The Systematics of the Anoa (Bubalus sp.)

Taxonomy, Biogeography and Conservation of Sulawesi's Endangered Buffalo

James A. Burton

Thesis submitted for the degree of PhD.

University of Edinburgh
May 2005

Declaration

I hereby declare that the work contained in this thesis was all my own work except for the following:

Chapter 4

The European cattle breed microsatellite allele data was supplied by P. Weiner / D. Burton (Roslin Institute). The Mongolian cattle and yak data was provided by T. Tsedev (The Institute for Biology of Mongolian Academy of Science).

Chapter 5

The PCR work for mtDNA sequencing was conducted by H. Greer under my supervision and the sequencing was completed by F. Murray of ARK Genomics, Roslin Institute. Assistance in constructing the neighbour joining tree was given by A. Hayward of the Institute of Evolutionary Biology (IEB), University of Edinburgh.

James Burton

May 2005

Acknowledgments

Firstly, thanks to my supervisors Alastair A. Macdonald, John Williams and Roy Wiles for their encouragement and support throughout my PhD. They were the people responsible for providing me with this opportunity by connecting the facilities and expertise to make this possible. For helping me to gain the laboratory techniques thanks to Deborah Burton, and all those in the lab. Also at Roslin, thanks to Anthea Springbett and Pan Weiner for statistical input, Andy Law and Dave Morrice for assisting in data preparation. Thanks to Hannah Greer, Alex Hastings (University of Edinburgh) and Frazer Murray (ARK Genomics, Roslin Institute) for their help in aspects of the mtDNA analysis. Comments about relevant statistical techniques were appreciated from Stephen Rossiter (Bristol University) and Simon Goodman (Zoological Society of London). For assistance in producing the maps I thank Ed Colijn and Colin Warwick.

There are a considerable number of people I am grateful to for their assistance in the acquisition of samples of the anoa and other species. Thanks goes to those people at European and Indonesian zoological collections who helped in sampling (Section 3.3.2). For their assistance, thanks goes to Phil Molyneux, Chester Zoo; Anna Plumb, Marwell Zoo; Dr Paul Vogt / Dr Wolfgang Dressen, Krefeld Zoo; Dr Ismianto, Ragunan Zoo; these samples have been analysed below. The help of those who have sent samples that are yet to be analysed is also appreciated: Charlotte Macdonald, Edinburgh Zoo; Les Radford, Whipsnade Zoo; Litasari Setiawan, Surabaya Zoo; Dr Roman Rehak, Decin Zoo. Samples from other species of Bovina (Section 3.3.3) were on loan from The National Museum of Scotland, Edinburgh by the kind permission of Dr A. Kitchener, donated from The University of Utrecht by Dr H. Lenstra / Dr I. Nijman and from Roslin Institute by Dr J. Williams.

The difficult work of collecting anoa samples from Sulawesi was made feasible by the collaborators from local conservation groups. The help of the following was greatly appreciated: Moh. Yasin; Andi Faisal Alwi; Yusran Zainudin and Jabar Lahadji; Kaharudin Kasim; Vitri and Piton; Yani Mile and Ridwan; Imran, Ading, A'o; La Ode Nafiu; Hally Day; Margaretha Pangau; Arfan Polontalo and Jemi Monoarfa. For their collaboration and logistical assistance in Indonesia thanks goes to numerous people, especially Steve and Puji Oliver, Duncan Neville, Edward Pollard, Ben Farrer, John Riley, and also in memory of Simon Dandun.

For permission and support to conduct this work I would also like to thank: The Indonesian Ministry of Forestry, Jakarta (PHKA), Sulawesi's Provincial Forestry Departments (BKSDA); The Indonesian Institute of Science (LIPI); Museum of Zoology, Research Center for Biology, Cibinong (LIPI); and the project's long-standing Indonesian sponsor, Ir. Harayanto MS, Bogor Agricultural University (IPB). The friendship, advice and enthusiasm of my counterpart, Abdul Haris Mustari, in our work on the anoa were great encouragement.

This work was supported by the University of Edinburgh Development Trust, the Roslin Institute, and the Stichting Dierentuin Helpen (Consortium of Dutch Zoos). The Rufford Small Grant, Royal Geographical Society, London and Royal Zoological

Society of Scotland and The University of Edinburgh Birrell-Gray Travel Award supplied additional funds.

Thanks to my friends who supported me through the writing phase, especially Sankari, Cathy, Bea, Steph, Ricardo and Duncan and also to my family for their encouragement.

Abstract

The endangered anoa, Sulawesi's dwarf buffalo, has suffered from unfocused conservation management because of taxonomic uncertainties. Analyses of skull morphology, karyotype, allozymes and mtDNA have produced conflicting results; the current classification identifies two species, the lowland and mountain anoa *Bubalus depressicornis* and *B. quarlesi,* respectively. To examine whether distinct phylogenies exist, 92 *in situ* individuals were sampled from throughout their range (Sulawesi and Buton Islands). Individuals in the zoo population are of unknown origin; eleven animals were included in the analysis to assess the level of genetic variation represented in this population relative to the wild. The assessment of *in situ* population genetic structure and assignment of *ex situ* animals was conducted using bovine microsatellite markers and a single mtDNA D-loop sequence. Of the 30 bovine microsatellite markers assessed, 13 amplified and were polymorphic in the anoa and were used in the study.

The analysis of microsatellite marker data indicated that anoa populations are structured geographically giving an F_{ST}= 0.3491. Additional methods corroborated the existence of four geographically distributed subpopulations, including assignment analysis, Hardy-Weinberg equilibrium tests, allelic variation and multivariate analysis. D-loop data corroborated the correlation of anoa population structure and biogeographic regions, but defined three clades.

Genetic distances calculated from microsatellite data between pairs of geographically differentiated anoa populations showed similar $(\delta\mu)^2$ distances or smaller Nei's distances than those observed between clearly different bovid species (domestic cattle *Bos taurus* and Mongolian yak *B. grunniens*). Therefore, using the indicator of dissimilarity species definition there was not clear support for the single or multiple species hypothesis. The genotypic cluster species definition (microsatellite alleles) supported the existence of a single species. In contrast the cohesion species concept (mtDNA) suggested multiple anoa species exist. Therefore, although the previous taxonomic classification was refuted by the lack of a correlation between genetic variation and either species morphotypes or altitude, a new taxonomic status could not be confirmed.

Individuals from the zoo population were found to be only partially representative of the wild population; greater genetic variation was found in the Indonesian zoo anoa than the European animals.

The population genetic structure was partially explained by isolation by distance (microsatellite and mtDNA) and barriers to gene flow. Microsatellite genetic variation was significantly correlated with the sea barrier between Buton and Sulawesi Islands. In this location the biogeographic barrier acted as a stronger dispersal barrier than predicted by geographic distance alone. However, on Sulawesi Island analyses showed that both known biogeographic barriers and geographic distance contributed to the genetic variation, but tests could not confirm which had the greater effect.

The anoa was defined as a single evolutionary significant unit (ESU) because although the genetic exchangeability has been limited there was a lack of regional ecological adaptation. Conservation strategies of *in situ* anoa should maintain gene flow between networks of protected populations to preserve a minimum viable population (MVP).

Contents

List of Figures

List of Tables

List of Abbreviations

ABI	-	Polyacrylamide gel
aDNA	-	Autosomal DNA
AOEs	-	Areas of endemism
APS	-	Ammonium persulfate, for polymerization of polyacrylamide gel
a_r	-	Statistic for assessment of effect of isolation by distance on genetic distance
BKSDA	-	Sulawesi's Provincial Forestry Departments, Office of Conservation Unit (Balai Konservasi Sumber Daya Alam)
bp	-	Base pair, number of base pairs present in microsatellite loci
CITES	-	Convention on International Trade in Endangered Species of Wild Fauna and Flora
D-loop		Displacement loop section of mitochondrial DNA
D_1	-	An SMM model
DMSO	-	Dimethyl sulphoxide
dNTPs	-	deoxynucleotide triphosphates
Dps	-	'Proportion of shared alleles' distance measure
EAZA	-	European Aquarium and Zoological Association
EDTA	-	ethylenediaminetetraacetic acid
ESU	-	Evolutionary significant unit
EtOH	-	Ethanol
ex situ	-	Biodiversity component outside their natural location
FAO	-	Food and Agriculture Organization of the United Nations

F_{IS}	-	F statistic to determine the probability that two alleles selected from the same individual are identical by descent
F_{ST}	-	F statistic to determine the probability that two alleles selected from the same sub-population are identical by descent, based on the IAM.
He	-	Expected heterozygosity
Ho	-	Observed heterozygosity
HWE	-	Hardy-Weinberg equilibrium
IAM	-	Infinite allele model, a microsatellite mutation model
IBD	-	Isolation by distance
in situ	-	Biodiversity component in their natural location
IPB	-	Bogor Agricultural University
IUCN	-	The World Conservation Union
K	-	Number of (unknown) populations in analysis with program STRUCTURE
LIPI	-	The Indonesian Institute of Science (Lembaga Institut Penelitian Indonesia)
MCMC	-	Markov chain Monte Carlo
m asl	-	Meters above sea level
$MgCl_2$	-	Magnesium chloride
mM	-	Millimolar
MU	-	Management unit
MVP	-	Minimum viable population
mtDNA	-	Mitochondrial DNA

Mya	-	Million years ago
n	-	Sample size
N	-	Population size
NaCl	-	Sodium chloride
nDNA	-	Nuclear DNA
Ne	-	Effective number of alleles
N_e	-	Effective population size
N_em	-	Effective number of migrants
NGO	-	Non-Governmental Organisation.
NJ	-	Neighbour-joining (phylogenetic tree)
Nm	-	Number of migrants per generation
OTUs	-	Operational taxonomic units
PC	-	Principal component
PCA	-	Principal components analysis
PCoA	-	Principal coordinates analysis
PCR	-	Polymerase chain reaction
PHKA	-	Forest Protection and Nature Conservation, Indonesian Ministry of Forestry, Jakarta (Perlindungan Hutan dan Konservasi Alam)
P(H)VA	-	Population (and habitat) viability analysis
ResGen	-	A permanent inventory of European farm animal genetic resources and of activities on characterization, conservation and utilization of these resources.
rpm	-	Revolutions per minute

R_{ST}	-	A statistic similar to the F_{ST}, but based on the SMM mutation rates
SDS	-	Sodium dodecyl sulfate, for hair lysis buffer
SMM	-	Stepwise mutation model, a microsatellite mutation model
TAMRA	-	A lane marker for polyacrylamide gel
TBE	-	Tris – borate – EDTA, electrophoresis buffer, for polyacrylamide gel
TEMED	-	N,N,N,'N'-tetramethylethylene-diamine, electrophoresis reagent, for polymerization of polyacrylamide gel
TRIS	-	Buffer for Proteinase K DNA extraction
q	-	Highest percentage of membership during analysis with program STRUCTURE

Chapter 1

Introduction: The Anoa, Species Concepts and the Applications of Genetics Tools

1.1 Overview

The goal of the study is to define the taxonomic status and the units of conservation management for the anoa, the wild bovid from Sulawesi, Indonesia, first described by Hamilton Smith (1827) as *Antilope depressicornis*. The use of microsatellite markers and mitochondrial DNA D-loop sequences, the preferred techniques to assess genetic variation will be central to this. Two species of anoa are presently recognized: the lowland anoa (*Bubalus depressicornis*) and the mountain anoa (*Bubalus quarlesi*) (Groves, 1969; Corbet and Hill, 1992; Wilson and Reeder, 1993). However, this classification has come under some scrutiny because various studies have suggested differing numbers of species (Groves, 1969; Weise, 1979; Schreiber, Nötzold and Held, 1993; Schreiber, Seibold, Nötzold et al., 1999). These two species are currently represented in the European zoo population as distinct breed lines. The International Anoa Studbook states 'they appear to comprise subspecies' because 'they appear to be borderline cases of true species' and so they are called a single species, *Bubalus [Anoa] depressicornis* (Nötzold, 1999). The zoo animals nevertheless show considerable genetic and chromosomal variation (Schreiber et al., 1993; Schreiber et al., 1999).

The primary focus of this thesis is to identify the degree of genetic structure of the *in situ* population. This will encompass assessments of the level of differentiation or introgression between any clusters of individuals identified. Information on the degree of genetic differentiation between clusters of individuals is essential for the accurate taxonomic classification of populations to species. This is also important for the identification of units of management for future population conservation measures.

The results of microsatellite marker analysis will define the representation of the genetic diversity maintained in the zoo population. The study will also address the classification of the number of species of the anoa by testing for a correlation of genetic variation with descriptive characteristics, geographic distance and biogeographic barriers.

A set of 13 microsatellite markers were used to analyse samples from 89 wild anoa individuals, ten zoo anoa individuals and an outgroup set of samples. Also mitochondrial DNA D-loop sequence analysis was conducted for 35 wild anoa and one zoo anoa. The origins of *in situ* individuals sampled were representative of the anoa's range, allowing assessment of genetic structure of the anoa population related to geographic distance and possible barriers to gene flow. This broad sampling is important when considering the complex phylogeographic history of the region (Whitten, Henderson and Mustafa, 2002).

The analysis of microsatellite data was conducted using traditional techniques such as genetic distance methods and new methods from the burgeoning field of likelihood analysis. Mitochondrial DNA sequence variation was assessed using phylogenetic analysis. This study incorporated *in situ* and *ex situ* individuals to allow interpretation of the management requirements for the future survival of the anoa.

The thesis is divided into seven chapters, two introductory chapters summarising species concepts, genetic techniques, and background on the anoa, biogeography and conservation in the anoa's range, the Wallacea region in Indonesia. Then the sampling strategy and genetic methodologies are outlined. The results have been divided into microsatellite analysis (Chapter 4), mitochondrial DNA analysis (Chapter 5) and combined analysis of all available informative data (Chapter 6). These results are discussed, including the implications for anoa conservation in the final chapter.

1.2 Introduction

The anoa or dwarf buffalo are the smallest of the Bovini; they stand about a metre tall at the shoulder and weigh 150-200kg. Anoa are forest dwelling browsers and are endemic to the two largest islands of Sulawesi and Buton in the Wallacea region of Indonesia. This region is one of 25 hotspots of high species endemism and species under a high degree of threat of extinction (Myers, Mittermeier, Mittermeier et al., 2000). Protected areas on Sulawesi and its offshore islands cover 2,223 square kilometres or 11.9% of the area, and 82% of the area under protection is reportedly still forested (Aden, Dore, Vincent et al., 2001). The anoa is classed as Endangered by the World Conservation Union (IUCN) (http://www.redlist.org/, August 2004) because of the reduced population size due to hunting and habitat degradation (Burton, Hedges and Mustari, 2005). Further information on the conservation status of the anoa, a summary of its taxonomic status and the biogeography of Sulawesi Island are presented in Chapter 2.

The maintenance of genetic diversity is the main aim of biodiversity conservation (Frankham, 1995). This is achieved by preventing the loss of species and the loss of diversity within species, which can be crucial to preventing extinction. The best surrogate indicator of extinction risk was found to be population size (mature individuals), along with population trend (% change in population size over 10 years / three generations) (O'Grady, Reed, Brook et al., 2004). The basic unit for grouping genetic variability for conservation has traditionally been the species, most frequently using the biological species concept (BSC), defined as the inability of individuals from different species to interbreed (Mayr, 1969). The recent development of molecular genetic techniques has facilitated direct assessment of genetic diversity, rather than indirect assessment of isolation between populations using methods such as morphological differentiation. For example, the red wolf was classed as a distinct species from the grey wolf and the coyote of North America until microsatellite markers and mitochondrial DNA (mtDNA) studies were conducted. Then the red wolf was described as a hybrid of the grey wolf and coyote (Wayne, 1996). However, a recent study using mtDNA and microsatellite markers suggested a further population in east Canada, the Algonquin group, was genetically more similar to the red wolf than the grey wolf or coyote, so adding further detail to the earlier hybrid theory (Wilson, Grewal, Lawford et al., 2000). The wealth of genetic information and other developments have led to the proposal for further species concepts and definition of units for managing groups of populations. This benefits conservation efforts by adding a more detailed scale and therefore increasing the effectiveness of management recommendations that can be made. However, there is much discussion about the most suitable species definition and conservation management units (Agapow, Bininda-Emonds, Crandall et al., 2004). A number of these are discussed below in relation to this study.

1.3 Species concepts

1.3.1 The variation of species concept definitions

As described by Agapow et al. (2004) 'species are the currency of biology'. The species as an evolutionary unit is used in the fields of systematics and evolutionary biology. Although conservation biology is based on species, until recently there has been little debate as to the definition of species related to this field (Rojas, 1992). The binomial naming system of Linnaeus initiated around 250 years ago, defines 'species' as units which remain the basis for classification of biological diversity (Linnaeus, 1753, 1758). The relevance of this system for use in conservation biology is currently being debated as genetic techniques replace morphological characteristics as the methods of choice for defining species (Godfray, 2002). This is accompanied by a shift in thinking from conserving the units of biodiversity to defining and conserving the evolutionary lineages (Avise and Wollenberg, 1997).

There are at least 22 definitions for species as a concept (Mayden, 1997), illustrating the considerable debate and the remaining uncertainty as to the most appropriate definitions (Claridge, Dawah and Wilson, 1997; Wheeler and Meier, 2000). The differences between these concepts have been reviewed (Claridge et al., 1997; Mayden, 1997; Wheeler et al., 2000). Conservation requires working definitions that are applicable to individuals with different breeding strategies and taxonomic affinities. However, to date, the concepts proposed have been assessed for their theoretical and operational qualities (Cracraft, 1989), without the establishment of one primary definition of what constitutes a species (Mayden, 1997).

The employment of contrasting concepts of two species means there is little relevance in comparing genetic diversity of the two species because they were defined by different operational parameters (Goldstein, Desalle, Amato et al., 2000). The assessment of different definitions for the same organisms have defined varying boundaries (Agapow et al., 2004). This has caused significant difficulties for conservation efforts. For example, the phylogenetic species concept, defined species as a group of individuals that share at least one uniquely derived characteristic (Nixon and Wheeler, 1990), has been criticized for the division of organisms into too many species in contrast to the biological species definition (Agapow et al., 2004). A review of 89 studies found that the total number of non-phylogenetic plant and animal species was between 1245 and 1282, whereas when the same species were reclassified under the phylogenetic species concept the number increased to between 1912 and 2112 species (Agapow et al., 2004). The implications are that this will cause a significant rise in the number of endangered species compared to previous classification. This is due to the decline in the number of individuals identified to each species and the likely range reduction of these species. The differences in the total number of species using the different criteria for species definition has been attributed to the over-division of genetic variation using the phylogenetic parameters for species definition and the clumping or underestimation of genetic divisions of distinct populations using the biological parameters to distinguish species (Agapow et al., 2004).

Alternatively, it may be possible to employ multiple criteria of describing species and thus reinforce species boundaries (Jarman and Elliott, 2000). When different parameters support the same species distribution, then consistent classifications strengthen the argument. The assessment of multiple species concepts is preferable to using a narrow definition, as variations in mammalian reproduction types can be classified in different ways (Corbert, 1997).

The difficulty of defining species is also due to the frequently blurred species boundaries, hybridisation, recent isolation, and horizontal gene transfer (Barton and Hewitt, 1989; Agapow et al., 2004). A clear definition of the parameters employed should be given, with interpretations of multiple definitions, where this could assist, or impact on management decisions.

1.3.2 Selected species definitions

The 22 definitions of species concepts previously described (Mayden, 1997), vary in their practical or theoretical applicability to this study. Certain concepts are omitted from discussion here because they were inapplicable or not relevant to the analysis conducted. For example, the evolutionary species concept (Waples, 1991) was also not employed because it is used to define a group of organisms that maintain their identity over time and space (Mayden, 1997). Although it has been suggested that this concept is the closest definition of entities that conservationists wish to protect, conservation requires an operational concept, unlike the primarily theoretical evolutionary species concept of Sites and Crandall (1997). The biological species concept also defines organisms close to the conservationists' goals. This is the first of a number of species concepts discussed below.

The biological species concept (BSC) (Mayr, 1963) has been the most widely used for mammalian studies. The US Endangered Species Act [USFWS] is based on this concept (1994). The definition has been recently described as 'groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups' (Mayr, 1991). Alternatively, other definitions have identified individual organisms as the individual units, not species (Templeton, 1989).

Though it has been widely used, the applicability of the biological species concept to conservation management has been hindered by the difficulty of testing isolation or interbreeding between populations, especially between parapatric species (Vogler and Desalle, 1994). Instead, an indirect method, using morphological characteristics, has been used to define whether populations are interbreeding; by the identification of intermediates with the morphological characteristics used in the definition. Invariably, different criteria have been used such as detailed skeletal measurements or colour and pattern variation (Cracraft, 1989; Templeton, 1989). It should be noted that the morphological characteristics used for identification of a species, might not be related to genetic or evolutionary processes. One of the greatest difficulties in accepting the biological species concept is that it involves a circular argument. Wallace (1865) defined this flaw when investigating the *Papilionidae* butterflies in Indonesia. His criticism was that this concept confuses the cause and effect. Theories used to define the process of speciation involve the reduction in ability or tendency to interbreed, while species themselves under the biological species concept are defined by the presence or absence of interbreeding. These significant disadvantages to the biological species concept mean that it was not used as the sole concept in this study.

The phylogenetic species concept is defined as a group of organisms that share a minimum of one unique derived characteristic, often including a shared pattern of ancestry or decent (Cracraft, 1983). The proposed advantages for this species definition are that it is more objective than others, and in so being, could identify important population divisions that are not defined by morphological differentiation. Though this has been a widely used definition, it has also been criticized. This is because the recent increase in the number of molecular techniques has allowed the identification of many more characteristics than before, such that a large number of 'new species' have been proposed that may have limited conservation significance (Agapow et al., 2004). Therefore, this species definition was not used for data interpretation below.

An alternative to the definitions described above is the genotypic cluster definition (Mallet, 1995). The null hypothesis states that there are multiple independent clusters of alleles defining each species for single and multiple loci. This is very similar to the 'genealogical concordance' method (Avise and Ball, 1990) that uses genetic information to address species classification. Molecular characteristics are used to identify where intermediate forms exist (single species) and where they do not (multiple species). The genotypic cluster definition has been shown to define groups of predominantly clonal individuals more accurately than gene flow or recombination methods (Cohan, 1994). This model may also be affected by chromosomal variation between organisms (Mallet, 1995). Molecular characterisation using this species definition has been included in the assessment of the anoa populations.

The cohesion species concept (Templeton, 1989) differs from the genotypic species definition because evolutionary history is incorporated as statistically significant lineage associations, whereas the genotypic cluster definition only considers current patterns of variation. The cohesion species concept is defined as 'the most inclusive group of organisms having the potential for genetic and/or demographic exchangeability' (Templeton, 1989). This species concept requires mtDNA monophyly, and has been used in the analysis presented here.

1.3.3 The scale of distinctiveness and conservation units

An alternative to species definitions is the use of genetic and morphological differentiation as indicators of dissimilarity between populations. Frequently, genetic distance, which is often defined using mtDNA sequence, has been used as an indicator of distinctiveness for evaluation of conservation status (Avise and Nelson, 1989). However, the use of this method as a scale of distinctiveness was found to be more applicable between closely related species than between distantly related species. The difficulty with assessment of the distantly related species is because there is an overlap in the level of sequence divergence, between populations and species from different taxonomic groups (Dizon, Lockyer, Perrin et al., 1992). Nevertheless, comparison between closely related species and populations, indicated that this method might be more applicable in these circumstances (Xu, Gullberg and Arnason, 1996b).

Molecular methods have been used to define new species with the indicator of dissimilarity criteria (Xu and Arnason, 1996a; Xu et al., 1996b), but there have also been criticisms of the approach because of the possibility of erroneous classification caused by the significant weight put on the species selected as a scale for comparison (Goldstein et al., 2000). Considering this criticism, the best use of the indicator of dissimilarity may be to allow comparison with the results of one of the species definitions above, when studying populations of closely related species.

The biological species concept has been criticized because it does not define taxonomic diversity to a high enough resolution (Cracraft, 1983). With the increase in detail provided by genetic techniques there has been greater interest in the intra-specific examination of biodiversity for the identification of groups of individuals in need of conservation (Avise, 1989). This has led to the use of mtDNA sequences to imply historical connections between populations, and has been defined as 'intra-specific phylogeography' (Avise, Arnold, Ball et al., 1987).

The minimal population 'units' for conservation prioritisation and management are now categorised using variation between nDNA as well as mtDNA. These units have been termed 'evolutionary significant units' (ESUs) sensu Ryder (1986) and 'management units' (MUs) (Moritz, 1994). One definition of ESUs by Moritz (1994) states that they are monophyletic. A disadvantage in using ESUs is that they could be inappropriately identified at the level of individuals (Moritz, 1994; Vogler et al., 1994). This is often due to an over-reliance on genetic data to the exclusion of morphological and other data types. This shift towards genetic data has also reduced the use of information on gene flow. Because of these shortfalls in the proposed definitions of ESUs there has been a lack of agreement and resulting confusion over their definition (Crandall, Bininda-Emonds, Mace et al., 2000); Vogler et al. (1994) even suggested defining ESUs as phylogenetic species due to the similarity in results produced by both criteria. Recently, a further definition of ESUs including both genetic and ecological data has been developed by Crandall et al. (2000) in an attempt to correct these shortfalls. This is described below.

The potential for species survival is maximized through the maintenance of adaptive diversity, by preserving the maximum diversity of functionally divergent gene copies across the geographic range. This is also the goal of conservation management. Crandall et al. (2000) proposed that both genetic and ecological exchangeability should be included in conservation units. These units would be best able to maintain evolutionary processes and so in the future permit adaptation to changing environmental conditions. This definition of conservation units should also include assessment of historical versus recent exchangeability. Multiple categories are required to represent the continuous habitat types, selective pressure and genetic diversity across populations, rather than the two categories of 'ESU' or 'not ESU' as previously proposed by Moritz (1994). In the definition suggested by Crandall et al. (2000), the null hypothesis of exchangeability is rejected or accepted for genetic and ecological data that is representative of a recent or historical timescale. These factors make this conservation unit the most practical for management purposes.

In summary, there is currently no single operational species concept applicable to all species; instead the use of multiple species definitions appears to be preferable to a single definition. Identified species are afforded legal protection, such as under the US Endangered Species Act [USFWS] (1994). The definition of other population units as distinct from species to define intra-specific phylogeny has the benefit of permitting management of smaller differentiated populations.

1.4 The use of genetics in taxonomy and conservation

1.4.1 Introduction

Molecular genetic techniques have now become a versatile tool for conservation biologists in their attempts to minimize the number of species going extinct. This includes resolving taxonomic uncertainties and population structure, identifying management units (ESUs or MUs), detecting hybridisation between populations, determining the effects of small population size (e.g. inbreeding) and examining the many implications of reintroducing species that have gone extinct in particular locations (Frankham, Ballou and Briscoe, 2002). A number of these uses for genetic techniques have been employed in the work described here.

Various genetic techniques have been designed for the study of evolution and population variation, such as karyotyping, and the use of molecular markers, including blood groups, allozymes, mitochondrial DNA (mtDNA) sequences and nuclear DNA (nDNA) variation. All of these markers have been used to estimate a date of evolutionary divergence of species by defining the mutation rate and hence the concept of a molecular clock (Nei, 1987). Karyotype analysis can be informative for exploring reproductive isolation or the level of gene flow. Variations, such as difference in chromosome number, inversions or translocations can prevent successful reproduction, or fertility of progeny (Benirschke and Kumamoto, 1991). The difficulty of applying this method to field surveys is the availability of suitable samples and the need for rapid karyotyping of samples after collection, which is often not possible. Protein electrophoresis uses the variation in the net charge and size of proteins that migrate through a gel at different rates. Examples of this are starch, agarose and polyacrylamide gel electrophoresis (PAGE) (Murphy, Sites, Buth et al., 1996), and isoelectric focussing. Also, SDS-PAGE compares apparent molecular mass of proteins, which are coated in SDS to remove charge variations. These methods have the advantage that they may define functional variation of proteins. However, the disadvantage of this method is that because only protein coding loci are surveyed, a limited amount of the genomic variation is identified.

More recently, many nuclear DNA (nDNA) markers have been described for assessing variation between individuals. Initially, these were based on DNA hybridisation methods e.g. DNA fingerprinting using multi-loci hypervariable mini-satellite repeats (Jeffreys, Wilson and Thein, 1985), and later the detection of allele size variation at microsatellite loci using the polymerase chain reaction (PCR). The first two techniques have limitations in interpretation from the analysis. For example, karyotyping may show variation in chromosome number, but the relationship of this and species definition is as yet undefined (Ryder, Kumamoto, Durrant et al., 1989). Although used for assessment of variation between European cattle breeds (Blott, Williams and Haley, 1998), blood group polymorphism does not provide the level of variation of hypervariable loci, which is vital for assessing closely related populations (Bowcock, Ruiz-Linares, Tomfohrde et al., 1994). Hypervariable mini-satellite loci have been extensively used in population studies, but less so for investigations of endangered species because of unresolved issues, such as correlations among loci due to linkage (Dowling, Moritz, Palmer et al., 1996). Microsatellite allelic variation has proved to be an increasingly popular technique as this is a PCR based approach. Additional advantages are their broad distribution in the genome and the characteristics of these repeat sequences, described in the section below, (Goldstein and Pollock, 1997; Goldstein and Schlötterer, 1999). A summary of microsatellite use is given in Section 1.4.2 below.

Mitochondrial DNA (mtDNA) sequence variation has been used in a considerable number of conservation studies, a major reason being that this region evolves exceptionally rapidly (Avise, 1994). The control region of mtDNA is frequently sequenced for a number of other reasons outlined in Sections 1.4.4 and 1.45.

