88)

[1] **Nuclear genetic markers:** Nei's unbiased gene diversities were calculated according to Nei (1987) for both nuclear SNP and microsatellite data using the MICROSATELLITE TOOLKIT (Park 2002). Standard deviations are shown in brackets.

[2] **mtDNA:** Within-region mtDNA sequence diversity estimates are given as the mean number of pairwise differences (MNPD) (Tajima 1983). Standard deviations are shown in brackets. Underlined values represent the MNPD values estimated for both the Near Eastern and Indian groups, when identified *Bos indicus* and *Bos taurus* introgressor sequences are removed.

Table 4.5: SNP-based genetic diversity estimates for each individual cattle population

Population	No. samples	Breed origin	Average gene diversity
Kerry (KE)	30	Ireland	0.358 (0.066)
Jersey (JE)	26	UK	0.296 (0.055)
Hereford (HE)	30	UK	0.325 (0.061)
Angus (AA)	29	UK	0.279 (0.059)
Highland (HL)	25	UK	0.393 (0.049)
East Finn (EF)	30	Finland	0.309 (0.052)
Vosges (VO)	27	France	0.332 (0.066)
Pinzgauer (PN)	27	Germany	0.338 (0.043)
Braunvieh (BV)	30	Germany	0.388 (0.051)
Fleckvieh (FV)	29	Germany	0.337 (0.043)
Holstein (SB)	35	Germany	0.307 (0.062)
Hungarian Grey (HG)	30	Hungary	0.290 (0.057)
Regional total	**348**		**0.377 (0.042)**
Anatolian Black (AB)	25	Central Turkey	0.332 (0.054)
South Anatolian Red (SAR)	30	South Turkey	0.299 (0.058)
Regional total	**55**		**0.318 (0.055)**
Guinean N'Dama (ND)	27	Guinea	0.218 (0.066)
Regional total	**27**		**0.218 (0.066)**
North Indian zebu (NIZ)	25	Indo-Pakistan border	0.123 (0.046)
Ongole (ON)	29	South India	0.109 (0.041)
Regional total	**54**		**0.118 (0.038)**

Within-population genetic diversity is given as Nei's unbiased gene diversity which was calculated according to Nei (1987) using the MICROSATELLITE TOOLKIT (Park 2002). Standard deviations are shown in brackets.

Genetic diversity as expressed as the total number of alleles and haplotypes unique to each regional group for all three genetic systems

Tables 4.6a, **4.6b** and **4.6c** summarise genetic diversity for all three genetic systems in terms of the total number of alleles/haplotypes per regional group and the number of alleles/haplotypes unique to regional groups (*i.e.* present in one group but not in the other groups).

A total of 193 mtDNA haplotypes were detected in all 504 sequences from all 4 regional groups. To ensure that sample sizes in each region were approximately equal, the European sample (n = 207), which is 1.93 and 2.18 times larger than the Near Eastern and African and Indian samples, respectively, was divided into two groups, namely Britain and West Fringe Europe (n = 116) and continental Europe (n = 91). Congruent with the mtDNA diversity estimates, the Near Eastern sample displays the highest number of haplotypes (66) and unique mtDNA haplotypes (44) of all regional groups surveyed.

A total of 214 microsatellite alleles were observed across all 18 loci in 806 samples from all 4 regional groups. The Near Eastern regional group exhibits both the highest total number of microsatellite alleles (199 alleles) and the greatest number of unique alleles (27 alleles) of all regional groups surveyed. Corresponding reductions in both the total number of alleles and the total number of unique alleles are seen for the remaining three regional groups.

Consistent with the SNP-based diversity estimates, the European regional group exhibits the greatest number of SNP alleles of all the regions surveyed. All possible 22 alleles for the 11 bi-allelic loci were encountered at varying frequencies in the European sample. The European group also displays the only unique SNP allele detected in the total data set. Notably, allele fixation or a predominance of one particular allele is observed within all

remaining regional groups. Twenty-one SNP alleles were encountered within the Near East (allele fixation detected at the 444-66 locus), 18 alleles were represented within the Indian sample (allele fixation at four loci, namely 444-66, CNTFR, 436-C10 and CAC) and 17 alleles were detected in the only African population surveyed, namely the N'Dama population from Guinea (allele fixation at the 444-66, 486-67, 421-10, 437 and CNTFR loci). SNP allele frequencies for each regional division are represented graphically in **Figure 4.3**.

Allele and haplotypic sharing among regional groups for all three genetic systems

Table 4.7 shows the pattern of haplotypic and allelic sharing among the four major regional divisions for each of the three genetic systems. MtDNA haplotypic sharing is evident from some pairwise regional comparisons, particularly between the Near East and Continental Europe and the Near East and Britain/West Fringe Europe. However, no mtDNA haplotypes are common to all of the major regional samples. Similarly for microsatellite data, substantial allelic sharing between regional groups is noted for all pairwise regional comparisons and an estimated 36.9% of all alleles are common to all four regional samples. In contrast, an estimated 68.2% of all detected SNP alleles (15 out of 22 alleles) are common all four geographical regional groups.

Table 4.6a: Intra-regional mtDNA genetic diversity as expressed in terms of the total number of haplotypes and number of unique haplotypes

Regional group	Sample size	Total no. haplotypes (*B. taurus* / *B. indicus*)	No. of unique haplotypes (*B. taurus* / *B. indicus*)
Britain /W. Fringe Europe	116	43 (43 / 0)	34 (34 / 0)
Continental Europe	91	32 (32 / 0)	18 (18 / 0)
Near East	107	66 (62 / 4)	44 (43 / 1)
Africa	95	43 (43 / 0)	39 (39 / 0)
India	95	37 (3 / 34)	34 (3 / 31)
Overall	**504**	**193 (158 / 35)**	**169 (137 / 32)**

Results were tabulated using the haplotypic data detailed in **Tables 4.2**. For this analysis, the European sample was divided into two groups to ensure that sample sizes for all regional groups were roughly equivalent. These European subgroups consisted of the following populations:

(1) Britain and West Fringe Europe (n = 116) - Icelandic (IC), Norwegian red (NO), Telemark (TE), Westland Fjord (WF), Kerry (KE), Aberdeen Angus (AA), Galloway (GA), Hereford (HE), Highland (HI).

(2) Continental Europe (n = 91) - Berrenda (BE), Charolais (CH), Friesian (FR), German Black (GE), Limousin (LI), Romagnola (RO), Simmental (SI), Sykia (SY).

Table 4.6b: Intra-regional microsatellite genetic diversity as expressed in terms of the total number of alleles and number of unique alleles

Regional group	Sample size	Total No. alleles	Mean number of alleles per region	Total No. of unique alleles
Europe *B. taurus*	383	161	7.28	10
Near East *B. taurus*	290	199	9.25	27
Africa *B. taurus*	76	143	7.36	1
India *B. indicus*	57	111	5.87	1
Overall	**806**	**214**		**39**

Results were tabulated using the genotypic data detailed in **Tables 4.3**.

Table 4.6c: Intra-regional SNP genetic diversity as expressed in terms of the total number of alleles and number of unique alleles

Regional group	No. samples	Total no. of alleles	No. of unique alleles
Europe *B. taurus*	348	22	1
Near East *B. taurus*	55	21	-
Africa *B. taurus*	27	17	-
India *B. indicus*	54	18	-
Overall	**484**	**22**	**1**

Results were tabulated using the genotypic data detailed in **Table 4.1**.

Table 4.7a: The number of mtDNA haplotypes shared between regional groups

Pairwise regional comparison	No. of shared haplotypes (*B. taurus*/ *B. indicus*)	Number of shared haplotypes as a percentage of the total number in each region
Near East & Britain/West Fringe Europe	7 (7 / 0)	10.6% / 16.3%
Near East & Continental Europe	12 (12 / 0)	18.2% / 37.5%
Near East & Africa	4 (4 / 0)	6.0% / 9.3%
Near East & India *B. indicus*	3 (0 / 3)	4.0% / 8.1%
Britain/West Fringe Europe & Cont. Europe	6 (6 / 0)	13.9% / 18.7%
Britain/West Fringe Europe & Africa	0	0
Continental Europe & Africa	0	0
Britain/West Fringe Europe & India	0	0
Continental Europe & India	0	0
Africa & India	0	0
Overall (sharing among all regional groups)	**0**	**0**

Values in the third column represent the number of shared haplotypes expressed as a percentage of the total number of haplotypes occurring within each region - underlined values correspond to the underlined regional group.

Table 4.7b: The number of nuclear microsatellite alleles shared between regional groups

Pairwise regional comparison	No. of shared alleles	Number of shared alleles as a percentage of the total number in each region
Near East & Europe	149	74.8% / 92.5%
Near East & Africa	139	69.8% / 97.2%
Near East & India	108	54.3% / 97.3%
Europe & Africa	128	79.5% / 89.5%
Europe & India	89	55.3% / 80.1%
Africa & India	94	65.7% / 84.7%
Overall (sharing among all regional groups)	**79**	**36.9% of all detected alleles**

Values in the third column represent the number of shared alleles expressed as a percentage of the total number of alleles occurring within each region. Underlined values correspond to the underlined regional group.

Table 4.7c: The number of nuclear SNP alleles shared between regional groups

Pairwise regional comparison	No. of shared alleles	Number of shared alleles as a percentage of the total number in each region
Near East & Europe	21	100% / 95.4%
Near East & Africa	17	80.9% / 100%
Near East & India	18	85.7% / 100%
Europe & Africa	17	77.3% / 100%
Europe & India	18	81.8% / 100%
Africa & India	15	88.2% / 83.3%
Overall (sharing among all regional groups)	**15**	**68.2% of all detected alleles**

Values in the third column represent the number of shared alleles expressed as a percentage of the total number of alleles occurring within each region. Underlined values correspond to the underlined regional group.

Figure 4.3: SNP allele frequencies for all regional groups

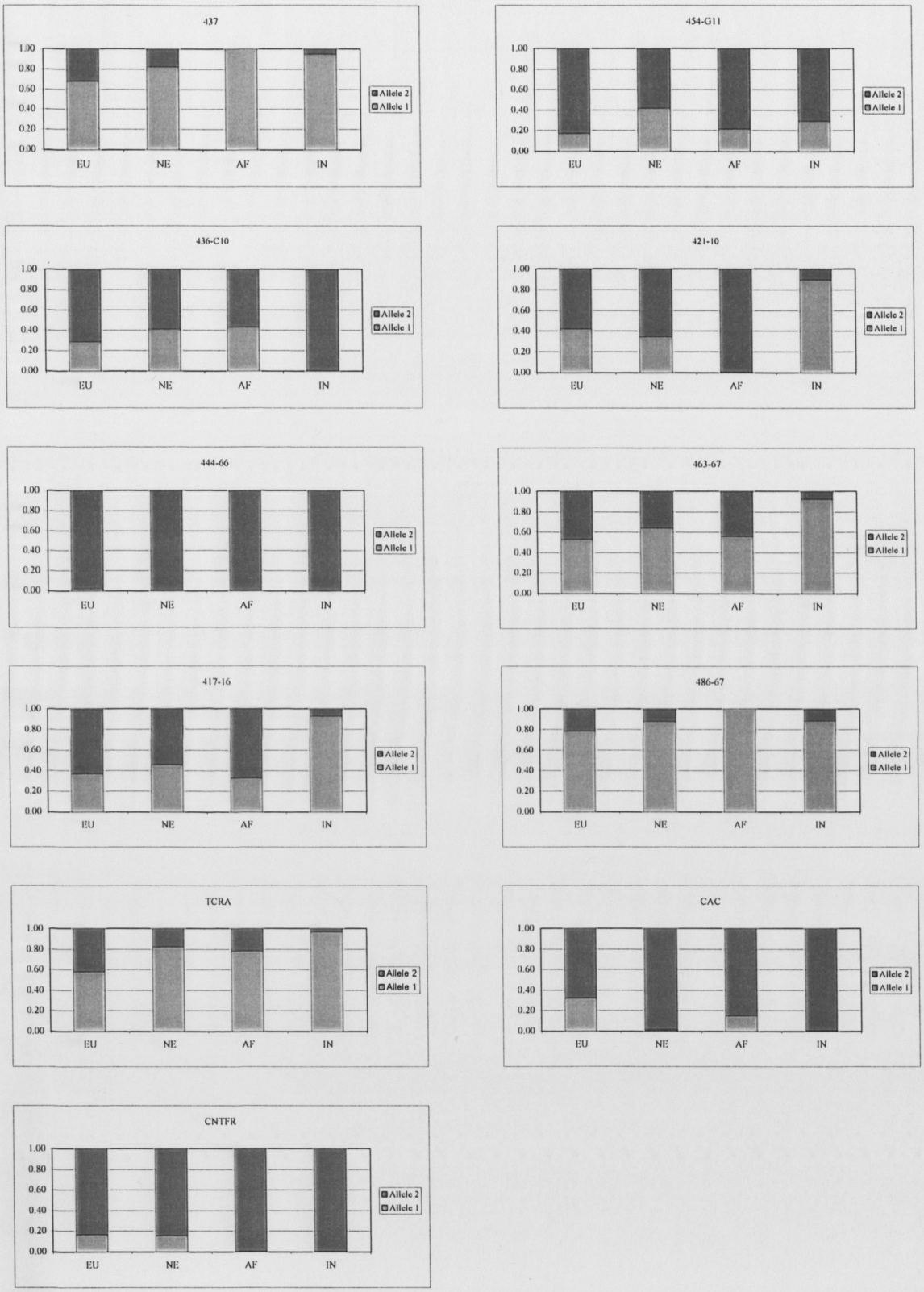


Figure 4.3. SNP allele frequencies for each locus within each assayed regional group. Bi-allelic variability is observed at most loci within all groups, however, allelic fixation is observed at a number of loci within the African *Bos taurus* and Indian *Bos indicus* groups. Group codes are as follows: EU (European *Bos taurus*), NE (Near Eastern *Bos taurus*), AF (African *Bos taurus*), IN (Indian *Bos indicus*). The nucleotide identities for alleles '1' and '2' are given in **Table 4.2**.

4.3.3 The apportionment of genetic diversity between and within regional groups

Genetic differentiation between regional groups

The proportion of the total genetic variance attributable to population subdivision for all three genetic systems was quantified using the F_{ST} statistic (Wright 1965). F_{ST} values estimated for the total Old World sample (*i.e.* over all four geographical regions) using the regional groups (Europe, Africa, the Near East and India) as subdivisions are given in **Table 4.8**. An F_{ST} value of 0.153 (or 15.3%) characterises the contribution of differences among regional groups to the total Old World nuclear SNP variance. A reduction is observed in the corresponding nuclear microsatellite-derived F_{ST} estimates (0.082). Collectively, these data suggest that the majority of the total nuclear variance is due to genetic differences between individuals within regional groups, rather than between regional groups. In contrast, an F_{ST} value of 0.689 generated for mtDNA, indicates that most Old World mtDNA variance is attributable to inter-regional genetic differentiation. When individual cattle populations, rather than the four regional populations, are used as units of subdivision, similar F_{ST} estimates are obtained for all three genetic systems (F_{ST} values of 0.199 for SNP data; 0.120 for microsatellite data; 0.643 for mtDNA data).

Genetic differentiation between populations within regional groups

Table 4.8 also shows F_{ST} estimates for each region separately using the individual populations as units of subdivision. Intra-regional F_{ST} values indicate that between 2%-13% of the total variance for all three genetic markers is accounted for by genetic differences between populations. For SNP data, European populations are the most differentiated (F_{ST} = 0.133), with Near Eastern (F_{ST} = 0.018) and Indian populations (F_{ST} = 0.044) displaying the least genetic differentiation. SNP-derived F_{ST} estimates were not generated for Africa as only one population was assayed for allelic variability.

African populations show the greatest differentiation for nuclear microsatellite data (F_{ST} = 0.126). European populations also show an appreciable level of differentiation for this genetic marker system (F_{ST} = 0.118). Similar microsatellite-based F_{ST} estimates have been generated for European populations in previous studies (MacHugh *et al.* 1998; F_{ST} = 0.112). In contrast, the populations of the Near East (F_{ST} = 0.024) and the Indian subcontinent (F_{ST} = 0.033) display increased genetic homogeneity. For mtDNA data, Near Eastern populations are the most highly differentiated (F_{ST} = 0.100) populations, with lower F_{ST} values noted for Europe (0.089), Africa (0.073) and India (0.039).

Table 4.8: The proportion of genetic variance attributable to population subdivision as represented by F_{ST} values

	F_{ST} value		
Regional group	**SNP**	**Microsatellite**	**mtDNA**
European *Bos taurus*	0.133 ($P \leq 0.001$)	0.118 ($P \leq 0.001$)	0.089 ($P \leq 0.001$)
African *Bos taurus*	n/a	0.126 ($P \leq 0.001$)	0.073 ($P \leq 0.001$)
Near Eastern *Bos taurus*	0.018 ($P \leq 0.030$)	0.024 ($P \leq 0.001$)	0.100 ($P \leq 0.001$)
Indian *Bos indicus*	0.044 ($P \leq 0.005$)	0.033 ($P \leq 0.001$)	0.039 ($P \leq 0.040$)
Old World: four regions	0.153 ($P \leq 0.001$)	0.082 ($P \leq 0.001$)	0.689 ($P \leq 0.001$)
Old World: individual pops.	0.199($P \leq 0.001$)	0.120 ($P \leq 0.001$)	0.643 ($P \leq 0.001$)

SNP-based F_{ST} values could not be generated for Africa as only one African population was assayed for variation. All F_{ST} values are significant at the $P \leq 0.05$ level.

Table 4.9 presents a comparison of the overall Old World F_{ST} values for all 3 genetic systems analysed in conjunction with the overall number of unique haplotypes/alleles (**Tables 4.6a, 4.6b, 4.6c**) and the overall number of shared haplotypes/alleles (**Tables 4.7a, 4.7b, 4.7c**). For this analysis, the number of unique haplotypes/alleles and the number of shared haplotypes/alleles are expressed as a percentage of the total number of haplotypes/alleles detected in the overall Old World sample.

Table 4.9: A comparison of the total number of unique haplotypes/alleles, shared haplotypes/alleles and F_{ST} values for the total Old World data set analysed in this study

Genetic system	Overall number of unique haplotypes/alleles expressed as a % of the total number of detected haplotypes/alleles	Overall number of shared haplotypes/alleles expressed as a % of the total number of detected haplotypes/alleles	Overall F_{ST} values
MtDNA	87.5	0	0.689 ($P \leq 0.001$)
Nuclear microsatellite	18.2	36.9	0.082 ($P \leq 0.001$)
Nuclear SNPs	4.5	68.2	0.153 ($P \leq 0.001$)

The data presented in this table combines the data presented in **Tables 4.6a-c, Tables 4.7a-c** and **Table 4.8**.

A comparison of inter-population genetic differentiation estimates for a number of large bodied species with large geographic ranges

A comparison between cattle F_{ST} values and F_{ST}-based derivatives, such as its multi-allelic analogue, G_{ST} (Nei 1973), the DNA sequence statistic, K_{ST} (Hudson *et al.* 1992) and the _ parameter (Weir & Cockerham 1984), generated for large-bodied mammalian species with large geographical ranges is presented in **Figure 4.4**. Notably, domestic cattle F_{ST} estimates derived from nuclear data are among the lowest observed while inter-regional genetic differentiation estimates based on mtDNA data are high.

Figure 4.4: Inter-population genetic differentiation for a number of large bodied species with large geographic ranges

Figure 4.4. The proportion of intra-specific genetic variance attributable to population subdivision for various large-bodied mammalian species with extensive geographical range. Information regarding the species, the genetic differentiation estimate employed and the units of population subdivision are summarised below:

(1) **Species:** domesticated cattle. **Genetic differentiation estimate:** F_{ST} values. **Units of subdivision:** Europe/Near East/Africa/India. **Reference:** data presented in this chapter.

(2) **Species:** domesticated goats. **Genetic differentiation estimate:** F_{ST} values. **Units of subdivision:** Europe/Africa/Asia. **Reference:** Luikart *et al.* (2001).

(3) **Species:** African buffalo. **Genetic differentiation estimate:** K_{ST} values. **Units of subdivision:** Each individual population assayed from Uganda, Kenya, Tanzania, Zimbabwe, Botswana and Namibia. **Reference:** Simonsen *et al.* (1998).

(4) **Species:** Humans. **Genetic differentiation estimate:** G_{ST} values. **Units of subdivision:** Europe/Africa/Asia. **Reference:** Jorde *et al.* (2001).

(5) **Species:** Impala. **Genetic differentiation estimate:** K_{ST} values. **Units of subdivision:** East Africa/Southwest Africa. **Reference:** Nersting & Arctander (1999).

(6) **Species:** Kudu. **Genetic differentiation estimate:** K_{ST} values. **Units of subdivision:** East Africa/Southwest Africa. **Reference:** Nersting & Arctander (1999).

(7) **Species:** Savannah elephant. **Genetic differentiation estimate:** F_{ST} values. **Units of subdivision:** Each individual population assayed from East Africa, West Africa and South Africa. **Reference:** Nyakaana *et al.* (2002).

(8) **Species:** Wildebeeste. **Genetic differentiation estimate:** K_{ST} estimates. **Units of subdivision:** South Africa/East Africa. **Reference:** Arctander *et al.* (1999).

(9) **Species:** Hartebeeste. **Genetic differentiation estimate:** K_{ST} estimates. **Units of subdivision:** East and South African regional populations **Reference:** Arctander *et al.* (1999).

(10) **Species:** Topi. **Genetic differentiation estimate:** K_{ST} estimates. **Units of subdivision:** East and South African regional populations. **Reference:** Arctander *et al.* (1999).

Analysis of molecular variation (AMOVA)

The hierarchical components of genetic variation (*i.e.* the proportion of variance attributable to differences among regional groups, differences between populations within regional groups and differences between individuals within populations) were estimated using an analysis of molecular variation (AMOVA) approach (Excoffier *et al.* 1992; **Table 4.10**).

Congruent with F_{ST} estimates discussed above, hierarchical AMOVA analysis reveals that only a small proportion of the total Old World nuclear SNP (11.37%) and microsatellite (6.68%) variation is accounted for by differences between regional groups. As with the F_{ST} estimates, the majority of the Old World mtDNA diversity (68.44%) is explained by inter-regional genetic differentiation. A substantial proportion of the total mtDNA variance (29.19%) is also accounted for by genetic difference between individuals within populations.

Table 4.10: Hierarchical AMOVA analysis showing the percentage of variation at each of the three levels of population hierarchy

	Percentage Variation (%)		
Population hierarchy	**SNP**	**Microsatellite**	**mtDNA**
Within populations	77.20 ($P \leq 0.001$)	86.05 ($P \leq 0.001$)	29.19 ($P \leq 0.001$)
Between populations within regional groups	11.42 ($P \leq 0.001$)	7.27 ($P \leq 0.001$)	2.37 ($P \leq 0.001$)
Between regional groups	11.37 ($P \leq 0.002$)	6.68 ($P \leq 0.001$)	68.44 ($P \leq 0.001$)

4.3.4 Admixture analysis

SNP data were used to estimate the relative proportions of zebu admixture within the populations of the Near East and Africa using the *m*R admixture coefficient (Bertorelle & Excoffier 1998). Zebu admixture proportions of 25.1% and 22.4% were estimated for the South Anatolian Red and Anatolian Black populations, respectively. These results corroborate those previously estimated for microsatellite and mtDNA data, supporting extensive zebu introgression in the populations from the Near East (Loftus *et al.* 1999; Chapter 2 of this thesis; **Figure 4.5**). This method also suggests that the Hungarian Grey population has also experienced appreciable levels of zebu influence (14.9%), which exceeds the zebu admixture proportions of 6.5% estimated from microsatellite data (Loftus *et al.* 1999) [Note: the zebu admixture proportions reported by Loftus *et al.* (1999) were

based on 20 microsatellite loci, using 7 European taurine breeds and 2 zebu breeds of Indian origin as surrogate *Bos taurus* and *Bos indicus* populations].

SNP-based zebu admixture results indicate that the N'Dama population of Guinea has not been subjected to zebu introgressive admixture (zebu admixture proportion of 0.0%). This finding agrees with the data of Hanotte *et al.* (2002), which estimate zebu admixture for this population as 3.0% (this figure was generated using the ADMIX program, based on data from 15 microsatellite loci and using 4 European taurine and 3 zebu populations of Indian origin as parental *Bos taurus* and *Bos indicus* populations), but contradicts the results of Loftus *et al.* (1999) which estimate zebu admixture for this population as 17.9% (this figure was generated using the ADMIX program, based on data from 20 microsatellite loci and using 7 European taurine and 2 zebu populations of Indian origin as parental *Bos taurus* and *Bos indicus* populations).

The SNP data are also suggestive of *Bos taurus* introgression in the populations of the Northwest of the Indian subcontinent. For SNP data, taurine admixture proportions were estimated as 11.1% in the North Indian zebu population. This result correlates with previous taurine admixture estimates in Northern Indian cattle based on mtDNA sequence data (6.2% - this estimate is based on the proportion of *Bos taurus* mtDNA haplotypes in a composite of 4 zebu populations from the Northwest of the Indian subcontinent, namely Hariana, Sahiwal, Red Sindhi and Tharparker as reported in **Table 4.2** of this thesis) and nuclear microsatellite data (6.2% - an estimate generated by Kumar *et al.* (in press) based on the average frequencies of statistically selected *Bos taurus* ancestry informative markers in Hariana, Red Sindhi, Sahiwal and Tharparker populations; R. Freeman, personal communication).

Figure 4.5: *Bos taurus* and *Bos indicus* admixture proportions for all 3 genetic systems

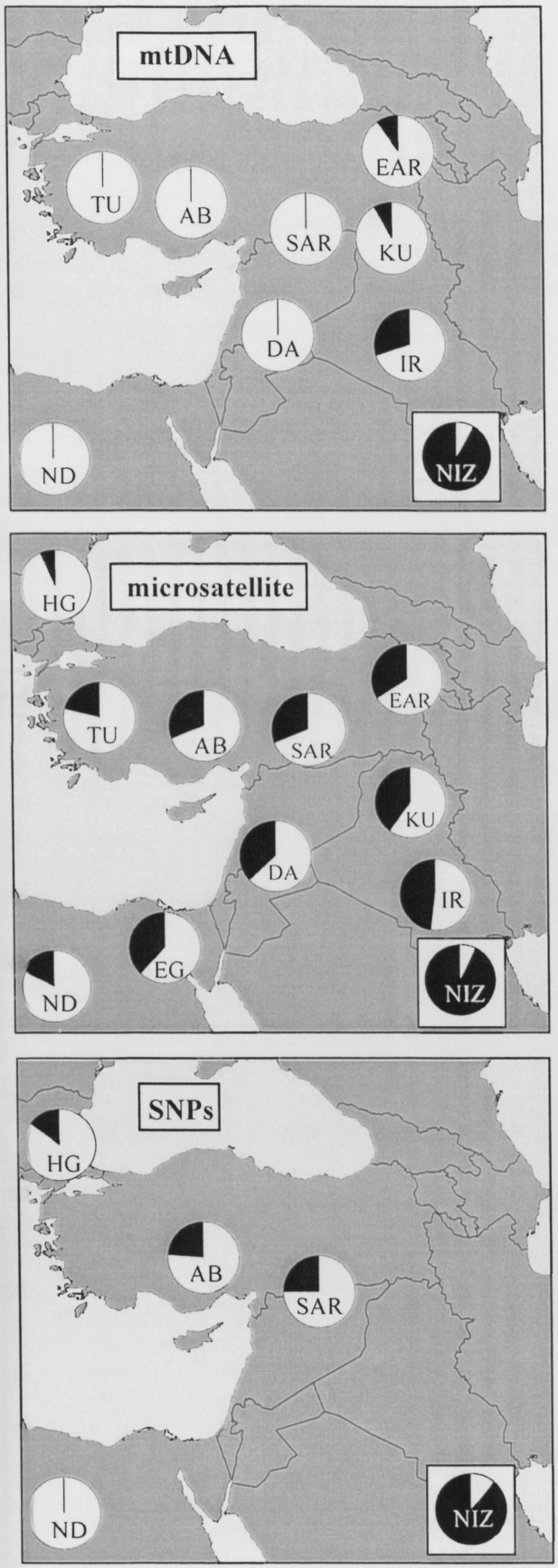


Figure 4.5. Zebu admixture proportions (shaded black) within morphologically taurine populations of the Near East. MtDNA estimates were based on the proportion of zebu introgressor sequencs within each population; nuclear microsatellite estimates were taken from Loftus *et al.* (1999) and Kumar *et al.* (in press).

Population codes:

AB (Anatolian Black), DA (Damascus), EAR (East Anatolian Red), EG (Egyptian), HG (Hungarian Grey), IR (Iraqi - for mtDNA data this population comprises the North, Middle and South Iraqi populations. KU (Kurdi), ND (Guinean N'Dama), SAR (South Anatolian Red), TU (Turkish Grey). Sample sizes and reference information for these populations are given in **Tables 4.1**, **4.2** and **4.3**.

NIZ (North Indian zebu) is a composite of Hariana, Sahiwal and Tharparker samples for SNP data. For both mtDNA and micorsatellite data this population is a composite of Hariana, Red Sindhi, Sahiwal and Tharparker samples.

4.3.5 Multi-dimensional scaling of inter-population genetic distances

Genetic distances between cattle populations were calculated for each genetic system (**Section 4.2.4**) and are represented in two dimensions following multi-dimensional scaling (MDS) (**Figure 4.6**). The primary feature of the MDS plot constructed for each genetic system is the marked divergence between *Bos taurus* and *Bos indicus* populations, which is clearly defined along dimension 1. Notably for each plot, some Near Eastern populations occupy intermediate positions between the European *Bos taurus* and Indian *Bos indicus* groups, which occur at the extremities. This undoubtedly reflects the partial zebu ancestry of these populations as indicated by admixture analysis.

Dimension 2 distinguishes African populations from all other regional groups. All African populations, regardless of morphology, share similar *Bos taurus* haplotypes, which are differentiated from those occurring in the Near East and Europe (Troy *et al.* 2001). Consequently, for mtDNA African populations constitute a discrete group, which is separated from the European *Bos taurus* cluster. Analysis of SNP data indicates that the Guinean N'Dama population of West Africa is differentiated from all other population groups. Similarly, for microsatellite data, the Guinean N'Dama population is separated from other regional populations, occurring at the periphery of the plot. Also, the microsatellite-derived plot reveals that the Guinean N'Dama population is clearly separated from the Egyptian population, which displays increased affinity with populations from the Near East. The positioning of the Egyptian population presumably reflects appreciable gene flow with geographically proximal populations of the Near East.

Figure 4.6: Two-dimensional MDS plots constructed for three genetic systems

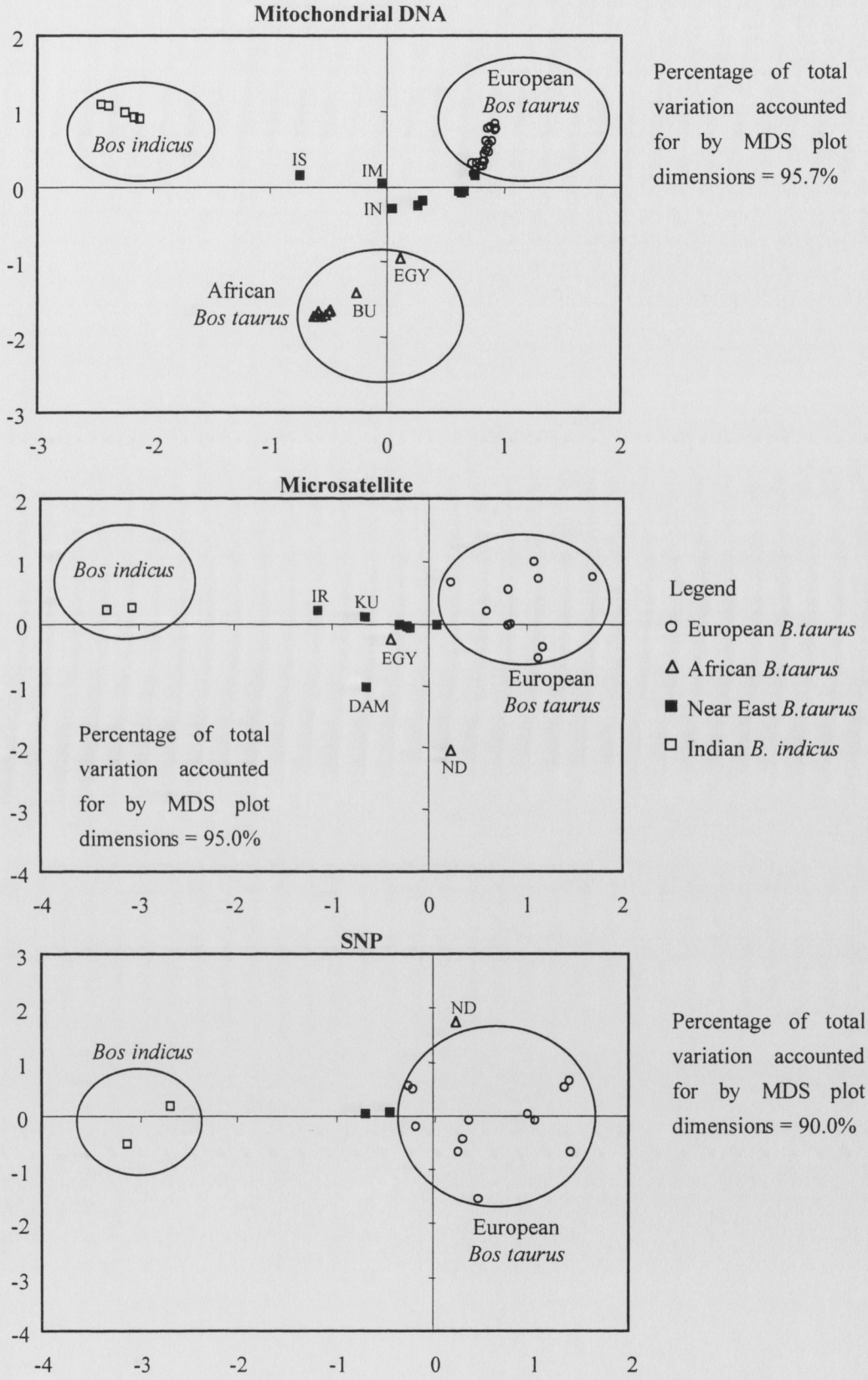


Figure 4.6. Two-dimensional MDS plots constructed for data obtained for three genetic systems, mtDNA, nuclear microsatellite allelic data and nuclear SNP allelic data. Dimension 1 and dimension 2 of each plot are represented by the horizontal and vertical axes, respectively. Symbols denote geographic regional groups. The divergence between *Bos taurus* and *Bos indicus* groups are evident for each individual plot.

4.4 DISCUSSION AND CONCLUSIONS

4.4.1 Intra-regional genetic diversity

Previous phylogenetic surveys have shown that *Bos taurus* samples from the Near East display elevated levels of mtDNA- and microsatellite-based genetic diversity, with corresponding reductions in African and European domestic populations. Furthermore, the levels of mtDNA diversity within the *Bos indicus* populations of the Indian subcontinent have also been shown to be high and directly comparable to those from the Near East. It has been argued that these patterns of variability are indicative of the increased retention of ancestral allelic and haplotypic variation associated with the initial domestications of *Bos taurus* and *Bos indicus* (Loftus *et al.* 1999; Troy *et al.* 2001; Chapter 2 of this thesis).

Contrasting patterns of diversity are observed for the nuclear SNP data presented here. Estimates of SNP-based diversity are highest for the European regional group, with lower levels detected in the Near East, Africa and India. Similarly, individual European cattle populations display increased levels of genetic diversity when compared to individual populations from the remaining three geographical regions. Elevated SNP-diversity levels within European samples is most probably explained by an ascertainment bias by which the majority of assayed SNP loci were initially identified as being polymorphic in European *Bos taurus* samples (Dr. Gregor Durstewitz, personal communication). This bias is particularly evident for the SNP allele frequency plots (**Figure 4.3**) in which bi-allelic variability is observed at every locus within the European group, whereas allele fixation is observed at 1 locus in the Near East group, 5 loci in the African group and 4 loci in the Indian group. Thus, SNP-based average heterozygosity levels are inflated in European cattle populations relative to other regional groups. As microsatellite markers exhibit higher levels of polymorphism when compared to SNP markers they are less subject to this ascertainment bias. This bias is not applicable to mtDNA polymorphisms as they are

acquired from DNA sequences ascertained uniformly in all cattle populations (Jorde *et al.* 1995; Jorde & Rogers 1996).

4.4.2 Population admixture analysis

Bos indicus introgression in Africa

Previous phylogenetic surveys have demonstrated that the original taurine genetic substrate of African cattle has been greatly modified by pre-dominantly male-mediated introgression of zebu into the continent over several millennia (Epstein & Mason 1984; Marshall 1989; Bradley *et al.* 1994; MacHugh *et al.* 1997; Hanotte *et al.* 2000; Hanotte *et al.* 2002). The N'Dama population of Guinea are widely regarded as the descendants of the first African taurine domesticates and previous data have shown them to be one of the West African populations that has been least affected by zebu influxes (MacHugh *et al.*1997; Hanotte *et al.* 2000). SNP-based estimates of zebu admixture for this population were estimated here as 0.0%. This figure is in good agreement with the recent findings Hanotte *et al.* (2002), which estimate zebu admixture proportions for this population as 3.0%. However, this figure opposes the microsatellite-derived admixture estimate of Loftus *et al.* (1999), which suggest that the Guinean N'Dama population has experienced an appreciable level of zebu admixture (17.9%).

The apparent discrepancy in microsatellite-based zebu admixture proportions for the Guinean N'Dama population is surprising. This may reflect the differences in the European taurine and Asian zebu breeds employed as parental populations for zebu admixture analysis in both studies. For example, Hanotte *et al.* (2002) employed three Asian zebu populations, namely Sahiwal, Ongole and Nelore as parental populations, whereas Loftus *et al.* (1999) used data from the Nelore and Ongole only. The data presented in Chapter 3 of this thesis show that although the Brazilian Nelore breed represents the descendants of

Ongole cattle brought to the Americas during the 19th century, they possess *Bos taurus* mtDNA haplotypes, presumably as a result of admixture with local Creole populations. It is therefore possible that such admixture events may have resulted in the introduction of autosomal taurine alleles into the Nelore genetic pool. Thus, the Nelore may not represent a pure zebu breed, even though it is treated as such by the ADMIX software, confounding zebu admixture estimates of Loftus *et al.* (1999). Alternatively, as the Asian zebu parental population used by Hanotte *et al.* (2002) also includes data from Sahiwal samples, this surrogate may be more appropriate than that employed Loftus *et al.* (1999). Indeed, Loftus *et al.* (1999) have stated that their 17.9% zebu admixture proportions estimated for the Guinean N'Dama population may be not be a true representation of zebu admixture and that their estimates based on the analysis of zebu-specific alleles (0.2%) are probably more reliable.

Bos indicus introgression in the Near East

In contrast to Africa, genetic imprints of zebu introgression have been detected within both the nuclear and mitochondrial genomes of Near Eastern *Bos taurus* populations (Loftus *et al.* 1999; Chapter 2 of this thesis). The patterns of *Bos indicus* admixture across the Near East as defined by nuclear SNP data show resemblance to those obtained for nuclear microsatellites. Loftus *et al.* (1999) estimated the nuclear microsatellite-based proportion of zebu ancestry for South Anatolian Red and Anatolian Black as 30.6% and 30.5%, respectively. Admixture proportions are estimated for the South Anatolian Red (25.1%) and Anatolian Black (22.3%) based the nuclear SNP presented here. Furthermore, these SNP-based estimates are consistent with an East-West distribution of zebu introgression across the Near East, as demonstrated for mtDNA and microsatellite data (Loftus *et al.* 1999; Chapter 2 of this thesis). The extent of zebu admixture within Near Eastern populations implies that *Bos indicus* introgression into the Near East has been continuous

since the emergence of the first domestic zebu herds, *circa* 7th millennium BC (Meadow 1993; Meadow 1996). Archaeological evidence points towards the presence of zebu cattle in Mesopotamia and Jordan some 3,000 YBP, providing opportunities for genetic exchange between the two domestic taxa (Zeuner 1963; Clason 1978; Naik 1978; Chapter 2 of this thesis).

Loftus *et al.* (1999) have demonstrated that the Hungarian Grey population of Southeast Europe have experienced appreciable levels of zebu introgression (admixture proportions of 6.5%). SNP-based admixture results suggest a substantially greater zebu genetic influence within the Hungarian Grey samples (14.9%). The partial zebu ancestry of Hungarian Grey cattle as suggested by genetic evidence may be representative of limited westward zebu introgression into Southern Europe. Bökönyi (1997) states that humped cattle may have been introduced to Europe by Roman military units stationed in the Near East and Africa, during the emergence of the Roman Empire. Indeed, archaeological data, in the form of artistic impressions, suggest that humped cattle may have been present in Plovdiv (Bulgaria) during the Roman colonisation of Europe (Bökönyi 1997).

Bos taurus introgression in the Indian subcontinent

Using microsatellite data, Kumar *et al.* (in press) have demonstrated *Bos taurus* introgression in the morphologically zebu populations of South Asia. With some symmetry to the patterns of zebu gene flow across the Near East, *Bos taurus* introgression appears to be stronger in North Indian populations compared to those of South Indian origin. MtDNA investigations have also demonstrated the existence of taurine haplotypes within more northerly zebu populations (Chapter 2 of this thesis). In agreement, nuclear SNP data suggest that northern *Bos indicus* populations have experienced appreciable levels of taurine introgression (11.6%), with the southern Ongole population employed as a

reference population. With some symmetry to the pattern of zebu gene flow across the Near East, the detection of taurine genetic influence in North India may be explained by historic introductions of *Bos taurus* animals into the Indian subcontinent. This genetic evidence is supported by the archaeological record, which has revealed the existence of humpless taurine cattle in the Indus Valley civilisations of Harappa and Mohenjo-daro, *circa* 3rd millennium BC (Zeuner 1963; Allchin & Allchin 1968).

It is important to also state that the high mutation rate, allelic size constraints and consequent allelic homoplasy of microsatellite loci implies that a considerable proportion of alleles are common to populations from different regional groups but which are not identical by descent. Alleles which display identity in state rather than descent may distort estimates of admixture proportions in hybridised populations. In contrast, the low mutation rate of SNPs suggests that they are not subject to allele homoplasy and therefore SNPs may provide more reliable estimates of admixture proportions than microsatellite data. However, it must also be emphasised that the zebu admixture proportions generated from the SNP data presented in this study are based only on 11 loci and that more definitive results may be obtained from a larger data set.

4.4.3 Inter-population genetic distances

Calculation of genetic distances using SNP data underlines the substantial genetic divergence between *Bos taurus* and *Bos indicus* populations, as described in previous studies of bovine biochemical and molecular variation (Naik 1978; Manwell & Baker 1980; Loftus *et al.* 1994a; Loftus *et al.* 1994b; MacHugh *et al.* 1997). Other notable features shared between the SNP plot and those constructed for mtDNA and microsatellite data are the intermediate positioning of the Near Eastern populations between the *Bos taurus* and *Bos indicus* groups and the differentiation of the African Guinean N'Dama

population from all other population groups. The former reflects the partial zebu ancestry of the Near Eastern populations as indicated by admixture analysis, while the latter is representative of the genetic differentiation of African *Bos taurus* from European *Bos taurus*, as suggested by previous genetic investigations (Bradley *et al.* 1996; MacHugh *et al.* 1997; Troy *et al.* 2001). The genetic distinction of African taurine populations from their European and Near Eastern counterparts supports archaeological assertions of an independent domestication of African *Bos taurus* from an aurochs strain that was distinct to the progenitors of Near Eastern, European and Indian cattle (Wendorf & Schlid 1994).

4.4.4 The apportionment of bovine genetic diversity

SNP-based F_{ST} estimates generated for the total Old World sample show that 15.3% of the total variance is attributable to differences between regional groups. This apportionment of Old World SNP variability is comparable to the F_{ST} estimates derived for nuclear microsatellite data (8.2%). These data indicate that the majority of cattle nuclear variability is due to genetic differences between individuals within regional groups rather that between regions. Furthermore, the Old World F_{ST} values for microsatellite systems generated here tend to be slightly lower than those based on SNPs. This observation most probably reflects the higher mutation rate and consequent allelic homoplasy of microsatellites, which increases within-region variation relative to between-region variation and thus decreases microsatellite-derived F_{ST} values (Jorde *et al.* 2000). In contrast to both nuclear genetic data sets, mtDNA-based F_{ST} estimates indicate the majority of Old World mtDNA variability (68.9%) is attributable to inter-regional differences. These findings are further attested to by the results obtained for the hierarchical AMOVA analysis (**Table 4.10**).

The apparent discordance between nuclear- and mtDNA-derived inter-regional bovine F_{ST} estimates can be explained by a number of contributory factors. The purported separate domestic origins of African *Bos taurus*, Near Eastern and derivative European *Bos taurus* and South Asian *Bos indicus* cattle from distinct aurochs populations has resulted in strong phylogeographic structuring of domestic cattle mtDNA haplotypes throughout the Old World (Loftus *et al.* 1994a; Bradley *et al.* 1996; Troy *et al.* 2001). Moreover, domestication would have also induced a severe bottleneck in genetic diversity, as it entails the recruitment of a subset of wild variation into the domestic pool. This bottleneck is more pronounced for mtDNA due to its haploid, maternal inheritance which decreases its effective population size (N_e) to one-quarter that of nuclear DNA. In turn, this leads to faster differentiation between regions due to genetic drift (Wilson 1985) and consequently, for mtDNA, cattle populations display low within-regional F_{ST} estimates and high between-region F_{ST} estimates. This scenario contrasts to domestic goat populations, where inter-regional genetic differentiation is extremely low, despite the fact that extant goat mtDNA lineages appear to have been derived from three divergent matrilines as result of multiple domestications (Luikart *et al.* 2001). The low phylogeographic structuring of domestic goat mtDNA lineages relative to domestic cattle may indicate that domestic goats have, over time, been a highly mobile species, presumably as a result of human trade (Luikart *et al.* 2001; MacHugh & Bradley 2001).

The low inter-regional genetic differentiation noted for bovine nuclear DNA systems can be explained by a number of contributing factors. Firstly, under managed cattle breeding programmes most genetic exchange between domestic populations is predominantly male-mediated (MacHugh & Bradley 2001). Thus, introgressing males have major effects on the nuclear genome of native breeds while having no detectable influence on indigenous mtDNA lineages. For example, comparative analyses of all three genetic systems presented

in this study reveal: (1) appreciable Near Eastern taurine and South Asian zebu influence on the nuclear genome of Egyptian taurine cattle; (2) a partial zebu ancestry within the nuclear genome of Hungarian Grey cattle and (3) substantial and continuous zebu influence on the nuclear genome of taurine populations dispersed throughout the Near East with zebu mtDNA influence restricted to the populations on the easternmost margins (**Figure 4.5**). Pre-dominantly male-mediated gene flow between regions has an overall effect of homogenising the nuclear genome of populations, thus reducing nuclear-derived inter-regional genetic differentiation estimates.

Secondly, differential evolutionary dynamics and allelic distribution of microsatellite and SNP loci may also contribute to low inter-regional estimates of genetic differentiation. Microsatellite allelic homoplasy resulting in an overlap in allelic size variance between regions has an overall effect of reducing inter-regional F_{ST} values and can distort true estimates of inter-regional genetic differentiation (Ritz *et al.* 2000). In contrast, the pronouncedly lower mutation rate and consequent identity by descent of SNPs ensures that the effects of allele homoplasy are largely negligible and therefore, SNP data may provide more dependable estimates of nuclear genetic differentiation due to population subdivision than microsatellites. A total of 68.2% of all detected SNP alleles (15 alleles out of 22) are common to all four regional populations. When the mutational stability of SNPs are considered, these data suggest that the 16 shared SNP alleles represent ancient mutational events that may predate the evolutionary divergence between *Bos taurus* and *Bos indicus*, resulting in low Old World SNP-derived F_{ST} estimates. Furthermore, these low F_{ST} estimates are also undoubtedly attributable to the fact that there are only two alleles per SNP locus.

Discordance between nuclear and mitochondrial based genetic structures has also been observed in a number of phylogeographic studies involving large bodied mammals with extensive geographic ranges such as the African savannah elephant (Nyakaana *et al.* 2002) and humpback whales (Palumbi & Baker 1994). As with domestic cattle populations, the discrepancies noted for these species are believed to involve increased differentiation of mtDNA due to accelerated random genetic drift and a male-biased geneflow, resulting in the homogenisation of the nuclear genome.

In contrast, estimates of genetic differentiation between African buffalo populations based on mtDNA and nuclear microsatellite allelic variation are congruent. It has been proposed that low genetic differentiation observed in the African cape buffalo is due to two possible factors: (1) fragmentation of a large panmictic group with homogenising gene flow between localities, into geographically isolated subpopulations with minimal or no gene flow between regions. Thus, the low genetic differentiation observed in African buffalo populations is due to a combination of founder effects and genetic drift since the development of these isolated subpopulations; or (2) a high female migration rate between subpopulations, which has not been so high that it has resulted in complete genetic mixing (Simonsen *et al.* 1998; van Hooft *et al.* 2002).

Similarly, analysis of human genetic data show that the majority of mitochondrial and nuclear variability exists within, rather than between, major regional populations (Jorde *et al.* 1995; Barbujani *et al.* 1997; Ziętkiewicz *et al.* 1997; Jorde *et al.* 1998; Jorde *et al.* 2000; Romualdi *et al.* 2002). Two possible explanations have been posited to account for this low inter-regional genetic differentiation. Firstly, according to the replacement hypothesis of human evolution, anatomically modern humans arose in Africa as descendants of African *Homo erectus* ancestors some 200,000 YBP and subsequently

colonised the rest of the Old World, replacing previously established archaic human forms (Kaessmann & Pääbo 2002). Therefore, the major regional Old World human populations separated relatively recently with little evolutionary time for sufficient inter-regional differentiation to occur (Harpending & Rogers 2000). Alternatively, the multi-regional hypothesis holds that modern *Homo sapiens* evolved from more than one archaic human form over the course of the last 1 million years in different parts of the Old World. Thus, the relative homogeneity observed within the human species was maintained by a combination of natural selection and substantial gene flow between the major regional groups (Jorde *et al.* 1998).

Interestingly, mtDNA-based phylogeographic analysis of several widely dispersed African herbivore species, such as the impala, greater kudu, hartebeeste, topi and wildebeeste, all exhibit considerable genetic differentiation between populations. It has been argued that prolonged periods of climatic instability throughout the Pleistocene, resulted in the fragmentation of these wide-ranging populations into geographically disparate refugia, which subsequently evolved in allopatry over long time periods. This allowed genetic divergence between isolated subpopulations to accumulate, thereby accounting for the differentiation observed between modern populations (Arctander *et al.* 1999; Nersting & Arctander 2001; Nyakaana *et al.* 2002).

CHAPTER 5

LABORATORY METHODS

5.1 THE SOURCES OF THE MOLECULAR GENETIC DATA USED FOR PHYLOGENETIC ANALYSIS IN THIS THESIS

5.1.1 The mitochondrial DNA sequence data used in this thesis

Table 5.1 details the mtDNA sequence data used for phylogenetic analysis in Chapters 2, 3 and 4 of this thesis. Information regarding the number of samples per population, the number of *Bos taurus* and *Bos indicus* mtDNA sequences detected within each population, the geographic origin of the population and the source of the mtDNA sequence data are listed.

5.1.2 The microsatellite genotypic data used in this thesis

Table 5.2 and **Table 5.3** detail the nuclear microsatellite data used for phylogenetic analysis in Chapter 3 and Chapter 4 of this thesis, respectively. Information regarding the number of samples per population, the geographic origin of the population and the source of the data are listed.

5.1.3 The SNP genotypic data used in this thesis

Table 5.4 details the nuclear microsatellite data used for phylogenetic analysis in Chapter 3 and Chapter 4 of this thesis, respectively. Information regarding the number of samples per population, the geographic origin of the population and the source of the data are listed.

Table 5.1: Mitochondrial DNA data used for phylogenetic analysis in Chapters 2, 3 and 4 of this thesis

European mtDNA sequence data

Population name (Population code)	Number of samples	Number *B. taurus* sequences	Number *B. indicus* sequences	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Icelandic (IC)	12	12	0	Iceland / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Norwegian red (NO)	12	12	0	Norway / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Telemark (TE)	5	5	0	Norway / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Westland Fjord (WF)	12	12	0	Norway / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Kerry (KE)	10	10	0	Ireland / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Aberdeen Angus (AA)	12	12	0	Scotland / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Galloway (GA)	18	18	0	Scotland / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Highland (HI)	12	12	0	Scotland / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Hereford (HE)	12	12	0	England / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Jersey (JE)	11	11	0	England / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Berrenda (BE)	11	11	0	Spain / *B. taurus*	Troy *et al.* (2001)	2, 3, 4
Charolais (CH)	12	12	0	France / *B. taurus*	Troy *et al.* (2001)	2, 3, 4

Table 5.1 (continued): European mtDNA sequence data

Population name (Population code)	*n*	Number *B. taurus* sequences	Number *B. indicus* sequences	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Friesian (FR)	12	12	0	France / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
German Black (GE)	10	10	0	German / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
Limousin (LI)	12	12	0	France / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
Romagnola (RO)	11	11	0	Italy / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
Simmental (SI)	11	11	0	Switzerland / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
Sykia (SY)	12	12	0	Greece / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
European total	**207**	**207**	0			

Table 5.1 (continued): Iberian mtDNA sequence data

Population name (Population code)	*n*	Number *B. taurus* sequences	Number *B. indicus* sequences	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Aroquesa (AR)	12	12	0	Portugal / *B. taurus*	Cymbron *et al.* (1999)	**3**
Alentejana (AL)	12	12	0	Portugal / *B. taurus*	Cymbron *et al.* (1999)	**3**
Barrosa (BA)	5	5	0	Portugal / *B. taurus*	Cymbron *et al.* (1999)	**3**
Maronesa (MA)	3	3	0	Portugal / *B. taurus*	Cymbron *et al.* (1999)	**3**
Mertolenga (ME)	12	12	0	Portugal / *B. taurus*	Cymbron *et al.* (1999)	**3**
Preta (PR)	5	5	0	Portugal / *B. taurus*	Cymbron *et al.* (1999)	**3**
Iberian total	**49**	**49**	**0**			

Note: for phyloghenetic analysis in Chapter 3, the Berranda mtDNA sequence data (Spain) was included with the total Iberian mtDNA pool.