1.4.2 Simple tandem repeats: microsatellite allele size

Microsatellites are simple tandem repeats of short sequence motifs of DNA. These are defined as being made of between two and five nucleotides. The most common dinucleotide repeat is CA/TG (Backmann and Weber, 1992). The significance of microsatellite loci as genetic markers is that the number of repeats of the sequence motif is highly variable in populations, but is inherited in a Mendelian fashion. These markers are distributed across all eukaryotic genomes that have been studied and are highly polymorphic (Weber and May, 1989). The significant levels of polymorphism allow the assessment of variation between populations of the same species as well as between different species. The process is relatively cost effective and rapid using the PCR method (Avise, 1994). The microsatellite markers have now been developed for many species and are usually the marker of choice for genetic diversity and genome mapping studies (Moore, Sargeant, King et al., 1991; Coote and Bruford, 1996; http://www.projects.roslin.ac.uk/cdiv/markers.html, January 2004).

The mutation mechanism of microsatellite loci is relevant to understanding the application of this genetic technique. The estimation of mutation rate of microsatellite loci varies between species and across different loci within a species, and ranged from 10^{-2} to 10^{-5} per generation (Levison and Gutman, 1987; Henderson and Petes, 1992). Human microsatellite loci were reported to have an average mutation rate of 10^{-3} per generation (Weber and Wong, 1993).

Two mutation mechanisms have been proposed to explain the variation in repeat number at a particular locus: slip strand mispairing (slippage) occurring during DNA replication; and recombination between DNA molecules (Levison et al., 1987). Therefore, with a higher number of repeats at a locus, the greater the opportunity for slippage and recombination, and hence the greatest possibility for variation in allele size.

The recombination mechanism is a second process that may explain microsatellite locus polymorphism. There are two ways that recombination could alter the lengths of microsatellite sequences: unequal crossing over or gene conversion. Of the two mechanisms, recombination is thought less likely to be responsible for microsatellite mutation. Levison et al. (1987) showed the length instability of tandem repeats in *E. coli* is unaffected by recombination deficient mutants.

The evolutionary dynamics of microsatellite loci have been assessed with a number of theoretical and empirical methods (Jarne and Lagoda, 1996). Although the single repeat sequence is the basis of microsatellite loci variation, other factors increase the evolutionary complexity of microsatellite sequences. These include differences in allele length (Bowcock et al., 1994) and the discovery that mutation sometimes involves multiple repeat units (Weber et al., 1993; Di Rienzo, Peterson, Garza et al., 1994). In other words, although the most frequent variation in allele length is by one repeat unit, there are other variations that occur occasionally.

To describe microsatellite polymorphisms, theoretical models were designed for comparison of microsatellite data (Goldstein, Ruiz-Linares, Cavalli-Sforza et al., 1995b; Slatkin, 1995). The infinite allele model (Kimura and Crow, 1964) and the stepwise mutation model (Kimura and Otha, 1978) are two examples. These methods are generally supportive of the slippage mechanism, rather than the recombination mechanism. However, there is considerable variation between the two models, because of this they are classed as two extremes of microsatellite mutation. These models have received wide use after the development of PCR technologies for microsatellites (Litt and Luty, 1989; Tautz, 1989; Weber et al., 1989). They are described in more detail in Section 4.3.4.

The function of microsatellites is still unclear. It has been suggested that they have a role in recombination (Pardue, Lowenhaupt, Rich et al., 1987). However, this has been debated (Tautz and Renz, 1984).

Statistical methods for the analyses of microsatellite data are developing rapidly, with the most commonly applied methods for retrieving information being maximum likelihood, coalescent and Bayesian approaches (Luikart and England, 1999). Statistical methods for analyzing population structures using microsatellite marker data are described in Chapters 4 and 6, and examples of the application of microsatellite loci are discussed below in Section 1.4.3.

1.4.3 The use of microsatellite markers

Since their initial discovery, microsatellite markers have been developed for a wide range of domestic species including cattle, pigs, horses and dogs (Laval, Iannuccelli, Legault et al., 2000; Cunningham, Dooley, Splan et al., 2001; Parker, Kim, Sutter et al., 2004). They have been used as highly informative markers in linkage studies, especially for localising qualitative trait loci (QTL) that are important in livestock improvement. For example, the QTLs have been identified for cattle milk quality (Wiener, Maclean, Williams et al., 2000), for pig fatness levels, ovulation rate and growth rate (Walling, Archibald, Cattermole et al., 1998), and fatness in chickens (Ikeobi, Woolliams, Morrice et al., 2002). The European cattle diversity database (http://www.projects.roslin.ac.uk/cdiv/markers.html, January 2004) contains data from a study of cattle using 30 microsatellite loci to explore the breed history among European cattle. These 30 loci were selected from a considerable database of hundreds of loci for certain characteristics; such as their position on different chromosomes, polymorphism in European cattle breeds, and other factors described in Chapter 3. The analysis of microsatellite data has allowed clustering of individuals into breed groups, and the definition of the relationship between the groups from the degree of allele sharing (MacHugh, Loftus, Bradley et al., 1994; MacHugh, Loftus, Cunningham et al., 1998; Wiener, Burton and Williams, 2004). Microsatellite and mitochondrial markers have also been used to assess the levels of introgression between cattle that have two distinct domestication origins and arrived in Africa via two separate routes: *Bos taurus* via the fertile crescent and North Africa and *B. indicus* via West Africa (MacHugh, Shriver, Loftus et al., 1997; Kumar, Freeman, Loftus et al., 2003), and between red and sika deer (*Cervus* *elaphus* and *Cervus nippon*) in Scotland (Goodman, Barton, Swanson et al., 1999). The bovine microsatellite markers have also been used to show the lack of genetic variation in a small closed herd of feral cattle, the Chillingham breed (Visscher, Smith, Hall et al., 2001), and the implications of disease on population distribution. For example, MacHugh (1996) identified a correlation between distribution of the disease resistant African N'dama cattle breed and the presence of tryposomiasis disease, while the non-disease resistant zebu cattle were found in areas where the disease was absent.

Microsatellite marker analysis is rapidly becoming an essential method for the taxonomic classification of species, often used in combination with sequence analysis of mtDNA. As discussed in the section above, a greater insight is obtained by combined microsatellite and mtDNA information. These two types of marker allow the assessment of varying timescales due to the differences in evolutionary rate of the two molecular markers. In addition differences in their mode of inheritance allow male and female dispersal to be explored. In this study both microsatellites and mtDNA datasets are analysed. One example of the use of combined datasets is the North American puma, which indicated a reduction in the number of proposed subspecies from eight to just a single distinct subspecies (Culver, Johnson, Pecon-Slattery et al., 2000). The genetic distance found between the wood and plains bison subspecies (*Bison bison athabascae* and *B. b. bison*) from microsatellite analysis was greater than that found between populations of the same subspecies (Wilson and Strobeck, 1999). In the study by Wilson et al. (1999) an 'indicator of dissimilarity' was their chosen method of assessing genetic differences between these groups of bison. It was proposed that plain bison individuals were introduced into wood bison populations, which the authors suggest would have reduced the genetic distance between the subspecies.

Conservation genetics has made wide use of the microsatellite markers developed in domestic species and transferred across to related species to investigate the structures of wild populations (Engel, Linn, Taylor et al., 1996; Nijman, Vastenburg, Williams et al., 1998; Slate, Coltman, Goodman et al., 1998). Some example applications of microsatellite markers are summarised below to demonstrate the type of analyses that are described in this thesis. Forbes and Hogg (1999) used microsatellite markers to assess the genetic distance between populations and between sub-species of Bighorn sheep (*Ovis canadensis canadensis*). They employed the genetic distance values as an indicator of dissimilarity between groups of sheep, and also to make comparisons with carnivore species, such as the brown bear (*Ursus arctos*), grey wolf (*Canis lupus*) and coyote (*Canis latrans*), distributed over the same range. The study's aim was to compare the genetic variation between populations and subspecies of the Bighorn sheep with that of other species inhabiting a continent wide range. The results showed that the sheep had a far lower genetic variation in F_{ST} than the carnivore species studied over the same geographic range, even though the distribution of the sheep was more patchy than the carnivores, suggesting greater gene flow in the sheep (Forbes et al., 1999).

The assessment of populations with unknown structure and the definition of management strategies have been identified by using genetic markers to assess the relative effects of migration and drift between populations. For example, this approach has been adopted for the analysis of island populations of the Komodo dragon (*Varanus komodoensis*) (Ciofi, Beaumont, Swingland et al., 1999a). Results showed gene flow between Rinca, Gili Motang and Flores west populations, compared to the historical isolation of Komodo Island and only recent isolation of Flores north due to increased human presence. This highlighted the management priorities to preserve the genetically distinct Komodo population and to maintain gene flow between the Flores populations. It has also been used across the ranges of subspecies with undefined migration barriers such as the American puma (*Puma concolor*) and the jaguar (*Panthera onca*) (Culver et al., 2000; Eizirik, Kim, Menotti-Raymond et al., 2001). On a larger taxonomic scale microsatellite markers were used to determine the phylogenetic relationships of the tribe Bovini (Ritz, Glowatzski-Mullis, MacHugh et al., 2000).

Fragmented populations are becoming increasingly common because of habitat fragmentation and hunting pressure, so more species require active conservation measures. The effective sizes of fragmented populations have been estimated using microsatellite markers to determine whether the fragmentation was historic or has occurred recently (O'Ryan, Harley, Bruford et al., 1998; Ciofi and Bruford, 1999b; Goodman, Tamate, Wilson et al., 2001; Cegelski, Waits and Anderson, 2003). For example, four populations of African buffalo sampled by O'Ryan et al. (1998) showed a correlation between population size and genetic variation. Using likelihood analysis O'Ryan et al. (1998) were able to predict that two populations would maintain genetic variability in the future, while two other smaller populations required regular introduction of individuals from an identified population. Results of these studies could be used to identify the more appropriate management strategies of either interference or non-interference in gene flow between the populations. In other words, whether it makes more sense to manage the fragments as a single or as multiple populations.

Microsatellite loci have also been used to assign individuals from unknown locales to their population of likely origin. This has been shown to be particularly useful for the identification of species or populations when detecting wildlife poaching. The median correct assignment was 61% of individuals to species using the assignment test, and accuracy of assignment increased to 99.9% for highly differentiated populations (Manel, Berthier and Luikart, 2002). This method can also be adopted to identify the origin of zoo individuals of unknown locales to instigate more effective conservation breeding programs, such as was used for the Whooping Crane (*Grus americana*) (Jones, Glenn, Lacy et al., 2001).

The use of microsatellite markers was chosen as the optimal technique for defining genetic variation within the anoa. These markers allowed assessment of the taxonomic status, population structure, gene flow levels, and possible recent fragmentation effects within the anoa.

1.4.4 Mitochondrial DNA sequence variation

The cattle mitochondrial DNA is a 16,338 base pair region of DNA (Anderson, Bruijn, Coulson et al., 1982), which is involved in the production of energy and its storage. Within this is the D-loop or control region (908bp sequence) that contains the signals that control RNA and DNA synthesis. It is also know as the hypervariable region because of the accumulation of point mutations at approximately ten times the rate of nuclear DNA (Palumbi, 1996). Other regions frequently sequenced include the cytochrome *b* region and 12S and 16S rDNA regions. The D-loop region was chosen for sequencing here, being the most variable portion of mtDNA it is more likely to define genetic variability between closely related populations (Lau, Drinkwater, Yusoff et al., 1998). Although the cytochrome *b* region had been used in previous studies of zoo anoa (Tanaka, Solis, Masangkay et al., 1996; Kikkawa, Yonekawa, Suzuki et al., 1997; Schreiber et al., 1999), the greater detail given to this study by using a more variable region was thought to be more important than comparison with the sequence of zoo animals from previous studies. The D-loop has been used in a number of studies of wild and domestic bovid species (Mannen, Tsuji, Loftus et al., 1998; Simonsen, Siegismund and Arctander, 1998; Ward, Bielawski, Davis et al., 1999; Troy, MacHugh, Bailey et al., 2001).

Mutation in mtDNA is due to the presence of a mutagen, possibly oxygen free radicals due to the proximity to the respiratory aspect of the cell. The resulting mutations are often a single nucleotide point variation (SNP). These variations can be compared between multiple individuals using phylogenetic analysis. This is only possible if positional homology exists, which occurs when the nucleotides that are being studied have originated from the same position in a common ancestor. The action of insertions and deletions during the mutation process requires alignment of sequences, prior to phylogenetic analysis (Swofford, Olsen, Waddell et al., 1996). Analysis using phylogenetic techniques is described in Section 5.3.

The importance of mtDNA sequence variation in genetic studies is largely due to the favourable attributes of this region of DNA. The fact that this is a sequence with a high mutation rate and so is highly variable, means that it is immediately attractive for assessing variation between closely related populations (Dowling et al., 1996). Sections of mtDNA sequence are commonly used because a complete nucleotide sequence of mtDNA is available from a number of species, which can be used to design universal PCR primer flanking regions of known variability (Kocher, Thomas, Meyer et al., 1989), such as the D-loop region (Attardi, 1985). The analysis of mtDNA is made less complex in most organisms because mtDNA is haploid (Avise et al., 1987), and hence the lack of recombination, which could obstruct the description of phylogenies. Finally, mtDNA is generally believed to be inherited through maternal lineages alone. This is transmitted in a non-recombining way, so sequences are common by descent. Therefore, this allows the separate exploration of male and female ancestry when combined with nuclear or Y chromosome specific markers. For example, the common warthog (*Phacochoerus africanus*) was shown to have a greater subdivision of the matrilineal line (F_{ST}=0.85) than the paternal lineage (F_{ST}=0.20) (Muwanika, Nyakaana, Siegismund et al., 2003).

The use of the mtDNA is limited in certain aspects. Firstly, the observed sequence differentiation may be a product of the lack of female dispersal, which may be in contrast to more widely dispersing males that maintain genetically undifferentiated populations. Therefore, without assessment of nDNA, incorrect assumptions can be made. A difference in the level of dispersal between the sexes of the common warthog mentioned above, could have been the reason for the variation in F_{ST} values between males and females (Muwanika et al., 2003). Selection may have varied effects in two populations due to unequal rates of evolution. Also, if the founder population is polymorphic before divergence, then genetic drift can mean that incorrect phylogenies are identified during analysis. Therefore, where taxonomic status is being defined, additional genetic analysis of nDNA is preferred (Frankham et al., 2002).

A number of fields have arisen or benefited from the development of this technology. For example, intra-specific phylogeography (Avise et al., 1987) is the study of population structure using mtDNA and geographic localities of samples. Also the field of coalescent theory developed to investigate gene flow among populations (Slatkin and Maddison, 1989). Inter-specific phylogenies are commonly assessed using mtDNA because of their relationships are important in the study of biogeography and speciation (Avise and Johns, 1999).

1.4.5 The application of mitochondrial DNA analysis

The application of mtDNA sequencing is widespread, from population studies, gene evolution to macroevolutionary studies for constructing species phylogenies, as reviewed by Avise (1994). The versatility of this method has been adapted to conservation applications and these are summarised below, with examples where relevant.

Firstly, mtDNA has had a large impact on taxonomic studies, being used to delimit proposed new species (Xu et al., 1996a), subspecies and the phylogenies of groups of species (Gatesy, Yelon, Desale et al., 1992). It has been proposed that the orang-utan *Pongo pygmaeus* of Borneo and Sumatra are two species, observed from the level of variation in mtDNA sequences compared across multiple primate species (Xu et al., 1996a). As a tool for inter-specific studies, cytochrome *b* sequence data was used to define the position of anoa relative to the genus *Bubalus*. The data presented in a maximum parsimony tree supported the status that the anoa should be was classified within the genus *Bubalus*, having a similar sequence to the Asian water buffalo *Bubalus bubalis* (Kikkawa et al., 1997). The anoa mtDNA data was not used here for taxonomic clarification.

The population structure of many species remains unknown, often for wild species this is because the structure is cryptic. In order to implement the correct conservation strategy, the genetic structure should be defined to allow allocation of units of management (Nersting and Arctander, 2001). For example, cytochrome *b* was sequenced to study the population structure of the impala *Aepyceros melampus* and greater kudu *Tragelaphus strepsiceros,* bush living African ungulates. The distinct genetic status of the South-western kudu population required recognition as being of greater conservation importance than was previously thought, as this formed a distinct clade, while the current population management for the impala management was sufficient.

Another study of the mtDNA structure of multiple African ungulates found new evidence for managing each species differently (Arctander, Johansen and Coutellec-Vreto, 1999). Three species, the hartebeest *Alcelaphus buselaphus*, the topi *Damaliscus lunatus* and the wildebeest *Connochaetes taurinus* inhabit a similar geographic region. However, their current population structures display two distinct phylogeographic histories. All these species had a pan-African distribution followed by isolation in small areas of their current range. The wildebeest was isolated in the southern regions of its range before the current population expansion, while the other two species were isolated in the north regions. An example of the opposite, concordance of population structure across diverse species was found in studies of Sulawesi species. Range fragmentation of multi-taxon has produced a similar phylogeographic pattern for macaques and toads (Evans, Morales, Supriatna et al., 1999; Evans, Supriatna, Andayani et al., 2003c). In this study analysis of the anoa data will assess the population structure of the *in situ* population.

Analysis of sequence variation has been successfully employed in forensics (Baker and Palumbi, 1996), in assessing introgression between domestic breeds (Cymbron, Loftus, Malheiro et al., 1999; Troy et al., 2001), and in determining coalescence times of populations. An example of the latter is the Asian water buffalo *Bubalus bubalis*. This is currently represented by two forms, the river and swamp type. These are now distributed throughout Asia, following domestication. Using D-loop and cytochrome *b* sequences the time of coalescence, location of the common ancestor and route of dispersal taken by domesticated animals of each type was estimated (Lau et al., 1998).

In this study, D-loop sequencing was selected to obtain a matrilineal perspective of the population structure of the anoa population, in Chapter 5. The mtDNA data was included in the analysis to view any deeper divergence than would be provided by microsatellite allelic data, by defining mtDNA population lineages.

1.4.6 The combined assessment of mitochondrial DNA and microsatellite analysis

As well as the additional matrilineal information that mtDNA gives there are other benefits to combining microsatellite and mtDNA analyses of the anoa. For example, a study of the warthog (*Phacochoerus africanus*) found that there was significant variation between individuals from 24 localities in Africa, while microsatellite data showed less differentiation between many population pairs (Muwanika et al., 2003). This difference between data types was explained by the high variability at microsatellite loci, meaning the mtDNA was a better representation of the population structure. Therefore, a second method has the advantage of being able to support or contradict the primary method for assessing genetic differentiation. The clustering of the Algonquin and red wolves (*Canis lupus*), distinct from the coyote and grey wolf was supported by both microsatellite and mtDNA data (Wilson et al., 2000). A final benefit of using microsatellite and mtDNA data is the feasibility of inference about population structures from historical and recent timescales because of differences in mutation rate between the two types of DNA (Palumbi, 1996). This can assist in defining ESUs (Crandall et al., 2000).

Chapter 2

Systematics and Conservation of the Anoa (*Bubalus* Sp.) and Sulawesi's Evolutionary History

2.1 Introduction

This chapter is divided into three sections. Firstly, previous taxonomic studies of the anoa will be summarized and related to the species concepts described in Chapter 1. This is followed by background information on the biogeographic status of the Wallacea region. The conservation status of the anoa will be described at the end of this chapter, including both *ex situ* and *in situ* populations followed by a statement of the aims of the study.

2.2 The taxonomic status of the anoa

2.2.1 Introduction

The anoa are currently described as two species: the lowland anoa *Bubalus depressicornis* and the mountain anoa *Bubalus quarlesi* (Groves, 1969; Corbet et al., 1992; Wilson et al., 1993). However, there is much debate over the taxonomic status and distribution of these two species (Mohr, 1921; Bohlken, 1958; Dolan, 1965; Groves, 1969; Weise, 1979; Kakoi, Namikawa, Takenaka et al., 1994; Pitra, Furbass and Seyfert, 1997; Schreiber et al., 1999). The designation as separate species by Groves (1969) was based on variation in skull morphological measurements and gross body characteristics with no intermediate forms observed (Figure 2.1). Even recent attempts to unravel the taxonomic mystery using genetic techniques with zoo animals have produced conflicting results (Schreiber et al., 1993; Schreiber et al., 1999). Therefore, conservation efforts by zoological institutions holding anoa and *in situ* managers and forestry departments are hindered by the lack of agreement about the species or taxonomic units to conserve (Burton et al., 2005).

The anoa has been proposed as a member of the most plesiomorphic clade of wild cattle (Bovini: oxen, bison, yak, and buffalo) (Geraads, 1992; McDougall, Brown, Cerling et al., 1992). The systematics of the anoa is also debated. The recent analysis of nuclear DNA (partial nucleotide sequences of the aromatase cytochrome P450 and lactoferrin genes) suggests that 'depressicornis' anoa are most closely related to *Boselaphus* not *Bubalus* and diverged around 12 to 14 Mya (Pitra et al., 1997). This contradicts the currently held classification, which indicates that *Anoa* diverged much more recently (*ca.* 2 Mya) from the ancestral *Proamphibos–Hemibos* lineage leading to the living Asiatic buffaloes (Groves, 1981).

Two of the concepts for defining conservation units, described in Chapter 1, have been used in previous taxonomic studies of the anoa. These are the biological species concept (Mayr, 1963) and the use of genetic distance as an indicator of dissimilarity (Avise et al., 1989). Differences in chromosomal complement have also been discussed as a form of possible speciation of anoa under the biological species concept.

The morphological characteristics of the anoa have been used to define species indirectly (Groves, 1969; Weise, 1979), according to the biological species concept (Mayr, 1963). These morphological characteristics were used to define isolated populations. However, they might not be related to genetic or evolutionary processes, meaning they may not be suitable as taxonomic indicators. Karyotypic variation has also been proposed as a way that species can be formed, as stasipatric speciation (White, 1978) or speciation due to chromosomal evolution. Considerable chromosomal variation has been reported between the two proposed species of anoa (Amano and Martojo, 1983; Schreiber et al., 1993; Sugiri and Hidayat, 1996).

MtDNA sequences and allozyme data have been used as indicators of dissimilarity to define the anoa taxonomy (Avise et al., 1989). This is achieved by comparison of genetic distance between individuals of the two proposed anoa species with the distance between individuals from clearly distinct bovid species. This method has come under some criticism recently because of the possibility of erroneous classification caused by variation in genetic distance between different species used as the scale (Goldstein et al., 2000).

2.2.2 Anoa morphological variation and the biological species concept

Skull morphological measurements and gross body characteristics used to define the anoa species were acquired from specimens in museum collections, often from unknown localities. Groves (1969) surveyed 22 skulls and additional skins. The phenotypic characters proposed to distinguish 'depressicornis' and 'quarlesi' anoa included horn cross-section, shoulder height, body markings, and hair characteristics. These are shown in Figure 2.1 and are described in Appendix 2, Table 1.

The English common names relate to a putative altitudinal separation (Groves, 1969), with the large 'depressicornis' form (lowland anoa) inhabiting low-lying areas and the smaller 'quarlesi' form (mountain anoa) living at higher elevations. The locality records of a maximum of 60 partial or complete skulls and an unknown number of skin specimens available to Groves (1969) suggested this altitudinal separation. However, no detailed comment on habitat variation was given to explain this supposed altitudinal boundary between the ranges of the two proposed species. Data collated since 1990, displayed in Figure 2.2 below, suggests that the altitudinal division between anoa morphotypes is less distinct than the above separation (Burton et al., 2005). Furthermore, it has also been suggested that altitude would cause a clinal variation within one species of anoa, rather than result in two parapatric species (Heller, 1889; Whitten et al., 2002).

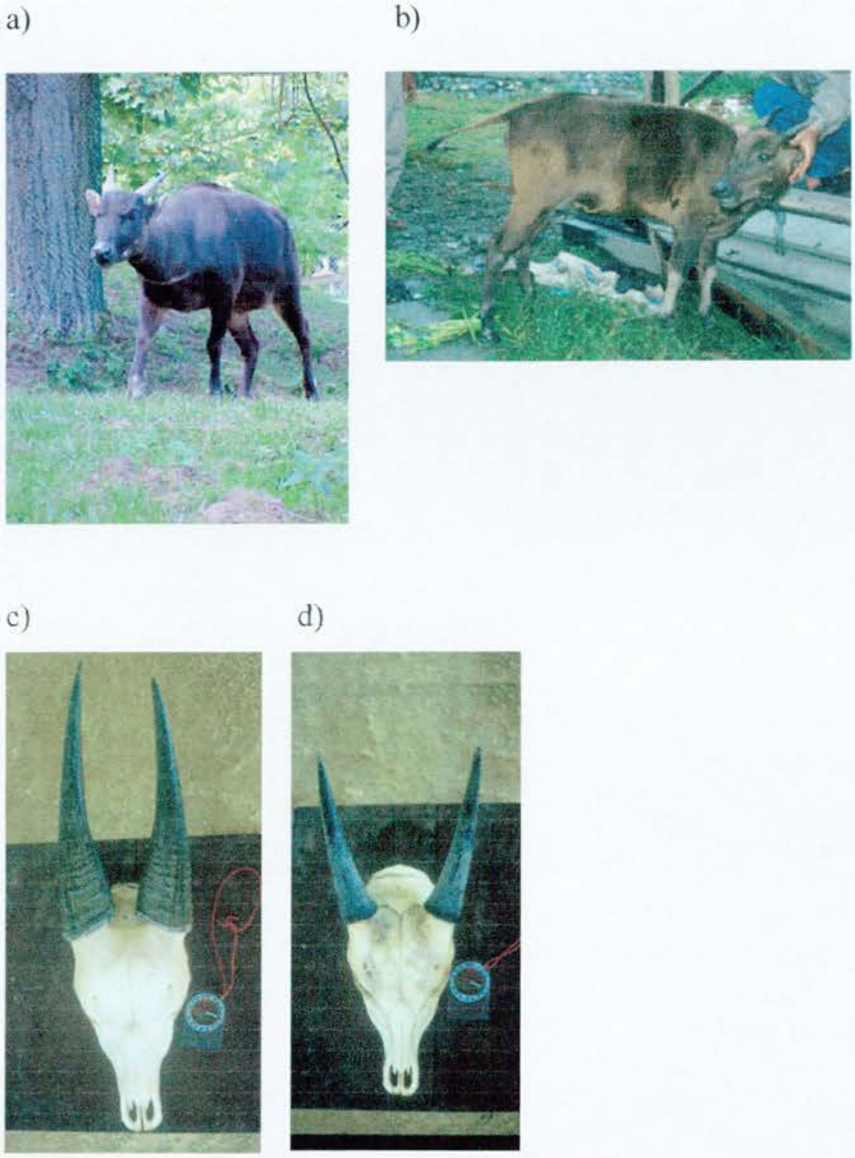


Figure 2.1. a) and b) Live anoa individuals *Bubalus sp.,* and c) and d) The morphological variation in anoa skulls.
Individuals represented are from a) Chester Zoo, UK and b) Central Sulawesi. Variation in skull morphology and horn dimorphism is shown in c) similar to the lowland anoa *Bubalus depressicornis* and d) similar to the mountain anoa *B. quarlesi*. Skulls originated from Buton Island, South-East Sulawesi.

Weise (1979) attempted a multivariate analysis using 26 measurements from 57 anoa skulls, compared to four measurements of 31 complete skulls by Groves (1969). However, Weise found insufficient data to show a statistically significant clinal range in skull size and also concluded that there were no biometric differences between 'depressicornis' and 'quarlesi' morphotypes that could not be explained as a result of allometric transformations due to the smaller (by 15%) size of the 'quarlesi' form. This is not a surprising finding because the *a priori* criteria he used to classify 36 of the 57 skulls of unknown locality into 'depressicornis' or 'quarlesi' morphotypes were based on horn and skull size characters. Therefore, the conclusions of this morphological analysis have little value (Burton et al., 2005). Weise (1979) points to inconsistencies in the characteristics used to identify the forms by Groves (1969). Weise (1979) also stated that the morphological data suggested that there may be contact between populations and possible admixture in the wild, suggesting a single species.

Intermediate forms in the wild and in zoo collections, have been reported by other observers (Schreiber et al., 1999; Burton et al., 2005) suggesting inter-breeding between the two species (Schreiber and Nötzold, 1995; Burton et al., 2005). Schreiber et al. (1993) concluded that the characteristics defined by Groves (1969) are ambiguous, following his examination of animals in zoological collections, particularly if only a few animals are available for comparison. Unfortunately, doubts have been raised about the location in Sulawesi from which the animals in zoos originated (Nötzold, 1999). Furthermore, consanguineous breeding and possible hybridization between lineages reduces the value of the conclusions about morphological variation made by Schreiber et al (1993). The possibility of hybridization between lineages has also hindered interpretation of the karyotype and genetic analysis described below. In conclusion, the morphological variation used to describe the two putative species by Groves (1969) has been questioned (Weise, 1979; Schreiber et al., 1993), and to date all analysis has been unable to clearly define the taxonomic status of the anoa.

2.2.3 Karyotypic variation

The published chromosome numbers of the anoa vary considerably, with individuals of 2n = 44, 45, 47, and 48 in European zoos (Koulischer, Tyskens and Mortelsmanns, 1972; Schreiber et al., 1993). These studies suggested that the two putative forms may have distinct karyotypes, with the 'depressicornis' morphotype having chromosome numbers of 2n = 47 and 48 and the 'quarlesi' morphotype of 2n = 44 and 45. Karyotypic variation in anoa has been attributed to centric fusions (Robertsonian translocations) and possible other forms of cytogenetic transformation (Schreiber et al., 1993). Schreiber et al. (1993) notes that the biological significance of the observed anoa chromosomal variation needs to be further investigated.

It has been reported that stasipatric speciation (White, 1978), or speciation due to chromosomal evolution can occur, for example in the dik-dik (*Madoqua* spp.) (Ryder et al., 1989). Under the biological species concept, if reproductive isolation was proved, the two populations of anoa, defined by chromosome number, could be classed as distinct species. However, karyotypic variation is not always a reliable indicator of distinct species because fertility and viability may be unaffected, or only a reduction in the number of progeny of matings between individuals from different chromosomal populations. This has been reported for other bovid species (Fischer, 1974). For example, a close relative of the anoa, the water buffalo (*Bubalus arnee*) has chromosome numbers of 2n = 48 and 50 (swamp and Murrah types) and fertile offspring are produced from crossing between the two types. Clearly, this issue requires further investigation to identify if the extent of the proposed variation in the anoa does exist (Koulischer et al., 1972; Schreiber et al., 1993).

In addition, anoa in Indonesian zoos are reported to have chromosome numbers of 2n = 36 and 38 (Amano et al., 1983) and 2n = 38, 42, 46 and 48 (Pranadewi, 1998; Marsono, Utami, Setiowati et al., 2001). This wide range of chromosome numbers reported from zoo individuals is highly uncommon for other bovid species (Gallagher and Womack, 1992), and so these results should be treated with suspicion and their accuracy questioned. Results from wild individuals have also found chromosomal variation with the identification of individuals of 2n = 44, 46 and 48 from Central Sulawesi (Sugiri et al., 1996), possibly putting into question the low chromosome number estimates (2n = 36 and 38). Two of the chromosome numbers reported from the wild have already been identified in zoo individuals, however 2n = 46 had not been previously reported. A cautionary note is worth sounding with respect to these observations; none of the karyotypic results from Indonesian studies have yet been published in peer reviewed journals.

If this karyotypic variation proves to be reliable, then it may show that the taxonomic diversity of anoa has been underestimated in the current classification of two putative species (Schreiber et al., 1999; Burton et al., 2005). Of equal taxonomic significance is the need for clarification of the level of reproductive isolation caused by the difference in karyotype. Genetic techniques have been employed to investigate this isolation using the level of dissimilarity between individuals of the two species held in European zoos. The results are summarized below.

2.2.4 Genetic distance as an indicator of dissimilarity

Recent genetic studies of anoa in zoo collections have lent support to the two species hypothesis. For example, analysis of partial mtDNA cytochrome *b* gene sequences (285bp) revealed clear differences between the putative 'quarlesi' form (from Krefeld Zoo) and 'depressicornis' form (from Berlin, Leipzig, and Antwerp zoos) (Schreiber et al., 1999) (Figure 2.2). Of the eight animals sampled, four haplotypes were identified, shown in Figure 2.2 as Anoa cyt b 1 to 4 and grouped as 'lowland' and 'mountain' forms. Between the four haplotypes 11 sites (3.9%) were variable. This level of nucleotide substitution was far higher than the intra-specific difference found in three other species (two dwarf zebus, *Bos [Bos] taurus;* two American bison, *Bison bison*; and four gaur, *Bos [Bibos] frontalis*); indeed these species showed no variation. Three further species: four African buffalo, *Syncerus caffer*; three European bison, *Bison bonanus*; and five banteng, *Bos [Bibos] javanicus* had 0.4% nucleotide substitutions, also lower than that observed in the anoa. The inter-specific variation within the anoa (between 'depressicornis' and 'quarlesi' anoa) was 3.1%. Comparing these data with the inter-species' variation between recognized species; the anoa have a higher nucleotide substitution rate than between domestic yak (*Bos [Poephagus] grunniens*) and American bison (2.1%); and a similar value to the gaur and banteng (3.2%); and gaur and yak (3.3%). These results seem to support the two species hypothesis.

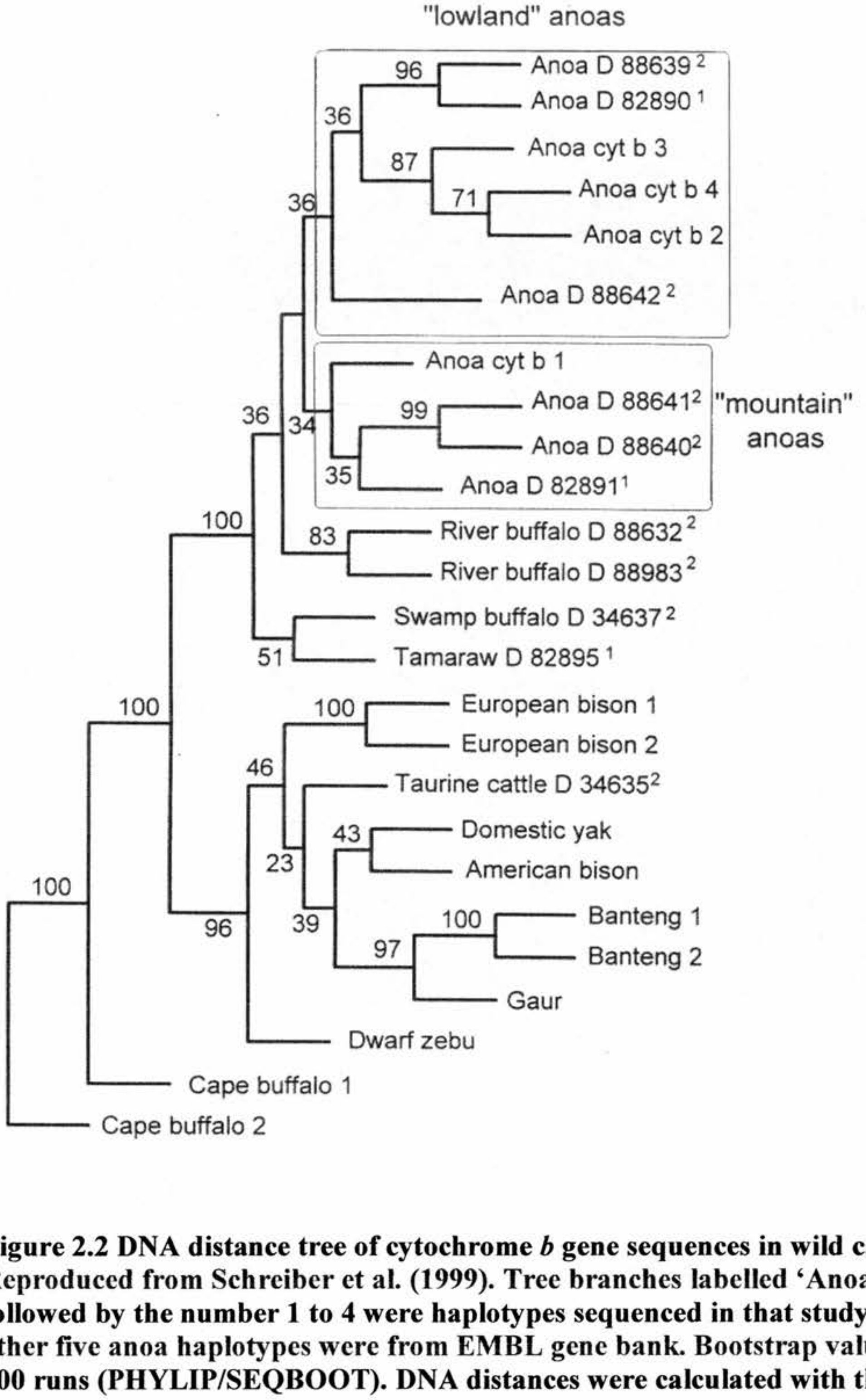


Figure 2.2 DNA distance tree of cytochrome *b* gene sequences in wild cattle. Reproduced from Schreiber et al. (1999). Tree branches labelled 'Anoa cyt b' followed by the number 1 to 4 were haplotypes sequenced in that study, the other five anoa haplotypes were from EMBL gene bank. Bootstrap values for 100 runs (PHYLIP/SEQBOOT). DNA distances were calculated with the Kimura 2 parameter model (PHYLIP/DNADIST), clustered by neighbour joining algorithm (PHYLIP/NEIBOR) combined using (PHYLIP/CONSENSE).

Two further studies using complete mtDNA cytochrome *b* gene sequences (1190bp) also lent support to the two species hypothesis (Tanaka et al., 1996; Kikkawa et al., 1997). These studies were conducted using samples from two and five anoa from Indonesian zoos, respectively. The results of the study by Tanaka et al. (1996), suggest that the 'depressicornis' and 'quarlesi' anoa should be regarded as two species, as they found the weighted genetic distance between them to be 0.0354, which is similar to the weighted genetic distance between the anoa ('depressicornis' and 'quarlesi') and water buffalo (both river and swamp types) that are clearly distinct species. The weighted genetic distance between anoa was greater than the intra-specific variation of domesticated water buffalo (swamp and river buffalo) (0.0256).

Kikkawa et al. (1997) sequenced part of the mtDNA cytochrome *b* gene from five anoa. Without clarifying criteria for describing subspecies, they identified these animals into three subspecies of anoa, 'lowland', 'mountain', and 'quarlesi'. The estimated sequence divergence of the mtDNA cytochrome *b* gene between their 'lowland' and 'mountain subspecies' was 3.6%, and between their 'lowland' and 'quarlesi subspecies' was 3.3%. These sequence divergence differences were closer to those found between acknowledged species such as the anoa and water buffalo of the swamp type (3.33%), than those found within water buffalo (i.e. between river and swamp type buffaloes) (2.67%). Even though the morphological descriptions of the different 'subspecies' were unclear, genetic variation between these animals was equivalent to that seen between other distinct species. Kikkawa et al. (1997) suggest that differences they found between anoa and water buffalo were not significant at the genus level, but that the results showed there are two lineages within anoa.

Earlier work that looked at allozyme distances and gross body morphology raised some doubts about the two species hypothesis. Schreiber et al. (1993) suggested that there was less variation between the anoa in European zoos than could be found between other closely related species. Genetic distances estimated from variations in allozymes were used to investigate differences between 25 anoa individuals. Pairwise comparisons of Nei's genetic allozyme distances for anoa were D=0.0206 to D= 0.0505, while for distinct wild bovid species the values ranged from D=0.1389 to D=0.7621 (Hartl, Göltenboth, Grillitsch et al., 1988). Therefore, these allozyme distances for anoa were similar to those found between geographic races or subspecies (Nei, 1987). However, Schreiber and his co-workers cautioned that the taxonomic relevance of these genetic distances was open to question because of the effect of bottlenecks and inbreeding. However, they concluded that the coincidence of differences encountered in karyotypes, polymorphic genes, body size, hair texture, horn shape, and other phenotypic characters indicated that the anoa comprise differentiated taxa (Schreiber et al., 1993).

There are at least two possible explanations that may be proposed for the difference in mtDNA and allozyme distance results discussed above. Firstly, it is conceivable that 'depressicornis' and 'quarlesi' anoa are well-separated species in the wild, but hybridization in the zoo population resulted in the observed pattern of protein variation (Schreiber et al., 1993), and the maternally inherited mtDNA sequence reflects only the genotype of the founder females (Tanaka et al., 1996; Kikkawa et al., 1997; Schreiber et al., 1999). Alternatively, the differences in the mtDNA sequences could be a chance effect arising from a sole female founder of the 'quarlesi' anoa form belonging to another mitochondrial lineage. Such polymorphism might have no taxonomic implications, but would merely indicate the presence of very deep mitochondrial lineages and testify to the great age of the taxon. Recent studies of well-differentiated hemoglobin sequences in anoa are in accordance with the second possibility (Schreiber and Goltenboth, 1990; Schreiber et al., 1993). Schreiber et al. (1999) suggest limited gene flow between populations could be sufficient to explain the lack of distinct variation in the wild, but would be insufficient to prevent differentiation within mtDNA over long periods. One explanation of this limited gene flow would be the reduction in fertility caused by karyotypic differences described above. Low levels of gene flow have been shown to occur in other species (Avise, 1994), including the Sulawesi macaques in zones of hybridization (Evans, Supriatna and Melnick, 2001). It would therefore be advantageous to investigate anoa nDNA because this would provide more detail than the allozyme analysis to date. Also combined with mtDNA sequence analysis this would help to clarify the difference between matrilineal and patrilineal genetic variation.

In conclusion, the morphological and karyotypic analyses, as well as mtDNA and allozyme results have produced conflicting species classifications for the anoa. This uncertain status is in urgent need of clarification in order to provide a basis for conservation efforts (Moritz, 1995). The application of two different species concepts has not been able to solve the uncertainty of the taxonomic status of the anoa. The low levels of polymorphism reported for the allozymes indicates that the use of a genetic technique suitable for assessment of closely related populations would be advisable. Microsatellite markers described in Section 1.4.2, would provide a greater level of polymorphism for better defining genetic differentiation. These markers have been widely and successfully used for investigations of intra and inter-species variation (Goldstein et al., 1997; Goldstein et al., 1999), such as the African buffalo (*Syncerus caffer*) (van Hooft, Groen and Prins, 2000) and in the tribe Bovini (Ritz et al., 2000).

The taxonomy of anoa remains uncertain from the research conducted to date. A final point is that the majority of the genetic and karyotype studies only assessed individuals from the European zoo population. These animals are from unknown locales on Sulawesi or Buton Islands (Nötzold, 1999). If the population structure of the anoa shows similarities to that of the Sulawesi macaques (Evans et al., 1999), the anoa's taxonomy may be related to the biogeographic regions of Sulawesi and Buton. Therefore, as the origin of individuals is uncertain, it is unknown if those animals analysed above are representative of the genetic variation of the wild population. The biogeography of the Wallacea region is described in the following section.

2.3 The biogeography of Sulawesi

2.3.1 The anoa's distribution in the Wallacea Region

Sulawesi and Buton Islands are situated within the Wallacea bioregion, renowned for its high level of endemism (Whitten et al., 2002). Of the mammal species found on Sulawesi, 61% (98% excluding bats) are endemic, compared to just 18% on the neighbouring island of Borneo and 58% in Papua New Guinea (Whitten et al., 2002). Being at the boundary of the Asian and Australasian biogeographic regions, Wallacea comprises a mix of the Asian and Australasian flora and fauna, including two cuscus species (*Ailurops ursinus* and *Strigocuscus celebensis*) from Australasia, and two suids (babirusa *Babyrousa babyrussa* and the Sulawesi warty pig *Sus celebensis*), and the tarsiers (*Tarsius* spp.) from Asia. The anoa is thought to have originated from Asia, as they are deemed close relatives of the Asian water buffalo *Bubalus arnee* (Groves, 1969). Deep oceanic trenches between the Wallacea region and the Asian and Australasian landmasses are thought to have caused a depauperate fauna (Whitten et al., 2002). The lines in Figure 2.3 define these barriers.

The colonisation of Sulawesi by the current fauna and the biogeographic units that make up the island are described in the following two sections. The anoa was thought to be present throughout the island's forests in historic times (Weber, 1890; Sarasin and Sarasin, 1901; Mohr, 1921; Harper, 1945; Groves, 1969). However, as the size of these forests is shrunk by human activities, the anoa's distribution has been reduced and fragmented. The effect of this on the conservation status of the anoa will be summarized in Section 2.3.

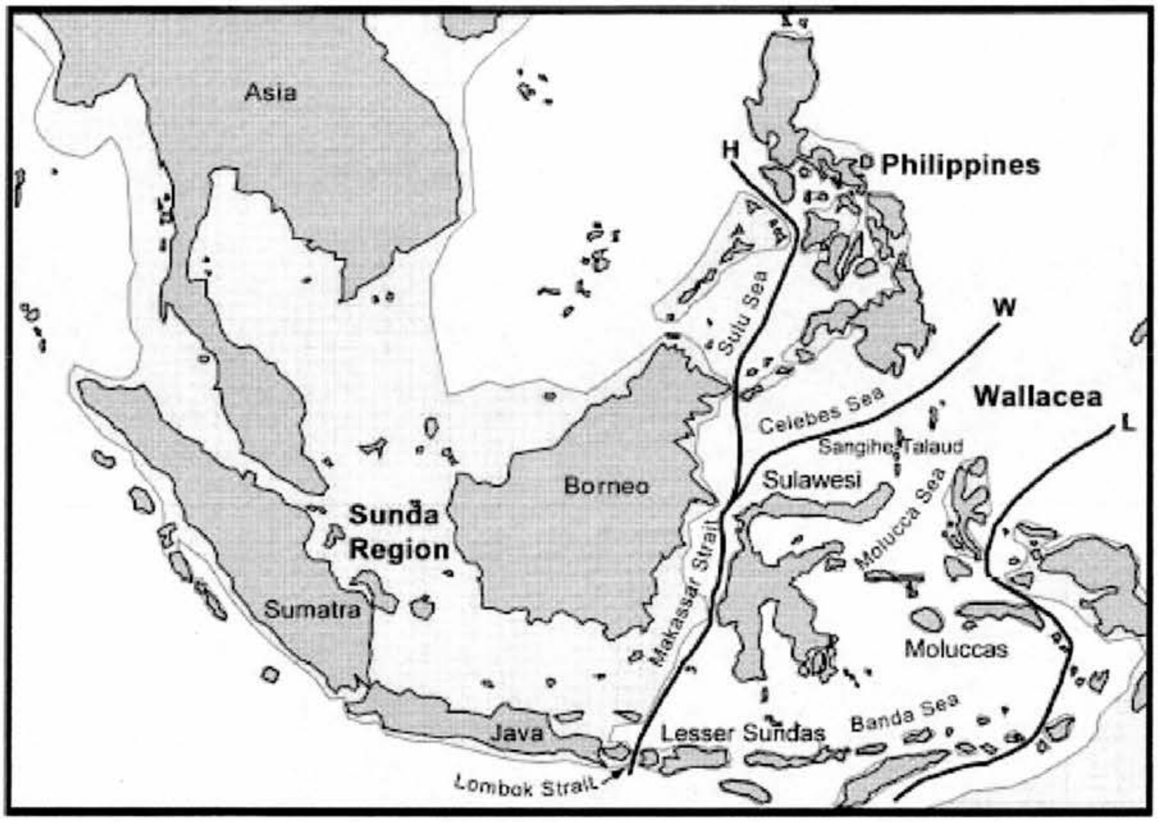


Figure 2.3. Characterization of multi-taxon break between Asian and Australasian biogeographic regions and change in sea level during the Late Pleistocene.
Biogeographic regions defined by lines labelled as Wallace 1963 – W, Huxley – H, Lydekker – L. Shading represents islands of South-East Asia in dark grey, sea level reduction of 120m. A modified version of the map from Voris (2000) was published by Evans et. al. (2003a) and displayed here.

2.3.2 The colonisation of Sulawesi Island

The arrival of Sulawesi Recent Faunas such as the anoa marks a chance event of crossing the oceanic barriers from neighbouring islands (Van den Berg, 1999). The available details concerning the anoa's colonisation of Sulawesi are few. Fossil and subfossil remains of anoa have been found in the Sulawesi Recent Faunas of Late Pleistocene to Holocene age (Van den Berg, de Vos and Sondaar, 2001), but not from earlier deposits as had been suggested previously (Hooijer, 1950). These fossils from the Recent Faunas were found in the south peninsula of the island. An early ancestor and possibly the closest relative of the anoa is thought to be the Asian water buffalo (*Bubalus arnee*) (Groves, 1969; Geraads, 1992; Tanaka et al., 1996; Schreiber et al., 1999). Though this species is not native to Borneo it occurred across the Asian mainland (Hedges, 1998). Borneo and the mainland may have been joined during the late Pleistocene era by a lowering of the sea level (Voris, 2000), so the anoa's ancestor could have reached Borneo without crossing a water body. The only other wild bovid species in South-East Asia are the distantly related banteng (*Bos javanicus*), which are native to Borneo (MacKinnon, Hatta, Halim et al., 1996) and the endemic tamaraw (*Bubalus mindorensis*) of the Philippines (Custodio, Lepiten and Heaney, 1996).

The analysis of geological history and plate tectonics suggest a number of routes for Asian species like the anoa to colonise Sulawesi (Van den Berg, 1999). A possible route was for animals to raft or swim across the shortened distance <40km between Borneo and Central Sulawesi as the width of the Makassar Strait would be much reduced, compared with the present day, by lowering of the sea level (Voris, 2000). The swimming ability of a range of mammals suggested that larger species had the ability to cross greater water bodies than lighter species, due to greater energy reserves they had available (Meijaard, 2001). Species of a body mass similar to that of the anoa may have been able to cover the distance between Borneo and Sulawesi possibly assisted by water currents. Thus, this is a feasible colonisation route for the anoa. Several shallow carbonate platforms south of the Makassar Basin have been reported which may have acted as stepping-stones from Java (or other islands in the Lesser Sundas) to Central Sulawesi (Van den Berg et al., 2001).

Hall (1998) and Moss and Wilson (1998) proposed colonisation to the north peninsula of Sulawesi by island hopping along the volcanic archipelago (Sulu and Sangie, Cagayan Islands) from the Philippines. It has even been proposed that two colonisations by the anoa may have occurred; this is based on variations in amino acid sequences of hemoglobin beta chains (Kakoi et al., 1994). However, Kakoi et al. (1994) did not discuss the biogeographic variation found on Sulawesi. They dismiss the possibility of isolation of parts of Sulawesi Island because of the lack of habitat differentiation, and do not consider the possible isolation by oceanic barriers. To date there is no evidence to conclusively prove the use of one colonisation route over another by the anoa (Burton et al., 2005).

Other Sulawesi species originating from Asia arrived by various routes, as indicated by defining lineages from mtDNA sequences. The macaques do seem to have colonised Sulawesi twice; although three regions (southwest, centralwest and/or northwest) have been proposed as their location of arrival by different authors (Fooden, 1969; Albrecht, 1978; Evans et al., 1999). It is proposed that the Celebes toad (*Bufo celebensis*) arrived in the northern peninsula of Sulawesi (Evans, Supriatna, Andayani et al., 2003b). Evans, Brown, Mcguire et al. (2003a) investigated the phylogenetic relationship between the fanged Ranid frogs (genus *Limnonectes*) of the region. They found that these species had arrived on Sulawesi once or twice from Borneo, and once again from the Philippines. The shrews *Crocidura* are also thought to have dispersed to Sulawesi on two occasions (Ruedi, Auberson and Savolainen, 1998). So, although the number of arrivals and the routes of arrival of some Asian species have been identified, colonisation by the anoa remains unclear.

2.3.3 The biogeography of Sulawesi and Buton Islands

The current anoa taxonomy classifies the anoa into two putative species, distributed according to elevation. More recently (post 1990) this altitudinal division has been reported to be less defined than was previously thought by Groves (1969), as displayed in Figure 2.4 below. However, the division of species along altitudinal boundaries is in contrast to other Sulawesi species, which have been phylogenetically defined into biogeographic areas of Sulawesi. These regions broadly represent the different peninsulas of the island. These peninsulas have been classified as biounits within a biogeographic province of the Dasmann-Udvardy system (MacKinnon and Wind, 1981; MacKinnon, 1997) and more recently were defined as Areas of Endemism (AOEs) (Evans et al., 2003b). The latter following the observed congruence of the ranges of the mtDNA phylogenies of macaque species and the Celebes toad populations, as well as those of the fanged frogs (Evans et al., 2003a). Additional species broadly fit the phylogeographic pattern of AOEs, showing there are factors that have affected multiple taxa. Considering that there is debate over the correlation of anoa phylogeny and altitude (Section 2.1.2), and that many Sulawesi species follow the phylogeographic pattern of AOEs, this makes an additional case for the reassessment of the anoa's phylogeography. Comparisons of the phylogenetic distribution of other Sulawesi species together with the theories concerning the phylogeny of the anoa are made below.

Between five and seven *Macaca* species are present on Sulawesi (Fooden, 1969; Albrecht, 1978; Bynum, 1999). Their distribution is thought to be parapatric (Bynum, 1999), though inter-breeding between some of these populations is known to occur in hybrid zones (Ciani, Stanyon, Scheffrahn et al., 1989; Watanabe and Matsumura, 1991; Evans et al., 2001). Distinct species or forms are found in another group of primates, the tarsiers of Sulawesi (Niemitz, Neitsch, Warter et al., 1991). At least two species *Tarsius dianae* and *T. spectrum*, occupy distinct regions within Sulawesi's lowland forests, while a further species *T. pumilus,* is found at high altitudes (Shekelle, Leksono, Ichwan et al., 1997; Groves, 1998). This altitudinal separation of tarsier species is likely to lead to a parapatric distribution due to the proposed uninhabited altitudinal band between low and high altitude species. The reason for this uninhabited zone is unknown. Further species such as squirrels and other small mammals, divided by biogeographic boundaries of unknown origin have been recorded on Sulawesi (Musser, 1987; Groves, 2001). A wide range of species on Sulawesi show genetic or morphological separation into similar biogeographic regions as the AOEs. These include birds such as the White-eye *Zosterops* (Holmes and Holmes, 1985), and insects such as cicadas *Cicadoidae* (Butlin, Walton, Monk et al., 1998; Holloway, 1998), Carpenter bees *Xylocopa* (van der Vecht, 1953) and pond skaters *Ptilonera* (Polhemus and Polhemus, 1986).

Numerous factors contribute to the phylogenetic relationships between AOEs of Sulawesi species including site of colonisation, routes of movement following colonisation, the effective population size, demography and time. Even though each of these factors vary between taxonomic groups, the AOEs are largely congruent for a number of species (Evans et al., 2003c). The effect of these biogeographic barriers was to compartmentalize different species into units at different locations on Sulawesi.

Although Sulawesi has a complex evolutionary history with much tectonic activity, the biogeographic areas do not appear to form boundaries at the sutures of the island's precursors (tectonic lines) (Evans et al., 2003b). Instead, the biogeographic regions are divided at quite different locations; certain boundary locations are due to defined barriers, but for others the reason for the division is unknown. The existence of previous historical barriers have been shown at two locations, the Tempe Depression in South Sulawesi (during the Pleistocene era) recorded from pollen records, and the second, near Gorontalo in the central area of North Sulawesi (Fooden, 1969). These barriers were formed because rising sea levels separated the landmasses, and as a result two macaque species evolved in each of these areas. The narrow sea barrier between Sulawesi and Buton Islands has also acted as an effective barrier for tarsiers (Nietsch and Burton, in prep.).

The explanation of divisions between other biogeographic regions is incomplete. Certain areas may have been separated by sea level rise and possibly by differences in soil type and vegetation cover (Evans et al., 2003b). For the anoa, Schreiber et al. (1999) suggested a number of possibilities to explain the presence of two morphologically similar species on Sulawesi. They also attempted to clarify the taxonomic uncertainty, as discussed above. Their first theory suggests that anoa colonised two geographically separated landmasses so becoming two populations. Then the landmasses and also the anoa populations joined as Sulawesi was formed (Hall, 1996, 1998). An alternative theory proposed by Schreiber et al. (1999) is that

that anoa arrived on Sulawesi after the island was spilt into smaller islands by rises in sea level. The latter is known to have occurred during the Pleistocene at two locations on Sulawesi, as mentioned above (Fooden, 1969). Alternatively, variation in climate may have reduced forest cover as well as leading to forest fragmentation during the Pleistocene period (Whitten et al., 2002). This could have caused the isolation of anoa populations and led to their subsequent divergence, thus giving rise to the two distinct populations thought to occur today. As already discussed above, additional theories to explain the variation within the anoa were, environmental variation due to altitude, (Heller, 1889; Hall, 1996, 1998; Whitten et al., 2002); and two separate immigrations (Kakoi et al., 1994; Kikkawa et al., 1997). Unfortunately, there is as yet insufficient biogeological data to choose between these alternative explanations (Van den Berg et al., 2001).

To conclude, the morphological, karyotypic and genetic differentiation of the anoa (described in Section 2.1 above) has not been attributed to the biogeographic variation of Sulawesi and Buton Islands. Instead, Groves (1969) suggested the differences between anoa were related to altitude. However, there is little evidence to support this. The distinct biogeographic regions of Sulawesi (AOEs) that are congruent for a number of species would strongly suggest that further assessment of anoa phylogeny should focus on geographically based genetic sampling of these regions. In order to assess all the proposed theories, data sampling should also attempt to incorporate morphological variation between anoa, and cover a range of altitudes in which the anoa are found.