Table 5.1 (continued): Japanese mtDNA sequence data

Population name (Population code)	*n*	Number *B. taurus* sequences	Number *B. indicus* sequences	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Japanese Black (JA)	32	32	0	Japan / *B. taurus*	Mannen *et al.* (1998)	**2**

Table 5.1 (continued): African mtDNA sequence data

Population name (Population code)	*n*	Number *B. taurus* sequences	Number *B. indicus* sequences	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Butana (BU)	12	12	0	Sudan / *B. indicus*	Troy *et al.* (2001)	**2, 3, 4**
Egyptian (EG)	11	11	0	Egypt / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
Guinean N'Dama (ND)	11	12	0	Guinea / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
Kapsiki (KA)	12	12	0	Cameroon / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
Kenana (KN)	7	7	0	Sudan / *B. indicus*	Troy *et al.* (2001)	**2, 3, 4**
Kuri (KU)	11	11	0	Lake Chad / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
Namchi (NA)	12	12	0	Cameroon / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
Somba (SO)	12	12	0	Benin / *B. taurus*	Troy *et al.* (2001)	**2, 3, 4**
White Fulani (FU)	7	7	0	Nigeria / *B. indicus*	Troy *et al.* (2001)	**2, 3, 4**
African total	**95**	**95**	**0**			

Table 5.1 (continued): Near Eastern mtDNA sequence data

Population name (Population code)	*n*	Number *B. taurus* sequences	Number *B. indicus* sequences	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Turkish Grey (TU)	11	11	0	Turkey / *B. taurus*	Troy *et al.* (2001)	**2, 4**
Anatolian Black (AB)	12	12	0	Turkey / *B. taurus*	Troy *et al.* (2001)	**2, 4**
East Anatolian Red (EAR)	10	9	1	Turkey / *B. taurus*	Troy *et al.* (2001)	**2, 4**
South Anatolian Red (SAR)	11	11	0	Turkey / *B. taurus*	Troy *et al.* (2001)	**2, 4**
Damascus (DA)	12	12	0	Syria / *B. taurus*	Troy *et al.* (2001)	**2, 4**
Kurdish (KUR)	12	11	1	North Iraq / *B. taurus*	Troy *et al.* (2001)	**2, 4**
North Iraq (NIR)	12	10	2	North Iraq / *B. taurus*	Troy *et al.* (2001); Chapter 2 of this thesis	**2, 4**
Troy *et al.* (2001) published 10 North Iraqi samples, 8 of which displayed *B. taurus* mtDNA sequences and 2 of which were *B. indicus* mtDNA types. 2 additional North Iraqi samples were sequenced as part of the analysis in Chapter 2 and both displayed *B. taurus* mtDNA sequences.						
Middle Iraq (MIR)	11	8	3	Central Iraq / *B. taurus*	Chapter 2 of this thesis	**2, 4**
South Iraq (SIR)	16	9	7	South Iraq / *B. taurus*	Troy *et al.* (2001); Chapter 2 of this thesis	**2, 4**
Troy *et al.* (2001) published 11 South Iraqi samples, 6 of which displayed *B. taurus* mtDNA sequences and 5 of which were *B. indicus* mtDNA types. 5 additional South Iraqi samples were sequenced as part of the analysis in Chapter 2 and 3 displayed *B. taurus* mtDNA sequences and 2 displayed *B. indicus* mtDNA sequences.						
Near East total	**107**	**93**	**14**			

Table 5.1 (continued): Asian mtDNA sequence data

Population name (Population code)	*n*	Number *B. taurus* sequences	Number *B. indicus* sequences	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Red Sindhi (RS)	22	2	20	Pakistan / *B. indicus*	Chapter 2 of this thesis	**2, 4**
Tharparker (TH)	21	1	20	India / *B. indicus*	Troy (1998); Chapter 2 of this thesis	**2, 4**
15 Tharparker samples were sequenced as part of the analysis in Chapter 2. Of these, 1 sample displayed a *B. taurus* mtDNA sequence, while the remaining mtDNA sequences were of *B. indicus* mtDNA types. The remaining 6 Tharparker mtDNA sequences (all *B. indicus*) were taken from Troy (1998).						
Sahiwal (SA)	12	0	12	India / *B. indicus*	Troy (1998); Chapter 2 of this thesis	**2, 4**
5 Sahiwal samples were sequenced as part of the analysis in Chapter 2 and all samples displayed *B. indicus* mtDNA sequences. The remaing 7 Sahiwal mtDNA sequences (all *B. indicus*) were taken from Troy (1998).						
Hariana (HA)	10	1	9	India / *B. indicus*	Troy (1998); Chapter 2 of this thesis	**2, 4**
3 Hariana samples were sequenced as part of the analysis in Chapter 2. Of these 1 sample displayed a *B. taurus* mtDNA sequences. The remaining 7 Sahiwal mtDNA sequences (all *B. indicus*) were taken from Troy (1998).						
Desi (DE)	19	0	19	South India / *B. indicus*	Chapter 2 of this thesis	**2, 4**
Ongole (ON)	11	0	11	South India / *B. indicus*	Troy (1998)	**2, 4**
Loatian (LS)	30	0	30	Laos	Prof. H. Mannen (University of Kobe)	**2**
Mongolian (MN)	9	0	9	Mongolia	Prof. H. Mannen (University of Kobe)	**2**
Asian total	**134**	**4**	**139**			

Table 5.1 (continued): South American and Caribbean Creole mtDNA sequence data

Population name (Population code)	*n*	Number *B. taurus* sequences	Number *B. indicus* sequences	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Guadeloupe Creole (GC)	25	25	0	Guadeloupe	Chapter 3 of this thesis	**3**
Antiguan Creole (AC)	21	21	0	Antigua	Chapter 3 of this thesis	**3**
St. Lucian Creole (SC)	13	13	0	St. Lucia	Chapter 3 of this thesis	**3**
Nelore (NE)	19	14	5	Brazil / *B. indicus*	Troy (1998); Chapter 3 of this thesis	**2, 3**
9 Nelore samples were sequenced as part of the analysis in Chapter 2. Of these, 1 sample displayed a *B. indicus* mtDNA sequence, while the remaining 8 displayed *B. taurus* mtDNA sequences. The remaining 10 Nelore sequences (4 *B. indicus* mtDNA sequences; 6 *B. taurus* mtDNA sequences) were taken from Troy (1998).						
Total S. America/Caribbean	**78**	**73**	**5**			

Table 5.2: The microsatellite allelic data used in Chapter 3 of this study

Population name (Population code)	*n*	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Charolais (CH)	36	France / *B. taurus*	MacHugh *et al.* (1997)	3
Friesian (FR)	40	France / *B. taurus*	MacHugh *et al.* (1997)	3
Simmental (SI)	36	Switzerland / *B. taurus*	MacHugh *et al.* (1997)	3
Alentejana (AL)	30	Portugal / *B. taurus*	T. Cymbron (Dept. of Genetics, TCD), pers. comm.	3
Aroquesa (AR)	34	Portugal / *B. taurus*	T. Cymbron (Dept. of Genetics, TCD), pers. comm.	3
Mertolenga (ME)	37	Portugal / *B. taurus*	T. Cymbron (Dept. of Genetics, TCD), pers. comm.	3
N'Dama (ND)	63	Guinea / *B. taurus*	MacHugh *et al.* (1997)	3
N'Dama (NGB)	54	Guinea Bissau / *B. taurus*	MacHugh *et al.* (1997)	3
N'Dama (NG)	58	Gambia / *B. taurus*	MacHugh *et al.* (1997)	3
N'Dama (NM)	44	Mali / *B. taurus*	MacHugh (1996)	3
N'Dama (NN)	19	Nigeria / *B. taurus*	MacHugh *et al.* (1997)	3
N'Dama (NS)	48	Senegal / *B. taurus*	MacHugh *et al.* (1997)	3

Table 5.2 continued

Population name (Population code)	*n*	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Gobra (GO)	59	Senegal / *B. indicus*	MacHugh *et al.* (1997)	3
Maure (MAU)	55	Mauritania / *B. indicus*	MacHugh *et al.* (1997)	3
Hariana (HA)	10	India / *B. indicus*	MacHugh *et al.* (1997)	3
Sahiwal (SA)	13	India / *B. indicus*	MacHugh *et al.* (1997)	3
Tharparker (TH)	10	India / *B. indicus*	MacHugh *et al.* (1997)	3
Guadeloupe Creole (GC)	52	Guadeloupe	Ciaran Meghan (Dept. of Genetics, TCD), pers. comm.	3
Antiguan Creole (AC)	70	Antigua	Chapter 3 of this thesis	3
St. Lucian Creole (SC)	39	St. Lucia	Chapter 3 of this thesis	3

Note: All microsatellite genotypic data used in Chapter 3 of this thesis comprised 3 nuclear loci: *BM2113, HEL1* and *ILSTS001.*

Table 5.3: The microsatellite allelic data used in Chapter 4 of this study

Population name (Population code)	*n*	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
Jersey (JE)	44	UK / *B. taurus*	Loftus *et al.* (1999)	4
Hereford (HE)	34	UK / *B. taurus*	T. Cymbron (Dept. of Genetics, TCD), pers. comm	4
Highland (HI)	37	UK / *B. taurus*	T. Cymbron (Dept. of Genetics, TCD), pers. comm	4
Simmental (SI)	50	Switzerland / *B. taurus*	Schmid *et al.* (1999)	4
Holstein (HO)	50	Holland / *B. taurus*	Schmid *et al.* (1999)	4
Charolais (CH)	36	France / *B. taurus*	Loftus *et al.* (1999)	4
Vosges (VO)	27	France / *B. taurus*	Edwards *et al.* (2000)	4
Pustertaler (PS)	31	Germany-Italy / *B. taurus*	Edwards *et al.* (2000)	4
Pinzgauer (PN)	30	Germany-Austria / *B. taurus*	Edwards *et al.* (2000)	4
Hungarian Grey (HG)	44	Hungary / *B. taurus*	Loftus *et al.* (1999)	4
Turkish Grey (TU)	43	Turkey / *B. taurus*	Loftus *et al.* (1999)	4
Anatolian Black (AB)	41	Turkey / *B. taurus*	Loftus *et al.* (1999)	4

Table 5.3: continued

Population name (Population code)	*n*	Population origin / Animal morphology	Source of data	Chapter in which data is used for analysis
South Anatolian Red (SAR)	43	Turkey / *B. taurus*	Loftus *et al.* (1999)	4
East Anatolian Red (EAR)	44	Turkey / *B. taurus*	Loftus *et al.* (1999)	4
Damascus (DA)	41	Syria / *B. taurus*	Loftus *et al.* (1999)	4
Iraqi (IR)	42	Iraq / *B. taurus*	Loftus *et al.* (1999)	4
Kurdish (KU)	36	Iraq / *B. taurus*	Loftus *et al.* (1999)	4
Guinean N'Dama (ND)	44	Guinea / *B. taurus*	Loftus *et al.* (1999)	4
Egyptian (EG)	32	Egyptian / *B. taurus*	Loftus *et al.* (1999)	4
Ongole (ON)	30	India / *B. indicus*	Loftus *et al.* (1999)	4
Nelore (NE)	27	India / *B. indicus*	Loftus *et al.* (1999)	4

The microsatellite allelic data was based on 18 nuclear loci: *BM1818, BM1824, BM2113, CSRM60, CSSM66, ETH10, ETH225, Hel1, Hel5, Hel9, Hel13, ILSTS005, ILSTS006, INRA005, INRA023, INRA032, INRA035, INRA063.*

Table 5.4: The SNP data used in Chapter 4 of this thesis

Population name (Population code)	*n*	Population origin / Animal morphology	Source of data	Chapter in which data is used
Kerry (KE)	30	Ireland / *B. taurus*	Chapter 4 of this thesis	4
Jersey (JE)	26	UK / *B. taurus*	Chapter 4 of this thesis	4
Angus (AA)	29	UK / *B. taurus*	Chapter 4 of this thesis	4
Hereford (HE)	30	UK / *B. taurus*	Chapter 4 of this thesis	4
Highland (HI)	25	UK / *B. taurus*	Chapter 4 of this thesis	4
East Finn (EF)	30	Finland / *B. taurus*	Chapter 4 of this thesis	4
Vosges (VO)	27	France / *B. taurus*	Chapter 4 of this thesis	4
Pinzgauer (PN)	27	Germany-Austria / *B. taurus*	Chapter 4 of this thesis	4
Braunvieh (BV)	30	Germany / *B. taurus*	Dr. G. Durstewitz (Technical University of Munich) pers. comm	4
Fleckvieh (FV)	29	Germany / *B. taurus*	Dr. G. Durstewitz (Technical University of Munich) pers. comm	4
Holstein (SB)	35	Holland / *B. taurus*	Dr. G. Durstewitz (Technical University of Munich) pers. comm	4
Hungarian Grey (HG)	30	Hungary / *B. taurus*	Chapter 4 of this thesis	4

Table 5.4 continued

Population name (Population code)	*n*	Population origin / Animal morphology	Source of data	Chapter in which data is used
Anatolian Black (AB)	25	Turkey / *B. taurus*	Chapter 4 of this thesis	4
South Anatolian Red (SAR)	30	Turkey / *B. taurus*	Chapter 4 of this thesis	4
Guinean N'Dama (ND)	27	Guinea / *B. taurus*	Chapter 4 of this thesis	4
North Indian zebu (NIZ; a composite of Hariana, Sahiwal, Tharparker	25	Indian *B. indicus*	Chapter 4 of this thesis	4
Ongole (ON)	29	South India / *B. indicus*	Chapter 4 of this thesis	4

5.2 DNA EXTRACTION METHODS

5.2.1 DNA extraction from Caribbean Creole cattle serum samples

2 mls of bovine serum were collected from animals at abattoirs located on the Caribbean Islands of Antigua and St. Lucia by Dr. Alex Morrow (University of Edinburgh) between June 1987 and June 1989. Upon collection, all samples were heat-treated at 55°C for 20-30 minutes to destroy any infectious agents (Dr. Alex Morrow, personal communication). Heat-treated serum samples were stored at -70°C.

All DNA extractions from serum samples were performed at the Centre for Tropical Medicine, University of Edinburgh, in January 2001. DNA was extracted from 200 μls serum samples using a slightly modified version of the standard phenol-chloroform procedure described by Sambrook *et al.* 1988. All consumable laboratory materials used during DNA extraction procedures were sterile. Prior to phenol-chloroform extractions, all serum samples were treated with Proteinase K-digestion buffer (**reagent A, Appendix 3**) to remove any extra-cellular protein or potential PCR inhibitors. This was performed by aliquoting 200 μls of each bovine serum sample into a 2 ml microcentrifuge tube containing 400 μls of Proteinase-K digestion buffer and were incubated overnight at 37°C.

Once protein digestion was complete, 450 μls of hydroxyquinoline-equilibrated phenol (pH 8.0, SIGMA-ALDRICH) was added to each sample. Tubes were inverted manually for 5 minutes at room temperature until an opaque solution was formed. Samples were centrifuged at 15,000 rpm for 10 minutes. The upper aqueous layer was removed and transferred in to a fresh 2 ml microcentrifuge containing 450 μls of a chloroform:iso-amyl alcohol (50:1) mix. Tubes were inverted manually for 5 minutes and centrifuged at 15,000 rpm for 10 minutes. The upper aqueous layer was removed and transferred to a fresh 2 ml microcentrifuge containing 40 μls of 3 M sodium-acetate solution.

To precipitate DNA from the phenol-chloroform extractions, 800 μls of cold (-20°C) 100% ethanol was added to each sample, vortexed and stored overnight at -70°C. Tubes were then centrifuged in a cold room (4°C) at 15,000 rpm for 15 minutes until glass-like DNA pellets formed at the base of the microcentrifuge tubes. The supernatant was discarded and the pellets were washed with 70% ethanol, air-dried in a fume hood and dissolved in 30 μls of 1X TE buffer. Samples were stored at -20°C and until required.

5.3 MITOCHONDRIAL DNA AMPLIFICATION AND SEQUENCING

5.3.1 Oligonucleotide primer information for mitochondrial DNA PCR amplification

A 375 bp fragment situated in hypervariable region 1 (HVR1) of the bovine mitochondrial genome (nucleotide positions 15,960 – 16,334) was amplified using the oligonucleotide primers described in **Table 5.5**. These primers were designed by Dr. Chris Troy (Department of Genetics, TCD) and have been used in previous bovine phylogenetic surveys (Cymbron *et al.* 1999, Troy *et al.* 2001). 1 ml of each primer was supplied in sterile ddH_2O by OSWEL SCIENTIFIC. Stock primers were prepared into 250 μls aliquots and stored at -20°C.

Table 5.5: Oligonucleotide primers used for bovine mtDNA amplification

Primer name	Primer sequence (5'- 3')	Modifications	Position in control region
AN3	CGA GAT GTC TTA TTT AAG AGG	none	L15960
AN4-bio	GGT AAT GTA CAT AAC ATT AAT G	5' Biotin	H16334

L15960 and H16334 refer to the start position of both primers along the light and heavy DNA strands of the bovine mitochondrial control region, respectively. Nucleotide positions are as per Anderson *et al.* (1982).

5.3.2 PCR conditions for the amplification of bovine mitochondrial DNA sequences

Optimum polymerase chain reaction (PCR (Mullis *et al.* 1986) amplification reagents and cycling parameters were previously established by Troy (1998) and are outlined in **Table 5.6** and **Table 5.7**. PCR amplifications were performed in 600 μls PCR certified microcentrifuge tubes (ROBBINS SCIENTIFIC) using a PERKIN ELMER CETUS thermal cycler.

Successful DNA amplifications were verified by loading 5 μls of PCR product together with 2 μls of 6X loading dye (**reagent B, Appendix 3**) on a 1.5% agarose gel stained with

ethidium bromide (0.5 μg/ml). All samples were electrophoresed at 100 Volts for 15 minutes using 1X TBE buffer (**reagent C, Appendix 3**) together with a 100 bp molecular size standard (GIBCO LIFE SCIENCES) to confirm the size of the resulting PCR products. Agarose gels were visualised using a UV light box and were photographed with a Polaroid camera. Once sufficient DNA yields were confirmed, single-stranded templates were prepared from the remaining 45 μls of PCR product.

Table 5.6: Mitochondrial DNA PCR reagents

Reaction component	Initial concentration	Volume	Final concentration
10X PCR reaction buffer[1]	15 mM $MgCl_2$	5 μls	1.5 mM $MgCl_2$
Primer: AN4-bio	3 mM	4.125 μls	0.247 μM
Primer: AN3	3 mM	4.125 μls	0.247 μM
dNTP mix	2 mM each dNTP	3 μls	0.12 mM
Taq DNA polymerase	200 Units	0.5 μls	2 Units
Sterile ddH_20	-	32 μls	-
DNA template		2 μls	10-25 ngs
Total reaction volume		50 μls	
Mineral oil overlay		30 μls	
[1]Reagent D, Appendix 3			

Table 5.7: Mitochondrial DNA PCR cycling parameters

Reaction step		Temperature (^{0}C)	Time	Number of cycles
Initial denaturation		94	3 minutes	1
Cycling	Denaturation	94	40 seconds	35
	Annealing	55	40 seconds	
	Extension	72	1 minute	
Final extension		72	4 minutes	1

5.3.3 Preparation of single stranded template DNA for sequencing

PCR products were purified using the standard Dynabead™ (DYNAL) protocol. This protocol takes advantage of the binding properties of two proteins, namely streptavidin and biotin. Dynabeads are comprised of magnetically charged beads coated with streptavidin molecules. Incubation of the beads with biotinylated PCR products at 43°C for 15 minutes results in the formation of a Dynabead-PCR product complex. This complex forms a pellet, which adheres to the wall of a microcentrifuge tube when placed on a magnetic particle concentrator (DYNAL). The pellet is then washed and purified using specified buffers and eventually rendered single-stranded. The resulting single-stranded DNA is used as a template in a manual DNA sequencing reactions.

Prior to the purification of biotinylated PCR products, Dynabeads were treated with 1X and 2X "Washing and Binding" buffers (**reagent E and reagent F, Appendix 3**) as per manufacturer's instructions. 40 μls of the pre-treated Dynabeads were placed into a 1.5 ml microcentrifuge tube together with an equal volume of PCR amplification product. The mixture was then incubated at 43°C for 15 minutes, gently vortexing each tube every five minutes to increase the efficiency of the complex formation. After incubation, the tubes were placed on the DYNAL magnetic particle concatenator and the supernatant was removed and discarded.

Duplex DNA was denatured by adding 50 μls of fresh 0.1 M NaOH to the complex and incubating at room temperature for 5 minutes, rendering the DNA single-stranded. Non-biotinylated PCR strands were subsequently removed and discarded. The single-stranded template was then washed with 50 μls of 1X TE buffer (**reagent G, Appendix 3**), re-suspended in 25 μls of sterile dd$H_2$0 and stored at –20°C until required.

5.3.4 Sequencing mitochondrial DNA control region sequences

Single-stranded, 2'-3' dideoxy-DNA sequencing (Sanger *et al.* 1977) was performed in this study using a modified protocol of the *Sequenase*™ VERSION 2.0 protocol (UNITED STATES BIOCHEMICAL), previously employed by Troy (1998).

Annealing reactions (**Table 5.8**) and labelling reactions (**Table 5.9**) were conducted in 600 μls PCR certified microcentrifuge tubes (ROBBINS SCIENTIFIC). Alternatively, termination reactions (**Table 5.10**) were performed in 60 well Terasaki low-profile tissue-typing trays (NUNC LIFE SCIENCES). Annealing reaction mastermixes were prepared and mixed thoroughly with the single-stranded DNA template. The reaction mix was heated at 65°C for 2 minutes on a heat block and allowed to cool at room temperature over a 30 minute period, vortexing gently every 5 minutes. This enhances the interaction of the reaction components and increases the annealing efficiency of the AN3 sequencing primer with the DNA template.

During the cooling period, both the labelling reaction mastermix and the 2'-3' di-deoxy-terminator solutions were prepared. 2.5 μls of each terminator solution was placed in their appropriate wells, designated 'G', 'A', 'T' and 'C', on a Terasaki low profile tissue typing tray. Both reaction mixes were then stored on ice until required.

Once the 30 minute cooling period of the annealing reaction was completed, 5 μls of the labelling reaction mastermix was mixed thoroughly with each annealing reaction and incubated at room temperature for 3 minutes. After incubation, 3.5 μls of the annealing-labelling reaction mix was added to the appropriate dideoxynucleotide terminator solutions. This was performed evenly over a 5 minute period at 37°C (the optimal temperature for the *Sequenase*™ enzyme). Termination reactions were interrupted by the

addition of 4 μls of a formamide-based loading buffer (**reagent H, Appendix 3**) to each well, which was again applied evenly over a 5 minute period. Sequenced samples were stored at 4°C until required for polyacrylamide gel electrophoresis (PAGE).

Table 5.8: DNA sequencing annealing reaction components (per sample)

Reaction component	Initial concentration	Volume	Final concentration
5'-Biotinylated single-strand DNA template	-	4 μls	-
Sequencing primer (AN3 primer)	1.5 μM	4 μls	0.55 μM
Sequenase buffer[1]	5X	2.2 μls	1X
DMSO	50%	1 μls	5%
Total reaction volume		**11 μls**	

[1] **reagent I, Appendix 3**

Table 5.9: DNA sequencing labelling reaction components (per sample)

Reaction component	Initial concentration	Volume	Final concentration
Labelling mix[1]	5X	1.0 μls	1X
DTT	0.1 M	1.5 μls	2 μM
DMSO	50%	0.5 μls	5%
$MnCl_2$ buffer[2]	0.1 M $MnCl_2$	0.1 μls	2 μM
Sequenase™ Version 2.0	13 Units / μl	0.25 μls	0.165 units
^{35}S-dATP (NEN LIFESCIENCES)	1 mCi / ml	0.5 μls	
Sterile ddH_2O	-	1.15 μls	
Total reaction volume		5 μls	

[1] **reagent J, Appendix 3**

[2] **reagent K, Appendix 3**

Table 5.10: DNA sequencing termination reaction components (per sample)

Terminator solution	Volume	Components
ddGTP-termination solution	2.5 μls	80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddGTP, 50 mM NaCl.
ddATP-termination solution	2.5 μls	80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddATP, 50 mM NaCl.
ddTTP-termination solution	2.5 μls	80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddTTP, 50 mM NaCl.
ddCTP-termination solution	2.5 μls	80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddCTP, 50 mM NaCl.

The volume refers to that used in each individual sequencing reaction. A final concentration of 5% DMSO was also added to all sequencing reactions

5.3.5 Polyacrylamide gel electrophoresis (PAGE) of sequenced cattle samples

Polyacrylamide gel electrophoresis (PAGE) was performed on all manually sequenced samples using a vertical 96 well S1BRL sequencing gel electrophoresis apparatus (GIBCO LIFE SCIENCES). 6% denaturing polyacrylamide sequencing gels were freshly prepared from stock solutions of Sequagel™ PAGE reagents (NATIONAL DIAGNOSTICS; **reagent L, Appendix 3**).

Prior to loading, all sequences were heat denatured for 3 minutes at 94°C. The samples were then snap-cooled by immediately placing them on ice. 1.5 μls of each termination reaction for each sequenced sample was loaded to each well on the pre-heated gel. Gels were electrophoresed in 1X TBE buffer at 70 Watts for approximately 165 minutes at room temperature. To ensure that sequencing bands remained compacted during electrophoresis 30 mls of 6 M ammonium acetate was added to the bottom buffer trough (Murphy & Kavanagh 1988). This procedure allowed all 240 bp to be observed in one single gel electrophoretic run.

Upon completion of electrophoresis, the gel was transferred to a sheet of chromatography paper (WHATMAN), covered in saran wrap and placed on a SAD 4050 slab gel vacuum drier (SAVANT) and dried for 55 minutes at 80°C. Once dried, the gel was placed in a film cassette and exposed to an autoradiographic X-ray film (AGFA) for periods ranging from 24 hours to 2 weeks, depending on the intensity of the radioactive signal. Autoradiographs were developed using an AGFA CURIX™ automatic film developer located in the departmental dark room.

5.3.6 Determination and recording of mitochondrial DNA sequence data

DNA sequences were entered manually into an Apple Macintosh© Performa 670 computer, using a DNA sequence data input package (GENEJOCKEY II; Taylor 1993). All entries were triple checked prior to further sequence analysis. In addition, all sequences were aligned using the multiple sequence alignment option on the GENEJOCKEY II package to identify nucleotide differences, which were subsequently confirmed against the sequence on the autoradiographs. All corrected sequences were complied into a single Microsoft Word ® text document and adapted into varying input formats for phylogenetic analysis.

5.4 MICROSATELLITE ALLELE GENOTYPING

5.4.1 The microsatellite loci genotyped in Chapter 3 of this thesis

Each Caribbean Creole population was assayed for allelic length variation at three autosomal microsatellite loci: *ILSTS001*, *Hel1* and *BM2113* (**Table 5.11**).

5.4.2 PCR amplification of microsatellites

PCR amplifications were carried out in 96 well microtitre plates (MJ RESEARCH) using a OmniGene thermal cycler (HYBAID). PCR reagent concentrations and volumes are shown in **Table 5.12.** Optimal PCR cycling parameters (**Table 5.13**) for each locus were previously established by Dr. David MacHugh (MacHugh 1996) and were modified to produce radioactively labelled amplification products via the random incorporation of standard ^{32}P-dCTP [3,000 Ci/μls] (NEW ENGLAND BIOSCIENCES). This permitted the visualisation of microsatellite alleles on autoradiographic film. Negative controls were included in all PCR amplifications to test for the presence of extraneous DNA contamination. Once completed, radiolabelled PCR products were stored appropriately at 4°C until required for electrophoresis

Table 5.11: Microsatellite loci genotyped in Creole cattle populations

Locus	Reference	Chromosome location	Primer sequences (5'-3')	Repeat motif	Allele sizes (bp)	$MgCl_2$ Conc.	Anneal. Temp (°C)
BM2113	Sunden *et al.* (1993)	BTA 2	GCT GCC TTC TAC CAA ATA CCC CTT CCT GAG AGA AGC AAC ACC	$(CA)_n$	123 - 147	20 mM	55°C
HEL1	Kaukinen & Varvio (1993)	BTA 15	AGT CCA TGG GAT TGA AAG AGT TGG CTT TTA TTC AAC AGC TAT TTA ACA AGG	$(CA)_n$	101 - 117	20 mM	55°C
ILSTS001	Brezinsky *et al.* (1992)	BTA 7	GGT GCT GTT ATC TAG AAT TTG G GGA GTC ATA CAC AAC TGA GC	$(TG)_n$	77 - 97	20 mM	58°C

Chromosomal locations were obtained from the ARK database (www.thearkdb.org). Primer sequences were taken from MacHugh *et al.* (1997). $MgCl_2$ and PCR annealing temperatures were modified from MacHugh (1996).

Table 5.12: Radiolabelled PCR amplification of microsatellite loci

Component	Volume	Final Concentration
10X *Taq* Polymerase PCR buffer[1]	1 μl	1X
Forward primer	1 μl	0.3 mM
Reverse primer	1 μl	0.3 mM
dNTPs	1 μl	0.2 mM dATP, dTTP, dGTP/10 μM dCTP
$MgCl_2$	1 μl	1.5 mM – 20 mM
^{32}P-dCTP [3,000 Ci/μl]	0.05 μls	
Platinum *Taq* Polymerase	0.1 μls	0.5 Units
DNA template	1 μl	~25 ngs
Sterile ddH_2O	3.85 μls	-
Total	10 μls	

[1] **reagent M, Appendix 3**

A mastermix containing all PCR components (excluding the DNA template) was prepared for each for the appropriate number of DNA samples. 9μls mastermix aliquots were added to 1 μl of each DNA sample contained on a 96-well polycarbonate microtitre plate, using a repeating multi-pipette (GILSON). A mineral overlay of 10μls was added to each well to prevent evaporation during the amplification reactions. All radiolabelled PCRs were performed on an OmniGene (HYBAID) DNA thermal cycler.

Table 5.13: Radiolabelled PCR amplification cycling parameters

Reaction step		Temperature (°C)	Time	Number of cycles
Initial denaturation		94	3 minutes	1
Cycling conditions	Denaturation	94	40 seconds	35
	Annealing	55-58	40 seconds	
	Extension	72	1 minute	
Final extension		72	4 minutes	1

5.4.3 Polyacrylamide gel electrophoresis (PAGE) of PCR-amplified microsatellites

Prior to electrophoresis, 10 μls of formamide denaturing buffer was added to each of the amplified radiolabelled PCR products, which were subsequently heat denatured at 94°C for

3 minutes and then snap-cooled by immediately placing them on ice. 3 μls aliquots of each sample were loaded into each well on a pre-warmed 6% denaturing polyacrylamide sequencing gel (NATIONAL DIAGNOSTICS) using a fine-tipped multi-channelled pipette (HAMILTON). All gels were run on an S1BRL (GIBCO LIFE SCIENCES) with 1X TBE electrophoresis buffer at 70 Watts for 1.5-3 hours at room temperature, depending on the size of the PCR amplicons. Standard samples of known allele size for each locus, as previously calibrated and determined by Dr. David MacHugh (Department of Genetics, TCD) were loaded in conjunction with the samples, in order to determine the size the microsatellite alleles present in each sample.

Gels were dried and developed as described in **Section 5.3.5**. Genotypes were determined by comparing the sizes of the alleles present in the samples with the pre-calibrated allele-size standards. Genotypes were read manually into a Microsoft Excel© spreadsheet on a Gateway© 2000 PC computer. This data was stored in a database and then printed out and double-checked against the original autoradiographs.

5.5 SINGLE NUCLEOTIDE POLYMORPHISM GENOTYPING

5.5.1 Multiplex genotyping of SNP loci

All SNP loci used for phylogenetic analysis in this study were discovered, characterised and mapped to the bovine genome by members of the Department of Animal Science at the Technical University of Munich, Germany (**Table 4.1**). SNP genotyping was performed using a modified protocol of the oligoligation assay (OLA) previously described by Nickerson *et al.* (1990) and Tobe *et al.* (1996). Multiplex PCR amplification conditions are described in **Table 5.14**. All multiplex PCR amplifications were performed in 96-well microtitre plates (MJ RESEARCH) on an OmniGene thermal cycler (HYBAID). In general, mulitplex PCR amplifications involved a maximum of 3 SNP loci. Negative controls were included in all PCR amplifications to test for the presence of extraneous DNA contamination. Successful amplifications were determined by agarose gel electrophoresis as described in **Section 5.3.2**.

OLA multiplex reaction parameters are detailed in **Table 5.15**. OLA products were stored at -20°C until required for polyacrylamide gel electrophoresis (PAGE). PCR and OLA oligonucleotide sequence information, together with OLA product and corresponding SNP alleles details, are presented in **Table 5.16**.

Table 5.14: Multiplex PCR conditions for SNP genotyping

PCR Component	Initial concentration	Volume
Platinum *Taq* Polymerase Buffer (GIBCO)	10X	2 μls
dNTPs	2 mM each	4 μls
Primers	20 mM	0.5 μls each
$MgCl_2$	25 mM	1.6 μls
Platinum *Taq* Polymerase (GIBCO)	5 Units/μl	0.1 μls
DNA Template	25 ng/μl	2 μls
ddH_2O	-	7.3 μls
Total	-	**20 μls**

Reaction mastermixes shown above are for a 3-SNP loci multiplex PCR. 18 μls from each mastermix were aliquoted to each DNA sample using a repeating multi-pipette (GILSON). All multiplex PCR amplifications were performed in 96-well microtitre plates (MJ RESEARCH). PCR cycling conditions were as follows: 94°C for 3 minutes; 35 cycles of 94°C for 1 minute, 54°C-58°C for 40 seconds, 72°C for 1 minute; 72°C for 10 minutes. All muliplex PCR conditions were established by Dr. G. Durtewitz (personal communication).

Table 5.15: Multiplex OLA reaction conditions for SNP genotyping

OLA Component	Concentration	Volume
Taq ligase buffer[1]	10X	2 μls
Primers	20 mM	0.5 μls each
Taq ligase (New England Biosciences)	40,000 Units/ml	0.1 μls
PCR product	-	3 μls
ddH_2O	-	10.4 μls
Total		**20 μls**

[1] **reagent N, Appendix 3**

OLA reaction volumes are given for a 3 SNP loci multiplex. OLA cycling conditions were as follows: 94 °C for 1 minute, 21 cycles of 94°C for 30 seconds, 45°C for 3 minutes; 99°C for 10 minutes. Muliplex reaction conditions were established by Dr. Gregor Durtewitz. All multiplex OLA reactions were performed in 96-well microtitre plates (MJ RESEARCH).

Table 5.16: Primer and allele information for SNP genotyping

Locus	Chromosome position	Primer sequence (5'-3')	Primer function	Amplicon size (bp)	SNP alleles	OLA product size	OLA product colour
TCRA	BTA 10q15	GTCACCTCCATTTATAGATCAGC	PCR forward	361	C	43 bp	Blue
		AACCAGAATGGCAGAAGTCTCC	PCR reverse				
		Pho-GCTGGAAGAGAAGAT-FAM	OLA common				
		5xPEO-GTAAGAGGAGGAC**C**	OLA specific allele 1		T	49 bp	
		7xPEO-TGTAAGAGGAGGAC**T**	OLA specific allele 2				
CAC	BTA 4q18	AATAACCACAGTACAGAGAGG	PCR forward	126	C	70 bp	Blue
		TTGTGCATAGATCAAAGGTTTATT	PCR reverse				
		Pho-TCTTCAAAAATATGTGTC-FAM	OLA common				
		15xPEO-GGACATACTTTCCA**C**	OLA specific allele 1		G	73 bp	
		16xPEO-GGACATACTTTCCA**G**	OLA specific allele 2				
421-10	BTA 1q43-44	GCTTCTTTTAATTTCCAGGTCTCA	PCR forward	125	C	93 bp	Blue
		CCATCTGAGGTGCTTCTTCG	PCR reverse				
		Pho-TTGTTGCCTTGAGGG-FAM	OLA common				
		26xPEO-CCCTGTTTCTGACT**C**	OLA specific allele 1		G	98 bp	
		28xPEO-CCCTGTTTCTGACT**G**	OLA specific allele 2				

Table 5.16: continued

Locus	Chromosome position	Primer sequence (5'-3')	Primer function	Amplicon size (bp)	SNP alleles	OLA product size	OLA product colour
417-16	BTA 4q22-24	TTGGAAGGAAGGAGGTGATG	PCR forward	145	G	62 bp	Yellow
		TTCACACCTGCTTTGACAGTCT	PCR reverse				
		Pho-TCACAGTTGATGTCA-TAMRA	OLA common				
		13xPEO-GAGCTGCATTTCAC**G**	OLA specific allele 1				
		14xPEO-GAGCTGCATTTCAC**A**	OLA specific allele 2		A	65 bp	
463-67	BTA 17q12-14	ACCACCAACCCATGAGAAGA	PCR forward	84	C	36 bp	Yellow
		CCCCACTTGGGGAAGATAGT	PCR reverse				
		Pho-CACGCCCTCTTTTAC-TAMRA	OLA common				
		6xPEO- TGTCCACAAGTTGA**C**	OLA specific allele 1				
		7xPEO-ATGTCCACAAGTTGA**T**	OLA specific allele 2		T	38 bp	
454-G11	BTA 17q24-26	TGGGAAAAATAAATAATAGAAACCAGA	PCR forward	107	C	53 bp	Blue
		TTTTCCTCAAATTTATTCTTACATTG	PCR reverse				
		Pho-TGCCTAATTGAGAACA-FAM	OLA common				
		8xPEO-TATAATGCCATCCAA**C**	OLA specific allele 1				
		9xPEO-TATAATGCCATCCAA**G**	OLA specific allele 2		G	55 bp	

Table 5.16: continued

Locus	Chromosome position	Primer sequence (5'-3')	Primer function	Amplicon size (bp)	SNP alleles	OLA product size	OLA product colour
436-C10	BTA 4q13-14	TGTAATCCAGACAACTGCAAGG	PCR forward	154	C	61 bp	Blue
		AAAATTACCTCCTCTTGGTGA	PCR reverse				
		Pho-ATAATTTAAATGTTGATCA-FAM	OLA common				
		9xPEO- GTATATATATTATACATA**C**	OLA specific allele 1		T	64 bp	
		10xPEO-TGTATATATATTATACATA**T**	OLA specific allele 2				
CNTFR	BTA 8q21	ATTTTGGCTGGGCCTGTAA	PCR forward	81	C	65 bp	Green
		AGGCACTCAGCAGCCTTC	PCR reverse				
		Pho-GGGAGGGGTGGAC-HEX	OLA common				
		14xPEO-TTGTAGCCAGCTG**C**	OLA specific allele 1		G	67 bp	
		15xPEO-TTGTAGCCAGCTG**G**	OLA specific allele 2				
437	BTA 14q14-16	TGCCAGAAACAGATGCTAGG	PCR forward	91	C	74 bp	Yellow
		GCCTCTTTGAACCAATTTTTCTT	PCR reverse				
		Pho-CATCAGATGATGGA-TAMRA	OLA common				
		17xPEO-CGCATCTTCAACTT**C**	OLA specific allele 1		A	77 bp	
		18xPEO-ACGCATCTTCAACTT**A**	OLA specific allele 2				

Table 5.16: continued

Locus	Chromosome position	Primer sequence (5'-3')	Primer function	Amplicon size	SNP alleles	OLA product size	OLA product colour
486-67	BTA 3q31-32	GACCTCATTCTCCTACTTACTGTAAAA	PCR forward	94	C	86 bp	Green
		CATCTGTGTGGTTTGTTTACCA	PCR reverse				
		Pho-TCAGAGTCTTCATTGA-HEX	OLA common				
		21xPEO-CTATCTTCTCTATTG**C**	OLA specific allele 1		T	90 bp	
		22xPEO-TCTATCTTCTCTATTG**T**	OLA specific allele 2				
444-66	BTA 8q18-22	CTGAACTTTGGCCAATGAAA	PCR forward	96	C	77	Blue
		TCCCAGTTGGTTGGTTGGTA	PCR reverse				
		Pho-TGCTCTTACATCTTCA-FAM	OLA common				
		18xPEO- TTGTCTGAGCTTCT**C**	OLA specific allele 1		T	81	
		19xPEO-TTTGTCTGAGCTTCT**T**	OLA specific allele 2				

All OLA primer sequences were designed and supplied by Dr. Gregor Durstewitz. OLA product sizes and corresponding SNP allele identity are colour coded. Smaller OLA products corresponding to 'Allele 1' of the numerical output system are denoted in red, larger OLA products corresponding to 'Allele 2' are denoted in blue. OLA product detection used three-colour fluorescent dye chemistry, namely 6-FAM (blue), HEX (green) and TAMRA (yellow). GENESCAN-ROX 500 (red) was used as internal size standard to size OLA products. Abbreviations: Pho, Phosphate group; PEO, Polyethylene oxide (oligonucleotide length modifier molecules). Nucleotide sites of ligation are shown in bold on the OLA-specific primers sequences.

5.5.2 Polyacrylamide gel electrophoresis (PAGE) of OLA products

Prior to gel electrophoresis, 2.5 μls of each OLA product was added to 1 μl ABI loading dye (**reagent O, Appendix 3**) and 0.6 μls of GENESCAN-ROX 500™ fluorescent internal size standard (APPLIED BIOSCIENCES). This mixture was heated at 94°C for 3 minutes and immediately placed on ice. All OLA products were loaded on 8% denaturing polyacrylamide (**reagent P, Appendix 3**) sequencing gels were freshly prepared from stock solutions of Sequagel™ PAGE reagents (NATIONAL DIAGNOSTICS). Electrophoresis was performed on an ABI373 automated sequencer (APPLIED BIOSCIENCES) using 1X TBE buffer. 3.5 μls of each OLA product was loaded in single lanes on pre-warmed gels (40 Watts for 10 minutes). OLA product detection used three-colour fluorescent dye chemistry, namely 6-FAM (blue), HEX, (green) and TAMRA (yellow). Each gel contained a maximum of 30 samples. Even numbered lanes were loaded first followed by a 5 minute pre-run after which odd numbered lanes were loaded and subsequently subjected to a 5 minute pre-run.

Once the 'pre-runs' were complete, all gels were electrophoresed for 5 hours at 40 Watts at room temperature. Raw genotype data were collected by used of GENESCAN software (APPLIED BIOSCIENCES) and gel files were analysed by use of the GENOTYPER software package (APPLIED BIOSCIENCES). SNP alleles, as determined by OLA product sizes (**Tables 4.1 and 5.16**) were given a simple numerical output denoted 'Allele 1' (corresponding to the smaller OLA product) or 'Allele 2' (corresponding to the larger OLA product) and were imported into a Microsoft Excel © worksheet. All recorded genotypes were double checked against the appropriate GENOTYPER analysis file. This numerical output format allowed SNP data to be evaluated by means of sophisticated phylogenetic software packages.

CHAPTER 6

REFERENCES

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

Allchin, F.R. (1963). *Neolithic cattle-keepers of the South India: A study of the Deccan ashmounds*. Cambridge University Press, Cambridge, U.K.

Allchin, B. and Allchin, R.A. (1968*). The birth of Indian civilization : India and Pakistan before 500 B.C.* Penguin.

Allchin, B. and Allchin, R.A. (1974). Some new thoughts on Indian cattle. In: *South Asian Archaeology 1973.* **van Lohuizen-de Leeuw, J.E. and Ubaghs, J.M.** (Editors). Pp, 71-77. E.J Brill, Leiden.

Allchin, B. and Allchin, R.A. (1982). *The rise of civilization in India and Pakistan.* Cambridge University Press, Cambridge, U.K.

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

Ankel-Simons, F. and Cummins, J.M. (1996). Misconceptions about mitochondria and mammalian fertilization: Implications for theories on human evolution. *Proceedings of the National Academy of Sciences USA* **93**, 13859-13863.

Arctander, P., Johansen, C. and Coutellec-Vreto, M.A. (1999). Phylogeography of three closely related African bovids (tribe *Alcelaphini*). *Molecular Biology and Evolution* **16**, 1724-1739.

Attardi, G. and Schatz, G. (1988). Biogenesis of mitochondria. *Annual Review of Cell Biology* **4**, 289-333.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., lamb, T., Neigel, J.E., Reeb, C.A. and Saunders, N.C. (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Avise, J.C. (1991). Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annual Review of Genetics* **25**, 45-69.

Avise, J.C. (1994). *Molecular Markers, Natural History and Evolution.* Chapman & Hall, New York.

Avise, J.C. (1998). The history and purview of phylogeography: A personal reflection. *Molecular Ecology* **7**, 371-379.

Avise, J.C.: (2000). *Phylogeography. The history and formation of a species.* Harvard University Press, Cambridge, Massachusetts.

Awadalla, P., Eyre-Walker, A. and Smith, J.M. (1999). Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* **286**, 2524-2525.

Bailey, J.F., Richards, M.B., Macaulay, V.A., Colson, I.B., James, I.T., Bradley, D.G., Hedges, R.E. and 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 B Biological Sciences* **263**, 1467-1473.

Baker, C.M.A. and Manwell, C. (1991). Population genetics, molecular markers and gene conservation of bovine breeds. In: *Cattle Genetic Resources Volume B7*. **Hickman, C.G.** (Editor). Pp. 221-304.

Balloux, F. and Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11**, 155-165.

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

Barbujani, G., Magagni, A., Minch, E. and Cavalli-Sforza, L.L. (1997). An apportionment of human DNA diversity. *Proceedings of the National Academy of Sciences USA* **94**, 4516-4519.

Bar-Yosef, O. and 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.* **Price, T.D. and Gebauer A.B.** (Editors). Pp. 39-95. School of American Research Press.

Bellwood, P. (1996). The origins and spread of agriculture in the Indo-Pacific regions: gradualism and diffusion or revolution and colonisation. In: In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. **Harris, D.R.** (Editor). Pp. 465-498. University College London, London, U.K.

Bernstein, F. (1931). Die geographische Verteilung der Blutgrippen und ihre anthropologische Bedeutung. In: *Comitato Italiano per o studio dei problemi della populazione*. Pp. 227-243. Instituto Poligrafico dello Stato, Rome, Italy.

Bertorelle, G. and Excoffier, L. (1998). Inferring admixture proportions from molecular data. *Molecular Biology and Evolution* **15**, 1298-1311.

Bertorelle, G. (1998). ADMIX VERSION 1.0: computing admixture coefficients from molecular data (Software package) [www.unife.it/genetica/Giorgio/giorgio.html]. Department of Integrative Biology, University of California, Berkeley, California.

Bertranpetit, J., Calafell, D., Comas, A., Pérez-Lezaun, A. and Mateu, E. (1996). Mitochondrial DNA sequences in Europe: an insight into population history. In: *Molecular Biology and Human Diversity*. **Boyce, A.J. and Mascie-Taylor, C.G.N.** (Editors). Pp. 112-129.

Binford, L.R. (1968). "Post-Pleistocene adaptations". In: *New perspectives in archaeology*. **Binford, S.R. and Binford, L.R.** (Editors) Aldine, Chicago, U.S.A.

Blumler, M.A. (1996). Ecology, evolutionary theory and agricultural origins. In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. **Harris, D.R.** (Editor). Pp. 25-49. University College London, London, U.K.

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

Bökönyi, S. (1997). Zebus and Indian wild cattle. *Anthropozoologica* **25-26**, 647-654.

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

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

Bradley, D.G. and Cunningham P. (1999). Genetic aspects of domestication, common breeds and their origins. In: *The genetics of cattle.* **Fries, R. and Ruvinsky, A.** (Editors). Pp. 15-31. CAB International.

Bradley, D.G. (2000). Mitochondrial DNA diversity and origins of domestic livestock. In: *Archaeogenetics: DNA and the population prehistory of Europe.* **Renfrew, C. & Boyle, K.** (Editors). Pp 315-320. McDonald Institute Publications.

Brezinsky, L., Kemp, S.J. and Teale, A.J. (1992). ILSTS001: a polymorphic bovine microsatellite. *Animal Genetics* **23**, 81.

Brookes, A.J. (1999). The essence of SNPs. *Gene* **234**, 177-186.

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

Brown, W.M. (1985). The mitochondrial genome of animals. In: *Molecular Evolutionary Genetics*. **MacIntyre, R.J.** (Editor). Pp. 95-130. Plenum, New York, U.S.A.

Bruford, M.W. and Wayne, R.K. (1993). Microsatellites and their application to population genetic studies. *Current Opinions in Genetics and Development* **3,** 939-943.

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

Byrne, R. (1987). Climatic change and the origins of agriculture. In: *Studies in the Neolithic and urban revolutions.* **Manzanilla, L.** (Editor). Pp. 21-34. B.A.R. International series **349**. British archaeological reports, Oxford.

Cann, R.L. (2001). Genetic clues to dispersal in, human populations: retracing the past from the present. *Science* **291**, 1742-1748.

Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1993). Demic expansions and cultural evolution. *Science* **259**, 639-646.

Cavalli-Sforza, L.L. (1996). The spread of agriculture and nomadic pastoralism: insights from genetics, linguistics and archaeology. In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia.* **Harris, D.R.** (Editor). Pp. 51-69. University College London, London, U.K.

Cavalli-Sforza, L.L. (1998). The DNA revolution in population genetics. *Trends in Genetics* **14**, 60-65.

Chakravarti, A. (1999). Population genetics--making sense out of sequence. *Nature Genetics* **21**, 56-60.

Chambers, G.K. and MacAvoy, E.S. (2000). Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology B* **26**, 455-476.

Charles, M.P. (1993). "Botanical remains'. In: *Abu Salabikh excavations. Volume 4. The 6G ash-tip and its contents: cultic and administrative discard from the temple?* **Green, A.** (Editor). Pp. 203-207. British school of archaeology in Iraq, London, U.K.

Chinnery, P.F., Thorburn, D.R., Samuels, D.C., White, S.L., Dahl, H.M., Turnbull, D.M., Lightowlers, R.N. and Howell, N. (2000). The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? *Trends in Genetics* **16**, 500-505.

Clason, A.T. (1978). Late Bronze-Iron Age zebu in Jordan? *Journal of Archaeological Science* **5**, 91-93.

Clayton, D.A. (1991). Replication and transcription of vertebrate mitochondrial DNA. *Annual Review of Cell Biology* **7**, 453-478.

Cload, S.T. and Schepartz, A. (1991). Polyether tethered oligonucleotide probes. *Journal of the American Chemical Society* **113**, 6324-6326.

Clutton-Brock, J. (1981). *Domestic animals from early times.* Heinemann (in association with the British Museum of Natural History), London, U.K.

Clutton-Brock, J. (1989). Cattle in ancient North Africa. In: *The walking larder: Patterns of Domestication, Pastoralism and Predation.* **Clutton-Brock, J.** (Editor). Pp. 201-206. Unwin Hyman, London, U.K.

Clutton-Brock, J. (1999). *A natural history of domesticated animals* (2nd edition). Cambridge University Press, Cambridge, U.K.

Collins, F.S., Guyer, M.S. and Charkravarti, A. (1997). Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580-1581.

Cotton, R.G. (1997). Slowly but surely towards better scanning for mutations. *Trends in Genetics* **13**, 43-46.

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

de Alba, J. (1978). Progress in the selection of the Latin American diary Criollo. *World Animal Review No. 28.* FAO, Rome, Italy.

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

Disotell, T.R. (1999), Human evolution: Sex-specific contributions to genome variation. *Current Biology* **9**, 29-31.

Edwards, C.J., Dolf, G., Looft, C., Loftus, R.T. and 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.

Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551-558.

Ellegren, H. (2002). Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics* **24**, 400-402.

Epstein, H. (1971). *The origin of the domestic cattle of Africa.* Africana, New York, U.S.A.

Epstein, H. and Mason, I.L. (1984). Cattle. In: *Evolution of Domesticated Animals.* **I.L. Mason** (Editor). Pp. 6-27. Longman, London, U.K.

Excoffier, L., Smouse, P.E. and 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.

Excoffier, L. and Schneider, S. (1999). Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proceedings of the National Academy of Sciences USA* **96**, 10597-10602.

Excoffier, L. and Schneider, S. (2000). The demography of human populations inferred from patterns of mitochondrial DNA diversity. In: *Archaeogenetics: DNA and the population prehistory of Europe.* **Renfrew, C. & Boyle, K.** (Editors). Pp 315-320. McDonald Institute Publications.

Excoffier, L (2001). Analysis of population subdivision. In: *Handbook of statistical genetics.* (**Balding, D.J., Bishop, M. and Cannings, C.** (Editors). Pp.271-307. Wiley, New York, U.S.A.

Eyre-Walker, A., Smith, N.H. and Smith, J.M. (1999). How clonal are human mitochondria? Proceeding of the Royal Society London B Biological Sciences **266**, 477-483

Eyre-Walker, A. and Awadalla, P. (2001). Does human mtDNA recombine? *Journal of Molecular Evolution* **53**, 430-435.

Felius, M. (1995). *Cattle breeds – an Encyclopedia.* Misset, Doetinchem, the Netherlands.

Felsenstein, J. PHYLIP VERSION 3.5 *computer package.* Washington University, Seattle, U.S.A.

Fries, R. and Durstewitz, G. (2001). Digital DNA signatures for animal tagging. *Nature Biotechnology* **19**, 508.

Fu, Y.X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-25.

Galton, F. (1865). The first steps towards the domestication of animals. *Transaction of the ethnological society of London* NS **3**, 122-138.

Garrard, A., Colledge, S., Martin, L. (1996). The emergence of crop cultivation and caprine herding in the "Marginal Zone" of the southern Levant. In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia.* **Harris, D.R.** (Editor). Pp. 204-227. University College London, London, U.K.

Gatesy, J., Yelon, D., DeSalle, R. and Vrba, E.S. (1992). Phylogeny of the *Bovidae* (*Artodactyla, Mammalia*), based on mitochondrial ribosomal DNA sequences. *Molecular Biology and Evolution* **9**, 433-446.

Gatesy, J., Amato, G., Vrba, E., Schaller, G. and DeSalle, R. (1997). A cladistic analysis of mitochondrial ribosomal DNA from the *Bovidae*. *Molecular Phylogenetics and Evolution* **7**, 303-319.

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

Gibbons, A. (1998). Calibrating the mitochondrial clock. *Science* **279**, 28-29.

Giovambattista, G., Ripoli, M.V., De Luca, J.C., Mirol, P.M., Liron, J.P. and Dulout, F.N. (2000). Male-mediated introgression of *Bos indicus* genes into Argentine and Bolivian Creole cattle breeds. *Animal Genetics* **31**, 302-305.

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

Gray, M.W. (1989). Origin and evolution of mitochondrial DNA. *Annual Review of Cell Biology* **5**, 25-50.

Gray, I.C., Campbell, D.A. and Spurr, N.K. (2000). Single nucleotide polymorphisms as tools in human genetics. *Human Molecular Genetics* **9**, 2403-2408.

Grompe, M. (1993). The rapid detection of unknown mutations in nucleic acids. *Nature Genetics* 5, 111-117.

Grigson, C. (1978). The craniology and relationships of four species of *Bos.* 4. The relationship between *Bos primigenius* Boj. and *Bos taurus* L. and its implications for the phylogeny of the domestic breeds. *Journal of Archaeological Sciences* **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. (1985). *Bos indicus* and *Bos namadicus* and the problem of autochthonous domestication in India. In: *Recent advances in Indo-Pacific prehistory.* **Misra, V.N. and Bellwood, P.** (Editors). Pp. 425-428. New Delhi: Oxford and IBH.

Gyllensten, U., Wharton, D., Josefsson, A. and Wilson, A.C. (1991). Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**, 255-257.

Hanotte, O., Tawah, C.L., Bradley, D.G., Okomo, M., Verjee, Y., Ochieng, J. and 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.

Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y., Hill, E.W. and Rege, J.E. (2002). African pastoralism: genetic imprints of origins and migrations. *Science* **296**, 336-339.

Harpending, C., Sherry, S.T., Rogers, A.R. and Stoneking, M. (1993). The genetic structure of ancient human populations. *Current Anthropology* **34**, 483-496.

Harpending, H. and Rogers A. (2000). Genetic perspectives on Human Origins. *Annual Review of Genomics and Human Genetics* **1**, 361-385.

Harris, D.R. (1996). Introduction: themes and concepts in the study of early agriculture. In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia.* **Harris, D.R.** (Editor). Pp. 1-9. University College London, London, U.K.

Hayashi, K. (1991). PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods and Applications* **1**, 34-38.

Heaton, M.P., Grosse, W.M., Kappes, S.M., Keele, J.W., Chitko-McKown, C.G., Cundiff, L.V., Braun, A., Little, D.P. and Laegreid, W.W. (2001). Estimation of DNA sequence diversity in bovine cytokine genes. *Mammalian Genome* **12**, 32-37.

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

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

Hillman, G. (1996). Late Pleistocene changes in wild plant-foods available to hunter-gatherers of the northern Fertile Crescent: possible preludes to cereal cultivation. In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia.* **Harris, D.R.** (Editor). Pp. 159-203. University College London, London, U.K.

Howell, N., Kubacka, I. And Mackey D.A. (1996). How fast does the human mitochondrial genome evolve? *American Journal of Human Genetics* **59**, 501-509.

Horwitz, L.K., Tchernov, E., Ducos, P., Becker, C., von den Driesch, A., Martin, L. and Garrard, A. (1999). Animal domestication in the Southern Levant. *Paléorient* **25**, 63-80.

Hudson, R.R., Boos, D.D. and Kaplan, N. (1992). A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* **9**, 138-151.

Isaac, E. (1962). On the domestication of cattle. *Science* **137**, 195-204.

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

Jarne, P. and Lagoda, P.J.L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**, 424-429.

Jarrige, J.-F. and Meadow, R.H. (1980). The Antecedents of civilisation in the Indus Valley. *Scientific American* **243**, 122-133.

Jobling, M.A. (2001). Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Science International* **118**, 58-62.