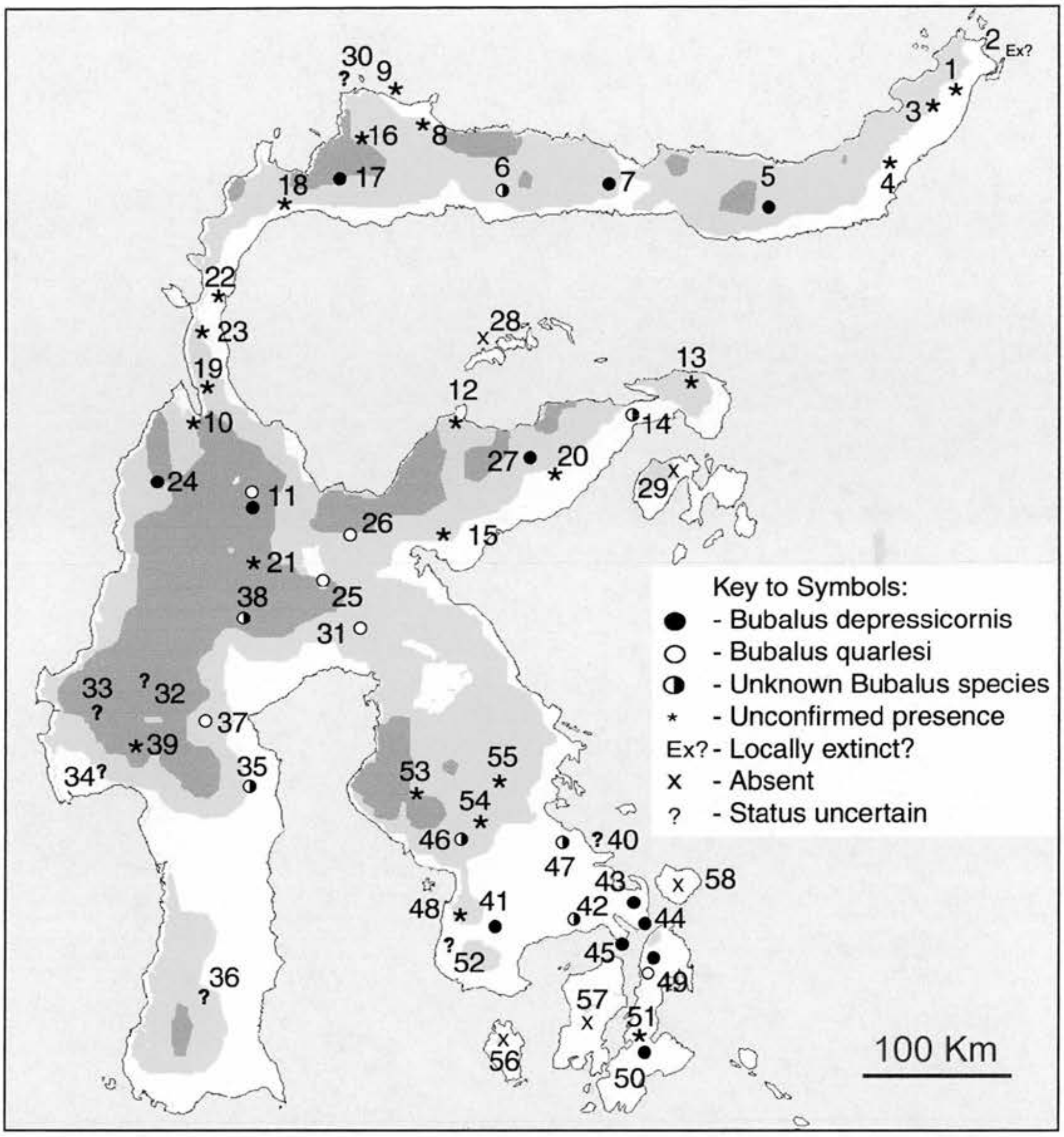


Figure 2.4. Distribution of lowland (*Bubalus depressicornis*) and mountain anoa (*B. quarlesi*) after 1990 and altitude (m asl).
Each symbol may represent between 1 and 19 data points from a single location. See Appendix 1, Table 1 for key to numbered areas; see Appendix 2 for morphological descriptions. Unconfirmed presence – secondary data indicating presence since 1990; Locally extinct? – anoa presumed extinct during the period since 1990, after extensive field surveys or area no longer suitable for anoa; Absent - anoa never reported from this area; Status uncertain – near local extinction. Altitude Key: White: 0 – 200 m asl; Light Grey 200-1000 m asl; Dark Grey +1000 m asl. Altitude measured in metres above sea level. Map data courtesy of E. Colijn.

2.4 Conservation status of anoa

2.4.1 Introduction

As the world enters the period of time known as the 'sixth extinction', the field of conservation has grown to encompass genetic, species and ecosystem diversity (Leakey and Lewin, 1995). In the context of this study the goal of conservation is to prevent species or population extinction by maintaining genetic diversity (Frankham, 1995). One current predicted rate of extinction, with a set of caveats attached, is 3.5% of all avifauna (350 species) and possibly a greater proportion of mammals could become extinct before 2050 (Jenkins, 2001). The IUCN Red Data List of Threatened Species has 11,046 animal and plant species recognised as threatened, facing a high risk of extinction in the near future (http://www.iucn.org/redlist/2000/news.html, September 2004). The effect of climate change is predicted to have a major impact on extinction rates. By 2150 it is predicted that 15-37% of species in the sampled regions and taxa will be 'committed to extinction' using mid-range climate-warming scenarios (De Siquera, Grainger, Hannah et al., 2004). This equates to a loss of one million species.

The decline in numbers of many large mammal populations, including the anoa is the result of direct human activities (Burton et al., 2005). Examples such as hunting and habitat loss are termed as deterministic factors of population decline. With the deforestation pressure predicted to remain high for Indonesia (Jepson, Jarvie, MacKinnon et al., 2001), it is likely that populations of species such as the anoa will only survive in heavily managed protected areas. Hunting and the trade in 'bushmeat' has long been established in Asia (Milner-Gulland, Bennett, Abernethy et al., 2002) and especially in parts of Sulawesi (O'Brien and Kinnaird, 1996; Avalard, 1999). The difference now is the increased intensity of hunting and the size of the human population consuming the meat.

The deterministic factors are only part of the cause of a species fall to extinction. The second group of factors that can lead to extinction are stochastic events. These often act concurrently on small or fragmented populations and are grouped into demographic, environmental and genetic events. This has led to the term 'extinction vortex' because of the increased chance to extinction induced by the compounded effect of these factors, especially when considering small populations (Gilpin and Soulé, 1986; Soulé and Mills, 1998). With a greater understanding of genetic diversity, the significance of maintaining the greatest possible variety as an integral part of population management has been realised (Humphries, Williams and Vane-Wright, 1995).

For conservation purposes a set of criteria for the prevention of population extinction has been devised. The aim is to retain evolutionary potential of a population. The estimated requirement is that 7000 adult vertebrate individuals be protected to ensure 99% probability of persistence for the 'long-term' or 40 generations (Reed, O'Grady, Brook et al., 2003). This value was defined by modelling variation in the drift, rate of mutation and initial genetic variation (Franklin and Frankham, 1998; Lynch and Lande, 1998).

Species conservation is now becoming a process of integrated management of *in situ* and *ex situ* populations as the threat to wild habitats increases. Conservation breeding facilities or zoological collections have an important role to play in providing an alternative gene pool in case deterministic or severe stochastic factors reduce or remove *in situ* populations (Ballou and Lacy, 1995).

2.4.2 Anoa *in situ* conservation status

The Wallacea region is one of 25 hotspots having high species endemism and containing species under a high degree of threat of going extinct (Myers et al., 2000). The anoa is one of the threatened species from this region. The World Conservation Union (IUCN) described the anoa as Endangered (Categories 1 and 2a) (http://www.redlist.org/, August 2004) because there was a continuing population decline of at least 20% in two generations and no subpopulation was estimated to contain more than 250 mature individuals. The anoa is legally protected under Indonesian law (Jahja, 1987) and is included in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (UNEP-WCMC, 2000). Despite this the two most significant threats to the anoa are hunting and habitat degradation by humans (Burton et al., 2005).

Estimates of the extinction risk of anoa populations from hunting pressure were made during the Anoa Population and Habitat Viability Analysis Workshop (PHVA) and conducted using the program VORTEX (Manansang, Hedges, Dwiatmo et al., 1996). This analysis predicted that with the 15% off-take hunting levels reported for North Sulawesi (O'Brien et al., 1996), the anoa would be extinct from all protected areas in that region in 100 years. With a 4% hunting off-take rate, the largest population would be reduced to only 232 individuals from an initial estimated population of over 2000 animals in the 100 year period (Manansang et al., 1996). Though recent surveys have shown hunting is still common, the annual off-take rate is now believed to be closer to the 4% level, with more data available (Burton et al., 2005). The economics of hunting are currently maintained with the price for one animal on average being £29.00–36.00 (Rp400,000–500,000), equivalent to several months of average wages (Burton et al., 2005). This severe threat is discussed further in Chapter 7 in relation to the results of the genetic analysis.

As indicated by the 'hotspot' status of the Wallacea region, many other species are under threat of extinction (Myers et al., 2000). With the rate of hunting reported for babirusa in North Sulawesi, survival of only small populations is predicted for areas with least vehicular accessibility (Clayton, Keeling and Milner-Gulland, 1997). One North Sulawesi urban market is reported to have sold 90,000 wild caught mammals in a year (Clayton and Milner-Gulland, 1999). This is part of the internationally deplored bushmeat trade, but is mainly supplying a local market demand (Milner-Gulland et al., 2002).

Habitat degradation and loss is the second most significant threat to the anoa. Many anoa populations are becoming isolated as the forest around protected areas is cleared or converted to plantations. The remaining forest areas in Sulawesi are patchily distributed. This is discussed in relation to conservation recommendations in Chapter 7 and indicated in Figure 7.2. The total forested area on Sulawesi in 1985 was estimated to be between 42 and 49% of the land area of the island (Holmes, 2000; Mathews, 2002). The loss of lowland forest between 1985 and 1997 was estimated to be 89% (Holmes, 2000). Recently it has been said, "Sulawesi's lowland forest has already been reduced to statistical insignificance" (Aden et al., 2001).

Although 11.9% of the island is protected, loss of suitable anoa habitat within protected areas is as much of a problem as in unprotected forest (Aden et al., 2001). For example, the expansion of agricultural and settled areas, logging, mining, and fires have encroached into many protected areas (Bogani Nani-Watabone, Lore Lindu and Rawa Aopa-Watumohai National Parks) (Burton et al., 2005). The fragmentation of these once contiguous forests has reduced the natural movement of anoa between forest areas. A study of protected areas in western America found that the area of the park was inversely correlated to the likelihood of mammalian extinction (i.e. species not observed in the previous 10 years) (Newmark, 1995). This is thought to be because of the loss of genetic diversity due to inbreeding in small populations, which will increase the chance of extinction (Gilpin, 1991; Wang and Caballero, 1999). The reduction in genetic diversity due to deterministic factors will reduce the ability of a population to positively respond to stochastic events. Thus, small populations enter an 'extinction vortex' (Gilpin et al., 1986; Soulé et al., 1998) that are anticipated to have serious consequences for the long-term survival of the anoa population (Burton et al., 2005).

In the wild, anoa benefit from only limited active protection from hunting and only sporadic patrolling of protected areas, due to lack of funds and motivation (Burton et al., 2005). The absence of an up-to-date species-specific management plan adds to the difficulties the local Forestry Departments face. If this plan were available it would be integrated into the current management strategy, with the help of ministerial backing. The lack of a species management document is the result of uncertainty as to the units for anoa conservation and their distribution in Sulawesi and Buton Islands.

2.4.3 Anoa *ex situ* conservation program

Ex situ conservation aims to support a viable population that retains genetic variability and is minimally affected by the negative effects of inbreeding, loss of diversity or outbreeding depression (Eldridge, King, Loupis et al., 1999; Groombridge, Jones, Bruford et al., 2000). An *ex situ* population have two purposes; as a genebank to represent the whole population and also for reintroduction or supplementation of declining wild populations (Ballou et al., 1995). In order to achieve this it was estimated that 90% genetic diversity of the wild population should be maintained for 100 years (Frankham et al., 2002). The United Nations predicts that the human population will peak at between seven and nine billion and start to decline by the year 2150 (UN, 1992). So the human population expansion will end in around 100 to 200 years, and reintroductions of threatened species should become more feasible (Soulé and Simberloff, 1986). The 100 year term was selected for genetic modelling because of the difficulties in maintaining small viable populations for many species in the longer-term (L. Bingaman Lakey pers. comm.).

A total of 125 (58.67) anoa were reported to be held in zoological collections on 31 December 1998. These animals were in 35 institutions: 17 in Europe, 10 in North America, and 8 in Asia (Schreiber et al., 1993; Schreiber et al., 1995; Nötzold, 1999). Of these 125 animals, 19 were listed as founders, i.e. obtained from the wild population. These founders were imported from zoos in Indonesia or Malaysia, or obtained from animal dealers, so their places of origin within Sulawesi are unknown.

To achieve the aims stated above for *ex situ* populations, a species studbook recommending breeding strategies is integral to anoa conservation management. Recommendations for the anoa have, of necessity, been rather general because of the uncertainty surrounding the taxonomic affiliations of the zoo animals. All 125 animals have been listed as lowland anoa or *Bubalus (Anoa) depressicornis* (Nötzold, 1999) in the International Studbook, in contrast to the taxonomic status proposed by Groves (1969) and Schreiber et al. (1993). Biometric and photographic assessment of the founders of the Antwerp, Rotterdam, and Leipzig lines has indicated their phenotypic similarity and concurrence with descriptions of the 'depressicornis' morphotype (Schreiber et al., 1993). These breed lines have subsequently been crossed. The anoa at Krefeld, and Decin Zoos, which are the descendants of three animals originally imported by Berlin Zoo, have been listed in a research publication as the 'quarlesi' morphotype (Schreiber et al., 1993).

Other difficulties with maintaining the current breeding population of anoa are described here. There is inbreeding within the Krefeld / Decin 'quarlesi' anoa breed line as all these animals are related, which could be causing the reported reduced fertility (Nötzold, 1999; Schreiber et al., 1999). Although the three European 'depressicornis' breed lines total a larger number of animals, they have been bred as three populations with little crossbreeding, and hence are potentially highly inbred.

Therefore, the 'depressicornis' population is only representative of a small proportion of the total number of founder individuals (Nötzold, 1999).

Three alternative breeding strategies were proposed by Schreiber and Nötzold (1995) to deal with these problems in different ways. They discussed the feasibility of separate breeding populations for the breed lines reported to have distinct karyotypes. In other ungulates, breeding between distinct chromosome races has resulted in reduced fertility (Ryder et al., 1989). Current anoa population management in zoos indicates that the chromosome races are being maintained as distinct populations, as was done previously. However, the current International Anoa Studbook groups all individuals as a single population, following an explanation of the taxonomic uncertainty (Nötzold, 1999). Therefore, it seems that none of the three breeding strategies proposed has been fully implemented.

To allow the initiation of an effective conservation management strategy the highest priority is for the clarification of the anoa's taxonomic status. Populations in the north-east and southern regions of Sulawesi are near to local extinction and the genetic diversity of the *ex situ* breeding population is declining. It is vital to base conservation efforts of *in situ* and *ex situ* populations on relevant genetic information relating to the anoa's taxonomy, and on what is achievable in the real world.

2.5 Aims

The taxonomic status of the anoa has to be clarified to permit the effective management of *ex situ* and *in situ* populations. Four null hypotheses, which arose from previous research are tested in this thesis:

H_{o1} = There are two distinct species of anoa.

H_{o2} = The genetic variation in the population is related to the morphological characteristics used to define the two species.

H_{o3} = The genetic structure of the population is correlated with altitude.

H_{o4} = The European and Indonesian zoo population represents the genetic variation of the wild anoa population, also forming two species.

In the case where any of the first three null hypotheses are rejected, then further areas of investigation should be included. Tests could be conducted to assess if the genetic structure of the population is correlated to the biogeographic areas of Wallacea as discussed for other species above. Also the degree of structuring should be determined to assist species management, such as identifying if there is a greater effect of isolation by distance or of barriers to gene flow.

To subject the above hypotheses to tests the aims in this thesis are:

1. To assess the genetic diversity of the anoa population, including the definition of any structuring and the existence of subpopulations. The tests used will include Hardy-Weinberg equilibrium, STRUCTURE analysis and allelic variation of microsatellites. Also sequence mismatch and phylogenetic trees of D-loop sequence variation will provide matrilineal data (Chapters 4 and 5).

2. To assess the taxonomic classification of the anoa using two species definitions and an 'indicator of dissimilarity'. The species definitions used are the genotypic cluster definition and the cohesion species concept. The 'indicator of dissimilarity' method defines a species using genetic distance measures. The genetic differences between subpopulations of anoa are compared to values between known species and known breeds, which act as a scale to determine the level of differentiation between anoa groups (Chapter 4).

3. To test for a correlation of genetic structure of the population with the biogeographic regions of Wallacea. These tests will assess the effect of isolation by distance and barriers to gene flow on anoa population genetics. Analyses will include PCA, and simple and partial Mantel tests with microsatellite and mtDNA data (Chapter 6).

4. To determine if the combined morphological characteristics employed to differentiate the two species are representative of the genetic structure of the population. Hypothesis testing will be conducted using partial Mantel tests (Chapter 6) to determine if there was a correlation between morphological factors for both microsatellite and mtDNA data. These results can be used to assess the accuracy of the current species classification, which uses morphological characteristics.

5. To determine if altitudinal variation is related to genetic structure of the population. Partial Mantel tests will again be used (Chapter 6).

6. To assess the genetic representation of individuals in the *ex situ* population relative to the *in situ* population. Methods to be employed are assignment analysis of individuals with the program STRUCTURE and visual representation using multivariate methods (Chapter 4).

Chapter 3

Materials and Methods for Microsatellite Loci and Mitochondrial DNA Analyses

3.1 Introduction

The genetic techniques of choice to assess genetic variation between individual anoa were to use microsatellite markers and mtDNA sequencing. The possible mechanisms for variation in the number of microsatellite sequence repeats at given loci have been summarised in Chapter 1, Section 1.4.2; and the variation in mtDNA sequences was described in Section 1.4.4. Microsatellite markers were selected for use in this study because of the attributes detailed in Chapter 1 and their effectiveness in determining the existence of population structure in a range of wild ungulate studies that employed domestic bovine markers (Engel et al., 1996; O'Ryan et al., 1998; Maudet, Luikart and Taberlet, 2001; Nijman, Otsen, Verkaar et al., 2003).

This chapter describes the materials and methods used. Initially the sampling strategies are discussed, followed by a summary of the samples to be analysed. Then the methodology for DNA isolation and extraction are given. In the next section the DNA amplification and allele size assessment are described. This involves the use of polymerase chain reaction (PCR) for DNA amplification, followed by running of polyacrylamide gels for discriminating between alleles, and finally the method for allele size estimation is given. The last section described the PCR and gel conditions used to determine the mtDNA sequences, followed by their alignment.

3.2 Sampling methodologies

3.2.1 Sampling techniques

In poorly known groups, initial inferences about species boundaries are usually based on a subset of characters and methods of analysis that by contemporary standards would be judged inadequate. The fact that the localities from which specimens originated are unknown or that particular locations are inadequately sampled, contribute to making these assessments less reliable. In cases where species are widely distributed across a range of altitudes and habitats, detailed systematic sampling is made nearly impossible. This leaves the option of initiating a sampling strategy from information on previous taxonomic classification (Hedin and Wood, 2002), or when looking at biogeographic differences, observing the distribution of other sympatric species inhabiting a similar range.

In the case of the anoa, where the taxonomic status is questioned, the sampling should include a range of altitudes to address the proposed division between the two debated species or morphotypes, lowland / 'depressicornis' and mountain / 'quarlesi' anoa. However, many other Sulawesi species are isolated into genetically distinct units clustered geographically into biogeographic areas (Whitten et al., 2002). So the anoa sampling strategy adopted in this thesis was designed to incorporate both regional and altitudinal variations.

As part of this sampling strategy the morphological characteristics used to originally describe the two debated species were recorded. Although both morphotypes were reported, no attempt was made to equally sample the two forms. The morphological variation of the anoa has recently been reviewed (Burton et al., 2005) and these descriptions are given in Appendix 2. This table and accompanying notes clearly show why the morphological characteristics of the two putative species have been debated, because of a lack of distinction between the characteristics present in the two forms. The anoa identified by these criteria are analysed in Section 6.4.

Methods of DNA sampling of wild species have included direct animal trapping or indirect approaches such as dung, scratch posts and sticky pads (Hoss, Kohn, Pääbo et al., 1992; Valdderama, Karesh, Wildman et al., 1999). In this study the chosen method had to cause no threat to the current anoa population. The collection of anoa DNA had to be conducted using indirect and non-invasive techniques because of the logistical difficulties of trapping low density, forest dwelling and reportedly aggressive species in the two year time frame given for the sampling. Indirect collection methods such as hair from scratch posts or dung were not used because of the low density at which these animals lived and the logistical difficulties of rapid collection of the hair before DNA degradation occurred. From a pilot study in 1999 it was observed that anoa were kept in villages as pet animals and there were also significant numbers of skulls, with tissue attached, kept in village houses. So these latter two sources were combined for the acquisition of DNA from known locations.

The sampling strategy was developed to encompass as many of the notional biogeographic units as possible (initially estimated to be between two and five). This had to be balanced with the need to obtain a representative sample within each region and to be able to broadly define the locations of any population division to explore differences from the biogeographic boundaries. The sampling also had to provide a sufficient number of samples in at least two regions (Buton and Central) to permit statistical analyses as described below.

3.2.2 Statistical requirements for sampling

Both the sample size per subpopulation and the number of loci used have to be adequate to prevent type I or type II errors of statistical analysis occurring (Baverstock and Moritz, 1996). These errors are described as when a true null hypothesis is incorrectly rejected (type I) and when a false null hypothesis is not correctly rejected (type 2). The two genetic methods of microsatellite and mtDNA analysis have different requirements. The differing statistical analyses used in the assessment of microsatellite allelic variation to answer the multiple hypotheses need varying sample size and number of loci. These are discussed in the rest of this section. For mtDNA sequence analysis the requirement of for a good geographic representation of sampling area, while there is only a single gene being sequenced and multiple samples per location are not required.

For microsatellite loci, in order to assess summary statistics from allele frequencies under Hardy-Weinberg equilibrium (Chapter 4), the standard error is estimated as the square-root of the variance (Nei, 1987). By assessing the standard error of cattle breed data and the effort expended with increasing sample size, it was suggested that 30 to 60 individuals are sufficient for analysis between breeds (MacHugh, 1996). The standard error of estimating genetic distance using an adjusted F_{ST} value was greatly reduced by the inclusion of a greater number of microsatellite loci, up to 20.

Addition of more loci did not cause considerable further reduction in standard error values (MacHugh, 1996). The use of allele frequencies to assess admixture in populations, employed in Chapter 5, is influenced by the presence of private alleles (Chakraborty, 1986). For these rare alleles to be observed the sample size should be large, which has the negative effect of increasing effort.

For the assessment of population structure when the population is in Hardy-Weinberg equilibrium, differences in allele frequency are assessed. The estimation of the minimum sample size required for the detection of given levels of allele frequency differences can be determined for various type I error levels and variations in power of the tests. Smaller differences in allele frequencies between populations will be indistinguishable unless very large sample sizes are surveyed from each population (Baverstock et al., 1996). However, using the Markov chain algorithms smaller differences can be identified with fewer individuals (Raymond and Rousset, 1995). Even so, a reasonable sample size is necessary to detect the smaller differences in allele frequencies. Weir (1996a) suggested that to allow statistical testing using re-sampling for population structure, at least five locations with more than 20 individuals from each should be sampled. For the estimation of linkage disequilibrium (i.e. non-random association of alleles between two loci), a large sample size is required for accurate rejection of the null hypothesis of no population structure (Guo and Thompson, 1992). The Hardy-Weinberg and linkage disequilibrium tests are discussed further in Chapter 4.

Specifically for microsatellite marker studies with various types of analyses, estimates of the number of individuals and loci required have varied considerably.

For example, in a study investigating genetic diversity and population structure of the African buffalo (*Syncerus caffer*) 14 markers and 162 samples from nine sites were used (van Hooft et al., 2000). Assessment of the location and characteristics of a hybrid zone between sika and red deer (Genus Cervus) in Scotland was conducted with 12 microsatellites and 246 individuals (Goodman et al., 1999). Sample size in a study of Asian water buffalo (*Bubalus bubalis*) varied from 14 and 25 individuals from 11 populations for the assessment of genetic diversity with 21 microsatellite markers (Barker, Moore, Hetzel et al., 1997). While for European cattle studies the sample sizes and numbers of microsatellite markers used were higher (Wiener et al., 2004), following the recommendations developed by the FAO (Food and Agriculture Organization of the United Nations) (http://dad.fao.org/en/refer/library/guidelin/marker.pdf, August 2004) for monitoring genetic diversity.

Therefore, a sample size 25 to 40 individuals from each region was set as a target number for the analysis to be carried out in this thesis. For sample collection an additional ten individuals was thought necessary due to the likely loss that might occur as the proportion of DNA in some samples was expected to be degraded before collection and preservation. The number of loci was initially the 30 markers that had been used in studies of domestic cattle, described in Section 3.4 below. It was appreciated that this number may be reduced significantly as sequence variations between species particularly in the primer sites would result in loci not being amplified by the PCR (Pemberton, Slate, Bancroft et al., 1995; Engel et al., 1996; Maudet et al., 2001).

Prior to initiation of the main work of the thesis, a pilot study with samples from wild anoa screened five individuals from two geographic regions divided by a sea barrier (South-East and Buton) to test for differences between regions (Baverstock et al., 1996). Distinct differences in allele sizes were found between individuals from the two regions as part of this work. So the full sampling and analysis programme described in this thesis was conducted.

3.3 Sample collection and DNA extraction

3.3.1 Sulawesi Island non-invasive sampling

DNA samples were collected from five biogeographic regions of Sulawesi: North, Central, East, South-East and Buton Island. Within these locations, sampling included a range of altitudes, and individuals displaying morphological characteristics representative of the two morphotypes. DNA was extracted from hair follicles and tissue found on skeletal remains. The biogeographic regions of south and north-east Sulawesi were not sampled because anoa were reportedly rare in these areas, making sampling nearly impossible (Burton et al., 2005). In each region, samples were collected from as many locations as possible, taking into account logistical constraints. Local assistants were trained in collection of samples, data recording, sample storage and coordination with the local Forestry Departments. These individuals were mainly members of local conservation orientated Non-Governmental Organisations (NGOs).

The collection and storage of these samples was conducted as described below. A few grams or 1cm^2 of tissue or 40 large follicles, e.g. tail hairs, were collected from each individual. These were stored in 5 or 10ml screw cap tubes with one of two transport solutions: either DMSO (dimethylsulfoxide (20%)) saturated with sodium chloride, which is reported to be effective for preservation of DNA for up to two years (Amos and Hoelzel, 1991); or SDS-Urea mix, which had the advantage that it degrades the tissue allowing more effective DNA extraction. Details of these solutions are presented in Table 3.1. Samples were then stored in a refrigerator

($\sim 5^0$C) where possible. Where available, data recorded with each sample included: collection location, morphological description and measurements, sex, age and details of any photographs taken.

Table 3.1. Transport solutions used for DNA preservation.

1. DMSO (dimethylsulfoxide)	
20% v/v DMSO	200mls
NaCl	280g
Made up to 1000ml with sdH_2O	
2. SDS (Sodium dodecyl sulphate)-Urea	
8M UREA	480g
10mM TRIS	10ml (1m)
0.1% SDS	10ml (10%)
1mM EDTA	2ml (0.5M)
Made up to 1000ml with sdH_2O	

A total of 177 samples were collected and from these, 116 samples successfully produced reasonable quality DNA. A significant number of microsatellite markers could not be amplified for some samples and mtDNA primers were only able to sequence incomplete sections for a number of individuals. The source of wastage could have been due to the poor quality of DNA collected, due to the degradation of tissue or hair follicles in the tropical climate. So a number of microsatellite markers and samples were omitted from the analysis. Therefore, 89 individuals summarised in Table 3.2, were used in the microsatellite analysis described in Chapter 4 and 6. For the mtDNA analysis only 35 individuals with sufficient sequence amplified were analysed, described in Chapter 5 and 6. The individuals in Table 3.2 are grouped according to the biogeographic regions that they were sampled from. There is an additional Atypical group, which showed anomalous genetic characteristics from the other individuals in the North region (see Section 4.1.3).

For the collection of anoa DNA, permissions were obtained from the Indonesian Institute of Science (LIPI), the Director General (Dir. Jen.) of the Indonesian Forestry Department (PHKA) and four local Forestry Departments (KSDA). Appropriate permits were obtained for export of samples to the UK from Balai KSDA offices, CITES Jakarta and UK, and the Scottish Executive.

Table 3.2. DNA from anoa on Sulawesi and Buton Islands, with adjustment of regions after initial analysis for a) microsatellite analysis and b) mitochondrial DNA analysis.

	Total number of samples producing good quality DNA	**Putative species ('depressicornis'/ 'quarlesi'/ unknown)**	**Sex (male/ female/ unknown)**	**Sample type (tissue / hair)**
a)				
Buton	28	19 / 2 / 7	15 / 11 / 2	27 / 1
South-East	12	6 / 0 / 6	2 / 9 / 1	6 / 6
Central	29	14 / 1 / 14	12 / 9 / 8	22 / 7
North	8	6 / 1 / 1	0 / 0 / 8	5 / 3
Atypical	12	9 / 0 / 3	0 / 0 / 12	2 / 10
TOTAL	**89**	**54/ 4 / 31**	**29 / 29 / 31**	**62 / 27**
b)				
Buton	15	7 / 3 / 5	6 / 6 / 3	15 / 0
South-East	6	3 / 0 / 3	0 / 5 / 1	0 / 6
Central	10	10 / 0 / 0	6 / 2 / 2	9 / 1
North	2	2 / 0 / 0	0 / 0 / 2	2 / 0
Atypical	2	1 / 0 / 1	0 / 0 / 2	0 / 2
TOTAL	**35**	**22 / 3 / 8**	**12 / 13 / 8**	**26 / 7**
Total No. Individuals	**92**	**56 / 5 / 31**	**29/ 31 / 32**	**63 / 29**

Some individuals were not common to both microsatellite and mtDNA datasets hence the total number of individuals sampled is 92.

3.3.2 European and Indonesian zoological collection sampling

A total of 16 anoa held in European and Indonesian zoological collections were sampled. DNA was extracted from follicles of hairs collected from live animals, or from tissue collected during post-mortem (one individual). Of these individuals ten were amplified using microsatellite primers and one different animal with the mtDNA primers (details are given in Table 3.3). This lack of DNA from many samples was due to the poor quality of DNA extracted from small hair follicles, with more time repeat sampling and extraction would have been conducted. From European zoos, individuals from three of the four breed lines were sampled, which included individuals from both putative species. DNA was extracted from two 'quarlesi' anoa individuals (K221 and K222) and from four 'depressicornis' anoa breed lines. There has been some cross breeding between these 'depressicornis' breed lines (Nötzold, 1999) and one individual (Ch225) was identified as a cross between animals from two 'depressicornis' breed lines according to the International Anoa Studbook (Nötzold, 1999). At the outset of the study the region of Sulawesi from which the anoa originated was unknown.

The same sample collection procedures were followed as for the *in situ* sampling, and appropriate permissions were acquired before movement of collection tubes. In a number of cases samples were collected by zoo staff and sent to Edinburgh.

Table 3.3 DNA for microsatellite and/or mitochondrial DNA analysis from anoa held in European and Indonesian zoological collections.