Jogelkar, P.P and Thomas, P.K. (1992). Ancestry of *Bos* species: myth and reality III. Biostatistical considerations. *Man and Environment* **16**, 109-114.

Jorde, L.B., Bamshad, M.J., Watkins, W.S., Zenger, R., Fraley, A.E., Krakowiak, P.A., Carpenter, K.D., Soodyall, H., Jenkins, T. and Rogers, A.R. (1995). Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *American Journal of Human Genetics* **57**, 523-538.

Jorde, L.B., Bamshad, M. and Rogers, A.R. (1998). Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* **20**, 126-136.

Jorde, L.B. and Bamshad, M. (2000). Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**, 1931.

Jorde, L.B., Watkins, W.S., Bamshad, M.J., Dixon, M.E., Ricker, C.E., Seielstad, M.T. and Batzer, M.A. (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *American Journal of Human Genetics* **66**, 979-988.

Jorde, L.B., Watkins, W.S. and Bamshad, M.J. (2001). Population genomics: a bridge from evolutionary history to genetic medicine. *Human Molecular Genetics* **10**, 2199-2207.

Joshi, N.R. & Phillips, R.W. (1953). *Zebu cattle of India and Pakistan.* FAO Agricultural Series, Number 19. FAO, Rome, Italy.

Kaessman, H. and Pääbo, S. (2002). The genetical history of humans and the great apes. *Journal of Internal Medicine* **251**, 1-18.

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

Kim, K.E., Gordon, D.F. and Maurer, R.A. (1988). Nucleotide sequence of the bovine gene for follicle-stimulating hormone beta-subunit. *DNA* **7**, 227-233.

Kivisild, T. and Villems, R. (2000). Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**, 1931.

Kruglyak, L. (1997). The use of a genetic map of biallelic markers in linkage studies. *Nature Genetics* **17**, 21-24.

Kruglyak, L. and Nickerson, D.A. (2001). Variation is the spice of life. *Nature Genetics* **27**, 234-236.

Kruskal, J. (1964). Multidimensional scaling by optimising goodness of fit to a non-metric hypothesis. *Psychometrika* **29**, 1-27.

Kumar, S., Hedrick, P., Dowling, T. and Stoneking, M. (2000). Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**, 1931.

Kumar, P., Freeman, A.R., Loftus, R.T., Gaillard, C., Fuller, D.Q. and Bradley, D.G. (2002). Admixture analysis of South Asian cattle. *In press.*

Kwok, P.-Y. (2001). Methods for genotyping single nucleotide polymorphisms. *Annual Review of Genomics and Human Genetics* **2**, 235-258.

Lagziel, A., Lipkin, E. and Soller, M.(1996). Association between SSCP haplotypes at the bovine growth hormone gene and milk protein percentage. *Genetics* **142**, 945-951

Lander, E.S. (1999). Array of hope. *Nature Genetics* **21**, 3-4.

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

Legge, A. (1996). The beginning of caprine domestication in Southwest Asia. In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. **Harris, DR** (Editor). Pp. 238-262. University College London, London, U.K.

Lenstra, J.A. and Bradley, D.G. (1999). Systematics and phylogeny of cattle. In: *The genetics of cattle.* **Fries, R. and Ruvinsky, A.** (Editors). Pp. 1-14. CAB International.

Lev-Yadun, S., Gopher, A. and Abbo, S. (2000). The Cradle of Agriculture. *Science* **288**, 1602-1603.

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

Loftus, R.T., MacHugh, D.E., Ngere, L.O., Balain, D.S., Badi, A.M., Bradley, D.G. and 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. and Bradley, D.G. (1999). A microsatellite survey of cattle from a centre of origin: the Near East. *Molecular Ecology* **8**, 2015-2022.

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

Macaulay, V.A., Richards, M.B., Forster, P., Bendall, K.E., Watson, E., Sykes, B. and Bandelt, H.J. (1997). MtDNA mutation rates--no need to panic. *American Journal of Human Genetics* **61**, 983-990.

Macaulay, V. (1998). CRED: credible regions for coalescence times. Computer program. University of Oxford, Oxford, U.K.

Macaulay, V., Richards, M. and Sykes, B. (1999). Mitochondrial DNA recombination-no need to panic. *Proceedings of the Royal Society London B Biological Sciences* **266**, 2037-2039.

MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, P. and 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. and 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., Eythorsdottir, E. and 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 London B Biological Sciences* **354**, 99-108.

MacHugh, D.E., Edwards, C.J., Bailey, J.F., Bancroft, D.R. and 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.

MacHugh, D.E. and Bradley, D.G. (2001). Livestock genetic origins: goats buck the trend. *Proceedings of the National Academy of Sciences USA* **98**, 5382-5384.

Maillard, J.C., Kemp, S.J., Naves, M., Palin, C., Demangel, C., Accipe, A., Maillard, N. and Bensaid, A. (1993). *Revue Élev Méd vét Pays trop* **46**, 283-290.

Maillard, J.C. and Maillard, N. (1998). Histoire du peuplement bovin et de l'introduction de la tique Amblyomma variegatum dans les Iles Françaises des Antilles: synthèse bibliographique. *Ethnozootechnie* **61**, 19-36.

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

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

Marth, G., Yeh, R., Minton, M., Donaldson, R., Li, Q., Duan, S., Davenport, R., Miller, R.D. and Kwok, P.Y. (2001). Single nucleotide polymorphisms in the public domain: how useful are they? *Nature Genetics* **27**, 371-372.

Mason, I.L (1988). *A World dictionary of livestock breeds*. CAB International.

Matthee, C.A. and Davis, S.K. (2001). Molecular insights into the evolution of the family *Bovidae*: a nuclear DNA perspective. *Molecular Biology and Evolution* **18**, 1220-1230.

McCorriston, J and Hole, F. (1991). The ecology of seasonal stress and the origins of agriculture in the Near East. *American Anthropologist* **93**, 46-69.

Mead, R., Curnow, R.N. and Hasted, A.M. (1993). Statistical methods in agriculture and experimental biology (2nd edition). Chapman & Hall, London.

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*. **Clutton-Brock, J.G. & Grigson, C.** (Editors). Pp. 309-337. Oxford University Press, Oxford, U.K.

Meadow, R.H. (1989). Osteological evidence for the process of animal domestication. In: *The walking larder: Patterns of Domestication, Pastoralism and Predation*. **Clutton-Brock, J.** (Editor). Pp. 80-90. Unwin Hyman, London, U.K.

Meadow, R.H. (1993). Animal domestication in the Middle East: A revised view from the Eastern Margin (2nd edition). In: *Harappan Civilisation*. **Possehl, G.** (Editor). Pp. 295-320. Oxford & IBH, New Delhi, India.

Meadow, R.H.: The origins and spread of pastoralism in northwestern South Asia. In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. **Harris, D.R.** (Editor). Pp. 390-412. University College London, London, U.K.

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

Mir, K.U. and Southern, E.M. (2000). Sequence variation in genes and genomic DNA: methods for large-scale analysis. *Annual Review Genomics and Human Genetics* **1**, 329-360.

Misra, V.N. (2001). Prehistoric human colonization of India. *Journal of Biosciences Supplemental* **26**, 491-531.

Moore, A.M.T. and 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.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbour Symposium of Quantitative Biology* **51**, 263-273.

Murphy, G. and Kavanagh, T. (1988). Speeding-up the sequencing of double-stranded DNA. *Nucleic Acids Research* **16**, 5198.

Naik, S. (1978). Origin and domestication of zebu cattle (*Bos indicus*). *Journal of Human Evolution* 7, 23-30.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences USA* **70**, 3321-3323.

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

Nersting, L.G. and Arctander, P. (2001). Phylogeography and conservation of impala and greater kudu. *Molecular Ecology* **10**,711-719.

Nickerson, D.A., Kaiser, R., Lappin, S., Stewart, J., Hood, L. and Landegren, U. (1990). Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. *Proceedings of the National Academy of Sciences USA* **87**, 8923-8927.

Nowak, R.M. (1991). In : *Walker's mammals of the world.* John Hopkins University Press, Baltimore, U.S.A.

Nyakaana S., Arctander P. and Siegismund H.R. (2002). Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci. *Heredity* **89**, 90-98.

Orita, M., Suzuki, Y, Sekiya, T. and Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874-879.

Orti, G., Hare, M.P. and Avise, J.C. (1997). Detection and isolation of nuclear haplotypes by PCR-SSCP. *Molecular Ecology* **6**, 575-580.

Ota, T. (1993). DISPAN software package. Pennsylvania State University, Pennsylvania, U.S.A.

Owen, F. and Jones, R. (1982). Statistics (2nd edition). Longman. London, UK.

Palumbi, S.R. and Baker, C.S. (1994). Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution* **11**, 426-435.

Park, S. (2002). Trypanotolerance in West African cattle and the population genetic effects of selection. Ph.D. thesis. Department of Genetics, University of Dublin, Trinity College, Ireland.

Parsons, T.J. and Irwin, J.A. (2000). Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**, 1931.

Payne, W. J. A. (1991). Domestication: a step forward in civilisation. In: *Cattle genetic resources* (1st edition). **Hickman, C.G.** (Editor). Pp. 51-72. World Animal Science, Volume B. Elsevier, Amsterdam.

Payne, W. J. A. (1999). *An Introduction to animal husbandry in the tropics*. Blackwell Science, U.K.

Penny, D., Steel, M., Waddell, P.J. and Hendy, M.D. (1995). Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*. *Molecular Biology and Evolution* **12**, 863-882.

Perkins, D.J. (1969). Fauna of Çatal Hüyük: Evidence for early cattle domestication in Anatolia. *Science* **164**, 177-178.

Porter, V. (1991). *Cattle: a handbook to the breeds of the world.* Christopher Helm, London, U.K.

Primo, A.T. (1992). The Iberic cattle in the Americas: 500 years later. *Arch Zootec* (extra) **41**, 421-432.

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

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

Przeworski, M., Hudson, R.R. and Di Rienzo, A. (2000). Adjusting the focus on human variation. *Trends in Genetics* **16**, 296-302.

Reed, C.A. (1977). *Origins of agriculture.* Monton, The Hague (Netherlands) and Paris (France).

Richards, M. and Macaulay, V. (2001). The mitochondrial gene tree comes of age. *American Journal of Human Genetics* **68**, 1315-1320.

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

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.

Rogers, A.R. and Jorde, L.B. (1996). Ascertainment bias in estimates of average heterozygosity. *American Journal of Human Genetics* **58**, 1033-1041.

Romualdi, C., Balding, D., Nasidze, I.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M.A. and Barbujani, G. (2002). Patterns of human diversity, within and among continents, inferred from bi-allelic DNA polymorphisms. *Genome Research* **12**, 602-612.

Rosenbaum, H.C. and Deinard, A.S. (1998). Caution before claim: an overview of microsatellite analysis in ecology and evolutionary biology. In: *Molecular approaches to ecology and evolution.* **DeSalle, R. and Schierwater B.** (Editors). Pps. 87-106. Birkhäuser, *Bos*ton/Berlin.

Rouse, J.E. (1973). *World cattle, III: Cattle of North America.* University of Oklahoma Press, Norman, Oklahoma, U.S.A.

Ruppert, S., Scherer, G. and Schutz, G. (1984). Recent gene conversion involving bovine vasopressin and oxytocin precursor genes suggested by nucleotide sequence *Nature* **308**, 554-557.

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

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning: A laboratory manual*. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York, U.S.A.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **74**, 5463-5467.

Schafer, A.J. and Hawkins, J.R. (1998). DNA variation and the future of human genetics. *Nature Biotechnology* **16**, 33-39.

Schlötterer, C. and Pemberton, J. (1998). The use of microsatellites for genetic analysis of natural populations-a critical review. In: *Molecular Approaches to Ecology and Evolution.* **DeSalle, R. and Schierwater, B.** (Editors). Pp. 71-86. Birkhäuser, *Bos*ton/Berlin.

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

Schneider, S., Roessli, D. and Excoffier, L. (1999). ARLEQUIN, VERSION 2.0: *A software for Population Genetic Data Analysis*. Department of Anthropology, University of Geneva, Switzerland.

Seyfert, H.M., Tuckoricz, A., Interthal, H., Koczan, D. and Hobom, G. (1994). Structure of the bovine lactoferrin-encoding gene and its promoter. *Gene* **143**, 265-269.

Sherratt, A. (1996). Plate tectonics and imaginary prehistories: structure and contingency in agricultural origins. In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. **Harris, D.R.** (Editor). Pp. 130-140. University College London, London, U.K.

Sherry, S.T., Rogers, A.R., Harpending, H., Soodyall, H., Jenkins, T. and Stoneking, M. (1994). Mismatch distributions of mtDNA reveal recent human population expansions. *Human Biology* **66**, 761-775.

Simonsen, B.T., Siegismund, H.R. and Arctander, P. (1998). Population structure of African buffalo inferred from mtDNA sequences and microsatellite loci: high variation but low differentiation. *Molecular Ecology* **7**, 225-237.

Slatkin, M. (1991). Inbreeding coefficients and coalescence times. *Genetical Research* **58**, 167-175.

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, U.S.A.

Steinhoff,U.M., Senft,B. and Seyfert,H.M. (1994). Lysozyme-encoding bovine cDNAs from neutrophile granulocytes and mammary gland are derived from a different gene than stomach lysozymes. *Gene* **143**, 271-276.

Stoneking, M. and Wilson, A.C. (1989). Mitochondrial DNA. In: The colonisation of the Pacific: a genetic trail. **Hill A.V.S. and Serjeanston, S.** (Editors). Pp. 215-245. Oxford University Press, U.K.

Stoneking, M. (1996). Mitochondrial DNA variation and human evolution. In: *Human Genome Evolution*. **Jackson, M., Strachan, T. and Dover, G.** (Editors). BIOS Scientific, Oxford, U.K.

Stoneking, M. and Soodyall, H. (1996). Human evolution and the mitochondrial genome. *Current Opinions in Genetics and Development* **6**, 731-736.

Stoneking, M. (2000). Hypervariable sites in the mtDNA control region are mutational hotspots. *American Journal of Human Genetics* **67**, 1029-1032.

Stoneking, M. (2001). Single nucleotide polymorphisms. From the evolutionary past. *Nature* **409**, 821-822.

Sunden, S.L., Stone, R.T., Bishop, M.D., Kappes, S.M., Keele, J.W. and Beattie CW. (1993). A highly polymorphic bovine microsatellite locus: *BM2113*. *Animal Genetics* **24**, 6.

Sunnucks, P., Wilson, A.C.C., Beheregaray, L.B., Zenger, K., French, J. and Taylor, A.C. (2000). SSCP is not so difficult: the application and utility of single-strand conformation polymorphism in evolutionary biology and molecular ecology. *Molecular Ecology* **9**, 1699-1710.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437-460.

Tanaka, K., Solis, C.D., Masangkay, J.S., Maeda, K., Kawamoto, Y. and 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.

Tautz, D. and Schlötterer, C. (1994). Simple sequences. *Current Opinions in genetics and Development* **4**, 832-837.

Taylor, P.L.T. (1990-1993). GENEJOCKEY II sequence processor software package.

Templeton, A.R. (1998). Human races: A genetic and evolutionary perspective. *American Anthropologist* **100**, 632-650.

Teres, V., Lester, D.H. and Williams, J.L. (1996). Mapping of bovine fibroblast growth factor receptor 3 (FGFR3) to the telomeric region of chromosome 6 by SSCP analysis. *Animal Genetics* **27**, 371.

The International SNP Map Working Group (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-933.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The CLUSTALX windows inference: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876-4882.

Thyagarajan, B., Padua, R.A. and Campbell, C. (1996). Mammalian mitochondrial possess homologous DNA recombination activity. *Journal of Biological Chemistry* **271**, 27536-27543.

Tobe, V.O., Taylor, S.L. and Nickerson, D.A. (1996). Single-well genotyping of diallelic sequence variations by a two-color ELISA-based oligonucleotide ligation assay. *Nucleic Acids Research* **24**, 3728-3732

Troy, C.S. (1998). *Mitochondrial DNA phylogeny and biogeography of domesticated cattle*. Ph.D thesis. Department of Genetics, University of Dublin, Trinity College, Ireland.

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

Uerpmann, H.-P. (1996). Animal domestication - accident or intention? In: Harris, D.R. (Ed.), In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. **Harris, D.R.** (Editor). Pp. 227-237. University College London, London, U.K.

Van Hooft, W.F., Groen, A.F. and Prins, H.H. (2002). Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: Pleistocene origin and population expansion of the Cape buffalo subspecies. *Molecular Ecology* **11**, 267-279.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. and Wilson, A.C. (1991). African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503-1507.

Vignal, A., Milan, D., SanCristobal, M. and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* **34**, 275-305.

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

Vishnu-Mittre (1978). Origins and history of agriculture in the Indian subcontinent. *Journal of Human Evolution* **7**, 31-36.

Wallace, D.C. (1995). 1994 William Allan Award Address. Mitochondrial DNA variation in human evolution, degenerative disease, and ageing. *American Journal of Human Genetics* **57**, 201-223.

Wang, D.G., Fan, J-B, Siao, C-J, Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., *et al.* (1998). Large scale identification, mapping and genotyping of single nucleotide polymorphisms in the human genome. *Science* **280**, 1077-1082.

Warren, W., Hovig, E., Smith-Sørensen, B., Børresen, A.-L., Fujimura, F.K., Liu, Q., Feng, J. and Sommer, S.S. (1997). Detection of mutations by single-strand conformation polymorphism (SSCP) analysis and SSCP-hybrid methods. In: *Current Protocols in Human Genetics* **Section 7.4.1-7.4.23**. Wiley, New York, U.S.A.

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

Watterson, G. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256-276.

Weir, B.S. and Cockerham, C.C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Wendorf, F. and 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*. Oxford University Press, Oxford, U.K.

Wilkins, J.V. (1984). Criollo cattle of the Americas. *Animal Genetic Resources Information No. 1*. FAO, Rome, Italy.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M.Jr., Gyllensten. U.B., Helm-Bychowski, K.M., Higuchi, R.G., Palumbi, S.R., Prager, E.M., Sage, R.D. and Stoneking, M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society* **26**, 375-400.

Wright, S. (1965). The interpretation of population structure by *F*-statistics with special regard to system of mating. *Evolution* **19**, 395-420.

Yu, N., Chen, F.C., Ota, S., Jorde, L.B., Pamilo, P., Patthy, L., Ramsay, M., Jenkins, T., Shyue, S.K. and Li, W.H. (2002). Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**, 269-274.

Zeuner, F.E. (1963). A history of domesticated animals. Hutchison, London, U.K.

Zietkiewicz, E., Yotova, V., Jarnik, M., Korab-Laskowska, M., Kidd, K.K., Modiano, D., Scozzari, R., Stoneking, M., Tishkoff, S., Batzer, M. and Labuda, D. (1997). Nuclear DNA diversity in worldwide distributed human populations. *Gene* **205**, 161-171.

Zohary, D. (1996). The mode of domestication of founder crops of Southwest Asian agriculture. In: *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. **Harris, D.R.** (Editor). Pp. 142-158. University College London, London, U.K.

APPENDIX 1

SINGLE NUCLEOTIDE POLYMORPHISM DISCOVERY IN THE BOVINE GENOME

INTRODUCTION

Research aims

The research presented in this Appendix describes a preliminary study of single nucleotide polymorphism (SNP) discovery with the bovine genome using PCR-SSCP analysis (polymerase chain reaction-single strand conformation polymorphism) in conjunction with automated DNA sequencing. PCR-SSCP offers a relative simple, inexpensive and sensitive method for investigating nucleotide variability between homologous DNA fragments, therefore greatly reducing the amount of sequencing necessary for SNP detection.

Single strand conformation polymorphism analysis

Single strand conformation polymorphism (PCR-SSCP) analysis detects genetic variation by taking advantage of the fact that single-nucleotide changes in DNA sequences alter the electrophoretic mobility of single stranded DNA in non-denaturing gels (Orita *et al.* 1989). This technique has been extensive applied in various fields of biomolecular research and when coupled with direct DNA sequencing offers an efficient and inexpensive method for the isolation of candidate SNP loci.

The traditional method of PCR-SSCP analysis entails the PCR amplification of regions of DNA, which are subsequently heat-treated in a denaturing buffer to render them single-stranded. The separated strands are then fractionated through a non-denaturing polyacrylamide gel matrix, followed by visualisation. Under optimal conditions, DNA strands adopt structures that migrate according to their morphology. DNA strands of different sequence rarely assume identical shapes and therefore have distinct gel mobilities allowing variant sequences to be identified (Sambrook *et al.* 1989; Warren *et al.* 1997; Mir & Southern 2000; Sunnucks *et al.* 2000).

In general, the sensitivity of PCR-SSCP is inversely proportional to the size of the DNA fragment under investigation. It is estimated that single-nucleotide differences are resolved 99% of the time when fragments of 100-300 bp are analysed and this decreases to 80% when the fragment size is 400 bp, however, successful PCR-SSCP analysis has been performed using fragments as large as 775 bp (Ortí *et al.* 1997). The degree of sensitivity in detecting sequence variation using PCR-SSCP may also vary depending on the experimental conditions employed. In certain cases where the detection of a given variant or all variants in a DNA fragment are required, strand separation using alterations in experimental conditions may be necessary. This can be achieved by changing the temperature, gel and buffer composition and the electrophoretic running time (Hayashi 1991).

PCR-SSCP amplicon selection and primer design

Twelve non-coding DNA regions from six cattle gene sequences deposited within the GENBANK database were selected for PCR-SSCP analysis (**Table 1**). Non-coding gene sequences (*i.e.* introns and 3' untranslated regions) were preferred as they are assumed to be less affected by natural selection and therefore expected to harbour more sequence variants compared with coding sequences (Yu *et al.* 2002).

Oligonucleotide primers were designed to amplify non-coding regions ranging between 378 bp – 550 bp within the selected gene sequences. Oligonucleotide primer pairs (**Table 2**) were designed by eye using the criteria outlined by MacHugh *et al.* (2000). The criteria for oligonucleotide primer design are summarised below:

- Oligonucleotide primer should be 18-25 nucleotides in length.
- Repetitive nucleotide stretches should be avoided.
- GC content should be between 40-60%.
- There should be no 3' complimentarity between primer pairs.
- Two cytosine or guanine resides (referred to as a 'GC clamp') should be present at the 3' end of each primer, where possible.
- Opportunity for internal secondary structure formation should be kept to a minimum.

All primers were produced and supplied by OSWEL SCIENTIFIC LTD in 1 ml of sterile ddH_2O. Upon arrival to the laboratory, stock primers were prepared in 250 μls aliquots and stored at -20°C.

Table 1: The amplicons designed for PCR-SSCP analysis

Amplicon name	Gene name	Chromosomal location	GENBANK accession	GENBANK submission reference	Amplicon location/ nucleotide posn.	Amplicon length (bp)
AGO-INT02a	Agouti	n/a	X99691	Oulmouden *et al.* (1996, unpubl.)	Intron 2; 123 - 614	492
AGO-INT02b	Agouti	n/a	X99691	Oulmouden *et al.* (1996, unpubl.)	Intron 2; 928 - 1420	493
AGO-INT03	Agouti	n/a	X99691	Oulmouden *et al.* (1996, unpubl.)	Intron 3; 3453 - 3999	547
FSHB-INT01	Follicle stimulating hormone beta-subunit	BTA 15	M83753	Kim *et al.* (1988)	Intron 1; 2048 - 2520	473
FSHB-INT02	Follicle stimulating hormone beta-subunit	BTA 15	M83753	Kim *et al.*(1988)	Intron 2; 2821 – 3370	550
LTF-INT07	Lactoferrin	BTA 22	Z93399	Seyfert *et al.* (1994).	Intron 7; 157 – 593	437
LTF-INT08	Lactoferrin	BTA 22	Z93399	Seyfert *et al.* (1994).	Intron 8; 954 – 1389	436
LTF-INT09	Lactoferrin	BTA 22	Z93399	Seyfert *et al.* (1994).	Intron 9; 2578 – 3065	488
LTF-INT011	Lactoferrin	BTA 22	Z93399	Seyfert *et al.* (1994).	Intron 11; 5915 – 6415	501
LYZ-3' UTR	Lysozyme	BTA 5	L19980	Steinhoff *et al.* (1994)	3' UTR; 965 – 1449	485
OXT-INT01	Pre-pro-oxytocin-neurophysin 1	BTA 15	X00502	Ruppert *et al.* (1984)	Intron 1; 296 – 674	378
PRC-INT02	Prostacyclin receptor	n/a	Z93039	Hasse & Schroer (1997, unpubl.)	Intron 2; 781 – 1193	413

Table 1. All chromosomal positions were obtained from the Roslin Ark database (www.thearkdb.org). Nucleotide position refers to the site of the PCR amplicon within the appropriate gene DNA sequence according to the GENBANK entry. n/a denotes chromosome position not assigned. Amplicon lengths includes primer sequences.

Table 2: PCR-SSCP oligonucleotide primer information

Amplicon name	GENBANK accession no.	Amplification primer sequences (5'-3')	Primer position
AGO-INT02a	X99691	CAT GAA CCT GTT GGA TTT CCC GAC AAA TTT TCC AAT GCT AGG C	123 614
AGO-INT02b	X99691	AAG AGG ATA TTC TAA ATA TAT GGG AAG G GAA GTG CAA AGA GCA GAG AGG	928 1420
AGO-INT03	X99691	AGT CGA TGA TGC CAT TCA GC ACG AAG ATC ATG GCA ACT GG	3443 3999
FSHB-INT01	M83753	ACA CTG TAA TCA ATG TTG CAA GC AAT CGT TTA TAT ATT CTG ACC AAA CC	2048 2520
FSHB-INT02	M83753	AAG TTT CCC AAC GAA TAG CC CCT GGG ATT AAT TCA AAG AAG C	2821 3370
LTF-INT07	Z93399	GCC AAG GAA GAC TTG ATC TGG CCA GGT TGA AAC AGA GTA TCT GC	157 593
LTF-INT08	Z93399	CCC TTG GTC ACT TTT CCA GCC CCA CAC GAC CCT GGT GTA CC	954 1389
LTF-INT09	Z93399	CCT TGG TTT GAT CAC TGT TTG G GTT CTT TTA CCC CGG ACT CG	2578 3065
LTF-INT011	Z93399	CAG AGA TGC AGC CAT AGA AGC CAC AGA GTG CAA CCT AGG CC	5915 6415
LYZ-3'UTR	L19980	TGA AGT GAC TTA GCA GCA GTA GC GCC TCT GAA TTT GTC ATA ATG C	965 1449
OXT-INT01	X00502	CGC CTG CTA CAT TCA GAA CTG C ACA CTG GTG GGA GGG AAG GG	296 674
PRC-INT02	Z93039	TTG ACC ACT TGA TTC TGC TGG AGG GAT TTC TAA ACC ACT GAG C	781 1193

The primer position given in the table refers to the first nucleotide residue of the primer as per the GENBANK entry. Sequences in red denote the forward primer sequences; sequences in blue denote reverse primer sequences.

Optimisation of conditions for PCR-SSCP analysis

Prior to PCR-SSCP analysis, optimal PCR amplification conditions for each amplicon were established using a standard PCR protocol (**Table 3**). Once the ideal amplification parameters had been established for each primer pair, radiolabelled amplifications were performed using a panel of samples consisting of representatives from a number of

domestic *Bos taurus* and *Bos indicus* populations from Europe, Africa, Near East and South Asia (**Table 4**).

Table 3: PCR amplification optimisation of each PCR-SSCP amplicon

PCR Component	Volume	Final concentration
10X Platinum *Taq* PCR Buffer (GIBCO)	1.0 μl	1X
PCR Primers	1.0 μl each primer	0.3 mM each primer
dNTPs	1.0 μl	0.2 mM of each dNTP
$MgCl_2$	1.0 μl	1.0-2.5 mM
Platinum *Taq* Polymerase	0.1 μls	0.5 Units
DNA template	1 μl	~25 ngs
ddH_2O	4.9 μls	
Total	10 μls	

Reaction volumes are given for 1 PCR amplification. A mastermix of all PCR components, excluding the DNA template, for the appropriate number of samples was prepared. 9 μls aliquots were added using a repeating multi-pipette (GILSON) to 1 μl of each DNA sample contained on a 96-well polycarbonate microtitre plate. A mineral overlay of 10 μls was added to each well to prevent evaporation during the amplification reactions. Optimal $MgCl_2$ concentrations were established by increasing the concentration in increments of 2.5 mM. Optimal annealing temperatures for PCR amplifications were determined by increasing the temperature in increments of 2°C. PCR cycling parameters consisted of an initial denaturation step at 94°C for 3 minutes followed by 35 cycles of 94°C for 40 seconds, 52-56°C for 40 seconds and 72°C for 1 minute, and a final extension at 72°C for 4 minutes. PCR amplifications were performed on an OmniGene thermal cycler (HYBAID).