Putative Species	**Sample code (sample type)**	**Amplified for microsatellite / mtDNA analysis**	**Supplied by, Studbook No.**
Mountain anoa (*Bubalus quarlesi*)	221 (hair)	Microsatellite	Krefeld Zoo, 275
Mountain anoa (*Bubalus quarlesi*)	222 (hair)	Microsatellite	Krefeld Zoo, 287
Lowland anoa (*Bubalus depressicornis*)	225 (hair)	Microsatellite	Chester Zoo, unknown
Lowland anoa (*Bubalus depressicornis*)	226 (hair)	Microsatellite	Chester Zoo, 399
Lowland anoa (*Bubalus depressicornis*)	235 (hair)	Microsatellite	Marwell Zoo, 328
Lowland anoa (*Bubalus depressicornis*)	236 (hair)	mtDNA	Marwell Zoo, (320 debated)
Unknown	70 (hair)	Microsatellite	Ragunan Zoo, unknown
Unknown	71 (hair)	Microsatellite	Ragunan Zoo, unknown
Unknown	77 (hair)	Microsatellite	Ragunan Zoo, unknown
Unknown	78 (hair)	Microsatellite	Ragunan Zoo, unknown
Unknown	229 (hair)	Microsatellite	Surabaya Zoo, 458

3.3.3 DNA from other species of Bovina

Outgroup species were sampled for comparison to the anoa individuals and to act as a root for tree-building analysis procedures. These are listed in Table 3.4.

Table 3.4. DNA for microsatellite and/or mitochondrial DNA analysis from outgroup species.

Species	**Sample code (sample type)**	**Amplified for microsatellite / mtDNA analysis**	**Supplied by**
Arabian oryx (*Oryx leucoryx*)	215 (tissue)	Microsatellite	Dr A. Kitchener
African buffalo (*Syncerus caffer*)	217 (DNA)	Microsatellite	Dr H. Lenstra
Asian water buffalo River type (*Bubalus bubalis*)	218 (DNA)	Microsatellite	Dr H. Lenstra
Asian water buffalo Swamp type (*Bubalus bubalis*)	219 (DNA)	Microsatellite	Dr H. Lenstra
Arabian Gazelle (*Gazella gazella*)	232 (tissue)	Microsatellite	Dr A. Kitchener
British cattle (*Bos taurus*)	260 (DNA)	Microsatellite and mtDNA (positive control in PCR)	Dr J. Williams
British cattle (*Bos taurus*)	DB2395 (DNA)	Microsatellite (size standard in ABI)	Dr J. Williams
Mongolian Cattle (*Bos taurus*)	237 (DNA)	Microsatellite and mtDNA	Dr J. Williams / Dr T. Tsedev
Mongolian Yak (*Bos grunniens*)	238 (DNA)	Microsatellite and mtDNA	Dr J. Williams / Dr T. Tsedev

3.3.4 Additional bovid microsatellite data for microsatellite analysis

Microsatellite allele data that was collected in other projects was also incorporated in the analysis stage (Table 3.5). These additional datasets were used for comparison of the anoa data with intra and inter-specific variation of other bovid species. This included data from European domestic cattle breeds from the UK (*Bos taurus*), Mongolian cattle (*Bos taurus*) and Mongolian Yak (*Bos grunniens*). It should be noted that although cattle and yak are taxonomically distinct species, they may not represent two pure species. This is because it has been reported that hybridisation of *Bos grunniens* and *Bos taurus* has occurred in China for at least 3 000 years and in northern India, Nepal and Bhutan with the cattle *Bos indicus* (Wiener, Jianlin and Ruijun, 2003). Therefore, cautious interpretation is required if this is to be used for comparison to suggest that genetic differentiation between anoa populations is greater than between two species.

The European cattle data was provided by Dr P. Wiener and Dr J. Williams (Roslin Institute) and has been analysed elsewhere (Wiener et al., 2004). Blood and semen samples were collected and laboratory work was conducted at Roslin Institute. The Mongolian cattle and yak data was provided by Dr J. Williams (Roslin Institute) and Dr. T. Tsedev (The Institute for Biology of Mongolian Academy of Science). DNA was extracted from blood samples at Roslin Institute by Dr. T. Tsedev (funded by The Royal Society). This data has not been described elsewhere.

Table 3.5. Sources of microsatellite allele data from other Bovina species.

Breed / Species*	Description	Number Sampled	Comments
Aberdeen Angus	Beef	50	Mainly UK bulls, including both 'traditional' and 'modern'
Ayrshire	Dairy	48	Mainly UK bulls
Dexter	Show	48	All UK animals
Friesian	Dairy	50	All UK bulls, none with North American Holstein ancestry
Guernsey	Dairy	64	UK, Guernsey Island and foreign bulls
Hereford	Beef	48	Mainly UK bulls
Highland	Beef	45	All UK animals
Jersey	Dairy	47	Both UK and foreign bulls
Mongolian Cattle		32	From Hovd province, Buyant sum, in western Mongolia
Mongolian Yak (*Bos grunniens*)		41	From Hovd province, Buyant sum, in western Mongolia, Range covers Tibet and central Asia, domesticated about 2,000 years ago. (http://www.animalinfo.org/species/artiperi/bos_mutu.htm#status November 2003)

* All *Bos taurus* species unless stated in brackets; includes Wiener et al. (2004) data.

3.3.5 Extraction of DNA from tissue samples

The protocol for extraction of DNA from tissue was adapted from Sambrook, Fritsch and Maniatis (1989). This is termed the phenol-chloroform method. To 0.5cm^2 of diced tissue 500ul blood lysis buffer (Table 3.6a) was added, these were then mixed using a 'vortex' mixer for five seconds. 50ul 10% SDS and 50ul Proteinase K were then added. The samples were then 'vortexed' for five seconds, followed by incubation for ten hours, or overnight, at 65^{0}C in a waterbath.

The samples were then centrifuged (five seconds at 13,000rpm) and 500ul phenol was added. The tubes were mixed by hand for ten minutes and centrifuged for two minutes to separate organic and aqueous phases. The aqueous (top) layer was removed to a new tube. Then 250ul phenol and 250ul chloroform were added to this aqueous layer, mixed for ten minutes as above, and centrifuged for two minutes. The lower layer was removed to waste. This was repeated again, followed by the addition of 500ul chloroform. The phenol, chloroform and aqueous samples were then mixed for ten minutes, centrifuged at 13,000rpm for two minutes and the upper aqueous layer removed and separated into two tubes containing DNA precipitated with two volumes (750ul) 100% EtOH. Tubes were left at -20^{0}C for a minimum of two hours to allow the DNA to precipitate.

Precipitated DNA was pelleted by centrifugation at 13,000rpm for ten minutes. The majority of the ethanol was aspirated off. The pellet was washed twice with 70% ethanol with centrifugation at 13,000rpm. Care was taken not to lose the DNA pellet when aspirating the ethanol. The pellet DNA was then air-dried and the DNA dissolved with 0.5ml of deionised H_2O. The DNA was dissolved by warming to 70^0C for two hours, before leaving at 40^0C overnight. Tubes containing the DNA were labelled with further details, such as date, replication number and my name. The amount of DNA recovered was very variable because of the variation in freshness of the sample that allowed different degrees of degradation of DNA, and the type of samples (e.g. tissue type). DNA samples were stored in a laboratory freezer at -20°C.

3.3.6 Extraction of DNA from hair samples

The methodology of hair sample DNA extraction followed Higuchi, von Beroldingen, Sensabaugh et al. (1988). Five hair follicles were cut from the hair samples, just above the root and placed follicle down-most in a 1.5ml Eppendorf tube. Then 200ul hair lysis buffer (Table 3.6b) was added, and this was incubated at 56–60^0C for 30 minutes in a waterbath. Following this 100ul Proteinase K (10mg/ml) was added. The contents were mixed using a 'vortex' mixer for five seconds and returned to 56–60^0C for a further two hours. The solution in the tubes was brought to the bottom (water having condensed on the lid during incubation) by centrifugation for five seconds (13,000rpm). Then 300ul phenol was added, mixed by gentle inversion by hand for four minutes and the organic and aqueous layers separated by centrifugation for five minutes at 13,000rpm. The top, aqueous, layer was removed into freshly labelled tubes and 150ul phenol and 150ul chloroform were added. These were mixed by hand for four minutes and then centrifuged for five minutes. Following removal of the lower organic layer, 300ul chloroform was added to the aqueous layer and mixed then centrifuged as above. The upper aqueous layer was divided into two fresh labelled tubes containing 700ul 100% EtOH. After inverting the tubes to mix the samples they were placed at -20^{0}C for a minimum of two hours to allow the DNA to precipitate. DNA was then collected by centrifugation for ten minutes at 13,000rpm and the supernatant was aspirated off. The DNA was washed twice in 70% EtOH and air dried. The DNA was re-suspended in ~25ul deionised H_2O. DNA was dissolved by heating to 70^{0}C for two hours then left overnight at 40^{0}C, before storing at -20^{0}C.

Table 3.6. Lysis buffer for Proteinase K DNA extraction from tissue and hair.

a) Tissue lysis buffer for Proteinase K digestion buffer

0.1M Tris-HCl (pH 8.0)	10mls (1M Tris)
0.25M NaCl	1.46g
0.01M EDTA	2ml
Made up to 100ml with sdH_2O	

b) Hair lysis buffer for Proteinase K digestion buffer

0.1M Tris-HCl (pH 8.0)	10mls (1M Tris)
0.5M NaCl	2.92g
0.01M EDTA	2ml
1% SDS	5ml
0.02% Mercaptoethanol	20ul
Made up to 100ml with sdH_2O	

3.4 Microsatellite markers and data collection

3.4.1 Bovine microsatellite markers

In an initial screen, 30 microsatellite markers were tested on a subset of anoa samples from two of the sampling regions (Buton and South-East). The panel of 30 markers (from the FAO recommended list) identified for the European Cattle Diversity Project were the starting point for selecting markers. These markers were selected in 1996 broadly using the following criteria: existence of prior population data, readability of marker, evenly spaced throughout the genome, number of alleles, suitability for use in an automated sequencer, possibility of multiplexing, and cross-species utility. Further details can be found about the ResGen project at http://www.projects.roslin.ac.uk/cdiv/markers.html (January 2004). The criteria used for selection of these markers for use with the anoa samples also included them being polymorphic in anoa, reliably amplified, ease of reading the allele sizes and the variance and effective number of alleles.

In total 13 markers proved to meet these criteria and were used with the complete set of anoa samples. The range of allele sizes for each of the 13 markers is shown in Table 3.7, column 4. The 17 unsuitable markers were omitted for two reasons. Firstly, after the initial trials, some markers were removed from the analysis because they did not amplify the anoa DNA. It was expected because of possible mutations at the primer locations as has been reported for other cross-species studies of bovine markers (Engel et al., 1996; Slate et al., 1998; Maudet et al., 2001). Pemberton et al. (1995) recognised that three of 16 microsatellite loci contained segregating nonamplifying alleles when conducting parentage assignment of red deer (*Cervus elaphus*) using sheep and cattle markers. These uninformative alleles should be taken into account during statistical analysis. Secondly, following tests with an increased number of anoa samples, it was found that some markers were either homozygous or showed little allele size variation between individuals. A lack of allele size variation was recognised as a possible problem because of the distinct ancestry of the domestic cattle and the anoa (Groves, 1981). This was expected because even though markers were identified as being informative for cattle the same markers may not be variable in the anoa. Polymorphic markers from cattle were also found to show no polymorphism when used with sheep and deer (Slate et al., 1998). Therefore, these awkward loci were removed from the analysis. In this study, redesigning primers was not feasible for markers where heterozygous loci may have been defined as homozygous, due to nonamplification of one allele. Also further markers were not added to replace those removed. Both of these possible improvement to the analysis were not conducted because a lack of time.

These same microsatellite markers were used in other studies of European domestic cattle and Mongolian cattle and yak. DNA from the European cattle had previously been amplified for the 30 markers. The allele size range is shown in Table 3.7, column 3. For Mongolian cattle and yak samples only 22 of these markers were successfully used.

Table 3.7. The 30 bovine microsatellite loci and the allele size range of the 13 selected for amplification of anoa DNA, including PCR conditions. Chromosome numbers for cattle are from a bovine genetic linkage map (Barendse, Vaiman, Kemp et al., 1997). Cattle and anoa product size was measured in base pairs and adjusted to the European Cattle Diversity Project sizes. References for these markers can be found at http://www.projects.roslin.ac.uk/cdiv/markers.html (January 2004).

Locus code & gene	**Chromosome**	**Cattle product size (bp)**	**Anoa product size (bp)**	**MgCl$_2$ (mM)**	**Anneal temp. (^{0}C)**
BM 1818	23	258-268	264-272	2	55
BM 1824	1	164-191		1.5	60
BM 2113	2	124-142		2	60
CSRM 60	10	89-105	91-99	1.5	55
CSSM 66	14	179-205		2	55
ETH 3	19	109-129		1.5	65
ETH 10-2	5	209-223		2	65
ETH 152	5	195-209	191-201	2	65
ETH 185	17	220-242		-	-
ETH 225	9	141-152		3	65
HAUT 24	22	104-124	104-120	3	55
HAUT 27	26	128-156	128-154	2	60
HEL 1	15	103-115		1	60
HEL 5	21	153-169		1.5	55
HEL 9	8	147-169		2	60
HEL 13	11	184-196	160-186	3	55
ILSTS 5	10	182-186	180-184	3	55
ILSTS 6	7	285-301		2	55
INRA 5-2	12	139-143		1.5	60
INRA 23	3	199-219		1.5	55
INRA 32	11	168-188		2	60
INRA 35	16	102-120	122-138	1.5	55
INRA 37	11	112-148	114-142	1.5	55
INRA 63	18	173-185		2	60
MM 12	9	115-137	83-99	2	60
SPS 115	15	236-256	241-254	1.5	65
TGLA 53	16	151-179		1.5	60
TGLA 122	21	138-184		1	65
TGLA 126	20	117-131	105-109	2	55
TGLA 227	18	79-105	69-83	1.5	55

3.4.2 Polymerase chain reaction (PCR) of microsatellite markers

The optimization of PCR reactions was carried out to identify the best combinations of annealing temperature and concentrations of $MgCl_2$ (Table 3.8a and c). Annealing temperature was varied from 55–65^{0}C (in 5^{0}C increments). Concentrations of $MgCl_2$ varied from 1-3mM (in 0.5 increments). PCR were performed using a Hybaid *OmniGene* thermal cycler or a Tetrad thermal cycler. The conditions used are shown in Table 3.8b. All PCR reactions included dye-labelled primers to facilitate analysis on ABI DNA fragment analysers. The reactions were carried out in 96-well microtitre plates with 1ul of DNA of unknown concentration. The reaction volume was 10ul. An oil overlay of 20ul was added and the plate was centrifuged at 700rpm before running the PCR. A positive and negative control was included on each 96 well plate and a size standard (cattle DB2395) was included in each reaction. The PCR product was stored at -20^{0}C covered with a Clingfilm wrap prior to analysis.

The PCR products were analysed on 1% agarose gel with 1X TBE buffer, 0.5µg/ml of ethidium bromide. Hyperladder I (3ul) was used as a size standard. Identification of the optimal PCR conditions was done visually using a UV transilluminator to photograph the gel, and comparing the result produced with different PCR conditions. The optimal conditions identified are shown in Table 3.7, column 5 and 6, above.

Table 3.8. PCR buffer and thermal cycler conditions.

a) PCR 20X Reaction Buffer

dNTPs	20ul
Buffer	20ul
Forward primer (2 pmol)	20ul
Reverse primer (2 pmol)	20ul
$MgCl_2$ (10mM/20mM)	20/30ul
Taq polymerase	1ul
sdH_2O	79/69ul

b) Cycling Parameters for PCR reactions

94 °C	3 minutes	
94 °C	45 seconds	)
55-65 °C	45 seconds	) 40 cycles
72 °C	45 seconds	)
72 °C	3 minutes	

c) PCR reaction buffer components

dNTPs	0.1mM dATP, dGTP, dTTP, dCTP
Buffer	Stock solution (Advanced Technologies, Epsom, UK)
Primers	Diluted from stock (Invitrogen, UK)
$MgCl_2$	Diluted from stock solution (Advanced Technologies, Epsom, UK)
Taq polymerase	Biotaq DNA polymerase (Bioline, UK)

3.4.3 Polyacrylamide gel

Polyacrylamide gel electrophoresis was used to identify allele sizes of DNA amplified for each of 36 individuals at a time. Formamide solution (3.5ul) (Table 3.9a) was mixed with 1ul of PCR product. The samples were heated at 92°C to denature DNA and placed on ice. Then 3.5ul was loaded onto 6% polyacrylamide sequencing gel, described in Table 3.9b. The gel was set to run on an ABI 373 (Applied Biosystems) for three hours in 1X TBE electrophoresis buffer. The resulting image was analysed using Genescan and Genotyper software (Applied

Biosystems) to reveal genotypes for each individual (see below). Analysis was carried out in comparison with ROX 500 as an internal standard. The cattle standard of known size and positive and negative controls for each PCR were also included. Relative concentrations of DNA were adjusted by trial and error, as these varied between individuals.

Table 3.9. Polyacrylamide gel preparation.

a) Formamide solution preparation		
	Deionised formamide	114ul
	Bromophenol blue 0.25%	5ul
	TAMRA	17ul
b) Polyacrylamide gel preparation		
	Acrylamide:Bis	6ml
	Acrylamide (19:1) 6%	
	Urea 50%	20g
	TBE	4ml
	sdH_2O	16ml
For polymerisation of gel:	TEMED	18ul
	APS (10%)	200ul
c) 10X TBE gel electrophoresis buffer		
	Tris-borate (pH 8.3)	1.00 M
	EDTA (pH 8.0)	0.02 M

3.4.4 Identification of microsatellite allele size

The polyacrylamide gel image was transferred to further programs for automatic analysis. The electrophoretogram was scored with the programs Genescan 3.0 (Applied-Biosystems, 1996a) and Genotyper 2.5 (Applied-Biosystems, 1996b). The program Genescan was used to check that the cattle internal standards had the correct allele sizes. Also the cattle size standards were measured at this stage to check that they agreed with the known allele size. Once completed, alleles carried by each individual were identified with the program Genotyper using target allele size ranges identified during the European Cattle Diversity Project. On occasion different allele sizes from those recorded in the ResGen project were observed for the other species studied. An allele was identified as a target microsatellite allele if it was observed as being of a similar size in other individuals of the same species and displayed the typical stutter pattern of the marker. The sizes for all alleles were then converted to tabular form and rounded to the European Standards to allow comparison with other datasets. This data was saved in Excel files, which allowed preparation and formatting for statistical analysis in Chapter 4 and 6.

3.5 Mitochondrial DNA sequencing and data collection

3.5.1 Bovine mitochondrial DNA D-loop primers

Initially two pairs of mtDNA D-loop primers were tested for their applicability with anoa DNA. These were designed to work on multiple bovine species and the sequences were supplied by Dr. D. Bradley, Trinity College, Dublin. These were: AN3 REV CGA GAT GTC TTA TTT AAG AGG and AN4 FOR GGT AAT GTA CAT AAC ATT AAT G (located between base 15960 to 16334 on the cattle D-loop) and AN5F TAT GCC CCA TGC ATA TAA and AN6R GCA GTT AAG TCC AGC TAC AA (located between bases 16019 to16460 on the cattle D-loop). The primer pair AN3 and AN4 was selected because of better amplification of anoa DNA shown on agarose gels. M13 tails added to the 5' end of the primers were M13 forward: 5'-TGT AAA ACG ACG GCC AGT G-3' and M13 reverse: 5'-GGA AAC AGC TAT GAC CAT G-3'.

3.5.2 Polymerase chain reaction (PCR) of mitochondrial DNA

Optimization of PCR conditions was conducted to determine the best annealing temperature and $MgCl_2$ concentration (1.5, 2, 3mM) for the D-loop primers. Temperature was varied according to Table 3.10b and variations for each column shown in Table 3.10c. PCR were performed using a Hybaid *OmniGene* thermal cycler or a Tetrad thermal cycler. The PCR reaction buffer components were as described for microsatellite loci in Table 3.8c except for the primers described above, which were supplied by Invitrogen UK. The reactions were carried out in 96-well microtitre plates with 1ul of DNA of unknown concentration, with a positive (a size standard sample: cattle DB2395) and negative control included on the plate. The reaction volume was 10ul. The plate was centrifuged at 700rpm after an oil overlay of 20ul was added and before running the PCR. The PCR product was stored at -20^0C covered with a Clingfilm wrap prior to sequencing, described in the section below. The optimal PCR conditions were 60°C with an $MgCl_2$ concentration of 2mM. The initial step was for 4 minutes at 94°C and a final step of 4 minutes at 72°C. The PCR work was conducted by H. Greer under my supervision.

Table 3.10. PCR buffer and thermal cycler conditions.

a) PCR 20X reaction buffer

dNTPs	20ul
Buffer	20ul
Forward primer (2ngµl^{-1})	20ul
Reverse primer (2ngµl^{-1})	20ul
MgCl$_2$ (10mM/20mM)	20/30ul
Taq polymerase	1ul
sdH$_2$O	79/69ul

b) Cycling parameters for PCR reactions using the 'Touch-Down program'

94 °C	4 minutes	
94 °C	40 seconds	)
See c) below	40 seconds	) 40 cycles
72 °C	40 seconds	)
72 °C	4 minutes	

c) Touch-Down program PCR Cycling part of b) above

Column Number	1st 5 Cycles	2nd 5 Cycles	30 Further Cycles
1	58.0	56.0	54.0
2	58.3	56.3	54.3
3	58.8	56.8	54.8
4	59.7	57.7	55.7
5	60.9	58.9	56.9
6	62.3	60.3	58.3
7	64.0	62.0	60.0
8	65.4	63.4	61.4
9	66.5	64.5	62.5
10	67.3	65.3	63.3
11	67.8	65.8	63.8
12	68.0	66.0	64.0

3.5.3 Sequencing preparation

Reagents listed in Table 3.11a (details in Table 3.11c) were added to each well containing the PCR product from Section 3.5.2, then mixed and spun briefly. A 10ul volume was required for the sequencing reactions. These were run on the thermal cycler according to the conditions in Table 3.11b.

Table 3.11. Sequencing reaction buffer and thermal cycler conditions.

a) Sequencing reaction buffer

	BigDye 3.1	0.5ul
	Buffer	1.0ul
	50% DMSO	1.0ul
	Primer	0.5ul
Template:	PCR product	3-10ng
	MilliQ	*q.s*
	Total volume	10ul

b) Cycling parameters for reaction

96 °C	3 minutes	1 cycle
96 °C	30 seconds	)
50 °C	20 seconds	) 20 cycles
60 °C	4 minutes	)
4 °C	Hold	

c) Reaction components

ABI BigDye 3.1 sequencing kit	Part No 4337455
ABI BigDye sequencing buffer	Part No 4336697
Primer (5pmol/ul)	Diluted from stock (Invitrogen, UK)
50% DMSO	Diluted from stock
MilliQ water	
96 well plate	ABI N8010560 or Abgene AB-1100
Purified PCR fragments	

Purification of the extension products was conducted by the addition of 2.5ul of 125mM EDTA to each well. Then 27ul of 96% ethanol was also added, making sure the EDTA and ethanol reached the bottom of the wells. The plates were sealed with tape and inverted four times. They were then incubated at room temperature for 15 minutes and centrifuged at 2,500rpm for 30 minutes. The plates were then inverted onto tissue and centrifuged up to 100rpm. Following this, 100ul of 70% ethanol was added to each well and the plates were centrifuged for 15 minutes at 1,650rpm. The plate was inverted onto tissue and centrifuged at 100rpm for one minute. The samples were then re-suspended in 10ul MilliQ water.

The ABI 3700 automatic DNA capillary sequencer (Applied-Biosystems, 1999) was used for sequencing these samples. A POP-6 polymer was used as the matrix. Interpretation of the sequencing was scored using three programs OrbixWeb™ Deamon software, 3700 DATA collection software and DATA Extractor software. The sequencing was conducted by F. Murray of the ARK Genomics Group at Roslin Institute.

3.5.4 Sequence alignment

The sequences were aligned by eye and using MegAlign program (DNASTAR, Madison, WI). A 298 base section was successfully sequenced that was located between 15960 to 16334 on the cattle D-loop (Anderson et al., 1982). Assistance was given by A. Hastings (Institute of Cell, Animal and Population Biology, University of Edinburgh). The sequence data was then stored in a text file for analysis, described in Chapters 5 and 6.

Chapter 4

Results: Population Structure and Taxonomic Status of the Anoa Assessed with Microsatellite Markers

4.1 Introduction and data preparation

4.1.1 Introduction

This chapter will focus on defining the presence or absence of population structure of the anoa sampled throughout Sulawesi and Buton Islands using microsatellite markers. The diversity within subpopulations and differentiation between them will be investigated. Phylogenetic analyses will use a scale of genetic distance as an indicator of distinctiveness for comparison of anoa to other bovid species. This may have implications for the taxonomic status of the anoa. Comparisons of microsatellite diversity to morphological, geographic and altitudinal variation will be made in Chapter 6.

The chapter is divided into seven sections, data preparation, five analysis sections and a summary. The analyses conducted are: assessment of Hardy-Weinberg equilibrium and population structure; allelic diversity; population differentiation; phylogenetic analysis; and assessment of diversity in the *ex situ* population. The first investigates population structure using the Hardy-Weinberg test and Bayesian analysis. The second analysis section (4.3) investigates intra-anoa variation of allele frequencies and heterozygosity and compares this with variation within and between other species. Thirdly, the F_{ST} test is used to assess for population differentiation and this is followed by phylogenetic analyses to calculate genetic distances between individuals and clusters of anoa and again makes comparisons with other species. The last analysis section investigates the representation of the genetic variation in the zoo population in relation to the wild population with Bayesian and multivariate analyses.

The preparation of data is covered in detail below. Firstly, possible causes of bias are assessed, such as missing values and possible sample duplication, and reduced where possible. Secondly, the geographic distribution of the individuals is presented, with these labelled according to their bio-geographic location and genetic clustering.

4.1.2 Missing values

The quality and variation in concentration of DNA, arising from the variable quality of samples obtained from the field, meant that many of the samples had missing values for at least one marker locus. While repeating DNA preparations, PCR amplification and ABI gel analysis may have allowed some missing values to be recovered, experience showed that the main problems arose from the quality of the original sample and not subsequent processing. Time constraints did not permit the exhaustive re-processing and testing of samples to produce a more complete dataset. The effect of missing values on genetic data analysis has been little investigated (Troyanskaya, Cantor, Sherlock et al., 2001). Some studies set a level of permitted missing values per individual that will be accepted, such as missing data for two of ten loci (Culver et al., 2000). A level of 10% was thought to be a conservative level that should not produce significant bias (A. Springbett pers. comm.). For the analyses using 13 loci a 10% missing data level was used. This equated to the inclusion of individuals with (a maximum of) missing data from one locus. In cases where only six loci were analysed, then only individuals with no missing values were included. This level of 10% missing values was found to have minimal effect on the data presented after comparison or results from a complete dataset. There has been no commonly used method developed to estimate missing values that has been tested for bias to the original data (A. Springbett pers. comm.).

4.1.3 Samples with identical genotypes, erroneous sampling or real individuals?

Duplication of samples might have occurred in the field or during laboratory analysis. Efforts were made to reduce the risk of this. The difficulty was in determining between two or more genotypically identical samples that were caused by duplication from those that were genuinely individuals with identical genotypes. During analysis it was therefore necessary to be aware that duplicated samples may be found, and that they would have to be removed if identified as such. For example, MacHugh et al. (1998) recorded two samples with identical allele sizes for 30 cattle microsatellite loci. This was attributed to laboratory error; the chance of this occurring in non-identical individuals was not mentioned. During the sampling clear instructions were given about the collection of the anoa samples to assistants in this study. Even so, there was no way to completely discard the possibility of duplication.

Following genotyping, seven samples from the North region were found to have identical genotypes for all 13 loci (sample numbers 111, 124, 125, 126,127, 136, 138). Also two pairs of samples were identical, found in the South-East and North regions (sample numbers 94 and 95, 114 and 115). The likelihood of two animals selected at random having the same genotype was assessed with mismatch curves between pairs of samples drawn for each region. If all other individuals showed high levels of variation this would suggest that these identical animals were duplicates, caused by an error of the sampling process. In the mismatch curve this would be represented by two isolated peaks, with one around zero alleles different, this would suggest the possibility of duplicate samples being present. On the other hand one peak, or two peaks that were incompletely isolated, gives less evidence that certain individuals are duplicates. These two peaks could occur for a number of reasons, such as the sampling of close relatives as well as repeat sampling of the same animal. Another possible explanation is that the sampled population could be a homogeneous subpopulation with little diversity observed. Although, in the case of the anoa there is little biogeographic evidence to support the existence of this subpopulation. A final scenario that could have caused this bimodal curve is the presence of two divergent genotypes, that showed significant variation, but within each genotype they could be homogeneous.

The regional mismatch curves for the anoa displayed two incompletely isolated peaks, suggesting the existence of real individuals with low genotypic variation, explained by one of the scenarios mentioned above. Therefore, this provided little evidence for removing these samples; instead they were included in the analysis. The seven samples from the North region with identical genotypes and five other individuals, with similar alleles and from the same small geographic locality, formed a genetic cluster during population structure analysis. This subpopulation did not conform to the biogeography of the region or the geographic distribution of the other anoa samples. It was therefore termed the 'Atypical' subpopulation in analyses below, and was interpreted cautiously where comment was made.

4.1.4 Geographic distribution of individuals analysed for microsatellite allele variation

The initial sample collection was aimed at examining geographic representation as well as obtaining genetic variation that may be related to altitudinal or morphological variation. As many other Sulawesi species have been shown to fall into a bio-geographic pattern defined by the peninsulas of the island, this was considered as a suitable initial grouping (Evans et al., 1999; Groves, 2001). However, for logistical reasons samples were somewhat clustered and did not represent a random distribution. Figure 4.1 shows the clustering of the 89 individuals sampled according to regional origins.

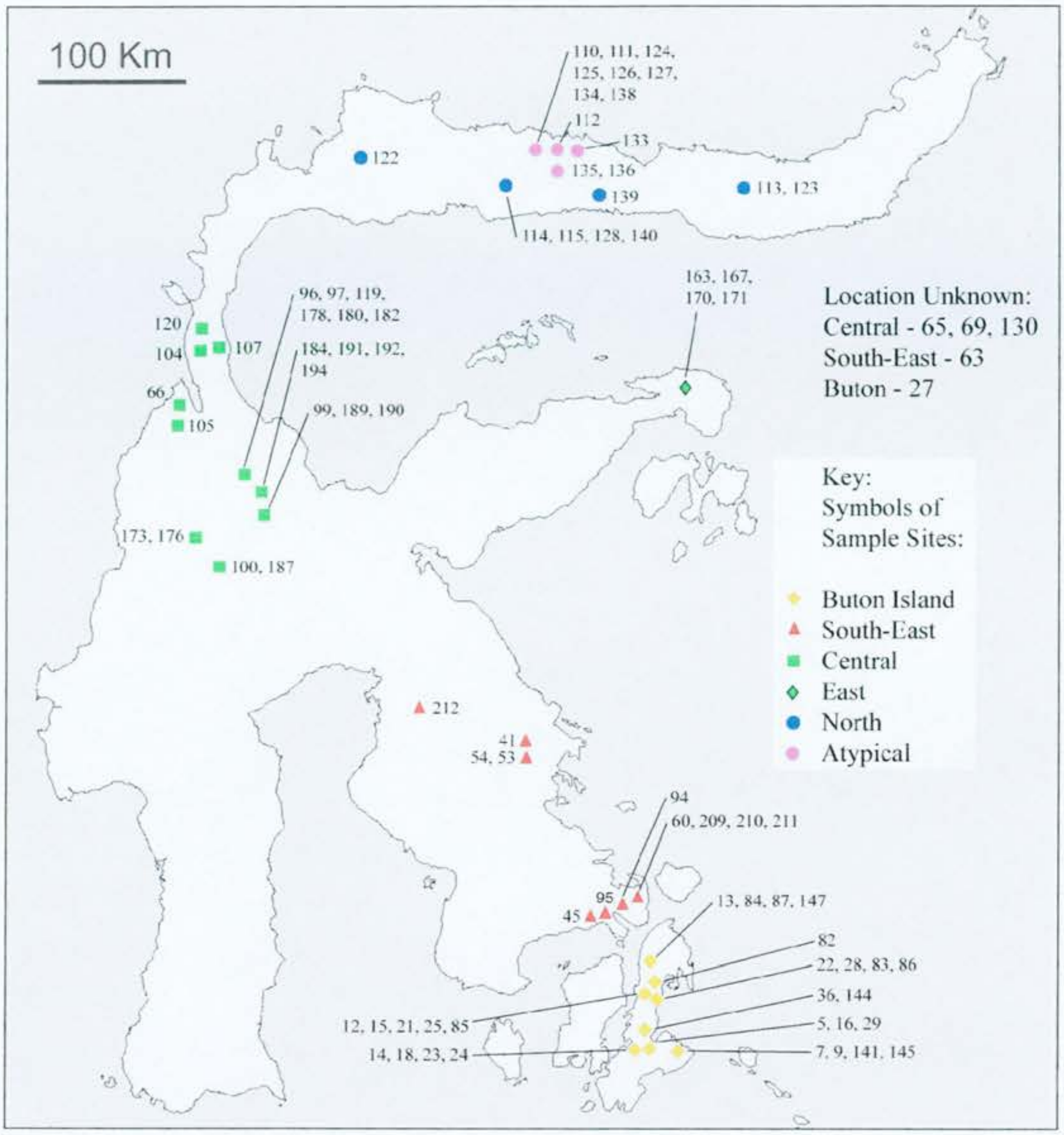


Figure 4.1. The geographic distribution of 89 individuals with symbols representing the four biogeographic regions and microsatellite genetic clusters and an additional 'Atypical' genetic cluster.

4.2 Population structure

4.2.1 Introduction

Genetic diversity and population structure are assessed using observed and expected heterozygosity, the Hardy-Weinberg equilibrium test (HWE test) (Nei, 1987) and later Bayesian analysis. The HWE is the state of allele and genotype frequencies in a population under random mating. This test provides a basis for assessing natural selection and modelling the effects of inbreeding or allele frequencies at dominant loci. In this study the test will be used to determine if there is random mating throughout the anoa's range, or if there are subdivisions present. This test is described further in Section 4.2.4. As only minor deviations from Hardy-Weinberg equilibrium were found genotypic disequilibrium and differentiation measures were not considered.

The results for various combinations of clustering of individuals into groups were analysed; these found that grouping anoa individuals based on genetic diversity produces clusters that correlate well with geographic regions of Sulawesi. The level of genetic variation in anoa was compared with the genetic diversity in other bovid species.

4.2.2 Observed and expected heterozygosity and Hardy-Weinberg equilibrium

If a population is at Hardy-Weinberg equilibrium, then this population can be completely characterised by allelic frequencies. If there is non-compliance with HWE then it may be that one of a number of biological or sampling factors may have caused this deviation. These can be summarised as: population subdivision (admixture), non-random mating (male dominance), natural selection, migration, temporal sampling effects, sampling closely related animals, presence of undetected null alleles during analysis, and small sample size (Frankham et al., 2002). The patterns of deviation may provide information that a locus is deviating from HWE across populations, suggesting possible linkage disequilibrium between loci within a population. Alternatively, a population may deviate for multiple loci due to subdivision or migration or undergoing non-random mating. The anoa population may be subdivided into demes or may have non-random mating, but it is unlikely that migration has an effect as nearly all the geographic range of the species was surveyed. In the case of a population deviating from Hardy-Weinberg equilibrium, genotypic frequencies are used for comparison of populations of a species (Weir, 1996a).

The actual number of observed heterozygote individuals and calculated expected heterozygote individuals were obtained using Program GENEPOP (Raymond et al., 1995). Observed (Ho) and expected heterozygosity (He) were calculated as the proportion of heterozygous individuals among all individuals from each locus / site combination. The difference between both Ho and He was used for assessment of HWE. To compare population variation, expected heterozygosity was used in preference to observed heterozygosity, as has been reported to be less affected by sampling bias (Weir, 1996a). Deviations from Hardy-Weinberg equilibrium were tested using an exact, two-tailed Hardy-Weinberg test (Weir, 1996a). This test investigated all possible sets of genotype frequencies for the observed set of allele frequencies to be analysed. If observed genotypic frequencies did not fit the HWE, then this hypothesis was rejected.

The Markov chain method was employed in this analysis to estimate without bias the exact P-value of this test (Guo and Thompson, 1992). Using a randomisation procedure the aggregate probability was calculated to determine the level of significance for the possible rejection of the HWE (Weir, 1996b). Estimation of exact P-values was implemented using settings: dememorization number = 1000, number of batches = 100, iterations per batch = 1000 (Guo et al., 1992). Assessment of alternative methods showed that this combination of analyses provided more power than other methods such as goodness of fit x^2 statistic (Weir, 1996b). Bonferroni corrections were conducted for these multiple tests (Rice, 1989). Bonferroni corrections were used to reduce type I errors, i.e. incorrectly declaring a deviation from the null model due to chance producing the observed state of events. This chance increases with multiple tests. So if the level was set at 0.05 for one test, then with multiple tests such as 10, a false positive was likely (0.50). Using the Bonferroni correction meant that the level of each individual test was adjusted downwards to ensure that for each test the level remains at 0.05.

4.2.3 Population Structure according to HWE

Results were calculated for the grouping of the samples into different numbers of populations, from a single 'global' population to dividing the population into between two and seven groups. In Table 4.5 the results presented are for the data divided according to the five groups and secondly treated as a single population (Global). These results show the existence of structuring in the anoa population and they do not support a single homogeneous population. The number of deviations was greatest at one population, all 13 loci deviated from HWE at P=0.05 level. The number and configuration of subpopulations was more difficult to determine. With an increase in the number of populations up to five this reduced the number of deviations from HWE to seven locus / site correlations of the 65 combinations at P=0.05 level. Four geographically important and distinct areas were defined as the population structure from this analysis; there were Buton, South-East, Central and North, with the Atypical individuals as a separate group. The number of deviations in each subpopulation was one in the Buton subpopulation (TGLA126), two in the South-East (BM1818 and SPS115), and four in the Central subpopulation (HAUT24, MM12, SPS115 and TGLA227). There were seven deviations if all samples from the North region were included in one population, however when these were separated into two populations there were no deviations from HWE.

The number of deviations was further reduced by two locus / site correlations following the removal of four individuals from the Central group that were sampled from the east peninsula. They were included in this population due to their proximity by the shortest land route and the reported biogeographic similarities between the Central area and east peninsula. Further alternative combinations such as dividing the data into two populations of Buton Island and Sulawesi Island did not further reduce the number of deviations from HWE.

Table 4.1. Anoa expected heterozygosity, observed heterozygosity, and Hardy-Weinberg test (P-values after Bonferroni correction) for clustering of anoa into five groups and for all individuals as a 'Global' single population. NS stands for 'non-significant'.

	5 Groups:					Global
	Buton	South-East	Central	North	Atypical	One Population
BM1818						
He	0.480	0.167	0.138	0.143	0.001	0.185
Ho	0.008	0.062	0.016	0.089	0.001	0.035
HWE	NS	0.046	NS	NS	NS	0.001
CSRM60						
He	0.250	0.167	0.241	0.250	0.917	0.365
Ho	0.008	0.056	0.010	0.075	0.069	0.016
HWE	NS	NS	NS	NS	NS	0.001
ETH152						
He	0.107	0.833	0.621	0.750	0.001	0.462
Ho	0.021	0.004	0.003	0.056	0.001	0.007
HWE	NS	NS	NS	NS	NS	0.001
HAUT24						
He	0.107	0.667	0.621	0.625	1.000	0.604
Ho	0.019	0.009	0.007	0.037	0.083	0.017
HWE	NS	NS	0.013	NS	NS	0.001
HAUT27						
He	0.538	0.500	0.778	0.875	1.000	0.738
Ho	0.002	0.013	0.004	0.017	0.083	0.022
HWE	NS	NS	NS	NS	NS	0.001
HEL13						
He	0.143	0.750	0.655	0.500	0.001	0.410
Ho	0.002	0.008	0.004	0.045	0.001	0.011
HWE	NS	NS	NS	NS	NS	0.001
ILSTS5						
He	0.107	0.001	0.429	0.625	0.001	0.232
Ho	0.001	0.001	0.004	0.021	0.001	0.005
HWE	NS	NS	NS	NS	NS	0.001
INRA35						
He	0.321	0.273	0.222	0.001	0.001	0.163
Ho	0.006	0.043	0.015	0.001	0.083	0.027
HWE	NS	NS	NS	NS	NS	0.001

continued overleaf

Table 4.5. continued.

	5 Groups:					Global
	Buton	South-East	Central	North	Atypical	One Population
INRA37						
He	0.464	0.800	0.778	0.143	0.001	0.437
Ho	0.009	-0.004	0.005	0.113	0.001	0.025
HWE	NS	NS	NS	NS	NS	0.001
MM12						
He	0.036	0.250	0.393	0.375	0.001	0.211
Ho	0.001	0.015	0.017	-0.013	0.001	0.004
HWE	NS	NS	0.001	NS	NS	0.001
SPS115						
He	0.538	0.001	0.107	0.571	0.167	0.277
Ho	0.006	0.083	0.029	0.045	-0.004	0.032
HWE	NS	0.033	0.001	NS	NS	0.001
TGLA126						
He	0.036	0.333	0.069	0.500	0.001	0.188
Ho	0.033	0.024	-0.001	-0.002	0.001	0.011
HWE	0.001	NS	NS	NS	NS	0.001
TGLA227						
He	0.179	0.167	0.276	0.250	0.001	0.174
Ho	0.017	0.061	0.018	0.074	0.001	0.034
HWE	NS	NS	0.001	NS	NS	0.001
Mean He	0.254	0.377	0.410	0.431	0.237	0.342
Mean Ho	0.009	0.026	0.009	0.023	0.012	0.011
HWE	0.001	0.001	0.001	0.001	0.001	0.001

4.2.4 Introduction to Bayesian methods

Bayesian and Markov chain Monte Carlo (MCMC) techniques have allowed the development of methods for a broad range of analyses (Beaumont and Rannala, 2004). These have included inference about partitioning into subpopulations and also assignment analysis of individuals of migrant ancestries (Pritchard, Stephens and Donnelly, 2000). Development of these methods has led to the assessment of short-term migration rates verses isolation (Ciofi et al., 1999a) and divergence estimates (Yang and Yoder, 2003). Examples of the current use of Bayesian methods include using program STRUCTURE for assignment analysis (Pritchard et al., 2000) and program '2mod' for assessment of admixture verses isolation (Ciofi et al., 1999a). The rapidly growing nature of this field is represented in the number of review papers recently produced on the subject (Holder and Lewis, 2003; Beaumont et al., 2004; Knowles, 2004).

This method has advantages over two difficulties found in traditional phylogenetic studies. The first difficulty is the stochasticity of genetic processes (Hudson, 1990). The second is the complexity and diversity of species' histories (Knowles and Maddison, 2002), which may have hindered the understanding of the population structure of the anoa. The interpretation of complex evolutionary history is made difficult by older events being perturbed by more recent events such as habitat fragmentation (Hedrick, 1999). Bayesian methods have greater power to resolve these issues than phylogenetic techniques.

The major advantage in the application of Bayesian methods over traditional phylogenetic methods is that Bayesian techniques assess the underlying evolutionary processes, not *ad hoc* explanations of genetic variation that have been reported. In this way interpretation of all the data occurs with often more accurate results, unlike when summary statistics are used (Cornuet, Piry, Luikart et al., 1999). Additionally, Bayesian methods allow assessment of the levels of confidence between different proposed evolutionary scenarios. The reduction in the computational load was provided by the development and inclusion of the Markov chain Monte Carlo (MCMC) technique. This coupled with the increase in speed of computers has facilitated rapid expansion in the use of Bayesian inferences especially in population genetics (Beaumont et al., 2004).

The basic methodology of Bayesian analysis can be described as initially specifying a model and a prior distribution of data. These are integrated during the analysis. The product is then used to determine the posterior probability for each tree. The posterior probability, which is proportional to the prior probability multiplied by the likelihood, defines the most reliable hypothesis. The MCMC algorithm is used to approximate these probability distributions. The posterior probability is then used in various forms for analyses as described above.

4.2.5 Introduction to program STRUCTURE

The program STRUCTURE (Pritchard et al., 2000) was selected because it provided a number of advantageous factors. Firstly, the lack of a distinctive pattern to the morphological variation of the anoa suggested a 'cryptic' population structure. This could be assessed in STRUCTRE by omitting prior grouping information. The program was then run with prior information related to geographic sampling location. This allowed assessment of admixture and migration between populations, which had been reported to greatly improve accuracy and assist interpretation of results (Pritchard et al., 2000). Assignment analysis with the program STRUCTURE was used to test the likelihood of assignment of an individual to more than one population simultaneously. This provided greater accuracy than the exclusion method (GENECLASS) (Rannala and Mountain, 1997). Program STURCTURE was found to be the most effective method for correctly identifying closely related populations, e.g. those with an F_{ST} value of 0.03, which was similar to some of the anoa values (Manel et al., 2002). A possible downside to using this program was that it assumes that the true population of origin was sampled. However, analyses such as HWE and the NJ tree of Dps with individuals as operational taxonomic units (OTUs), described later, placed all anoa individuals into clusters. Both these results suggest there were no single animals from a distinct population. An alternative analysis technique, the exclusion method (Rannala et al., 1997) that overcame this problem was not used as the assignment analysis performed better in assigning individuals from a variety of species to their known populations (Manel et al., 2002).

4.2.6 Clustering of anoa samples using program STRUCTURE

The dataset analysed was as above and this was prepared as described by (Pritchard and Wen, 2002). Each allele was identified by a code between one and the maximum number of alleles identified for that locus, according to the length of allele. Missing values were identified in the input file and those individuals with omitted values were included in the analysis.

A Bayesian clustering procedure was implemented in program STRUCTURE version 2.0 (Pritchard et al., 2000). This two-phase procedure involved the identification of the number of (unknown) populations (*K*). The assumptions of this model were that all loci analysed were at Hardy Weinberg equilibrium and at linkage equilibrium. The model subdivided the individuals in order to minimize departures from HWE and linkage equilibrium, which might have been the result of population division followed by admixture, hybridization or migration. The program was run with 200,000 repetitions for the burn-in period, followed by 200,000 MCMC repetitions. These burn-in and run lengths were chosen after assessment of summary statistics and repeats of different numbers of replications to check for consistency. Five iterations were conducted for each of K=1 to 15, assuming admixture and correlated allele frequencies, as this solved complex structuring of populations more realistically than other analysis methods (Manel et al., 2002). The most representative number of K populations in relation to the actual data was defined by the posterior probability, estimated using the log-likelihood of each K option.

The second step of the analysis assigned individuals to one population according to the highest percentage of membership (q). More than one parental population was permitted if admixture was recorded for an individual, shown by the percentage of membership divided between two populations. The percentage of membership was derived from the posterior probability of an individual's genotype originating from the sampled populations. For assignment analysis, the threshold value for the highest percentage of membership (q) ≥0.90 was chosen. This level was proposed in earlier literature and showed concordant results with other assignment methods (Manel et al., 2002). The threshold value of percentage of membership was interpreted as ≥90% of that individual's ancestry having been attributed to the selected subpopulation. The data was analysed using a range of values for v (migration) from 0.1 to 0.001, because there was limited information to assist in making a choice about the level of migration between groups. In this way it was possible to test if the model was robust to various levels of v (Pritchard et al., 2000).

4.2.7 Assessment of population structure

Structuring occurred in the anoa population, observed from the Bayesian analysis, but the number of genetic clusters was not clear-cut. The values of log P(k/x) reached an asymptote at a greater number of clusters than *K*=7. This is presented in Figure 4.2, which shows the difficulty in selecting a most likely number of subpopulations in the data set due to the possible effect of isolation by distance. Therefore, two factors were used to determine the number of genetic clusters from the STRUCTURE analysis; these were the geographic pattern of clusters and the confidence of cluster assignment at the different *K* values.

The assignment of individual animals from *K*=2 to 7 (Figure 4.3.) show the anoa individuals ordered according to their geographic origin, from the most southerly (left) to the most northerly (right). Each vertical bar represents one individual. Each colour represents one cluster; the length of each coloured segment represents the proportion of membership to each of the clusters for that individual. Results presented are those averaged over five runs.

The most genetically differentiated cluster was the Buton cluster that was present in *K*=3 onwards, as was the Atypical cluster (Figure 4.3.). The clustering of *K*=4 differentiated the South-East individuals that remained robust with increasing *K*. The remaining samples separated with greater *K* values, which identified a Central and North cluster at *K*=5. All these genetic clusters, except samples from the Atypical group, formed discreet and cohesive geographic clusters that broadly related to biogeographic regions of Sulawesi.

After the *K*=5 level, the confidence of assignments fell considerably. The assessment of the confidence of assignment of all individuals to their most probable cluster (mean q) at *K*=5 was 0.934, compared to 0.897 for *K*=6. Therefore, we chose to assign individuals to four subpopulations and the Atypical samples (K=5).

Although these individuals generally fit the south to north pattern of genetic clustering, one exception was a group of four individuals from the eastern peninsula. This can be seen as an anomaly in Figure 4.3. Three of the east peninsula samples (to the left of the Central cluster) showed greatest affinity to the North cluster (K=5), rather than the Central cluster.

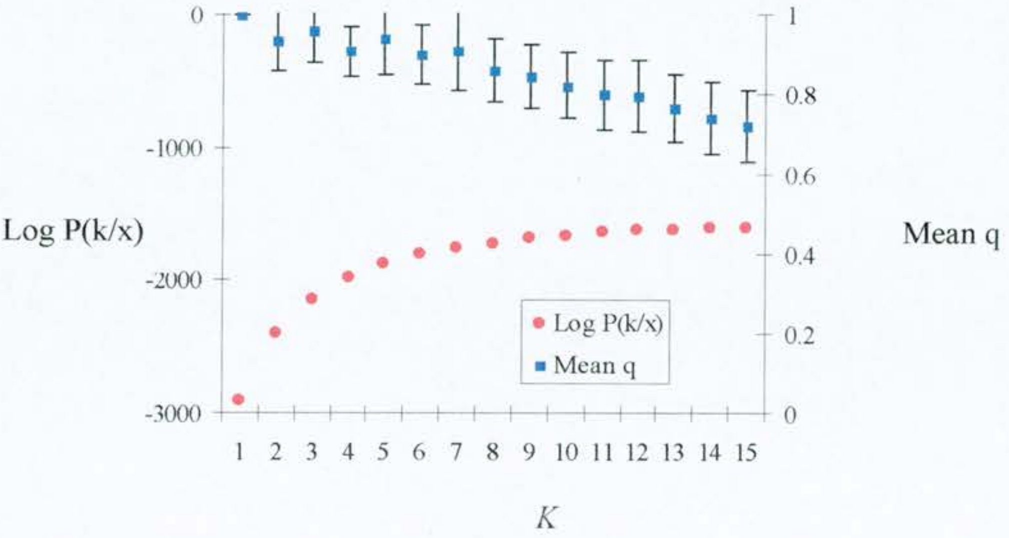


Figure 4.2. Likelihood plot of program STRUCTURE results and mean confidence assignment of all individuals to their most probable cluster indicating robustness of assignment (shown with standard error).

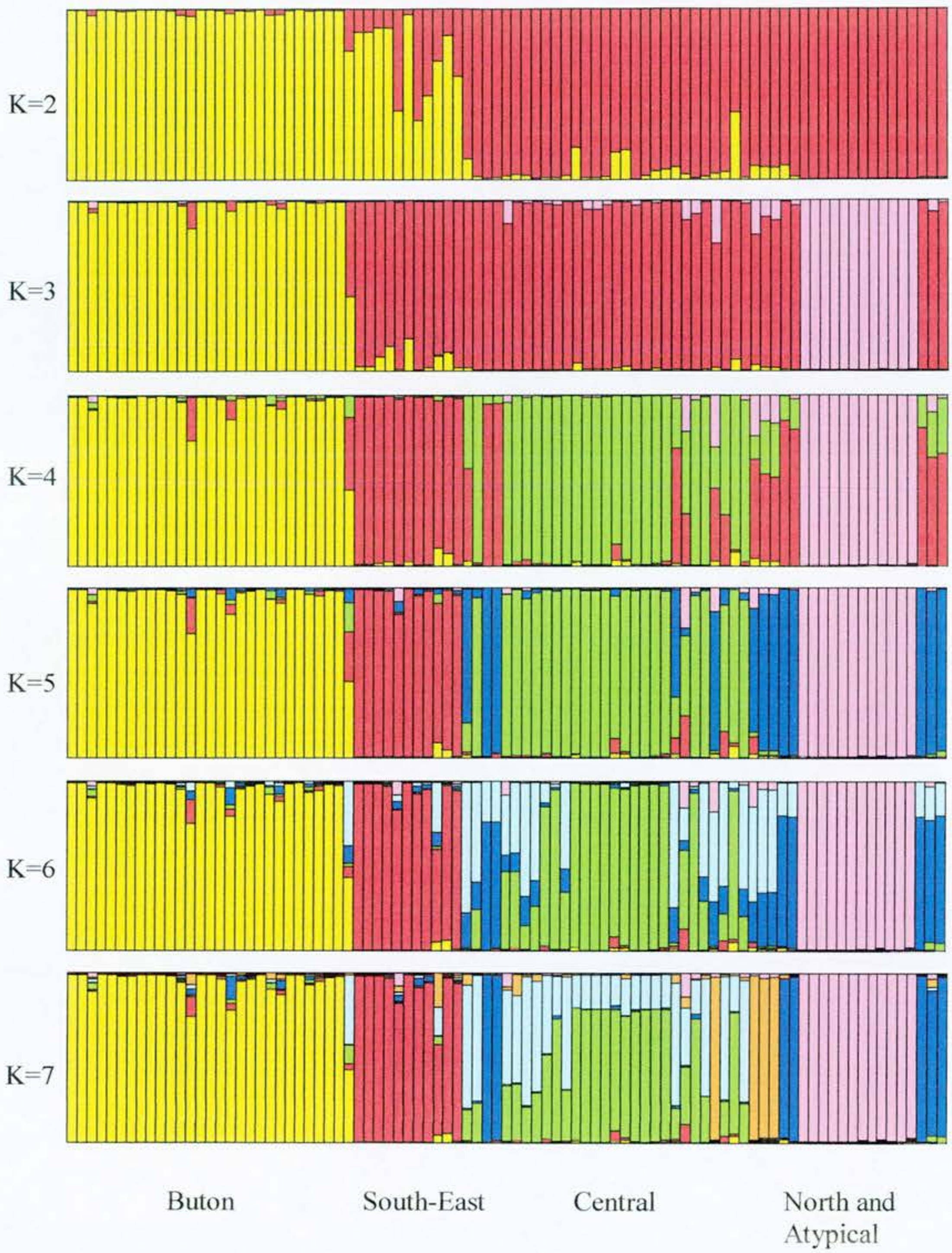


Figure 4.3: Population structure of individuals clustered according to their affinity to between two and seven clusters (*K*=2 to 7) using program STRUCTURE.

In summary, from the structure assessment above, the least robust division occurred between the Central and North clusters. These results suggest that this represents the least genetic differentiation between any pair of clusters defined here. The reverse of this is that the differentiation between Buton, South-East and all other samples was the most pronounced. On comparison of the results from program STRUCTURE with those of the HWE, there was support from both analyses for the selection of four groups of genetically and geographically clustered individuals.

4.2.8 Assignment methods for identification of *in situ* individuals to origin in Sulawesi and identification of immigrant individuals.

The results of repeated analysis of the data with various levels of v (migration) did not cause differences in results, so the model was robust and conclusions drawn from this data should be valid. Assignment analysis showed clear clustering of individuals according to the geographical groups defined by the HWE analysis. The STURUCTURE analysis identified very few admixed individuals, with the majority of individuals having membership values (q) ≥ 0.90 for their site of origin (Table 4.2). The anoa that showed values below this arbitrary level have been described individually below. Individual SE45 had the highest association to individuals from its sampling location, the South-East cluster (0.618), but also showed some affiliation to the Buton cluster (0.324), suggesting that this animal was an admixed individual. One individual (C66) was not associated with individuals from its reported origin in the Central group (0.000). Instead, this individual had the highest affiliation with the North cluster (0.993). In other words the origin of this individual had been miss-classified during sampling or it was a migrant. Individuals E170 and E171 were collected from the east peninsula, with two other samples, and included in the Central group. They showed a high percentage of membership to the North cluster (1.000 and 1.000). As there was not a large enough sample size from this east peninsula region it as not possible to determine if the origin of these individuals has been miss-classified or the east peninsula anoa are genetically similar to the North cluster.

Table 4.2. Bayesian assignment analysis of anoa using program STRUCTURE, including geographic population information for *K*=5.

Population (*n*)	Cluster (*K*) 1	2	3	4	5	Unassigned individuals	% of assigned individuals
Buton	0.997	0.001	0.000	0.001	0.000	0	100
(28)	(28)	(0)	(0)	(0)	(0)		
South-East	0.014	0.982	0.003	0.001	0.001	0	100
(12)	(0)	(12)	(0)	(0)	(0)		
Central	0.000	0.001	0.893	0.104	0.001	3	89.7
(29)	(0)	(0)	(26)	(3)	(0)		
North	0.000	0.000	0.001	0.996	0.002	0	100
(8)	(0)	(0)	(0)	(8)	(0)		
Atypical	0.000	0.000	0.000	0.000	1.000	0	100
(12)	(0)	(0)	(0)	(0)	(12)		

The assignment analysis confirmed the existence of discrete clusters of individuals, and cohesion within each of these clusters, which supported the division of the anoa population into four subpopulations (Figure 4.3 *K*=5). The presence of an apparently admixed individual and three misclassified individuals suggests introgression and possible migration either by natural events or transportation by man.

4.3 Allelic diversity of anoa

4.3.1 Introduction and allelic diversity of the loci

The number of alleles in a population is an indication of its genetic diversity. Due to the small sample size of some clusters of individuals (North n=8), the effective number of alleles is calculated (*Ne*). This value is calculated as the number of alleles as if all alleles are equally frequent, which result in the observed heterozygosity. The actual allele frequency does not assume alleles are equally frequent. This is because the effective number of alleles is less sensitive to sample size and also less affected by rare alleles. Allele frequencies can be plotted as frequency histograms to assess the distribution of allele sizes. A unimodal pattern suggests demographic expansion (Reich and Goldstein, 1998), whereas alternatives such as bimodal distribution of allele frequencies can suggest structuring within a species or multiple species.

All analysis was conducted using GENEPOP Version 3.4 (Raymond et al., 1995). Allele data was analysed to give effective number of alleles, allele frequencies (in GENEPOP) as 1- the sum of the squared allele frequencies at each locus. Private alleles, or those found in only one cluster, were identified for each site, when the data was divided.

Actual number of alleles and effective number of alleles for the 13 loci were calculated for the 89 anoa individuals as one population (Table 4.3). The total number of alleles was 83, with a range at individual loci from three (TGLA126 and ILSTS5) to 15 alleles (INRA37). Loci that had a high number of alleles had an effective number of alleles far below the number of alleles observed. This suggests that a number of rare alleles were present.

Table 4.3. Actual number of alleles and effective number of alleles for the 13 loci and the 89 anoa individuals as one population.