Table 4: The cattle samples used for PCR-SSCP analysis

Population name	Geographic origin	Regional group	No. samples
Hereford (HE)	England	European *Bos taurus*	3
Jersey (JE)	England		2
Friesian (FR)	Holland		3
Simmental (SI)	Switzerland		2
Total			**10**
Anatolian Black (AB)	Anatolia	Near Eastern *Bos taurus*	**10**
Guinean N'Dama (ND)	Guinea	African *Bos taurus*	**5**
Sahiwal (SA)	Indo-Pakistan border	Indian *Bos indicus*	3
Hariana (HA)	Northwest India		2
Tharparker (TH)	Northwest India		2
Nelore (NE)	India		2
Ongole (ON)	South India		2
Total			**11**
Overall Total			**36**

All samples were procured as DNA previously extracted by various members of the Molecular Population Genetics laboratory (TCD) prior to the initiation of this project. Population codes are given in brackets.

PCR-SSCP experimental conditions – PCR amplification

The experimental conditions used for PCR-SSCP analysis were modified from the protocols published by Lagziel *et al.* (1996) and Teres *et al.* (1996). Radiolabelled PCR amplifications involved the random incorporation of ^{32}P-dCTP into the newly generated amplification products. All radiolabelled PCR amplifications were performed in 96-well microtitre plates using a OmniGene DNA thermal cycler (HYBAID). All PCR amplifications included negative controls. PCR reaction components and cycling parameters are outlined below in **Table 5**. $MgCl_2$ concentrations and annealing temperatures for each PCR-SSCP amplicon are given in **Table 6**.

Table 5: PCR-SSCP reagents and cycling parameters

PCR Component	Volume	Final concentration
10X Platinum *Taq* Buffer	0.55 μls	1X
PCR Primers	0.55 μls each primer	0.3 mM each primer
dNTPs	0.55 μls	0.2 mM dATP, dTTP, dGTP / 10 μM dCTP
$MgCl_2$	0.55 μls	1.25-1.5 mM
^{32}P-dCTP (3,000 Ci/μl)	0.015 μls	
Platinum *Taq* Polymerase	0.055 μls	0.5 Units
DNA template	1 μl	~25 ngs
ddH_2O	1.68 μls	
Total	5.5 μls	

A mastermix of all PCR components, excluding the DNA template, for the appropriate number of samples was prepared. 4.5 μls aliquots were added using a repeating multi-pipette (GILSON) to 1 μl of each DNA sample contained on a 96-well polycarbonate microtitre plate. A mineral overlay of 10 μls was added to each well, to prevent evaporation during the amplification reactions. PCR cycling parameters consisted of an initial denaturation step at 94°C for 3 minutes followed by 35 cycles of 94°C for 40 seconds, 52-56°C for 40 seconds and 72°C for 1 minute, and a final extension at 72°C for 4 minutes.

PCR-SSCP experimental conditions-gel electrophoresis

Upon completion of PCR amplification, 10 μls of formamide-based loading buffer was added to each of the PCR products, which were subsequently heat-denatured at 94°C for 3 minutes and immediately placed on ice. 3 μls aliquots of each product was loaded on a 5% AccuGel™ non-denaturing polyacrylamide gel (NATIONAL DIAGNOSTIC; **solution Q, Appendix 3**) and electrophoresed using a vertical 96 well S1BRL sequencing gel electrophoresis apparatus. Gels were run using standard 1X TBE buffer.

Electrophoresis was performed at both room temperature and at 4°C in order to detect all possible DNA sequence variants within each amplicon, however most PCR-SSCP amplicons produced best results at 4°C. The optimal electrophoretic conditions for each PCR-SSCP amplicon are detailed in **Table 6**.

Once electrophoresis was complete, gels were dried, placed in a cassette with autoradiographic film and developed after 24-72 hours of exposure. Autoradiographs were analysed for DNA mobility shifts. PCR products displaying sequence variation were categorised and selected for automated sequencing. Replicate individuals were amplified using a non-radiolabelled standard PCR protocol (**Table 3**). Successful amplifications were confirmed by agarose gel electrophoresis (**Section 5.3.2**) and resulting products were purified using a Concert^{TM} PCR purification kit as per manufacturer's instructions. Commercial automated double-stranded DNA sequencing was supplied by MWG BIOTECHNOLOGIES (UK) using the PCR primers presented in **Table 2**.

Table 6: $MgCl_2$, annealing temperatures and electrophoretic conditions for each PCR-SSCP amplicon

Amplicon Name	Gene Name	Chromosome	Amplicon Location	Amplicon size (bp)	Electrophoretic conditions	PCR T_A	$MgCl_2$ Conc.
AGO-INT02a	Agouti	n/a	Intron 2	492	8W @ 4°C for 18 hours	58°C	12.5mM
AGO-INT02b	Agouti	n/a	Intron 2	493	8W @ 4°C for 18 hours	58°C	12.5mM
AGO-INT03	Agouti	n/a	Intron 3	547	8W @ 4°C for 18 hours	58°C	15mM
FSHB-INT01	Follicle stimulating hormone-beta subunit	BTA 15	Intron 1	473	8 - 12 W @ 4°C for 18 hours	58°C	12.5mM
FSHB-INT02	Follicle stimulating hormone-beta subunit	BTA 15	Intron 2	550	12 W @ 4°C for 16 hours	58°C	12.5mM
LTF-INT07	Lactoferrin	BTA 22	Intron 7	437	8W @ 4°C for 18 hours	58°C	12.5mM
LTF-INT08	Lactoferrin	BTA 22	Intron 8	436	8W @ 4°C for 18 hours	58°C	15mM
LTF -INT09	Lactoferrin	BTA 22	Intron 9	488	8W @ 4°C for 17 hours	58°C	12.5mM
LTF -INT11	Lactoferrin	BTA 22	Intron 11	501	12W @ 4°C for 17 hours	58°C	12.5mM
LYZ-3'UTR	Lysozyme	BTA 5	3'UTR	485	18W @ 4°C for 16 hours	58°C	12.5mM
OXT-INT01	Prepro-oxytocin-neurophysin 1	BTA 13	Intron 1	378	8W @ 4°C for 18 hours	58°C	15mM
PRC-INT02	Prostacyclin receptor	n/a	Intron 2	413	8W @ 4°C for 18 hours	58°C	15mM

All chromosome locations were obtained from the Roslin ARK data base (www.thearkdb.com).

Selection of samples for DNA sequencing and discovery of nucleotide variation

A select number of samples displaying electrophoretic mobility band shifts (indicative of nucleotide variation) were selected and re-amplified via non-radiolabelled PCR (**Table 3**). The resulting PCR products were purified using Concert © PCR purification kits as per manufacturer's instructions and were subsequently exported for commercial automated DNA sequencing provided by MWG-BIOTECHECHNOLOGY (U.K.).

Putative SNPs and other nucleotide variants (*i.e.* insertions and deletions) were identified by aligning the resulting DNA sequences using the muliple sequence alignment option incorporated in the GENEJOCKEY II software package (Taylor 1993). Nucleotide variants were subsequently confirmed by comparing the DNA sequence output traces provided by MWG-BIOTECHECHNOLOGY (U.K.).

Nucleotide sequence variation detected via PCR-SSCP

The results from PCR-SSCP analysis are presented in **Table 7**. Electrophoretic mobility shifts were observed at 4 PCR-SSCP amplicons. Three amplicons displayed differentiation between *Bos taurus* and *Bos indicus* samples, one revealed intra-taurine variability and one showed diversity within and between both domestic taxa. A select number of samples showing sequence variability were used to identify potential nucleotide variants using automated DNA sequencing.

Table 7: A summary of PCR-SSCP analysis results for the 12 amplicon loci assayed in this study

Amplicon name	Gene name	Chromosomal location	Amplicon Location	Amplicon size (bp)	Mobility shifts
AGO-INT02a	Agouti	n/a	Intron 2	492	-
AGO-INT02b	Agouti	n/a	Intron 2	493	-
AGO-INT03	Agouti	n/a	Intron 3	547	-
FSHB-INT01	Follicle stimulating hormone-beta subunit	BTA 15	Intron 1	473	T-Z
FSHB-INT02	Follicle stimulating hormone-beta subunit	BTA 15	Intron 2	550	T-Z
LTF-INT07	Lactoferrin	BTA 22	Intron 7	437	-
LTF-INT08	Lactoferrin	BTA 22	Intron 8	436	-
LTF-INT09	Lactoferrin	BTA 22	Intron 9	488	-
LTF-INT11	Lactoferrin	BTA 22	Intron 11	501	T-Z; IZ
LYZ-3'UTR	Lysozyme	BTA 5	3'UTR	485	IT
OXT-INT01	Prepro-oxytocin -neurophysin 1	BTA 13	Intron 1	378	-
PRC-INT02	Prostacyclin receptor	n/a	Intron 2	413	-

T-Z denoted difference between taurine and zebu samples; IT denotes diversity within *Bos taurus*; IZ denotes diversity within *Bos indicus*.

DNA sequencing and SNP detection

Figure 1 shows a typical sequence PCR-SSCP autoradiograph displaying DNA electrophoretic mobility shifts for the FSHB-INT01 amplicon. The DNA sequence of 4 of these samples (Guinean N'Dama 11, 16, 21 and Sahiwal 8) is shown below and putative SNPs are highlighted. Notably, the nucleotide variants observed at this amplicon locus illustrate differences between *Bos taurus* and *Bos indicus*. Candidate SNP markers for each of the polymorphic PCR-SSCP amplicons are summarised in **Table 8**.

Figure 1: Autoradiograph illustrating DNA sequence variability as detected via PCR-SSCP analysis

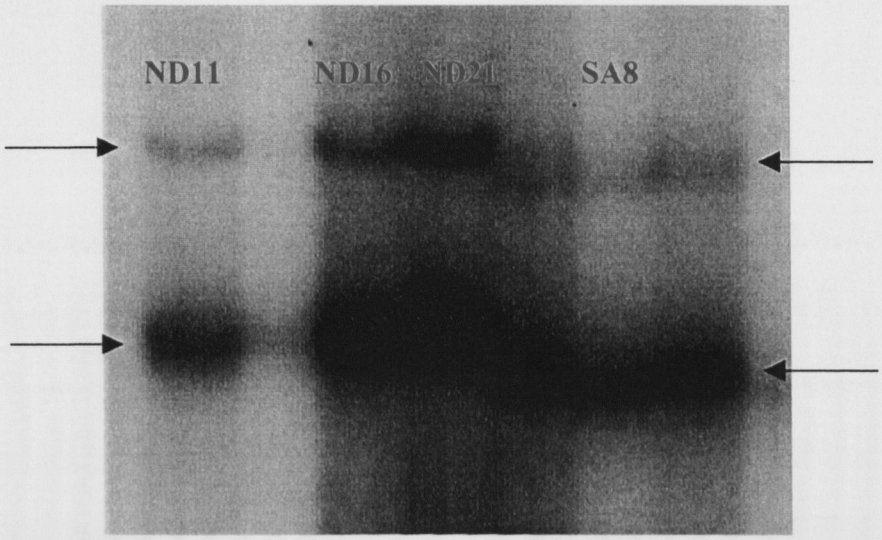


Figure 1. Nucleotide variability at the FSH-INT01 PCR-SSCP amplicon locus. Band shifts are indicted by arrows and illustrate differences between 3 African N'Dama *Bos taurus* samples (ND11, ND16, ND21) and 1 Indian *Bos indicus* Sahiwal sample (SA8).

Figure 1 continued

```
                          120        130        140        150        160
                           |          |          |          |          |
Contig# 1                 GGGGGCAAAAGCAAATGATANNNTTTATGAGATGCTAAAAATGATGAACTAATTT
FSH int 1 consensus seq   GGGGGCAAAAACAAATGATA---TTTATGAGATGCTAAAAATGATGAACTAATTT
N"D 16 for                GGGGGCAAAAGCAAATGATA---TTTATGAGATGCTAAAAATGATGAACTAATTT
ND 16 rev. inv.           GGGGGCAAAAGCAAATGATA---TTTATGAGATGCTAAAAATGATGAACTAATTT
N'D 11 for                GGGGGCAAAAGCAAATGATA---TTTATGAGATGCTAAAAATGATGAACTAATTT
ND 11 rev inv             GGGGGCAAAAGCAAATGATA---TTTATGAGATGCTAAAAATGATGAACTAATTT
N"D 21 for                GGGGGCAAAAGCAAATGATA---TTTATGAGATGCTAAAAATGATGAACTAATTT
ND21 rev inv              GGGGGCAAAAGCAAATGATA---TTTATGAGATGCTAAAAATGATGAACTAATTT
Sah 8 for                 GGGGGCAAAAGCAAATGATAATATTTATGAGATGCTAAAAATGATGAACTAATTT
Sah 8 rev. inv.           GGGGGCAAAAGCAAATGATAATATTTATGAGATGCTAAAAATGATGAACTAATTT
```

278

```
                          340        350        360        370        380
                           |          |          |          |          |
Contig# 1                 TATGTTTCTACTTTAACAACAACAACAAAAAATTGAGGAGCATAATTAGAAATGC
FSH int 1 consensus seq   TATGTTTCTACTTTAACAACAACAACAAAAAATTGAGGAGCATAATTAGAAATGC
N"D 16 for                TATGTTTCTACTTTAACAACAACAACAAAAAATTGAGGAGCATAATTAGAAATGC
ND 16 rev. inv.           TATGTTTCTACTTTAACAACAACAACAAAAAATTGAGGAGCATAATTAGAAATGC
N'D 11 for                TATGTTTCTACTTTAACAACAACAACAAAAAATTGAGGAGCATAATTAGAAATGC
ND 11 rev inv             TATGTTTCTACTTTAACAACAACAACAAAAAATTGAGGAGCATAATTAGAAATGC
N"D 21 for                TATGTTTCTACTTTAACAACAACAACAAAAAATTGAGGAGCATAATTAGAAATGC
ND21 rev inv              TATGTTTCTACTTTAACAACAACAACAAAAAATTGAGGAGCATAATTAGAAATGC
Sah 8 for                 TATGTTTCTACTTTAACAACAACAACAACAAATTGAGGAGCATAATTAGAAATGC
Sah 8 rev. inv.           TATGTTTCTACTTTAACAACAACAACAACAAATTGAGGAGCATAATTAGAAATGC
```

Figure 1 continued. DNA sequence at the bovine FSH-INT01 PCR-SSCP amplicon. Multi-base pair insert and a putative SNP showing differences between 3 African Guinean N'Dama *Bos taurus* and 1 Indian *Bos indicus* Sahiwal samples are shown in red.

Table 8: The location and character of the SNP loci discovered in this study

Amplicon name	Gene name	Amplicon position	SNP position	SNP alleles	Possible SNP variation
FSHB-INT01	Follicle-stimulating hormone beta-subunit	2048 - 2520	2374	'A' allele in 3 African *B. taurus* samples and GENBANK *B. taurus* sample. 'C' allele in 1 Indian *B. indicus* sample.	Between *Bos taurus/Bos indicus*
			2148	3 bp insert in 1 Indian *B. indicus* sample; absent in 3 African *B. taurus* samples and GENBANK *Bos taurus* sample.	Between *Bos taurus/Bos indicus*
FSHB-INT02	Follicle-stimulating hormone beta-subunit	2821 - 3370	3081	'T' allele in 1 European, 1 Near Eastern *B. taurus* sample and GENBANK *B. taurus* sample. 'C' allele in 1 Indian *B. indicus* sample.	Between *Bos taurus/Bos indicus*
			3119	5 bp insert in 1 Indian *B. indicus* sample; absent in 1 European, 1 Near East and GENBANK *B. taurus* samples.	Between *Bos taurus/Bos indicus*
LYZ-3' UTR	Lysozyme-3'UTR	965 - 1449	1064	'G' allele in 1 Near East *B. taurus* sample; 'A' allele in 1 Near East *B. taurus* sample; 'G/A' heterozygote in 1 Near East *B. taurus* sample.	Within-*Bos taurus* diversity
			1280	'C' allele in 2 Near East *B. taurus* samples; 'C/T' heterozygote in 1 Near East *B. taurus* sample.	Within-*Bos taurus* diversity

Table 8: continued

Amplicon name	Gene name	Amplicon position	SNP position	SNP alleles	Possible SNP variation
LACTF-INT11	Lactoferrin	5915 - 6415	5996	'G' allele in 1 Near East *B. taurus* sample, GENBANK *B. taurus* sample and 1 Indian *B. indicus* sample. 'A' allele in 1 Indicus *Bos indicus* sample.	Within-*Bos indicus* diversity
			6326	'C' allele in 1 Near East *B. taurus* samples and GENBANK *B. taurus* sample. 'T' allele in 2 Indian *B. indicus* samples.	Between *Bos taurus/Bos indicus*

The nucleotide positions of each PCR amplicon and SNP locus are reported according to the position given in the relevant GENBANK submissions. In some cases the diagnostic potentiality of each SNP allele is difficult to interpret and further analysis is required. All samples were homozygous unless stated.

A description of the SNPs discovered in this study

In this study, a total of 6 putative SNP loci were detected within 4 PCR-SSCP amplicon loci. Furthermore, two multiple nucleotide insertions (3 bp and 5 bp in length) were also detected within two PCR-SSCP amplicons. Of the 6 putative bovine SNP loci discovered, three display genetic differences between *Bos taurus* and *Bos indicus* cattle samples. The two multiple nucleotide insertions also exhibit genetic differentiation between the two domestic cattle taxa.

Two SNP loci illustrated nucleotide variability in a panel of 3 Near Eastern *Bos taurus* samples (both loci detected in the LYZ-3'UTR locus). Based on this limited data, it is possible that these loci may represent within-*Bos taurus* polymorphisms, however it is necessary to screen a larger panel of samples for allelic variability at these loci to fully evaluate this possibility. Also, 1 SNP locus (within the LACTF-INT11 locus) displayed variation between 2 Indian *Bos indicus* samples. Again, further analysis is required to fully establish the nature of this SNP locus.

Estimated SNP frequencies

Table 9 presents a summary of SNP loci frequencies over all assayed PCR-SSCP amplicons based on the DNA sequence data presented in this study, in accordance with the analysis of Heaton *et al.* (2001). Although the results presented in this study are based on a limited number of samples, preliminary results yield an average of 1 SNP per 379 bp sequenced. This figure decreases to 1 SNP per 346 bp when the primer sequences are excluded from the each amplicon locus as per Heaton *et al.* (2001).

Table 9: Estimated SNP frequencies across all polymorphic PCR-SSCP amplicon loci

Amplicon name	Amplicon length[a]	SNPs per amplicon	SNP frequency (bp per SNP)[b]
FSHB-INT01	473 (424)	1	473 (424)
FSHB-INT02	550 (508)	1	550 (508)
LYZ-3'UTR	501 (456)	2	250 (228)
LACTF-INT11	485 (444)	2	243 (222)
Totals	2009 (1832)	6	n/a[c]
Averages[d]	502 (458)	1.5	379 (346)

[a] Figures in brackets represent amplicon lengths excluding the primer sequences.

[b] A comparison of the total number of samples assayed for each individual locus (including GENBANK reference sequences) as indicated in **Table 5.4**.

[c] n/a implies not applicable.

[d] The average over all assayed amplicon loci (*i.e.* total values divided by the number of amplicons).

Table 10 summarises the number and frequency of the SNPs reported in this study between *Bos taurus* samples only (*i.e.* within *Bos taurus*); between *Bos indicus* samples only (*i.e.* within *Bos indicus*) and between *Bos taurus* and *Bos indicus*. Notably, SNPs illustrating differentiation between *Bos taurus* and *Bos indicus* are more frequently encountered (1 SNP per 670 bp) than those displaying variation within-*Bos taurus* (1 SNP per 1004 bp) and within-*Bos indicus* (1 SNP per 2009 bp), respectively. This most probably reflects the divergent ancestries of *Bos taurus* and *Bos indicus* cattle (Loftus *et al.* 1994a; MacHugh *et al.* 1997; Bradley *et al.* 1998).

Table 10: A summary of the SNP loci displaying within-*Bos taurus* differences, within-*Bos indicus* differences and differences between *Bos taurus* and *Bos indicus*

Amplicon name	Amplicon length (bp)	SNPs within *Bos taurus*[a]	SNPs within *Bos indicus*[b]	SNPs between *B. taurus-B. indicus*[c]
FSHB-INT01	473 (424)	0	0	1
FSHB-INT02	550 (508)	0	0	1
LYZ-3'UTR	501 (456)	2	0	0
LACTF-INT11	485 (444)	0	1	1
Totals	2009 (1832)	2	1	3
Frequency[d] (bp per SNP)		1004 (916)	2009 (1832)	670 (610)

[a] The number of SNPs per amplicon showing differences within *Bos taurus* only.

[b] The number of SNPs per amplicon showing differences within *Bos indicus* only.

[c] The number of SNPs per amplicon showing differences between *Bos taurus* and *Bos indicus*.

[d] The number of bp per SNP within *Bos taurus*; within *Bos indicus* and between *Bos taurus* and *Bos indicus*. Figures in brackets represent the number SNPs per bp excluding the amplicon primer sequences.

APPENDIX 2

GENETIC DISTANCE MATRICES AND MULTI-DIMENSIONAL SCALING CO-ORDINATES

Appendix 2.A: *Bos taurus* mtDNA sequence diversity observed in 4 morphologically zebu samples from the Indian subcontinent

	51	57	58	185	247	255	260
T3	T	G	C	G	C	T	C
Har8	C	.	.	.	T	C	.
Rsi13	.	.	.	.	.	C	T
Rsi35	.	.	.	.	.	C	T
Thar21	.	C	T	A	.	C	.

Sequence codes are shown in the first column. Only variable sites are shown. The position of the nucleotide variants from the T3 haplotype are given above each column. (.) denotes identity.

Appendix 2.B: *Bos taurus* mtDNA sequence diversity observed in 13 cattle samples from the Near East

	57	58	66	67	68	73	74	85	109	133	185	247	255	260
T3	G	C	A	A	T	A	T	T	T	T	G	C	T	C
MIrq19	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MIrq20	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MIrq30	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MIrq37	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MIrq27	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MIrq29	.	.	.	.	.	.	C	.	.	C	.	.	.	.
MIrq56	.	.	G	.	.	.	.	.	.	.	.	.	.	.
SIrq21	.	.	G	.	.	.	.	C	C	.	.	.	.	.
NIrq52	C	.	.	.	.	.	.	.	.	.	A	T	C	.
NIrq64	C	.	.	.	.	.	.	.	.	.	A	.	C	.
MIrq18	C	.	.	.	.	.	.	.	.	.	A	.	C	.
SIrq23	C	.	.	.	C	.	.	.	.	.	A	.	C	.
SIrq24	C	.	.	.	.	.	.	C	.	.	A	.	C	T

Sequence codes are shown in the first column. Only variable sites are shown. The position of the nucleotide variants from the T3 haplotype are given above each column. (.) denotes identity. Nirq, Mirq and Sirq refer to North, Middle and South Iraq, respectively.

Appendix 2.C: Matrix of mtDNA pairwise F_{ST} genetic distances for the Asian zebu populations analysed in this study

	NEZ	RS	SA	HA	TH	DE	ON	NE	LS	MN
NEZ	0									
RS	-0.0222	0								
SA	-0.0493	-0.0265	0							
HA	0.30112	0.18606	0.24349	0						
TH	0.19874	0.13709	0.15919	0.0343	0					
DE	0.06337	0.02747	0.03248	0.04182	0.01742	0				
ON	0.20796	0.12977	0.16732	0.01705	0.04455	0.02699	0			
NE	0.52662	0.33841	0.46929	0.03374	0.0707	0.12805	0.11707	0		
LS	0.56154	0.42434	0.51631	0.15555	0.1708	0.23546	0.25713	-0.0461	0	
MN	0.57763	0.39788	0.53204	0.09451	0.11571	0.17416	0.14569	0.27419	0.18139	0

Population codes as per **Table 2.1**

Appendix 2.D: Two-dimension MDS plot co-ordinates calculated from mtDNA pairwise F_{ST} values in Appendix 2.C

	Dim 1	Dim 2
NEZ	2.0423	0.0545
RS	1.4017	0.0749
SA	1.6685	0.0041
HA	-0.6019	0.048
TH	-0.1446	-0.3658
DE	0.3147	0.0299
ON	-0.1174	-0.4519
NE	-1.3234	0.8588
LS	-1.6863	0.9819
MN	-1.5536	-1.2344

Population codes as per **Table 2.1**

Appendix 2.E: Population pairwise F_{ST} genetic distances for European, African and Creole populations