Loci	No Alleles	Effective No. Alleles
BM1818	4	1.644
CSRM60	5	3.372
ETH152	6	3.692
HAUT24	8	4.954
HAUT27	13	5.487
HEL13	11	4.672
ILSTS5	3	1.866
INRA35	4	1.792
INRA37	15	5.838
MM12	6	2.491
SPS115	6	3.516
TGLA126	3	1.472
TGLA227	7	2.276
Total:	83	43.071
Mean:	6.385	3.313

4.3.2 Variation in numbers of actual and effective allele within anoa clusters

Allelic data was then used to determine if there were any differences in genetic diversity within the clusters defined above. The total number of alleles varied between subpopulations (Table 4.4). For example, the number of alleles observed in the Central group (n=29) was greater than twice that found in the Buton cluster (n=28) (not significantly different, Wilkoxon's signed rank test). The subpopulations with smaller sample sizes, South-East (n=12) and North (n=8), had similar allelic diversities of 3.692 and 3.154, respectively (not significantly different, Wilkoxon's signed rank test). The private or distinct alleles, in other words those only present in individuals from one group, were identified for each group (Table 4.4). However, few of these were common enough within the particular populations to be useful for regional identification.

Allele frequency distributions for each of the four regions and the Atypical group are presented in Figure 4.4. Seven loci had unimodal allele distributions, two approximately unimodal and four had departed from this pattern. The latter supported the proposed structuring in the population. The distribution of the loci ETH152 and MM12 clearly shows a geographic cline from north to south (Figure 4.4 c & j). The inclusion of the Atypical individuals caused some loci to produce a non-unimodal pattern of allele sizes (ILSTS5 and INRA35) (Figure 4.4 g & h).

Table 4.4. Actual number of alleles (above) and effective number of alleles (below) for five clusters of anoa individuals (A is Actual Allele, *Ne* is Effective Alleles).

Site:	Buton	South-East	Central	North	Atypical	Distinct Alleles
N	28	12	29	8	12	
BM1818	4	4	4	2	1	0
	2.36	2.47	1.33	1.51	1.00	
CSRM60	2	3	5	4	2	0
	1.46	1.88	1.49	2.29	1.99	
ETH152	4	5	4	2	1	1
	1.35	4.23	2.99	2.00	1.00	
HAUT24	4	4	8	3	2	3
	1.29	3.47	4.04	1.85	2.00	
HAUT27	3	5	13	4	2	7
	2.27	1.71	6.76	3.65	2.00	
HEL13	2	6	8	4	1	4
	1.15	4.72	3.78	3.56	1.00	
ILSTS5	2	1	3	3	1	0
	1.11	1.00	1.62	2.03	1.00	
INRA35	2	4	3	1	2	1
	1.37	1.92	1.57	1.00	1.39	
INRA37	3	6	13	4	1	7
	2.57	3.77	8.71	2.52	1.00	
MM12	2	3	6	3	1	3
	1.04	1.41	3.48	1.47	1.00	
SPS115	4	2	3	5	2	2
	2.61	1.95	2.13	4.09	1.18	
TGLA126	2	2	2	2	1	1
	1.73	1.80	1.07	1.85	1.00	
TGLA227	3	3	6	4	1	3
	1.48	2.32	2.25	2.24	1.00	
Total A	37	48	78	41	18	
Mean A	2.846	3.692	6.000	3.154	1.385	
Total *Ne*	21.805	32.657	41.229	30.064	16.551	
Mean *Ne*	1.677	2.512	3.171	2.313	1.273	

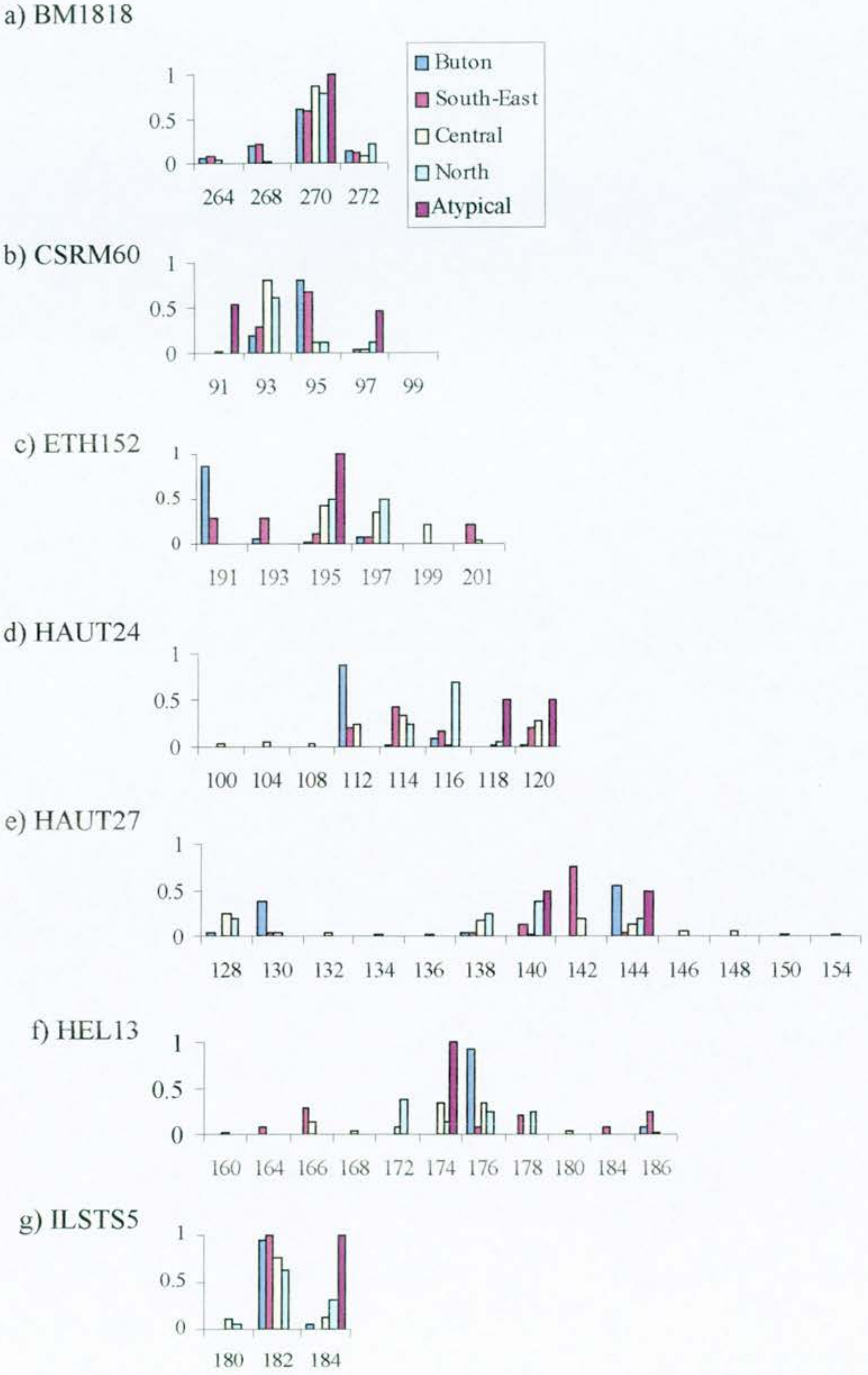
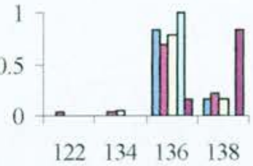
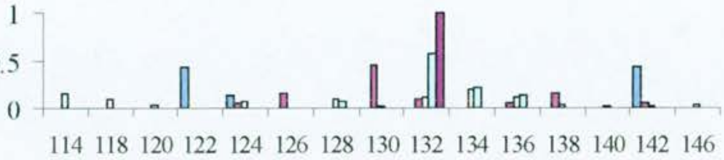


Figure 4.4. Allele length frequency histograms for five clusters of individuals with 13 loci for 89 individuals.
(continued overleaf)

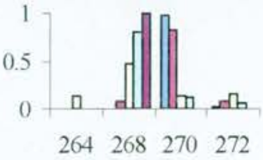
h) INRA35



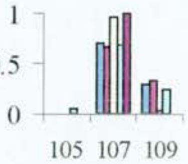
i) INRA37



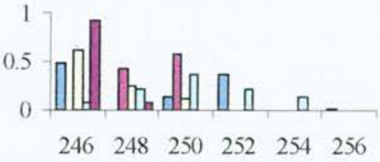
j) MM12



k) SPS115



l) TGLA126



m) TGLA227

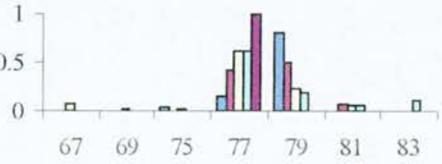


Figure 4.4. continued.

4.3.3 Comparison of allelic variation with Mongolian cattle breeds (*Bos taurus*) and Mongolian yak (*Bos grunniens*) and other outgroup species

To relate the allelic variation found in the anoa, comparison to other species is required. The 'scaling' of allelic variation of anoa was conducted using a comparison of cattle breeds as a measure of within species variation and cattle to Mongolian yak as a measure of between species variation. Previous authors have used this scaling method for different genetic data (Xu et al., 1996a; Xu et al., 1996b). For example, changes to the level of taxonomic differentiation between Bornean and Sumatran orang utans *Pongo pygmaeus pygmaeus* and *P.p. abelii* were proposed, using mtDNA sequence differentiation as a scale (Xu et al., 1996a). The results in this section have been summarised for comparison to the anoa data. Additionally, comparison was made of the allele frequency plots from six loci common to the cattle breeds and Mongolian yak. Mallet (1995) suggested that the presence of distinct allele peaks indicated distinct populations that were reproductively isolated. Therefore, this was compared in the anoa and between Mongolian cattle (*Bos taurus*) and Mongolian yak (*Bos grunniens*) (inter-species) to determine if under this genotypic cluster definition (1995), the anoa formed a single or multiple species.

Assessment of the level of informativeness of loci for Mongolian species was made using the number of alleles and the number of expected alleles for each locus. Values for allelic variance and standard deviation were higher than for the anoa. The effective number of alleles used to define informative and uninformative loci for the anoa (above and below 1.45), was compared to this data. All loci for Mongolian cattle and all but two (HEL5, INRA35) for the yak had a greater number of effective alleles than 1.45. Therefore, 20 loci have been analysed unless stated otherwise. Ideally it would have been appropriate to compare the results for the same loci; however this was not possible because only six of the 13 loci that were most variable for the anoa amplified in the Mongolian yak. This was understandable as the loci were originally designed for cattle and here have been tested on two other bovid species for which they were not designed. The number of individuals analysed was 10 Mongolian cattle and 19 Mongolian yaks, after removal of individuals with more than one missing value.

Results of allelic variation of these data were summarised in Table 4.5 along with the anoa data. This permitted the comparison between species and between anoa groups. Table 4.5 shows similar levels of mean A, *Ne* and He for anoa global values with the Mongolian cattle and yak for all three variation estimates listed, with no significant difference between populations (Wilkoxon's signed rank test). Population values of mean A and He were higher for the anoa (Central for mean A and Central and North for He) than Mongolian cattle and Mongolian yak. The Buton and Atypical anoa had lower values of mean A, *Ne* and He than Mongolian cattle and yak species. All loci screened in the Mongolian species showed more variation than anoa except ETH152 and INRA37 for which the anoa had higher variation (not shown).

Table 4.5. Comparison of allelic variation between anoa populations / global values and Mongolian cattle (*Bos taurus*) and Mongolian yak (*Bos grunniens*)

Anoa population / global value & other species	N	Number of loci	Mean A	Mean *Ne*	Mean He
Buton population	28	13	2.846	1.677	0.254
South-East population	12	13	3.692	2.512	0.377
Central population	29	13	6.000	3.171	0.410
North population	8	13	3.154	2.313	0.431
Atypical population	12	13	1.385	1.273	0.237
Anoa global value	89	13	6.385	3.313	0.342
Mongolian cattle	10	20	5.550	3.748	0.358
Mongolian yak	19	20	4.750	2.691	0.295

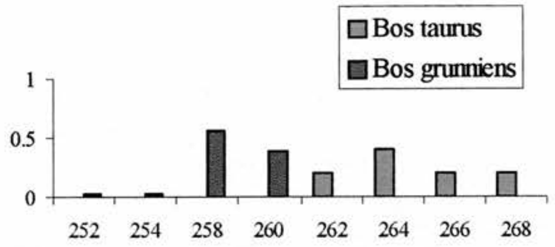
Mean A (mean number of alleles for all loci); Mean *Ne* (mean number of effective alleles for all loci); Mean He (mean heterozygosity for all loci).

Therefore, allelic diversity in anoa appeared to be comparable to that found within these other bovid species. However, intra-specific comparison of allele frequency plots between the Mongolian cattle and yak were made for comparison with those of the subpopulations of anoa. This would determine the level of differentiation between populations. Five loci that were highly variable in all three species were selected (BM1818, ILSTS5, ETH152, INRA35, INRA37). Distribution of alleles were investigated for Mongolian cattle and yak (Figure 4.5), which allowed comparion with the five anoa subpopulations (Figure 4.4).

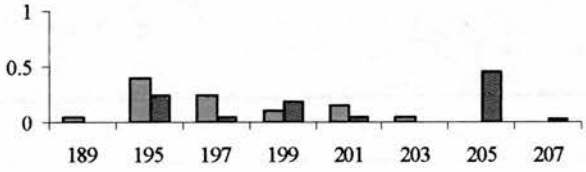
As described above no loci could be used uniquely to assign individual anoa to a particular cluster. However, informative loci that could be used to discriminate between yak and cattle were HEL5 and INRA32. A very high proportion of the alleles from these two loci were private alleles to either species. BM1818 also appeared to have several alleles unique to one or other species, but on inclusion of further samples with greater proportions of missing values this loci was found to be less informative.

Distribution of allele sizes between yak and cattle differed from the distribution between the anoa subpopulations. For yak and cattle there were a greater number of distinct alleles (23), compared to ten found in the anoa subpopulations for the five loci. There was less overlap in allele size between the yak and cattle than the anoa populations, suggesting that the anoa populations were less differentiated than between these two distinct species. Differentiation of locus ILSTS5 between Mongolian cattle and yak was less clear than for other loci. Therefore, for this single comparison with just two other species Mallet's (1995) genotypic cluster definition does provide evidence to reject the null hypothesis of two anoa species. However, a greater number of comparative species, including pairs that are more closely related should be included in future analysis to verify this.

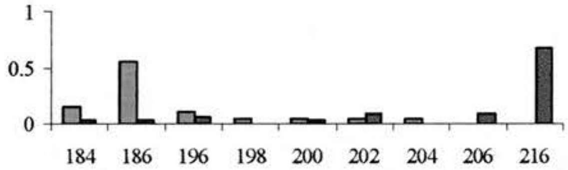
a) BM1818



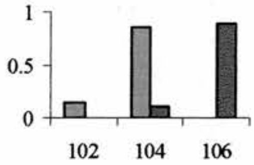
b) ETH152



c) INRA35



d) INRA37



e) ILSTS5

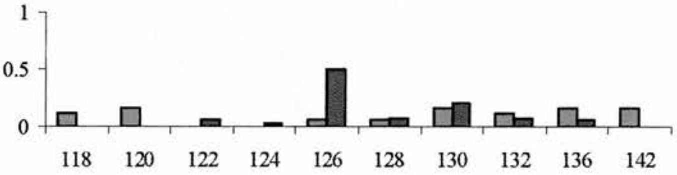


Figure 4.5. Allele length frequency histograms for 5 loci for 10 Mongolian cattle (*Bos taurus*) and 19 Mongolian yak (*Bos grunniens*).

4.3.4 Summary

In summary, anoa subpopulations on mainland Sulawesi had similar values for the allelic data and expected heterozygosity (Table 4.4). These clusters had more genetic variation, while the Buton and Atypical group had lower numbers of alleles and a narrow distribution of allele sizes. The highest number of private alleles, mean number of alleles and effective alleles was found in the Central group (Table 4.4). However, allele frequencies showed a pattern in the form of a cline for at least two loci from north to south (Figure 4.4). Different alleles were common to one group, but became less frequent as geographical distance from that group increased.

Comparison of the relative level of genetic differentiation between anoa groups compared with Mongolian cattle and yak showed that the number of alleles, effective number of alleles, and expected heterozygosity in the global anoa population was similar to that found in these other bovid species (Table 4.5). The allele size distribution was less differentiated between anoa groups than between the two other bovid species (Figure 4.5). This comparison of anoa data with a single pair of bovid species suggests the anoa may form a single species under the genotypic cluster definition (Mallet, 1995). However, this comparison is insufficient to conclusively clarify the anoa's taxonomic status. So, further comparisons of allelic variation with closely related bovid species are required. These results show a pattern of allelic diversity related to geographic regions in the anoa's range. Therefore, the analysis will continue by assessing the degree of clustering of individuals related to geographical distribution.

4.4 Population differentiation

4.4.1 Introduction

To investigate the extent of this population structure between the five clusters, genetic distance measures between the sub-divisions of the population were compared. Various models have been employed in attempts to understand the mutation of microsatellites, which was described in Section 1.4.2. Two possible mutation models are the infinite allele model IAM (Kimura et al., 1964) and the stepwise mutation model SMM (Kimura et al., 1978). These differ in a number of ways and could be said to be at opposite extremes of possible mutation scenarios. New mutations according to the IAM create a new allele, so homoplasy is not permitted. For the IAM, alleles that are 'identical by state' are 'identical by decent'. The IAM estimates the number of migrants when mutation is negligible. However, for microsatellite loci mutation often occurs and could have more impact than population migration in some cases.

The second model, the SMM differs in that new alleles created by mutation are defined according to the previous allele. The SMM assumes that microsatellite alleles evolve by the addition or deletion of one microsatellite repeat unit at each mutation. The probability of either insertion or deletion is the same. This proposed model of allelic evolution has two important implications. Firstly, alleles that are 'identical by state' under the SMM, may not be 'identical by decent', unlike the IAM. This is because one allele that has a repeat unit inserted followed by later deletion, will maintain the same length as another allele with no insertion or deletion. Secondly, under this model variation in allele length has greater significance. For example, between human families most mutations result in alleles varying by one or two repeat units (Di Rienzo et al., 1994). It has been suggested that larger variations may occur, but this is debated and not frequently observed (Goldstein, Ruiz-Linares and Cavalli-Sforza, 1995a). This means that alleles of a similar size are likely to be more closely related than alleles with greatly differing allele sizes in the majority of cases. In effect, the SMM has a memory, unlike the IAM, which does not take into account the previous allele size. The SMM conforms more closely to the mutation rate of microsatellites, which may be higher than the level of migration for microsatellites (Slatkin, 1995). Therefore, the stepwise mutation model (SMM) may follow the pattern of microsatellite allele evolution more closely than the IAM (Balloux and Lugon-Moulin, 2002). However, results from R_{ST} (the SMM statistic) often have higher levels of variance than F_{ST} (the IAM statistic) (Gaggiotti, Lange, Rassmann et al., 1999). This has the disadvantage that reliability of results may be lower when the SMM model is used. This can even be true for R_{ST} under the strictest SMM, where F_{ST} results are more reliable than R_{ST} (Gaggiotti et al., 1999), in certain cases. The general applicability of the SMM, underlying the R_{ST} measure, and its use with data from natural populations has also been questioned (Goodman, 1997). Therefore, only the F_{ST} will be calculated here.

Established measures of genetic distance were designed around these models. For example the F_{ST} measure (Wright, 1965) was based on the IAM. $F_{ST,}$ the fixation index, is one of the F statistics designed to assess population differentiation or subdivision. Global F_{ST} was used to identify the scenarios of population subdivision defined by microsatellite data. Population structure can be further examined with pairwise F_{ST} to identify clustering of populations.

High F_{ST} values occur when there is a high level of inbreeding and sub-populations exist. The F_{ST} is calculated as the probability that two alleles selected from the same sub-population are identical by descent. While F_{IS} is the probability that two alleles selected from the same individual are identical by descent (Wright, 1978). The F_{ST} measure was redefined by Nei (1977) who showed that F_{ST} could be assessed using observed and expected heterozygosities (Weir and Cockerham, 1984). F_{ST} is a function of the number of migrants when mutation is negligible. However, for microsatellites, mutation is often not negligible. Another difficulty with this method is when the mutation model cannot be assumed to be in agreement with the IAM.

F_{ST} values range from 0 (no differentiation between sub-populations) to 1 (different alleles in sub-populations). Experience in the interpretation of F_{ST} data has broadly suggested that values in the range of 0 to 0.05 show little genetic differentiation; between 0.05 and 0.15 moderate differentiation; 0.15 to 0.25 great genetic differentiation; greater than 0.25 very great genetic differentiation (Wright, 1978; Hartl and Clark, 1997). For the 0.05 value the general rule is that structuring between populations is weak. However, this is not a negligible level of differentiation; it may be quite important. This is because although the upper value is 1, even much differentiated populations may not reach this. There are numerous possible reasons for this. One explanation is that there is a finite rate of mutation possible so that not all alleles will show variation. Also in any study, the number of markers and the number of alleles at those markers is finite, so even totally unrelated populations will share a certain number of alleles. There is also the chance that convergent evolution of alleles could occur between individuals in distinct population. All these factors can lead to reduced F_{ST} expectations (Wright, 1978; Charlesworth, 1998; Hedrick, 1999).

F_{ST} is the most commonly used statistics for investigating population structure. This statistic is important for investigating populations of conservation significance, as these are often small. Such small isolated populations experience the highest level of the effect of genetic drift, because individuals have a significant influence on the population structure. This affects their evolutionary potential because of two factors: the potential for fixation of deleterious mutations; and the lack of individuals where mutation has the opportunity to create advantageous alleles. The F_{ST} value for small populations should increase over time more rapidly than large populations, as the structure of smaller populations is likely to diverge more rapidly. The knowledge of structuring gained from F_{ST} values of populations can add to the effectiveness of population management (Rossiter, Jones, Ransome et al., 2000).

Global F_{ST} analysis was conducted using Slatkins' (1995) method in (Schneier, Roessli and Excoffier, 2000). 1000 permutations were used for each analysis. The level of missing data included in the analysis was set at 0.1; this meant that 1 locus (BM1818) was omitted from the analysis. When this locus was included, using the mean values global F_{ST} setting, this did not vary the results.

Combining data from multiple loci has been conducted by a) averaging the numerator and denominator of fractions (Slatkin, 1995) and using this ratio, or b) averaging across loci. The latter is an arithmetic average. Both approaches would lead to the same result with no missing data. However, usually values were missing, which led to variation in degrees of freedom between loci caused the global variance to differ, so the F_{ST} would be different. Both of these methods were compared and where differences occurred due to missing values then the former method was used. To calculate pairwise F_{ST}, methods of Cockerham and Weir (1984) in ARLEQUIN were used. Other settings were as above.

4.4.2 Global and pairwise F_{ST} values

The global F_{ST} values for anoa showed that there was relatively high differentiation within the anoa population as a whole. Comparing the genetic differentiation between individuals within each population (F_{IS} value 0.2626) to that between groups (Buton, South-East, Central, North and Atypical) (F_{ST} value 0.3491) showed that there was greater differentiation between the groups. This supports the theory of some genetic structuring in the anoa population.

The pairwise F_{ST} values ranged from 0.0970 (Central and North) to 0.4090 (Buton and North) (Table 4.6), although values for the Atypical group were higher (0.653). The F_{ST} pairwise differentiation values were highly significantly differentiated for all pairwise combinations ($P< 0.001$). The high F_{ST} values showed considerable differentiation between groups at the extremities of the anoa's range. However, neighbouring groups had lower values, which although highly significantly different gave less support to high levels of structuring within the anoa population.

Table 4.6. Pairwise F_{ST} measure of differentiation of five groups for anoa, by an AMOVA with microsatellites.
(all values are highly significantly differentiated, $P<0.001$)

	Buton	South-East	Central	North
South-East	0.279			
Central	0.320	0.187		
North	0.409	0.212	0.097	
Atypical	0.653	0.566	0.363	0.473

4.4.3 Comparison to British cattle breeds (*Bos taurus*)

To quantify the differentiation shown within the anoa population, comparison with the known subdivision in another bovid species was investigated. This was most accurately achieved using data from the same loci. Data for 30 microsatellite loci, including those used here was available for eight British cattle breeds (*Bos taurus*) (Wiener et al., 2004). The anoa and cattle datasets were analysed independently. This was due to reported suggestions that differentiation between (sub) species could cause an underestimation of the actual F_{ST} value (Hedrick, 1999; Muwanika et al., 2003).

The global F_{ST} for the eight cattle breeds was 0.126, with an F_{IS} value of -0.027. The pairwise values varied from 0.196 between Jersey and Hereford to 0.050 and 0.077, respectively between Ayrshire paired with Friesian and Dexter breeds (Table 4.7). These were all highly significantly differentiated.

A comparison of the differentiation between anoa populations and between cattle breeds showed that the global F_{ST} for anoa was over twice that within British cattle, while F_{IS} was also much larger. The pairwise variation was also greater for the anoa than British cattle, with populations of both species being highly significantly differentiated (P<0.001). The greatest anoa value (0.409 Buton and North) was higher than the greatest value between two British cattle breeds (0.196 Hereford and Jersey).

Previous analyses of these samples by Wiener et al. (2004) included 30 microsatellite loci. The global F_{ST} 0.126 calculated from 13 loci was similar to that for the 30 loci of 0.13 (Wiener et al., 2004). The pairwise values were also similar. Therefore, although fewer loci were analysed than previously (Wiener et al., 2004), it appeared that the selected markers gave a good representation of the genetic variation in British cattle.

Table 4.7. Pairwise F_{ST} measure of eight British cattle breeds, with an AMOVA method with 13 microsatellite loci.
(all values are highly significantly differentiated, P<0.001)

	Aberdeen Angus	Ayrshire	Dexter	Friesian	Guernsey	Hereford	Highland
Ayrshire	0.106						
Dexter	0.095	0.077					
Friesian	0.115	0.050	0.088				
Guernsey	0.147	0.078	0.129	0.093			
Hereford	0.161	0.093	0.110	0.139	0.122		
Highland	0.176	0.121	0.117	0.134	0.133	0.118	
Jersey	0.168	0.154	0.114	0.136	0.160	0.196	0.183

In conclusion, the anoa population showed higher inter-group than intra-group differentiation. Following the general rule as stated by Balloux, Brünner, Lugon-Moulin et al. (2000), genetic differentiation of the *in situ* anoa population was considered to be high with seven of the ten anoa population pairs having F_{ST} values above 0.25. Neighbouring populations had lower values than distant populations.

The population differentiation of distant anoa populations was greater than that found in British cattle breeds.

4.5 Phylogenetic analysis

4.5.1 Introduction

Phylogenetic analysis of *in situ* anoa populations had two aims; the first was to examine the genetic diversity and confirm any structuring of the anoa population, the second was to define any intra-specific phylogeny of the anoa in relation to other species. The first aim was addressed by assessing the proportion of shared alleles (Dps) using individuals as operational taxonomic units (OTUs). The second aim used Nei's standard and $(\delta\mu)^2$ genetic distances, which allowed comparison with other known taxonomic levels.

Initially, the analysis investigated groupings of individual animals, which identified the clustering of animals from their geographic locales. The level of this genetic differentiation between neighbouring populations was low relative to that between distant populations. This showed population structure that could be explained by either isolation by distance or geographical separation, which produced the initial founder effects, followed by variation due to genetic drift. Comparing the data for the anoa with data from cattle and yak suggested that the genetic distances observed between anoa populations represented phylogenetic variation on a scale ranging from that between breeds to just below that of this pair of bovid species.

4.5.2 Data preparation

In addition to anoa data, datasets of eight British cattle breeds and Mongolian yak and cattle and four outgroup individuals (African Buffalo *Syncerus caffer*, Asian Buffalo Swamp-type *Bubalus bubalis* and British cattle *Bos taurus* (two individuals)) were used. The cattle and yak individuals were used as a 'scaling' comparison of genetic distance. The four other individuals were selected to reflect the variation between three subgenus of the bovid tribe.

4.5.3 Phylogenies using individuals as operational taxonomic units (OTUs)

The high variability of microsatellite loci allows the calculation of allele sharing statistics between individuals. This approach uses individuals as operational taxonomic units for phylogenetic reconstructions (OTUs) and was first used by Bowcock et al. (1994). The statistic calculates the proportion of shared alleles between two individuals. Then pairwise comparisons between all individuals are calculated to produce a matrix between individuals. The possible range of resulting values is between 0 and 1 depending on the similarity of allele sizes between individuals. This method was able to group human individuals with 87.7% correct clustering to racial and geographic origins (Bowcock et al., 1994). More recently the statistic has been used in a number of studies in several species including: honey bee (Estoup, Garnery, Solignac et al., 1995); Zebu and Taurine cattle breeds, (MacHugh, 1996); jaguar subspecies, (Eizirik et al., 2001); Spanish donkey breeds, (Aranguren-Mendez, Jordana and Gomez, 2002). All these studies investigated the distances between individuals of the sub-specific level or below (populations / breeds).

For genetic differentiation between individuals, 1 minus the proportion of shared alleles statistic was used (Dps) (Bowcock et al., 1994). The program MICROSAT (Minch, Ruiz-Linares, Goldstein et al., 1995) was used for calculation of the genetic distance matrix. From the distance matrix produced, neighbour-joining (NJ) phylogenetic trees were constructed using the program NEIGHBOUR [in PHYLIP 3.5 (Felsenstein, 1993)]. This method is based on the principle of minimal evolution (Saitou and Nei, 1987). Bootstrapping of the data was repeated 100 times. For calculation of bootstrap values, repetitions were converted into a single tree with bootstrap values using CONSENSE [in PHYLIP 3.5]. Trees were displayed using the program TREEVIEW (Page, 1996).

The Neighbour-joining (NJ) tree, shown in Figure 4.6, suggests a clustering of individuals according to geographical sampling origin. The tree branch lengths were constructed to a scale of the proportion of shared allele units. The tree was rooted to the outgroup species *S. caffer.* This clustering broadly agrees with the four geographical locations identified in the HWE and STRUCTURE analyses. However, this structure was not supported by high regional bootstrap values, with 2% between the Central and North branch, and between Buton and South-East branch, and 39% between the Central and Atypical branch. The only bootstrap between more than two individuals that was greater than 90% was within the Atypical group (no other group of three individuals had a value >50%). These low bootstrap values between regions showed a lack of strong evidence for geographical structure or a relatively recent separation/isolation of the geographical populations.