	IC	KE	HL	GL	AA	HE	JE	BE	WF	NO	TE	LI	FR	CH	SI	GR	RO	SY	KN	FU	ND	SO	KU	KA	NA	BU	BA	MA	AR	AL	PR	ME	AC	GC	SC	NE
IC	0	0.232	0.093	0.102	0.155	0.17	0.154	0.276	0.19	0.114	0.129	0.084	0.123	0.201	0.107	0.202	0.203	0.241	0.679	0.659	0.642	0.655	0.673	0.662	0.629	0.523	0.184	0.122	0.148	0.195	0.202	0.118	0.168	0.429	0.147	0.275
KE	0.232	0	0.063	0.018	0.109	0.074	0.067	0.289	-0.05	0.073	0.114	0.017	0.08	0.101	0.057	0.184	0.125	0.21	0.817	0.784	0.732	0.747	0.775	0.755	0.707	0.573	0.206	0.221	0.068	0.183	0.217	0.1	0.11	0.416	0.156	0.251
HL	0.093	0.063	0	-0	0.057	-0	0.085	0.144	0.027	0.012	-0.06	0.006	0.031	0.03	0.015	0.096	0.018	0.122	0.567	0.541	0.533	0.541	0.568	0.567	0.529	0.426	0.06	-0.07	-0.02	0.048	0.068	-0.02	0.079	0.327	0.094	0.137
GL	0.102	0.018	-0	0	0.072	-0.01	0.09	0.117	0.023	0.034	-0.08	0.007	0.016	8E-04	0.014	0.082	-0.02	0.126	0.573	0.557	0.552	0.553	0.578	0.579	0.548	0.469	0.024	-0.05	0.014	0.111	0.103	0.018	0.089	0.383	0.099	0.193
AA	0.155	0.109	0.057	0.072	0	0.069	0.079	0.169	0.092	0.102	-0	0.007	0.068	0.091	0.028	0.132	0.092	0.153	0.522	0.497	0.491	0.502	0.536	0.533	0.504	0.407	0.078	-0.04	0.036	0.079	0.073	0.029	0.041	0.348	0.052	0.17
HE	0.17	0.074	-0	-0.01	0.069	0	0.101	0.173	0.029	0.053	-0.07	-0.02	0.033	-0	0.029	0.079	-0.01	0.065	0.661	0.636	0.608	0.62	0.647	0.638	0.593	0.471	0.058	-0.08	-0.01	0.084	0.094	0.006	0.077	0.363	0.104	0.164
JE	0.154	0.067	0.085	0.09	0.079	0.101	0	0.175	0.072	0.1	0.018	0.026	0.098	0.105	0.079	0.15	0.114	0.172	0.539	0.519	0.53	0.54	0.563	0.563	0.534	0.448	0.048	-0.02	0.095	0.154	0.11	0.099	0.077	0.371	0.079	0.208
BE	0.276	0.289	0.144	0.117	0.169	0.173	0.175	0	0.201	0.176	0.203	0.123	0.168	0.19	0.143	0.233	0.233	0.254	0.761	0.739	0.698	0.709	0.742	0.728	0.687	0.566	0.235	0.214	0.076	0.222	0.245	0.15	0.143	0.427	0.208	0.28
WF	0.19	-0.05	0.027	0.023	0.092	0.029	0.072	0.201	0	0.061	-0.04	0.008	0.057	0.045	0.043	0.135	0.045	0.161	0.713	0.686	0.657	0.67	0.695	0.683	0.645	0.527	0.12	0.002	0.039	0.151	0.152	0.07	0.103	0.402	0.136	0.225
NO	0.114	0.073	0.012	0.034	0.102	0.053	0.1	0.176	0.061	0	-0.02	0.008	0.047	0.062	0.043	0.064	0.076	0.164	0.629	0.609	0.594	0.604	0.631	0.625	0.593	0.49	0.077	0	0.05	0.138	0.131	0.054	0.111	0.381	0.127	0.2
TE	0.129	0.114	-0.06	-0.08	-0	-0.07	0.018	0.203	-0.04	-0.02	0	-0.07	-0.05	-0.09	-0.07	0.074	-0.05	0.106	0.827	0.775	0.706	0.731	0.763	0.736	0.672	0.506	0.083	0.189	-0.05	0.079	0.125	-0.02	0.011	0.356	0.051	0.157
LI	0.084	0.017	0.006	0.007	0.007	-0.02	0.026	0.123	0.008	0.008	-0.07	0	0.019	-0	-0.02	0.042	0.018	0.067	0.563	0.541	0.539	0.55	0.574	0.572	0.538	0.434	-0.01	-0.1	0.012	0.084	0.043	0.015	0.032	0.346	0.035	0.158
FR	0.123	0.08	0.031	0.016	0.068	0.033	0.098	0.168	0.057	0.047	-0.05	0.019	0	0.026	0.024	0.113	0.046	0.164	0.672	0.649	0.63	0.642	0.667	0.659	0.624	0.518	0.07	-0.01	0.046	0.152	0.162	0.063	0.096	0.404	0.111	0.212
CH	0.201	0.101	0.03	8E-04	0.091	-0	0.105	0.19	0.045	0.062	-0.09	-0	0.026	0	0.021	0.137	0.026	0.19	0.791	0.762	0.719	0.729	0.758	0.742	0.699	0.572	0.107	0.104	0.032	0.168	0.207	0.065	0.088	0.412	0.129	0.232
SI	0.107	0.057	0.015	0.014	0.028	0.029	0.079	0.143	0.043	0.043	-0.07	-0.02	0.024	0.021	0	0.115	0.029	0.146	0.637	0.617	0.603	0.61	0.639	0.632	0.599	0.498	0.032	-0.04	0.032	0.123	0.147	0.046	0.03	0.393	0.06	0.214
GR	0.202	0.184	0.096	0.082	0.132	0.079	0.15	0.233	0.135	0.064	0.074	0.042	0.113	0.137	0.115	0	0.16	0.159	0.594	0.573	0.559	0.581	0.6	0.595	0.56	0.457	0.099	-0.02	0.077	0.139	0.109	0.107	0.158	0.377	0.156	0.207
RO	0.203	0.125	0.018	-0.02	0.092	-0.01	0.114	0.233	0.045	0.076	-0.05	0.018	0.046	0.026	0.029	0.16	0	0.164	0.789	0.758	0.712	0.727	0.755	0.737	0.686	0.559	0.155	0.114	0.031	0.154	0.236	0.06	0.095	0.417	0.139	0.243
SY	0.241	0.21	0.122	0.126	0.153	0.065	0.172	0.254	0.161	0.164	0.106	0.067	0.164	0.19	0.146	0.159	0.164	0	0.581	0.561	0.554	0.571	0.588	0.584	0.543	0.447	0.159	-0	0.105	0.141	0.148	0.125	0.18	0.383	0.188	0.216
KN	0.679	0.817	0.567	0.573	0.522	0.661	0.539	0.761	0.713	0.629	0.827	0.563	0.672	0.791	0.637	0.594	0.789	0.581	0	0.037	0.035	0.025	0.034	0.048	0.09	0.051	0.679	0.683	0.496	0.247	0.3	0.477	0.566	0.312	0.56	0.283
FU	0.659	0.784	0.541	0.557	0.497	0.636	0.519	0.739	0.686	0.609	0.775	0.541	0.649	0.762	0.617	0.573	0.758	0.561	0.037	0	-0.03	0.014	-0.04	0.004	0.066	0.036	0.643	0.623	0.473	0.225	0.28	0.455	0.553	0.281	0.542	0.259
ND	0.642	0.732	0.533	0.552	0.491	0.608	0.53	0.698	0.657	0.594	0.706	0.539	0.63	0.719	0.603	0.559	0.712	0.554	0.035	-0.03	0	0.003	0.007	0.012	0.077	0.025	0.627	0.578	0.466	0.239	0.287	0.44	0.541	0.3	0.534	0.275
SO	0.655	0.747	0.541	0.553	0.502	0.62	0.54	0.709	0.67	0.604	0.731	0.55	0.642	0.729	0.61	0.581	0.727	0.571	0.025	0.014	0.003	0	-0	0.032	0.087	0.068	0.645	0.62	0.468	0.225	0.285	0.446	0.542	0.303	0.542	0.26
KU	0.673	0.775	0.568	0.578	0.536	0.647	0.563	0.742	0.695	0.631	0.763	0.574	0.667	0.758	0.639	0.6	0.755	0.588	0.034	-0.04	0.007	-0	0	-0.01	0.069	0.046	0.67	0.646	0.505	0.266	0.321	0.483	0.575	0.318	0.571	0.293
KA	0.662	0.755	0.567	0.579	0.533	0.638	0.563	0.728	0.683	0.625	0.736	0.572	0.659	0.742	0.632	0.595	0.737	0.584	0.048	0.004	0.012	0.032	-0.01	0	0.07	0.03	0.659	0.623	0.505	0.274	0.322	0.479	0.574	0.326	0.568	0.307
NA	0.629	0.707	0.529	0.548	0.504	0.593	0.534	0.687	0.645	0.593	0.672	0.538	0.624	0.699	0.599	0.56	0.686	0.543	0.09	0.066	0.077	0.087	0.069	0.07	0	0.065	0.61	0.552	0.47	0.246	0.294	0.447	0.554	0.335	0.542	0.3
BU	0.523	0.573	0.426	0.469	0.407	0.471	0.448	0.566	0.527	0.49	0.506	0.434	0.518	0.572	0.498	0.457	0.559	0.447	0.051	0.036	0.025	0.068	0.046	0.03	0.065	0	0.471	0.368	0.373	0.176	0.159	0.341	0.474	0.263	0.449	0.216
BA	0.184	0.206	0.06	0.024	0.078	0.058	0.048	0.235	0.12	0.077	0.083	-0.01	0.07	0.107	0.032	0.099	0.155	0.159	0.679	0.643	0.627	0.645	0.67	0.659	0.61	0.471	0	0.017	0.018	0.116	0.07	0.081	0.099	0.342	0.1	0.153
MA	0.122	0.221	-0.07	-0.05	-0.04	-0.08	-0.02	0.214	0.002	0	0.189	-0.1	-0.01	0.104	-0.04	-0.02	0.114	-0	0.683	0.623	0.578	0.62	0.646	0.623	0.552	0.368	0.017	0	-0.1	-0.04	-0.06	-0.06	0.004	0.274	0.016	0.047
AR	0.148	0.068	-0.02	0.014	0.036	-0.01	0.095	0.076	0.039	0.05	-0.05	0.012	0.046	0.032	0.032	0.077	0.031	0.105	0.496	0.473	0.466	0.468	0.505	0.505	0.47	0.373	0.018	-0.1	0	0.014	0.007	-0.02	0.073	0.3	0.087	0.105
AL	0.195	0.183	0.048	0.111	0.079	0.084	0.154	0.222	0.151	0.138	0.079	0.084	0.152	0.168	0.123	0.139	0.154	0.141	0.247	0.225	0.239	0.225	0.266	0.274	0.246	0.176	0.116	-0.04	0.014	0	-0.06	-0	0.147	0.213	0.123	0.032
PR	0.202	0.217	0.068	0.103	0.073	0.094	0.11	0.245	0.152	0.131	0.125	0.043	0.162	0.207	0.147	0.109	0.236	0.148	0.3	0.28	0.287	0.285	0.321	0.322	0.294	0.159	0.07	-0.06	0.007	-0.06	0	-0.02	0.155	0.182	0.115	-0.01
ME	0.118	0.1	-0.02	0.018	0.029	0.006	0.099	0.15	0.07	0.054	-0.02	0.015	0.063	0.065	0.046	0.107	0.06	0.125	0.477	0.455	0.44	0.446	0.483	0.479	0.447	0.341	0.081	-0.06	-0.02	-0	-0.02	0	0.08	0.305	0.083	0.101
AC	0.168	0.11	0.079	0.089	0.041	0.077	0.077	0.143	0.103	0.111	0.011	0.032	0.096	0.088	0.03	0.158	0.095	0.18	0.566	0.553	0.541	0.542	0.575	0.574	0.554	0.474	0.099	0.004	0.073	0.147	0.155	0.08	0	0.364	-0.03	0.209
GC	0.429	0.416	0.327	0.383	0.348	0.363	0.371	0.427	0.402	0.381	0.356	0.346	0.404	0.412	0.393	0.377	0.417	0.383	0.312	0.281	0.3	0.303	0.318	0.326	0.335	0.263	0.342	0.274	0.3	0.213	0.182	0.305	0.364	0	0.331	0.085
SC	0.147	0.156	0.094	0.099	0.052	0.104	0.079	0.208	0.136	0.127	0.051	0.035	0.111	0.129	0.06	0.156	0.139	0.188	0.56	0.542	0.534	0.542	0.571	0.568	0.542	0.449	0.1	0.016	0.087	0.123	0.115	0.083	-0.03	0.331	0	0.182
NE	0.275	0.251	0.137	0.193	0.17	0.164	0.208	0.28	0.225	0.2	0.157	0.158	0.212	0.232	0.214	0.207	0.243	0.216	0.283	0.259	0.275	0.26	0.293	0.307	0.3	0.216	0.153	0.047	0.105	0.032	-0.01	0.101	0.209	0.085	0.182	0

Population codes: IC, Iceland; KE, Kerry; HL, Highland; GL, Galloway; AA, Angus; HE, Hereford; JE, Jersey; BE, Berranda; WF, Westland Fjord; NO, Norwegian red; TE, Telemark; LI, Limousin; FR, Friesian; CH, Charolais; SI, Simmental; GR, German Black; RO, Romagnola; SY, Sykia; KN, Kenana; FU, White Fulani; ND, Guinean N'Dama; SO, Somba; KU, Kuri; KA, Kapsiki; NA, Namchi; BU, Butana; BA, Barrosa; MA, Maronesa; AR, Arouquesa; AL, Alentejana; PR, Preta; ME, Merolenga; AC, Antiguan Creole; GC, Guadeloupe Creole; SC, St. Lucian Creole; NE, Nelore.

Appendix 2.F: Two-dimensional MDS co-ordinates for European, African and Creole populations presented in Appendix 2.E

	Dim 1	Dim 2
IC	0.9547	0.8684
KE	1.2695	-0.5095
HL	0.7075	0.0866
GL	0.7721	0.1693
AA	0.5408	0.2194
HE	0.9179	-0.1568
JE	0.6976	0.2019
BE	1.1327	1.0709
WF	1.1225	-0.052
NO	0.9075	-0.0559
TE	1.2091	-0.2343
LI	0.736	0.06
FR	1.0258	0.1476
CH	1.2853	-0.0674
SI	0.9316	0.1936
GR	0.7595	-0.524
RO	1.286	-0.1979
SY	0.6628	-0.8294
KN	-2.5067	0.2377
FU	-2.4318	0.1154
ND	-2.3797	0.1347
SO	-2.4051	0.2186
KU	-2.4984	0.1822
KA	-2.4822	0.1844
NA	-2.3625	0.1313
BU	-1.9056	-0.0755
BA	0.9304	-0.3696
MA	0.7825	-0.4092
AR	0.4032	0.1059
AL	-0.3095	0.0141
PR	-0.2306	-0.0592
ME	0.3634	0.1603
AC	0.7572	0.2223
GC	-0.9314	-1.2772
SC	0.7056	0.3261
NE	-0.4177	-0.2326

Population codes: IC, Iceland; KE, Kerry; HL, Highland; GL, Galloway; AA, Angus; HE, Hereford; JE, Jersey; BE, Berranda; WF, Westland Fjord; NO, Norwegian red; TE, Telemark; LI, Limousin; FR, Friesian; CH, Charolais; SI, Simmental; GR, German Black; RO, Romagnola; SY, Sykia; KN, Kenana; FU, White Fulani; ND, Guinean N'Dama; SO, Somba; KU, Kuri; KA, Kapsiki; NA, Namchi; BU, Butana; BA, Barrosa; MA, Maronesa; AR, Arouquesa; AL, Alentejana; PR, Preta; ME, Merolenga; AC, Antiguan Creole; GC, Guadeloupe Creole; SC, St. Lucian Creole; NE, Nelore.

Appendix 2.G: Microsatellite D_A genetic distance estimates for European, African and Creole populations

	GC	AC	SC	ALN	MRT	ARQ	HAR	SAH	THR	NDM:	NDN	NDGM	NDS	NDB	NDG	GOB	MAUR	CHR	FRI	SIM
GC	0	0.1367	0.1824	0.305	0.2945	0.2952	0.2611	0.2402	0.2695	0.1862	0.182	0.1221	0.1666	0.214	0.2657	0.1058	0.0923	0.3113	0.3107	0.439
AC	0.1367	0	0.0698	0.1833	0.1451	0.1632	0.3237	0.2759	0.3221	0.3285	0.3093	0.2911	0.3429	0.3713	0.4372	0.2349	0.2102	0.1967	0.1562	0.2844
SC	0.1824	0.0698	0	0.1879	0.1615	0.1538	0.4016	0.2991	0.3342	0.3708	0.3661	0.3302	0.3859	0.4184	0.4406	0.2747	0.2511	0.1699	0.1032	0.3029
ALN	0.305	0.1833	0.1879	0	0.0378	0.0917	0.6522	0.5652	0.526	0.366	0.3938	0.3653	0.3557	0.3385	0.3591	0.5029	0.4572	0.152	0.187	0.1905
MRT	0.2945	0.1451	0.1615	0.0378	0	0.0698	0.6666	0.5698	0.5453	0.379	0.388	0.367	0.3549	0.3469	0.3628	0.4828	0.4477	0.081	0.1296	0.1321
ARQ	0.2952	0.1632	0.1538	0.0917	0.0698	0	0.6579	0.5639	0.5083	0.3692	0.3741	0.3526	0.3335	0.3614	0.3756	0.4417	0.4071	0.103	0.1424	0.2211
HA	0.2611	0.3237	0.4016	0.6522	0.6666	0.6579	0	0.0611	0.1963	0.5166	0.5146	0.4559	0.5689	0.6084	0.754	0.1777	0.1993	0.7083	0.6871	0.8192
SA	0.2402	0.2759	0.2991	0.5652	0.5698	0.5639	0.0611	0	0.1737	0.4656	0.4591	0.3916	0.5012	0.5338	0.6627	0.1555	0.1758	0.6046	0.5532	0.7237
TH	0.2695	0.3221	0.3342	0.526	0.5453	0.5083	0.1963	0.1737	0	0.4446	0.4183	0.4377	0.4824	0.5773	0.6468	0.2217	0.1973	0.589	0.5865	0.6852
NDM	0.1862	0.3285	0.3708	0.366	0.379	0.3692	0.5166	0.4656	0.4446	0	0.0814	0.0473	0.0611	0.0557	0.0782	0.2125	0.1585	0.4105	0.4466	0.4908
NDN	0.182	0.3093	0.3661	0.3938	0.388	0.3741	0.5146	0.4591	0.4183	0.0814	0	0.0713	0.0645	0.1055	0.1154	0.197	0.1448	0.4458	0.4637	0.5193
NDGM	0.1221	0.2911	0.3302	0.3653	0.367	0.3526	0.4559	0.3916	0.4377	0.0473	0.0713	0	0.0346	0.0445	0.0708	0.1584	0.1287	0.3781	0.4209	0.5007
NDS	0.1666	0.3429	0.3859	0.3557	0.3549	0.3335	0.5689	0.5012	0.4824	0.0611	0.0645	0.0346	0	0.0463	0.0657	0.2407	0.1787	0.3716	0.4383	0.4584
NDB	0.214	0.3713	0.4184	0.3385	0.3469	0.3614	0.6084	0.5338	0.5773	0.0557	0.1055	0.0445	0.0463	0	0.0391	0.2971	0.2522	0.3919	0.4429	0.4733
NDG	0.2657	0.4372	0.4406	0.3591	0.3628	0.3756	0.754	0.6627	0.6468	0.0782	0.1154	0.0708	0.0657	0.0391	0	0.3758	0.3243	0.3753	0.4242	0.4703
GOB	0.1058	0.2349	0.2747	0.5029	0.4828	0.4417	0.1777	0.1555	0.2217	0.2125	0.197	0.1584	0.2407	0.2971	0.3758	0	0.0323	0.486	0.501	0.623
MAUR	0.0923	0.2102	0.2511	0.4572	0.4477	0.4071	0.1993	0.1758	0.1973	0.1585	0.1448	0.1287	0.1787	0.2522	0.3243	0.0323	0	0.4452	0.4611	0.5824
CH	0.3113	0.1967	0.1699	0.152	0.081	0.103	0.7083	0.6046	0.589	0.4105	0.4458	0.3781	0.3716	0.3919	0.3753	0.486	0.4452	0	0.1293	0.1584
FR	0.3107	0.1562	0.1032	0.187	0.1296	0.1424	0.6871	0.5532	0.5865	0.4466	0.4637	0.4209	0.4383	0.4429	0.4242	0.501	0.4611	0.1293	0	0.2553
SM	0.439	0.2844	0.3029	0.1905	0.1321	0.2211	0.8192	0.7237	0.6852	0.4908	0.5193	0.5007	0.4584	0.4733	0.4703	0.623	0.5824	0.1584	0.2553	0

Appendix 2.H: Two-dimensional MDS co-ordinates constructed from the data presented in Appendix 2.G

	Dim 1	Dim 2
GC	0.4466	-0.1
AC	-0.0466	0.7833
SC	-0.1911	0.9676
ALN	-1.2207	0.5447
MRT	-1.2387	0.5642
ARQ	-1.1337	0.5914
HAR	2.1863	0.7136
SAH	1.7136	0.8312
THR	1.6545	0.9064
NDM	0.2042	-1.1494
NDN	0.2862	-1.1463
NDGM	0.3047	-0.9768
NDS	0.003	-1.1783
NDB	-0.2089	-1.3507
NDG	-0.5954	-1.5213
GOB	1.2283	-0.2117
MAUR	0.9952	-0.2666
CHR	-1.3045	0.6295
FRI	-1.1969	0.8851
SIM	-1.8862	0.4842

GC	Guadelopue Creole
AC	Antiguan Creole
SC	St. Lucian Creole
ALN	Alentejana
MRT	Mertolenga
ARQ	Arouquesa
HAR	Hariana
SAH	Sahiwal
THR	Tharparker
NDM	N'Dama (Mali)
NDN	N'Dama (Nigeria)
NDGM	N'Dama (Gambia)
NDS	N'Dama (Senegal)
NDB	N'Dama (Guinea Bissau)
NDG	N'Dama (Guinea)
GOB	Gobra
MAUR	Maure
CHR	Charolais
FRI	Friesian
SIM	Simmental

Appendix 2.I: MtDNA pairwise F_{ST} genetic distances computed for European, African, Near Eastern and Indian populations

	IC	KE	HI	GA	AN	HE	JE	BE	WF	NO	TE	LI	FR	CH	SI	GE	RO	SY	KN	FU	ND	SO	KU	KA	NA	BU	EG	DE	SA	HA	TH	RS	ON	TU	AB	SAR	DA	KUR	EAR	IS	IN	IM
IC	0	0.2318	0.0931	0.1056	0.1553	0.1705	0.1528	0.2763	0.1897	0.1145	0.1289	0.0842	0.1229	0.201	0.1068	0.2021	0.2026	0.2406	0.6795	0.659	0.6447	0.6546	0.6729	0.6622	0.6294	0.5231	0.3515	0.868	0.9134	0.7959	0.822	0.7686	0.8778	0.1529	0.1561	0.1087	0.1457	0.1089	0.1436	0.3048	0.1487	0.1584
KE	0.2318	0	0.0627	0.0205	0.1095	0.074	0.0668	0.2895	-0.053	0.0729	0.1139	0.0166	0.0795	0.1012	0.057	0.184	0.1253	0.2102	0.6173	0.7839	0.7316	0.7474	0.7747	0.7551	0.7073	0.5734	0.3681	0.887	0.9401	0.822	0.84	0.7852	0.9041	0.0749	0.1895	0.0949	0.155	0.0867	0.1422	0.3208	0.1706	0.1822
HI	0.0931	0.0627	0	-0.001	0.0571	-0.004	0.0953	0.1443	0.0275	0.012	-0.061	0.006	0.0307	0.0303	0.0154	0.0955	0.0178	0.1222	0.5675	0.5411	0.5393	0.5409	0.5684	0.5668	0.5288	0.4263	0.2144	0.8629	0.9067	0.7888	0.8172	0.7636	0.8705	-0.008	0.0944	0.006	0.0576	0.032	0.0742	0.2901	0.1143	0.1358
GA	0.1056	0.0205	-0.001	0	0.0744	-0.012	0.0924	0.1223	0.0242	0.0357	-0.075	0.008	0.0168	0.0018	0.0155	0.0845	-0.016	0.1295	0.5828	0.567	0.5606	0.5611	0.5868	0.5876	0.5553	0.4756	0.2917	0.8689	0.9055	0.8081	0.8285	0.7818	0.8753	0.0144	0.138	0.0445	0.0962	0.0729	0.1293	0.3365	0.1689	0.1805
AN	0.1553	0.1095	0.0571	0.0744	0	0.0686	0.0795	0.1686	0.092	0.1016	-0.001	0.0069	0.0679	0.0909	0.0281	0.1317	0.092	0.1534	0.522	0.4968	0.4915	0.5022	0.5356	0.5332	0.504	0.4073	0.2467	0.8534	0.8935	0.7733	0.8091	0.7551	0.8569	0.0431	0.139	0.0532	0.1031	0.0306	0.0898	0.2836	0.1165	0.1363
HE	0.1705	0.074	-0.004	-0.012	0.0686	0	0.1009	0.173	0.0291	0.0535	-0.068	-0.024	0.0329	-0.005	0.0289	0.0794	-0.007	0.0655	0.6811	0.6357	0.6076	0.6196	0.647	0.6383	0.5934	0.4714	0.2651	0.8763	0.9235	0.8096	0.8309	0.7781	0.8885	-0.05	0.0857	-0.007	0.0151	0.0256	0.0698	0.305	0.1195	0.162
JE	0.1538	0.0668	0.0853	0.0924	0.0795	0.1009	0	0.175	0.0722	0.0998	0.0179	0.026	0.0978	0.1052	0.0786	0.15	0.1143	0.1717	0.5386	0.5192	0.5304	0.5399	0.5625	0.5629	0.534	0.4481	0.294	0.8428	0.8808	0.7562	0.798	0.7453	0.841	0.0842	0.1636	0.0883	0.1331	0.0666	0.107	0.2824	0.1425	0.1402
BE	0.2763	0.2895	0.1443	0.1223	0.1686	0.173	0.175	0	0.2007	0.1763	0.2033	0.1227	0.1681	0.1905	0.1431	0.2333	0.2333	0.254	0.7612	0.7387	0.6967	0.7085	0.7417	0.7277	0.6869	0.5658	0.4	0.8804	0.9304	0.8159	0.8344	0.7777	0.8956	0.15	0.1908	0.1517	0.1796	0.1013	0.1316	0.3145	0.1725	0.1663
WF	0.1897	-0.053	0.0275	0.0242	0.092	0.0291	0.0722	0.2007	0	0.0613	-0.036	0.0079	0.0574	0.0455	0.0433	0.135	0.0451	0.1611	0.7133	0.6862	0.6566	0.6704	0.695	0.6835	0.6449	0.5271	0.3291	0.8816	0.9289	0.8174	0.8368	0.785	0.8948	0.0365	0.1614	0.0579	0.1096	0.0757	0.1186	0.3282	0.165	0.1886
NO	0.1145	0.0729	0.012	0.0357	0.1016	0.0535	0.0998	0.1763	0.0613	0	-0.017	0.0075	0.0473	0.062	0.0429	0.0638	0.0761	0.1645	0.6289	0.6086	0.5943	0.6036	0.631	0.6252	0.5928	0.4902	0.3092	0.8681	0.9112	0.7966	0.8235	0.771	0.8763	0.0525	0.1342	0.0671	0.1155	0.0812	0.1234	0.3112	0.1563	0.1565
TE	0.1289	0.1139	-0.061	-0.075	-0.001	-0.068	0.0179	0.2033	-0.036	-0.017	0	-0.072	-0.052	-0.093	-0.07	0.0738	-0.047	0.1059	0.8275	0.7752	0.7059	0.7313	0.7629	0.7359	0.6725	0.5055	0.2954	0.8696	0.9312	0.7753	0.8158	0.7593	0.8834	-0.047	0.0873	-0.016	0.049	-0.011	0.0357	0.2353	0.0723	0.0799
LI	0.0842	0.0166	0.006	0.008	0.0069	-0.024	0.026	0.1227	0.0079	0.0075	-0.072	0	0.0193	-0.005	-0.016	0.042	0.0184	0.0665	0.5632	0.5406	0.5397	0.55	0.5764	0.5717	0.5378	0.4345	0.2441	0.8525	0.8956	0.7671	0.8055	0.7509	0.8559	-0.018	0.0655	0.006	0.0111	0.0011	0.0476	0.2659	0.0913	0.1166
FR	0.1229	0.0795	0.0307	0.0168	0.0679	0.0329	0.0978	0.1681	0.0574	0.0473	-0.052	0.0193	0	0.0255	0.0243	0.1132	0.0457	0.1636	0.6723	0.6486	0.63	0.6417	0.6671	0.6591	0.6238	0.5182	0.3351	0.8756	0.9198	0.8083	0.8316	0.7795	0.8857	0.0514	0.1566	0.0673	0.1173	0.082	0.1339	0.3242	0.1701	0.178
CH	0.201	0.1012	0.0303	0.0018	0.0909	-0.005	0.1052	0.1905	0.0455	0.062	-0.093	-0.005	0.0255	0	0.0214	0.137	0.0262	0.1897	0.7912	0.7623	0.7183	0.7294	0.7577	0.7421	0.6988	0.5721	0.3795	0.8887	0.9391	0.8287	0.8432	0.79	0.9055	0.0356	0.1867	0.08	0.1166	0.0798	0.1349	0.331	0.176	0.1882
SI	0.1068	0.057	0.0154	0.0155	0.0281	0.0289	0.0786	0.1431	0.0433	0.0429	-0.07	-0.016	0.0243	0.0214	0	0.1148	0.0291	0.1455	0.6373	0.6171	0.6031	0.6104	0.6391	0.6322	0.5989	0.4981	0.3198	0.869	0.9122	0.7987	0.825	0.7726	0.8776	0.028	0.1426	0.0478	0.1139	0.0568	0.1208	0.3134	0.1573	0.1635
GE	0.2021	0.184	0.0955	0.0845	0.1317	0.0794	0.15	0.2333	0.135	0.0638	0.0738	0.042	0.1132	0.137	0.1148	0	0.16	0.159	0.5942	0.573	0.5592	0.5807	0.6	0.595	0.5603	0.4572	0.2725	0.8647	0.9081	0.7691	0.8201	0.7673	0.8714	0.0546	0.1297	0.0703	0.0722	0.0689	0.0789	0.3044	0.1331	0.1691
RO	0.2026	0.1253	0.0178	-0.016	0.092	-0.007	0.1143	0.2333	0.0451	0.0761	-0.047	0.0184	0.0457	0.0262	0.0291	0.16	0	0.1636	0.7886	0.7585	0.7125	0.7273	0.7546	0.7367	0.6865	0.5594	0.3737	0.8851	0.9369	0.8207	0.8382	0.7844	0.9015	0.0203	0.1578	0.0595	0.1302	0.0709	0.1262	0.3168	0.1596	0.1759
SY	0.2406	0.2102	0.1222	0.1295	0.1534	0.0655	0.1717	0.254	0.1611	0.1645	0.1059	0.0665	0.1636	0.1897	0.1455	0.159	0.1636	0	0.5809	0.5613	0.5537	0.5709	0.5863	0.5838	0.5432	0.4473	0.269	0.8674	0.9085	0.7961	0.8241	0.7734	0.8736	0.0372	0.0379	0.0106	0.0289	0.0272	0.0551	0.3036	0.0996	0.1838
KN	0.6795	0.6173	0.5675	0.5828	0.522	0.6611	0.5386	0.7612	0.7133	0.6289	0.8275	0.5632	0.6723	0.7912	0.6373	0.5942	0.7886	0.5809	0	0.037	0.0263	0.0249	0.0337	0.048	0.0698	0.0509	0.1513	0.8736	0.9278	0.7924	0.8265	0.7676	0.8864	0.532	0.425	0.4209	0.4388	0.2626	0.2373	0.3259	0.1619	0.2898
FU	0.659	0.7839	0.5411	0.567	0.4968	0.6357	0.5192	0.7387	0.6862	0.6086	0.7752	0.5406	0.6486	0.7623	0.6171	0.573	0.7585	0.5613	0.037	0	-0.026	0.014	-0.037	0.0043	0.0657	0.0361	0.1071	0.8685	0.9216	0.7829	0.8211	0.7606	0.8798	0.5096	0.4112	0.3995	0.4156	0.2269	0.2217	0.3111	0.1518	0.2742
ND	0.6447	0.7316	0.5333	0.5606	0.4915	0.6076	0.5304	0.6967	0.6566	0.5943	0.7059	0.5397	0.63	0.7183	0.6031	0.5592	0.7125	0.5537	0.0263	-0.026	0	-0.002	0.0125	0.0219	0.0804	0.031	0.1295	0.8753	0.9222	0.8067	0.832	0.7768	0.8877	0.4998	0.4203	0.3996	0.4218	0.27	0.256	0.3517	0.1898	0.3165
SO	0.6546	0.7474	0.5409	0.5611	0.5022	0.6196	0.5399	0.7085	0.6704	0.6036	0.7313	0.55	0.6417	0.7294	0.6104	0.5807	0.7273	0.5709	0.0249	0.014	-0.002	0	-9E-05	0.0321	0.0866	0.0684	0.1487	0.8812	0.9283	0.8179	0.8382	0.7842	0.8949	0.509	0.4395	0.4207	0.438	0.2864	0.2815	0.3671	0.2041	0.3284
KU	0.6729	0.7747	0.5684	0.5868	0.5356	0.647	0.5625	0.7417	0.695	0.631	0.7629	0.5744	0.6671	0.7577	0.6391	0.6	0.7546	0.5883	0.0337	-0.037	0.0125	-9E-05	0	-0.012	0.0692	0.0462	0.1364	0.8803	0.9287	0.8144	0.8368	0.7826	0.8939	0.5367	0.4476	0.4404	0.4489	0.2985	0.2836	0.3659	0.2013	0.3367
KA	0.6622	0.7551	0.5668	0.5876	0.5332	0.6383	0.5629	0.7277	0.6835	0.6252	0.7359	0.5717	0.6591	0.7421	0.6322	0.595	0.7367	0.5838	0.048	0.0043	0.0219	0.0321	-0.012	0	0.0698	0.03	0.1705	0.8819	0.9277	0.8184	0.8399	0.788	0.8945	0.5346	0.4517	0.4341	0.4528	0.3142	0.2996	0.3814	0.2135	0.3541
NA	0.6294	0.7073	0.5288	0.5553	0.504	0.5934	0.534	0.6869	0.6449	0.5928	0.6725	0.5378	0.6238	0.6988	0.5989	0.5603	0.6865	0.5432	0.0698	0.0657	0.0804	0.0866	0.0692	0.0698	0	0.0646	0.1577	0.8747	0.9188	0.8077	0.8328	0.7806	0.8847	0.4917	0.4148	0.4029	0.421	0.2896	0.27	0.3672	0.1879	0.3351
BU	0.5231	0.5734	0.4263	0.4756	0.4073	0.4714	0.4481	0.5658	0.5271	0.4902	0.5655	0.4345	0.5182	0.5721	0.4981	0.4572	0.5594	0.4473	0.0509	0.0361	0.021	0.0684	0.0462	0.03	0.0646	0	0.0787	0.8502	0.9898	0.7697	0.8078	0.7554	0.8528	0.3868	0.3394	0.3105	0.3232	0.2272	0.2002	0.3243	0.1432	0.2777
EG	0.3515	0.3881	0.2144	0.2917	0.2467	0.2651	0.294	0.4	0.3291	0.3092	0.2954	0.2441	0.3251	0.3795	0.3198	0.2725	0.3737	0.269	0.1313	0.1071	0.1295	0.1487	0.1364	0.1705	0.1577	0.0787	0	0.9391	0.8789	0.7508	0.7939	0.7385	0.8395	0.1799	0.1771	0.1218	0.1136	0.0786	0.0523	0.2655	0.0592	0.1713
DE	0.868	0.887	0.8629	0.8689	0.8534	0.8763	0.8428	0.8804	0.8816	0.8681	0.8696	0.8525	0.8756	0.8887	0.869	0.8647	0.8851	0.8674	0.8738	0.8685	0.8753	0.8812	0.8803	0.8819	0.8747	0.8502	0.8391	0	0.0325	0.0416	0.0099	0.0147	0.027	0.8632	0.8377	0.8478	0.8494	0.7662	0.7647	0.4066	0.6874	0.615
SA	0.9134	0.9401	0.9067	0.9055	0.8935	0.9235	0.8808	0.9304	0.9289	0.9112	0.9312	0.8956	0.9198	0.9391	0.9122	0.9081	0.9369	0.9085	0.9278	0.9216	0.9222	0.9283	0.9287	0.9277	0.9188	0.8898	0.8789	0.0325	0	0.1649	0.1122	-0.005	0.1673	0.9068	0.8706	0.8869	0.888	0.7891	0.7905	0.4032	0.7005	0.6206
HA	0.7959	0.822	0.7888	0.8081	0.7733	0.8096	0.7562	0.8159	0.8174	0.7966	0.7753	0.7671	0.8083	0.8287	0.7987	0.7891	0.8207	0.7961	0.7924	0.7839	0.8067	0.8179	0.8144	0.8184	0.8077	0.7697	0.7508	0.0416	0.1649	0	-0.006	0.0412	0.0001	0.7867	0.7517	0.7627	0.7686	0.6498	0.6392	0.2451	0.54	0.4513
TH	0.822	0.84	0.8172	0.8285	0.8091	0.8309	0.798	0.8344	0.8368	0.8235	0.8158	0.8055	0.8316	0.8432	0.825	0.8201	0.8382	0.8241	0.8265	0.8211	0.832	0.8382	0.8368	0.8399	0.8328	0.8078	0.7939	0.0099	0.1122	-0.006	0	0.0482	0.0174	0.8181	0.7948	0.803	0.8053	0.7209	0.7174	0.3598	0.6402	0.5633
RS	0.7686	0.7852	0.7636	0.7818	0.7551	0.7781	0.7453	0.7777	0.785	0.771	0.7533	0.7509	0.7795	0.79	0.7726	0.7673	0.7844	0.7734	0.7676	0.7606	0.7768	0.7842	0.7826	0.788	0.7806	0.7554	0.7385	0.0147	-0.005	0.0412	0.0482	0	0.0489	0.7653	0.7431	0.7502	0.7531	0.6647	0.6591	0.2997	0.5654	0.4927
ON	0.8778	0.9041	0.8705	0.8753	0.8565	0.8885	0.841	0.8956	0.8948	0.8763	0.8834	0.8559	0.8857	0.9053	0.8776	0.8714	0.9015	0.8736	0.8864	0.8798	0.8877	0.8949	0.8929	0.8945	0.8847	0.8528	0.8395	0.027	0.1673	0.0001	0.0174	0.0489	0	0.8695	0.834	0.8486	0.8509	0.7453	0.7415	0.354	0.6467	0.5672
TU	0.1529	0.0749	-0.008	0.0144	0.0431	-0.05	0.0842	0.15	0.0365	0.0525	-0.047	-0.018	0.0514	0.0356	0.028	0.0546	0.0203	0.0372	0.532	0.5096	0.4998	0.509	0.5367	0.5346	0.4917	0.3868	0.1799	0.8632	0.9068	0.7667	0.8181	0.7653	0.8695	0	0.0431	-0.036	-0.011	0.0039	0.0308	0.2873	0.085	0.1479
AB	0.1561	0.1895	0.0944	0.138	0.139	0.0857	0.1636	0.1908	0.1614	0.1342	0.0873	0.0655	0.1566	0.1867	0.1426	0.1297	0.1578	0.0379	0.425	0.4112	0.4203	0.4395	0.4476	0.4517	0.4148	0.3394	0.1771	0.8277	0.8706	0.7517	0.7948	0.7431	0.834	0.0431	0	0.0173	-0.006	0.0045	0.006	0.2575	0.0353	0.1344
SAR	0.1087	0.0949	0.006	0.0445	0.0532	-0.007	0.0883	0.1517	0.0579	0.0671	-0.016	0.006	0.0673	0.08	0.0478	0.0703	0.0595	0.0106	0.4209	0.3995	0.3996	0.4207	0.4404	0.4341	0.4029	0.3105	0.1218	0.8478	0.8869	0.7627	0.803	0.7502	0.8486	-0.036	0.0173	0	-0.023	-0.008	0.0084	0.2653	0.0549	0.1286
DA	0.1457	0.155	0.0576	0.0962	0.1031	0.0151	0.1331	0.1796	0.1096	0.1155	0.049	0.0111	0.1173	0.1166	0.1139	0.0722	0.1302	0.0289	0.4388	0.4156	0.4218	0.438	0.4489	0.4528	0.421	0.3232	0.1136	0.8494	0.888	0.7686	0.8053	0.7531	0.8509	-0.011	-0.006	-0.023	0	-0.015	-0.015	0.2685	0.0404	0.1377
KUR	0.1089	0.0867	0.032	0.0729	0.0306	0.0256	0.0666	0.1013	0.0757	0.0812	-0.011	0.0011	0.082	0.0798	0.0568	0.0689	0.0705	0.0272	0.2626	0.2369	0.27	0.2864	0.2985	0.3142	0.2896	0.2272	0.0786	0.7662	0.7891	0.6498	0.7209	0.6647	0.7453	0.0039	0.0045	-0.008	-0.015	0	-0.062	0.1446	-0.03	0.0124
EAR	0.1436	0.1422	0.0742	0.1293	0.0898	0.0698	0.107	0.1316	0.1186	0.1234	0.0357	0.0476	0.1339	0.1349	0.1208	0.0789	0.1262	0.0551	0.2373	0.2217	0.256	0.2815	0.2836	0.2996	0.27	0.2002	0.0523	0.7647	0.7905	0.6392	0.7174	0.6591	0.7415	0.0308	0.006	0.0084	-0.015	-0.062	0	0.136	-0.054	0.0185
IS	0.3048	0.3208	0.2901	0.3365	0.2836	0.305	0.2824	0.3145	0.3282	0.3112	0.2353	0.2659	0.3242	0.331	0.3134	0.3044	0.3168	0.3036	0.3259	0.3111	0.3517	0.3671	0.3659	0.3814	0.3672	0.3243	0.2655	0.4066	0.4032	0.2451	0.3598	0.2997	0.354	0.2873	0.2575	0.2653	0.2685	0.1446	0.136	0	0.0701	-0.007
IN	0.1487	0.1706	0.1143	0.1689	0.1165	0.1198	0.1425	0.1725	0.165	0.1563	0.0723	0.0913	0.1701	0.176	0.1573	0.1331	0.1596	0.0996	0.1619	0.1518	0.1898	0.2041	0.2013	0.2135	0.1879	0.1432	0.0592	0.6874	0.7005	0.54	0.6402	0.5854	0.6467	0.085	0.0353	0.0549	0.0404	-0.03	-0.054	0.0701	0	-0.011
IM	0.1584	0.1822	0.1358	0.1805	0.1363	0.162	0.1402	0.1663	0.1886	0.1565	0.0799	0.1166	0.178	0.1882	0.1635	0.1691	0.1759	0.1838	0.2898	0.2742	0.3165	0.3284	0.3367	0.3541	0.3351	0.2777	0.1713	0.615	0.6206	0.4513	0.5633	0.4927	0.5672	0.1479	0.1344	0.1286	0.1377	0.0124	0.0185	-0.007	-0.011	0

Population codes as per **Table 4.2.**

Appendix 2.J: MtDNA two-dimensional MDS plot co-ordinates for the data presented in Appendix 2.I

	Dim 1	Dim 2
IC	0.8544	0.5929
KE	0.9279	0.8226
HI	0.793	0.2666
GA	0.8375	0.3335
AN	0.7416	0.1582
HE	0.8697	0.449
JE	0.734	0.298
BE	0.8925	0.7777
WF	0.9013	0.6005
NO	0.8401	0.4376
TE	0.8583	0.7658
LI	0.7709	0.2985
FR	0.8815	0.5521
CH	0.935	0.7662
SI	0.8457	0.4762
GE	0.8198	0.3238
RO	0.9241	0.7492
SY	0.819	0.2687
KN	-0.6272	-1.7203
FU	-0.5805	-1.6704
ND	-0.4969	-1.6646
SO	-0.5244	-1.6991
KU	-0.5907	-1.7285
KA	-0.5633	-1.7289
NA	-0.4747	-1.6494
BU	-0.259	-1.406
EG	0.1148	-0.9542
DE	-2.3942	1.0543
SA	-2.4547	1.0846
HA	-2.1697	0.9067
TH	-2.255	0.9809
RS	-2.1251	0.8813
ON	-2.3981	1.0516
TU	0.7658	0.148
AB	0.6197	-0.0802
SAR	0.6464	-0.0876
DA	0.6678	-0.0686
KUR	0.3125	-0.2
EAR	0.2721	-0.2513
IS	-0.7406	0.1407
IN	0.0542	-0.3107
IM	-0.0456	0.0346

Population codes as per **Table 4.2**

Appendix 2.K: Microsatellite-derived D_A genetic distances for European, African, Near Eastern and Indian populations

	JE	HE	HL	SI	HO	CH	HG	TG	AB	EAR	SAR	DA	IR	KUR	ND	EG	ON	NE	VO	PS	PN
JE	0	0.221	0.223	0.193	0.168	0.166	0.196	0.166	0.194	0.192	0.206	0.26	0.274	0.233	0.251	0.222	0.624	0.584	0.197	0.188	0.218
HE	0.221	0	0.138	0.187	0.177	0.221	0.203	0.181	0.217	0.215	0.211	0.315	0.283	0.229	0.374	0.238	0.614	0.574	0.167	0.167	0.218
HL	0.223	0.138	0	0.202	0.22	0.255	0.239	0.215	0.252	0.239	0.249	0.347	0.321	0.264	0.324	0.271	0.682	0.628	0.23	0.203	0.267
SI	0.193	0.187	0.202	0	0.126	0.141	0.161	0.158	0.203	0.195	0.215	0.294	0.283	0.232	0.34	0.226	0.61	0.57	0.117	0.158	0.171
HO	0.168	0.177	0.22	0.126	0	0.144	0.172	0.131	0.152	0.161	0.169	0.249	0.253	0.187	0.281	0.19	0.57	0.536	0.108	0.139	0.166
CH	0.166	0.221	0.255	0.141	0.144	0	0.184	0.154	0.184	0.191	0.198	0.254	0.292	0.241	0.285	0.233	0.586	0.548	0.139	0.177	0.189
HG	0.196	0.203	0.239	0.161	0.172	0.184	0	0.131	0.153	0.153	0.154	0.237	0.241	0.2	0.285	0.19	0.569	0.521	0.172	0.148	0.186
TG	0.166	0.181	0.215	0.158	0.131	0.154	0.131	0	0.057	0.056	0.066	0.148	0.114	0.088	0.238	0.096	0.389	0.363	0.137	0.129	0.14
AB	0.194	0.217	0.252	0.203	0.152	0.184	0.153	0.057	0	0.049	0.059	0.132	0.091	0.064	0.239	0.083	0.348	0.307	0.178	0.141	0.168
EAR	0.192	0.215	0.239	0.195	0.161	0.191	0.153	0.056	0.049	0	0.06	0.138	0.08	0.054	0.237	0.081	0.315	0.288	0.186	0.167	0.169
SAR	0.206	0.211	0.249	0.215	0.169	0.198	0.154	0.066	0.059	0.06	0	0.103	0.098	0.071	0.248	0.078	0.353	0.326	0.185	0.164	0.174
DA	0.26	0.315	0.347	0.294	0.249	0.254	0.237	0.148	0.132	0.138	0.103	0	0.158	0.154	0.275	0.117	0.379	0.361	0.257	0.245	0.247
IR	0.274	0.283	0.321	0.283	0.253	0.292	0.241	0.114	0.091	0.08	0.098	0.158	0	0.059	0.316	0.103	0.253	0.228	0.246	0.218	0.222
KUR	0.233	0.229	0.264	0.232	0.187	0.241	0.2	0.088	0.064	0.054	0.071	0.154	0.059	0	0.274	0.085	0.284	0.26	0.216	0.195	0.201
ND	0.251	0.374	0.324	0.34	0.281	0.285	0.285	0.238	0.239	0.237	0.248	0.275	0.316	0.274	0	0.213	0.565	0.525	0.316	0.266	0.291
EG	0.222	0.238	0.271	0.226	0.19	0.233	0.19	0.096	0.083	0.081	0.078	0.117	0.103	0.085	0.213	0	0.321	0.292	0.192	0.173	0.196
ON	0.624	0.614	0.682	0.61	0.57	0.586	0.569	0.389	0.348	0.315	0.353	0.379	0.253	0.284	0.565	0.321	0	0.099	0.551	0.534	0.497
NE	0.584	0.574	0.628	0.57	0.536	0.548	0.521	0.363	0.307	0.288	0.326	0.361	0.228	0.26	0.525	0.292	0.099	0	0.535	0.478	0.473
VO	0.197	0.167	0.23	0.117	0.108	0.139	0.172	0.137	0.178	0.186	0.185	0.257	0.246	0.216	0.316	0.192	0.551	0.535	0	0.128	0.157
PS	0.188	0.167	0.203	0.158	0.139	0.177	0.148	0.129	0.141	0.167	0.164	0.245	0.218	0.195	0.266	0.173	0.534	0.478	0.128	0	0.125
PN	0.218	0.218	0.267	0.171	0.166	0.189	0.186	0.14	0.168	0.169	0.174	0.247	0.222	0.201	0.291	0.196	0.497	0.473	0.157	0.125	0

Population codes as per **Table 4.3**

Appendix 2.L: Two-dimensional MDS plot co-ordinates for the data presented in Appendix 2.K

	Dim 1	Dim 2
JE	1.133	-0.57
HE	1.086	0.986
HL	1.688	0.74
SI	1.126	0.718
HO	0.854	0.002
CH	1.176	-0.37
HG	0.827	-0.04
VO	0.829	0.542
PS	0.603	0.19
PG	0.225	0.677
TG	0.098	-0.04
AB	-0.21	-0.07
EAR	-0.3	-0.03
SAR	-0.19	-0.08
DA	-0.64	-1.04
IR	-1.13	0.199
KUR	-0.65	0.086
ND	0.236	-2.03
EG	-0.39	-0.27
ON	-3.31	0.182
NE	-3.05	0.238

Population codes as per **Table 4.3**

Appendix 2.M: SNP-derived D_A genetic distance estimates for European, African, Near Eastern and Indian populations

	KE	JE	HER	ANG	HL	EF	VO	PN	SB	FV	BV	HGR	AB	SAR	ND	ONG	BI
KE	0	0.0453	0.0466	0.0659	0.0379	0.0242	0.0095	0.0244	0.051	0.0368	0.0631	0.0331	0.0395	0.0577	0.0751	0.1433	0.1292
JE	0.0453	0	0.0795	0.0775	0.0296	0.0499	0.0403	0.0252	0.0331	0.0319	0.0484	0.0351	0.0279	0.0361	0.0577	0.0959	0.0906
HER	0.0466	0.0795	0	0.0767	0.0666	0.0233	0.0637	0.0343	0.0526	0.0924	0.0748	0.0575	0.0596	0.0666	0.0931	0.1756	0.1516
ANG	0.0659	0.0775	0.0767	0	0.047	0.0415	0.0746	0.0397	0.0801	0.0819	0.0693	0.0358	0.0594	0.0767	0.1181	0.1411	0.1322
HL	0.0379	0.0296	0.0666	0.047	0	0.0356	0.0338	0.0261	0.0476	0.0385	0.0294	0.031	0.0225	0.0461	0.0929	0.1096	0.1013
EF	0.0242	0.0499	0.0233	0.0415	0.0356	0	0.0421	0.0148	0.052	0.0706	0.0683	0.0167	0.0244	0.0309	0.0811	0.1131	0.1014
VO	0.0095	0.0403	0.0637	0.0746	0.0338	0.0421	0	0.0351	0.0499	0.0262	0.0642	0.0496	0.0491	0.0787	0.0809	0.1459	0.1271
PN	0.0244	0.0252	0.0343	0.0397	0.0261	0.0148	0.0351	0	0.0286	0.0341	0.0405	0.012	0.0216	0.033	0.0601	0.1137	0.1041
SB	0.051	0.0331	0.0526	0.0801	0.0476	0.052	0.0499	0.0286	0	0.0486	0.0553	0.0524	0.0326	0.0524	0.0502	0.1023	0.0771
FV	0.0368	0.0319	0.0924	0.0819	0.0385	0.0706	0.0262	0.0341	0.0486	0	0.0475	0.064	0.0587	0.0897	0.0931	0.1679	0.159
BV	0.0631	0.0484	0.0748	0.0693	0.0294	0.0683	0.0642	0.0405	0.0553	0.0475	0	0.0545	0.0409	0.0573	0.0631	0.1789	0.1522
HGR	0.0331	0.0351	0.0575	0.0358	0.031	0.0167	0.0496	0.012	0.0524	0.064	0.0545	0	0.0191	0.0227	0.0609	0.0946	0.0873
AB	0.0395	0.0279	0.0596	0.0594	0.0225	0.0244	0.0491	0.0216	0.0326	0.0587	0.0409	0.0191	0	0.0094	0.0473	0.0861	0.065
SAR	0.0577	0.0361	0.0666	0.0767	0.0461	0.0309	0.0787	0.033	0.0524	0.0897	0.0573	0.0227	0.0094	0	0.0506	0.0879	0.0713
ND	0.0751	0.0577	0.0931	0.1181	0.0929	0.0811	0.0809	0.0601	0.0502	0.0931	0.0631	0.0609	0.0473	0.0506	0	0.1595	0.1107
ONG	0.1433	0.0959	0.1756	0.1411	0.1096	0.1131	0.1459	0.1137	0.1023	0.1679	0.1789	0.0946	0.0861	0.0879	0.1595	0	0.016
IZ	0.1292	0.0906	0.1516	0.1322	0.1013	0.1014	0.1271	0.1041	0.0771	0.159	0.1522	0.0873	0.065	0.0713	0.1107	0.016	0

Population codes as per **Table 4.1**

Appendix 2.N: Two-dimensional MDS plot co-ordiantes for the data presented in Appendix 2.M

	Dim 1	Dim 2
KE	0.9502	0.0334
JE	-0.215	0.4908
HER	1.3789	-0.674
ANG	0.4506	-1.561
HL	0.2907	-0.447
EF	0.2525	-0.683
VO	1.0281	-0.083
PN	0.3625	-0.088
SB	-0.256	0.5515
FV	1.3646	0.6503
BV	1.3219	0.5197
HGR	-0.185	-0.201
AB	-0.453	0.0647
SAR	-0.693	0.0331
ND	0.2147	1.7347
ONG	-3.138	-0.518
IZ	-2.676	0.1763

Population codes as per **Table 4.1**

APPENDIX 3

BUFFERS AND REAGENTS

BUFFERS AND REAGENTS

(A) Proteinase K digestion buffer

4 M NaCl	13.8 mls
1 M Tris-Cl (pH 8.0)	27.5 mls
0.5 M EDTA (pH 8.0)	1.1 mls
10% SDS	27.5 mls
Sterile ddH_2O	430.1 mls

To extract DNA from 120 samples 5 mls of Proteinase K (20mg/ml) was added to 45 mls of the above solution.

(B) 6X loading dye (for agarose gel electrophoresis)

Bromophenol blue	0.25%
Xylene cyanol FF	0.25%
Glycerol	30%

(C) 10X TBE electrophoresis buffer

Tris-borate (pH 8.3)	100 M
EDTA (pH 8.0)	0.02 M

(D) 10X 15 mM $MgCl_2$ PCR buffer (50mls)

1M KCl	25 mls
1M Tris-HCl (pH8.0)	5 mls
0.5M $MgCl_2$	1.5 mls
1% Triton X-100	0.5 mls
ddH_2O	18 mls

(E) 1X Dynabead Washing & Binding buffer (DYNAL)

Tris-HCl (pH8.0)	10 mM
EDTA (pH 8.0)	1 mM
NaCl	1 M

(F) 2X Dynabead Washing & Binding buffer (DYNAL)

Tris-HCl (pH 8.0)	10 mM
EDTA (pH 8.0)	1 mM
NaCl	2 M

(G) 100X TE Buffer (SIGMA)

Tris-Cl (pH 8.0)	1 M
EDTA	0.1 M

(H) Formamide denaturing buffer

Deionised formamide	98%
EDTA (pH8.0)	10 mM
Bromophenol blue	0.025%
Xylene cyanol FF	0.025%

(I) 5X *Sequenase* buffer (UNITED STATES BIOCHEMICAL)

Tris-HCl	200 mM
$MgCl_2$	100 mM
NaCl	250 mM

(J) *Sequenase* labelling mix (UNITED STATES BIOCHEMICAL)

dGTP	7.5 μM
dCTP	7.5 μM
dTTP	7.5 μM

(K) $MnCl_2$ buffer (UNITED STATES BIOCHEMICAL)

Sodium Isocitrate	0.15 M
$MnCl_2$	0.1 M

(L) 6% denaturing polyacrylamide gel (100 mls) (NATIONAL DIAGNOSTICS)

Sequagel™ buffer	10 mls
Sequagel™ concentrate	24 mls
Sequagel™ diluent	66 mls

Gels were made according to manufacturer's instructions. The above volumes produced gels containing acrylamide: bis-acrylamide (19:1), 5% Urea and 1X TBE buffer. Acrylamide gels were polymerised with 400 mls of 10% APS and 40 μls of TEMED for a period of 1 hour.

(M) Platinum *Taq* Polymerase buffer (GIBCO LIFE SCIENCES)

Tris-HCl (pH 8.4)	200 mM
KCl	500 mM

(N) 10X *Taq* ligase buffer (NEW ENGLAND BIOLABS)

Tris-HCl (pH 7.6)	200 mM
Potassium acetate	250 mM
Magnesium acetate	100 mM
DTT	100 mM
NAD	10 mM
Triton X-100	1%

(O) ABI loading buffer (APPLIED BIOSCIENCES)

Blue dextran	50 mg / ml
EDTA	25 mM

For sample loading, ABI loading dye is mixed with deionised formamide (1:5).

(P) 8% denaturing polyacrylamide gel (100 mls) (NATIONAL DIAGNOSTICS)

Sequagel™ buffer	10 mls
Sequagel™ concentrate	24 mls
Sequagel™ diluent	66 mls

Gels were made according to manufacturer's instructions. The above volumes produced gels containing acrylamide: bis-acrylamide (19:1), 5% Urea and 1X TBE buffer. Acrylamide gels were polymerised with 400 mls of 10% APS and 40 μls of TEMED for a period of 1 hour.

(Q) 5% Non-denaturing polyacrylamide gel (100 mls) (NATIONAL DIAGNOSTICS)

AccuGel™ solution	12.5 mls
Glycerol (100%)	5 mls
10X TBE buffer	10 mls
ddH_2O	72.5 mls

Gels were made according to manufacturer's instructions. The above volumes produced gels containing acrylamide: bis-acrylamide (29:1), 1 X TBE buffer and 5% glycerol. Acrylamide gels were polymerised with 400 mls of 10% APS and 40 μls of TEMED for a period of 1 hour.