The four outgroup samples acted as the tree root. All samples purporting to be from anoa individuals clustered together and did not cluster with outgroup species. It was possible that samples from other species may have been included in samples collected by assistants, but the tight clustering of anoa samples showed that this has not occurred. The individuals from the east peninsula clustered within the North branch, in agreement with the STRUCTURE analysis.

There was a good match between the geographical sites and the relative branching order. For example Buton Island is geographically closest to the South-East, and these animals formed a branch within the South-East branch. This type of relationship also occurred for North individuals within the Central branch.

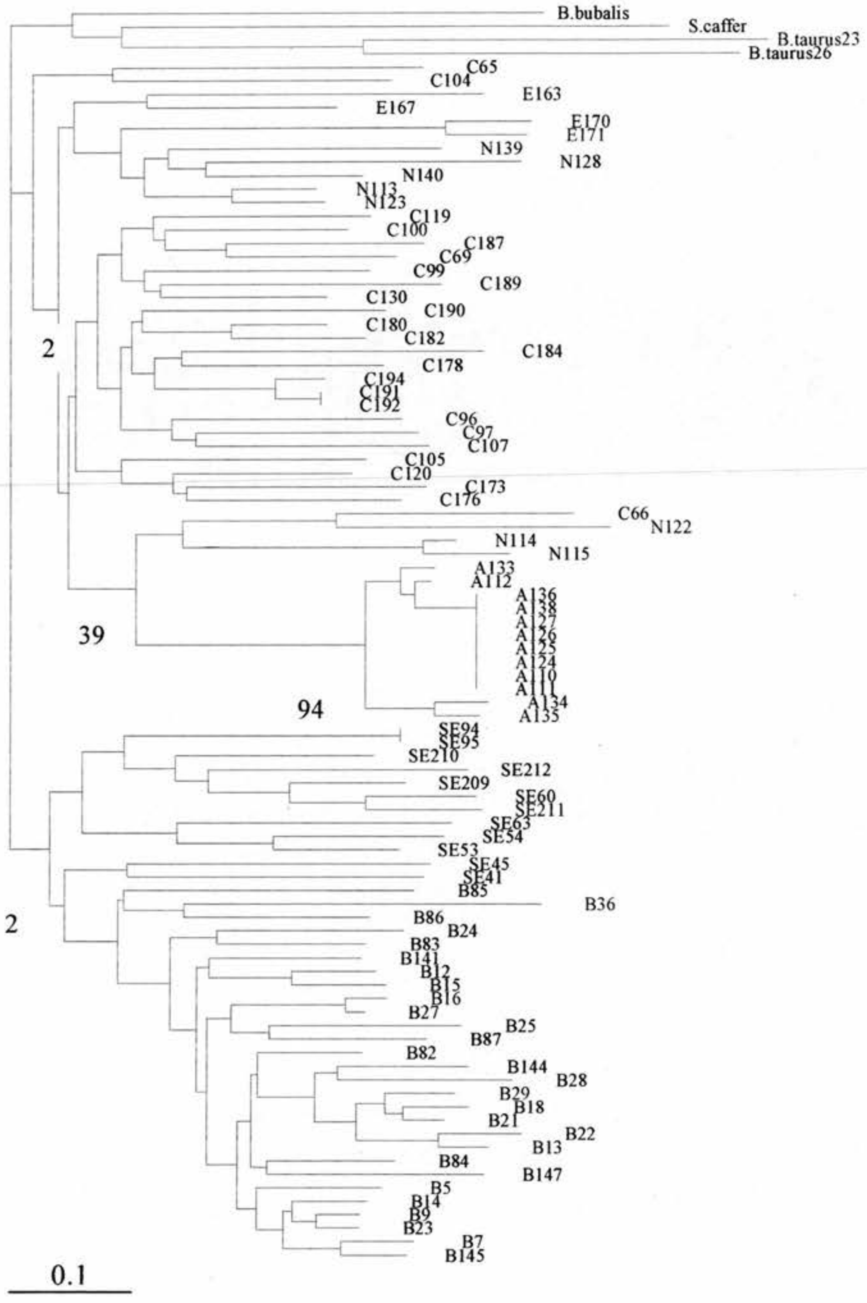


Figure 4.6. Rooted NJ tree produced from 'proportion of shared alleles' for 89 anoa samples and outgroup species *S. caffer*, *B. taurus* (two) and *B. bubalis*. Bootstrap values are shown. The sampling sites were identified as: B - Buton; SE - South-East; C – Central; N - North; E - East, A -Atypical.

4.5.4 Genetic distance between populations

The genetic distance between populations was used to quantify the relationship between groups of anoa individuals. This also allowed the assessment of correlation between genetic distance and geographic positioning of the populations. The second aim was to establish a scale of taxonomic differentiation by analysis of other bovid species.

Two genetic distance measures were used for comparison of distance between groups of anoa: Nei's standard genetic distance (D_S) (Nei, 1987) and $(\delta\mu)^2$ (Goldstein et al., 1995b). Nei's distance (an IAM model) were selected because this has been identified as appropriate for comparisons among closely related populations (Takezaki and Nei, 1996; Goldstein et al., 1997), such as the cattle breeds. The $(\delta\mu)^2$ measure (an SMM model) was said to be preferable, for more distantly related populations, such as between cattle, Mongolian yak and anoa data. An earlier version of this model, D_l was tested by simulating evolution with three-taxon trees separated by different numbers of generations. The phylogenetic reconstruction of trees for greater than 500 generations were more reliable than using an IAM model (Goldstein et al., 1995a). The $(\delta\mu)^2$ version of the model reduced the variance of results (Goldstein et al., 1995b) and was found to be more suitable for clarification between deeper ancestral nodes (Ritz et al., 2000). By comparison of results from two methods, a more detailed picture may be presented allowing better interpretation of the data.

Distance measures were calculated using the program MICROSAT (Minch et al., 1995). The construction of NJ trees was conducted with the programs in PHYLIP as described above, which were used for the analysis of individuals as OTUs. Bootstrapping was conducted 1000 times (Weir, 1996b).

4.5.4.1 Comparison of genetic distance within anoa

Two unrooted NJ trees constructed for Nei's and $(\delta\mu)^2$ distance measures (Table 4.8) show the relationships of the five groups of anoa (Figure 4.7 a and b). The tree constructed using Nei's genetic distance was consistent with the geographic locations of the individuals on Sulawesi. The geographically neighbouring subpopulations clustered together, such as Buton and South-East, as well as Central and North. The $(\delta\mu)^2$ genetic distance defined Buton as the most genetically distinct population (a distance of 3.081 and 5.707 for South-East and Central, respectively), with 64.2% bootstrap support. Bootstrap values, which showed consistent replicates supporting most branches, were >50% in both NJ trees for all nodes except one in Nei's NJ tree. Branch lengths were truncated between a number of groups, such as the South-East and North in both NJ trees. This is an indication of hybridisation having occurred between these clusters of individuals. This would suggest the lack of complete barriers to gene flow between these subpopulations. Admixture or hybridisation was also found during STRUCTURE analysis (Figure 4.3) in a single individual, which showed partial affinity to two populations.

Table 4.8. Pairwise Nei's standard (below diagonal) and $(\delta\mu)^2$ (above diagonal) genetic distance measure of population differentiation for five groups of anoa.

	Buton	South-East	Central	North	Atypical
Buton		3.081	5.707	3.647	3.768
South-East	0.505		0.829	0.017	1.655
Central	1.082	0.359		0.369	2.087
North	0.681	0.424	0.161		1.221
Atypical	1.408	1.199	0.461	0.495	

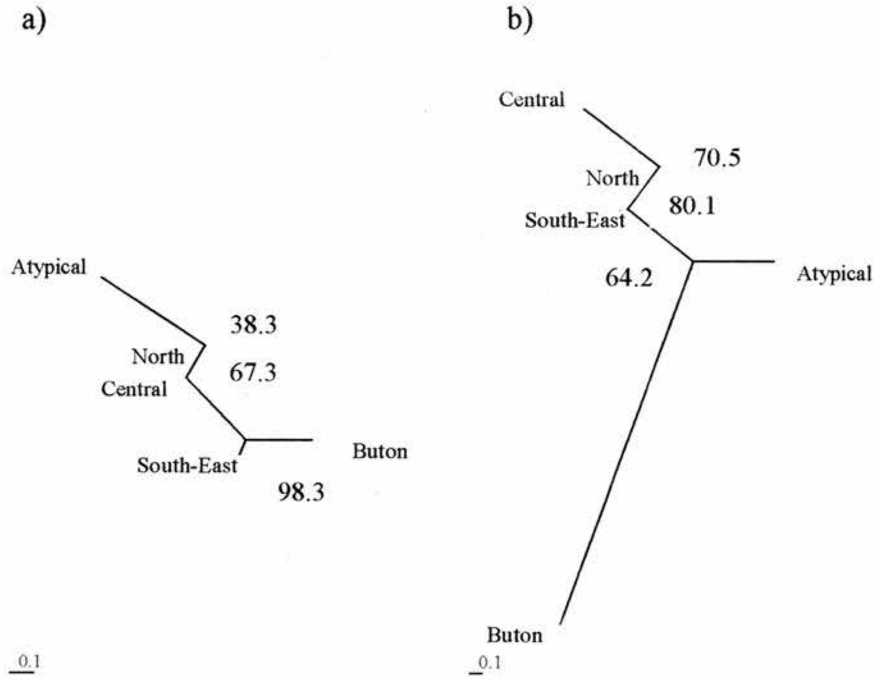


Figure 4.7. NJ tree produced from a) Nei's standard and b) $(\delta\mu)^2$ genetic distances between groups, for 89 anoa samples. (bootstrap values given as the percentage of consistent replicates).

4.5.4.2 Comparison to British and Mongolian cattle breeds (*Bos taurus*) and Mongolian yak (*Bos grunniens*)

This section aims to clarify if the genetic variation in the anoa population reported by distance measures was similar to the inter or intra-specific level of other bovid species. The same two distance measures (Nei's and $(\delta\mu)^2$) were used as above.

Data from the markers used to investigate the anoa was available from 8 British cattle breeds (*Bos taurus*) (n=45 to 64), Mongolian cattle (*Bos taurus*) (n=14) and Mongolian yak (*Bos grunniens*) (n=18). None of these animals had more than one missing data value. The Mongolian cattle and yak are classified as distinct species, even though they produce fertile offspring and it has been reported that introgression has occurred (Wiener et al., 2003). Complete data was available for the loci selected below in a total of 72 anoa individuals.

Of the 13 loci found to be variable in anoa, six were analysed across this dataset (BM1818, ILSTS5, ETH152, INRA35, INRA37 and TGLA126). The remaining seven loci were omitted due to missing data or low variance and standard deviation of alleles in the Mongolian cattle and yak individuals. The level of allelic variance and standard deviation was established by comparison with other loci of these two species and by comparison with loci from British cattle and the 13 anoa loci. The loci that remained were not the loci with the highest values in the other species, but had similar values to the loci informative in the anoa and most importantly were the loci that were the most informative for the anoa.

The results of both Nei's and $(\delta\mu)^2$ distance measures clustered populations / breeds from the same species together: the five anoa groups; nine cattle breeds; and the Mongolian yak as a single population. The NJ trees of the four distance measures are displayed in Figure 4.4. Branch lengths of Nei's distance tree between clusters were considerably larger than those within the three species (anoa, cattle and yak). In the $(\delta\mu)^2$ NJ tree the branch length between the South-East and Buton subpopulation of anoa was similar to that between two distinct species (Mongolian yak and Guernsey cattle). This was the greatest genetic distance between all of the anoa subpopulation. However, the other anoa subpopulations were located midway between them, suggesting intermediate populations existed. The meaning of this was unclear as the locations of these populations in the tree did not conform to previous structuring, related to geographic localities. These anomalous results may be explained as the $(\delta\mu)^2$ measure was developed for topological investigations and was not designed to cope with a microevolutionary scale (Ritz et al., 2000).

The branching between all populations for both NJ trees was not supported by high bootstrap values; all values were less than 50%. The bootstrap method tested the reliability of the results produced from a dataset with the creation of pseudo-replicate datasets by re-sampling (Weir, 1996b). These bootstrap results suggested that there was lower reliability for all structuring of the dataset.

The greatest and least genetic distances within the anoa subpopulations and within other study species populations have been summarised to allow comparison, using the scale of genetic distance as an indicator of distinctiveness. Table 4.9 shows these distances. The least genetic distance for both measures between anoa subpopulations is similar to the least distance between the cattle breeds (intra-species comparison). The greatest Nei's genetic distance is below that found between two species (inter-species comparison), the Mongolian yak and Mongolian cattle and is similar to that between two cattle breeds. However, the largest $(\delta\mu)^2$ distance between anoa groups (8.904) was greater than between cattle breeds (2.861), and only slightly below that between the Mongolian yak and Mongolian cattle (9.442). Therefore, the Nei's measure alone suggests the null hypothesis of two species can be rejected for the anoa. However, two reasons prevent this simple conclusion being made. Firstly, the similarity of the greatest $(\delta\mu)^2$ distance of a pair of anoa groups and a pair of bovid species does not support the conclusions of the former distance. Secondly, it would be advisable to include further data from other bovid species to act as multiple comparisons for both distance measures before defining the number of anoa species.

The relationships between groups of anoa, defined by the data from the six loci considered here, and 13 loci above (Figure 4.8 a and b) were similar. It was therefore reasonable to assume that genetic variation within anoa was well represented in this data set. This comparison was also made for the British cattle with 30 loci selected (Wiener et al., 2004). Representation of the branching was similar for Nei's distance, only the Friesian breed was found in an alternative branch. Therefore, the use of six loci was considered to give a reasonable representation of the genetic differentiation and structure of the British cattle samples.

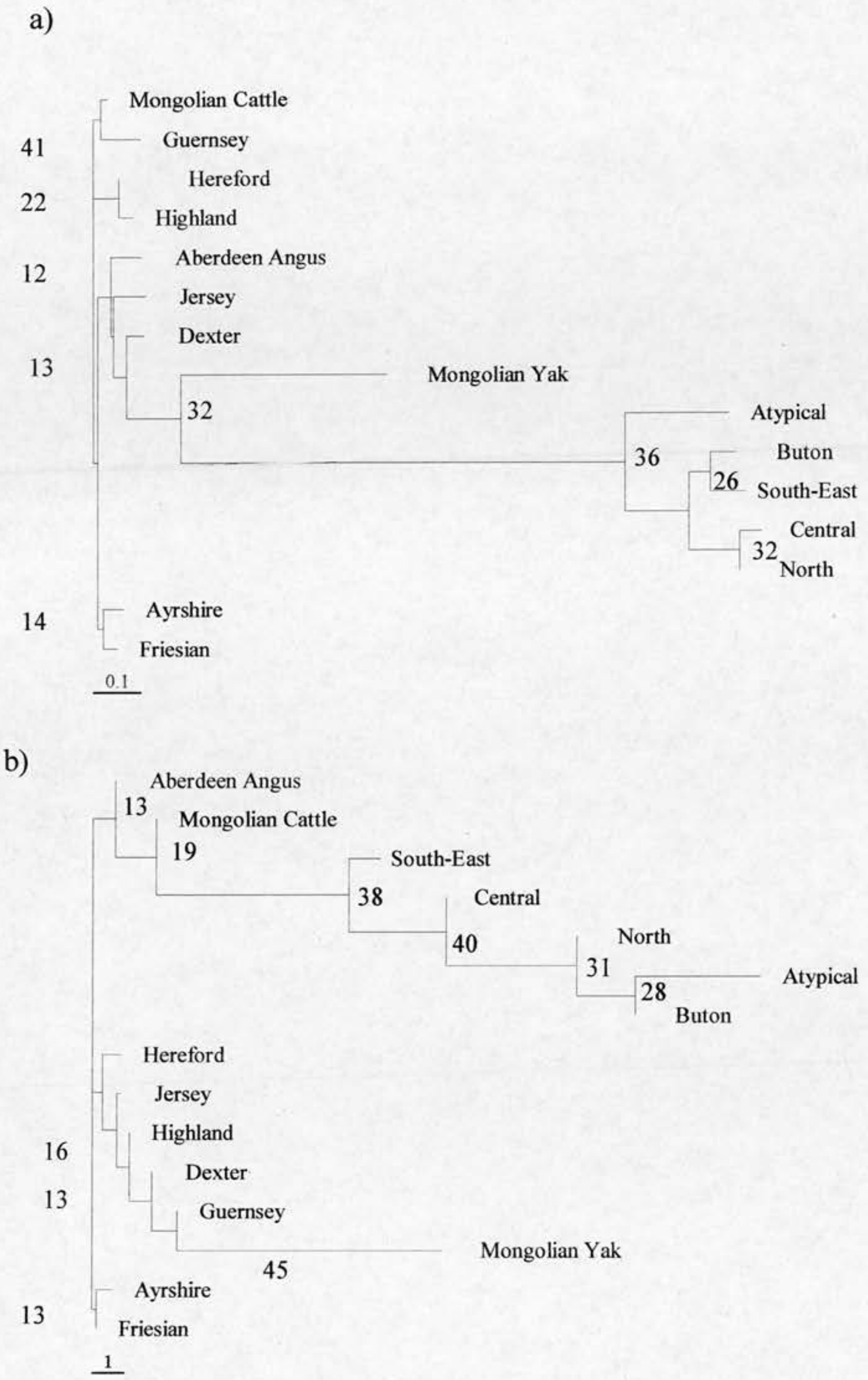


Figure 4.8. NJ tree produced from a) Nei's standard and b) $(\delta\mu)^2$ genetic distance between populations, for six loci.
The Mongolian yak is labelled, as are the anoa groups by their regional names (Buton, South-East, Central, North and Atypical); all other names relate to cattle breeds. Bootstrap values >10% are shown as the percentage of consistent replicates. The tree branch lengths were constructed to a scale of each distance measures' units. Note the difference in scale the two trees. All trees were unrooted.

Table 4.9. Nei's standard and $(\delta\mu)^2$ genetic distance measures of anoa, nine cattle breeds and Mongolian yak, for six loci.

Genetic distance measure	Greatest distance / inter-specific			Least distance / intra-specific	
	Anoa	Mongolian yak and cattle breed	British cattle	Anoa	British cattle
Nei's	0.302 (Buton – North)[1]	0.604	0.262 (Guernsey – Aberdeen Angus & Dexter)	0.040 (Central – North)	0.072 (Ayrshire – Friesian)
$(\delta\mu)^2$	8.904 (Buton – South-East)	9.442	2.861 (Aberdeen Angus – Dexter)	-0.059 (Central – North)	0.401 (Highland – Guernsey)

[1] All $(\delta\mu)^2$ values and two of the Nei's values for the Atypical group were greater than the greatest values stated here, reaching 0.637 and 14.524, respectively.

It is important to put these distances into a biological context when using them as a comparison with the anoa. The cattle samples are drawn from pedigreed individuals of distinct breeds, which have been selected for particular phenotypes for about 100 years with limited gene flow between breeds. The yak is a distinct species, which may have been affected by introgression of cattle genes (Wiener et al., 2003). The distances found between combinations of grouped individual anoa were comparable to genetic separation of cattle breeds and also slightly below the separation between cattle and yak. However, as stated above, these were only groups of individuals and from the analyses above have not been identified as discrete populations.

To conclude, phylogenetic analysis using the proportion of shared alleles showed the clustering of individuals was according to their geographic origin on Sulawesi and Buton Islands. The genetic distance measures used above showed that the level of differentiation between groupings varied greatly. Neighbouring groups such as Central and North had much lower differentiation than between more distant groups. The Buton group was especially differentiated from the other subpopulations. Even so these genetic distance measures, when compared with other bovid species, suggested anoa groups differentiation was less than or similar to that between this pair of bovid species. With the limited comparable data available the null hypothesis of two anoa species could not be rejected.

Although there were high bootstrap values between subpopulations of anoa, differentiation of groupings of individuals was not supported by high bootstrap values (Dps). These results suggest that there may only be partial geographical division possibly due to introgression, so weakening the structure of the anoa population. This would suggest the effect of isolation by distance, which is investigated in Chapter 6.

4.6 Assignment of *ex situ* anoa individuals

4.6.1 Introduction

The International Anoa Studbook states that a population of 125 anoa existed in zoos in 1998 (Nötzold, 1999). These animals were maintained as two populations ('depressicornis' / lowland and 'quarlesi' / mountain anoa), however, there is uncertainty about the genetic status of the zoo population, even following genetic studies (Schreiber et al., 1993; Schreiber et al., 1999). Previous work only investigated the conservation breeding population, so there was a lack of information on the relative genetic representation found in zoo individuals compared to the wild population. As stated previously there was no information on the origin of zoo animals or their breed lines from Sulawesi. The two breeding populations were reported to have morphological differences (Schreiber et al., 1993), for the majority but not all individuals. This variation was also reported from Sulawesi, but without clear distribution patterns of the two morphotypes (Schreiber et al., 1995; Burton et al., 2005). Therefore, breeding was hampered by these uncertainties.

In this section the *in situ* anoa microsatellite data was compared to that found in anoa individuals sampled from zoological collections. This allowed comparisons with the genetic variation reported previously (Schreiber et al., 1993; Schreiber et al., 1999) and the relative variation compared to the wild population.

The methods employed here were assignment analysis using program STRUCTURE, described in Section 4.2 and Principal Components Analysis (PCA) described below. Both methods attempted to show similarities of zoo individuals to the subpopulations in Sulawesi.

4.6.2 Multivariate statistical analysis

The advantage of using multivariate analyses is that they can represent the data where considerable genetic exchange between geographic neighbours has occurred, but where genetic distance methods are less suitable. Multivariate techniques are more flexible than traditional phylogenetic techniques and provide more realistic representation where population history is unknown. They are more suitable in situations where there is a relative poverty of data (Cavalli-Sforza, Menozzi and Piazza, 1994). In addition multivariate analyses do not assume the stepwise evolution of a population, which is the central assumption for building phylogenetic trees.

Two techniques were used, Principal Coordinates Analysis (PCoA) in Chapter 6 and Principal Components Analysis (PCA) here. The PCoA method assumes a fixed number of principal components (PCs). PCA is a special case of PCoA, which uses the ordination method: "Ordination is the collective term for multivariate techniques that arrange sites along axes on the basis of data on species composition" (ter Braak, 1987). PCoA was used to test for any genetic affinity of individuals from geographic locales (Section 6.2).

4.6.3 Principal Components Analysis (PCA)

The most commonly used multivariate technique in genetic studies is Principal Components Analysis (PCA). This method uses the linear transformation of allele frequencies. The aim of PCA using eigenvector analysis-based methods was to place species/individuals along gradients, whereas other multidimensional scaling methods relate distance to difference. The location of each individual in relation to the origin was described by that individual's eigen values. Most of the variation between a group of individuals would be accounted for by using two or three principal components, which were plotted on two or three dimensional plots (Manly, 1994).

PCA of distantly related populations has been shown to be effective for zebu and taurine cattle by MacHugh et al. (1997). Definition of relationships among closely related populations also have been successful using allele frequencies (Diez-Tascon, Littlejohn, Almeida et al., 2000; Canon, Alexandrino, Bessa et al., 2001; Maudet, Luikart and Taberlet, 2002). However, PCA seemed less suitable when using canonical data of individuals. The PCA results of MacHugh et al. (1998) did not support the clustering of certain cattle breeds as found in previous studies using different techniques. PCA plots of individuals using categorical data were less informative than dendograms. However, from the analyses above, which included some of the same cattle breeds and the anoa, it was apparent that the amount of genetic variation was greater in the anoa. This suggested that PCA might be a suitable technique to analyse anoa variation.

4.6.4 Data preparation

The 89 wild anoa individuals, analysed above were used in addition to the ten individuals from zoos. These zoo animals were: two individuals from the 'quarlesi' / mountain anoa breed line in Europe- Krefeld Zoo (Code K221 and K222); three individuals from 'depressicornis' / lowland breed lines in Europe- Chester and Marwell Zoos (Ch225, Ch226 and M235); and four individuals from Ragunan Zoo, Indonesia (R70, R71, R77 and R78) and one from Surabaya Zoo, Indonesia (S229). Another six individuals were not included in the analysis due to incomplete data for all 13 loci.

In order to compare this study with information from previous studies of the European zoo population (Schreiber et al., 1993; Schreiber et al., 1999), attempts were made to identify individual animals used in previous studies. Unfortunately, anoa individuals sampled here could not be identified directly as the animals sampled by Schreiber et al. (1993). This was because the International Studbook numbers quoted in Schreiber et al.'s publication were not in agreement with the 1998 Anoa Studbook (Nötzold, 1999). Also Schreiber et al. (1999) did not include data on studbook numbers. Instead, comparisons of microsatellite variation were made by assessment of individuals from the same breed lines as those studied previously (Schreiber et al., 1993; Schreiber et al., 1999).

The 'quarlesi' anoa population was reported by Schreiber et al. (1993) to consist of a breeding pair and their offspring, totalling seven individuals (5:2). As no new individuals had been added to this population, it was reasonable to assume that the individuals sampled here were related to those in the previous studies. Comparison of the three 'depressicornis' anoa analysed here with the previously sampled individuals was more complex. The Marwell Zoo animal was identified as originating from the Leipzig breed line. The Chester Zoo individual Ch226 was found to have originated from the Rotterdam line. The second Chester Zoo animal (Ch225) was a cross bred animal from the Rotterdam line, with one quarter originating from the Leipzig line. Both these breed lines were sampled by Schreiber et al. (1993) and Schreiber et al. (1999). These three animals represented a mix of two of the three 'depressicornis' breed lines in Europe. The four individuals from Ragunan Zoo were classified as 'depressicornis' anoa by the zoo, but were not included in the studies mentioned above.

4.6.5 Assignment of zoo individuals using program STRUCTURE

Using the program STRUCTURE, two different analyses were performed. Initially, analysis with all individuals was conducted without data on the origin of individuals, as recommended (Pritchard et al., 2000). This was to check that the defined clusters actually existed. Once the groupings had been confirmed analysis was conducted that included geographic labels for individuals collected from Sulawesi, having selected the 'use population information' setting in program STRUCTURE. Zoo individuals were not classified because the aim was to define them according to the four Sulawesi subpopulations and the Atypical group. The settings for program STRUCTURE were as in the analysis above in Section 4.2 (Pritchard et al., 2000). The program was run with *K* set at five populations and 200,000 repetitions for the burn-in period, followed by 200,000 MCMC repetitions.

The results of both analyses (without geographic data and assignment testing) agreed in their results. However, the results displayed in Figure 4.9 and Table 4.10 of proposed ancestry did not clearly assign the individuals into particular populations with high probability values. All the values were below the 0.90 threshold level (Manel et al., 2002). The individual with the highest assignment value of 0.713, from Ragunan Zoo, showed affinity to individuals from the North of Sulawesi. The highest values for other individuals are shown in bold in Table 4.10. The individuals from the two zoo populations defined as 'depressicornis' and 'quarlesi' did now cluster with individuals from the same subpopulation in Sulawesi. These assignment results should be interpreted cautiously as assignment of individuals showed low probability levels.

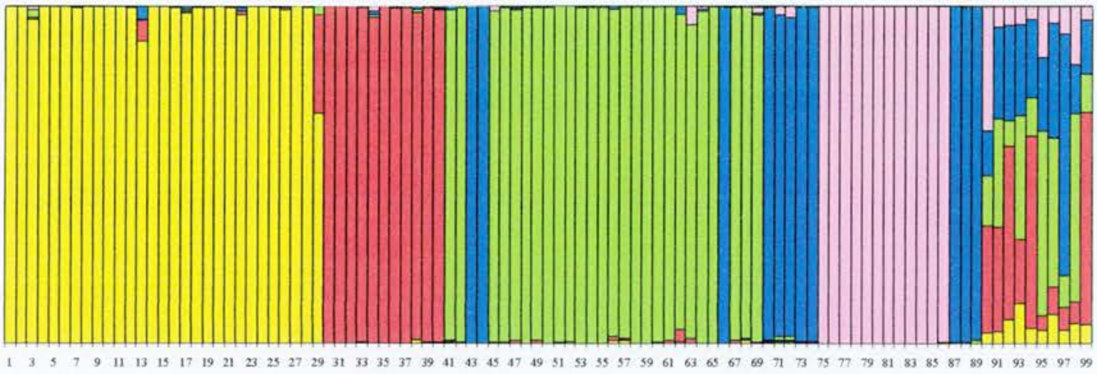


Figure 4.9. Assignment of ten zoo individuals (far right, numbers 90-99) to five genetic clusters (K=5) using program STRUCTURE.

Table 4.10. The inferred ancestry of six zoo anoa in five subpopulations with program STRUCTURE (with geographic data for Sulawesi individuals).

	Inferred ancestry of individuals in 5 subpopulations:				
Individual	Buton	South-East	Central	North	Atypical
Krefeld Zoo (K222) 'quarlesi' type	0.032	0.316	0.150	0.133	**0.369**
Krefeld Zoo (K221) 'quarlesi' type	0.037	0.309	**0.321**	0.272	0.062
Chester Zoo (Ch225) 'depressicornis' type	0.073	**0.512**	0.076	0.281	0.058
Marwell Zoo (M235) 'depressicornis' type	0.119	0.190	**0.364**	0.273	0.053
Chester Zoo (Ch226) 'depressicornis' type	0.046	**0.569**	0.112	0.234	0.039
Ragunan Zoo (R70)	0.038	0.047	**0.544**	0.219	0.153
Ragunan Zoo (R71)	0.086	0.081	**0.440**	0.34	0.052
Ragunan Zoo (R77)	0.039	0.069	0.096	**0.713**	0.083
Ragunan Zoo (R78)	0.059	0.066	**0.557**	0.144	0.174
Surabaya Zoo (229)	0.058	**0.626**	0.113	0.161	0.042

4.6.6 Identification of zoo individuals using PCA

Another approach to assess the genetic similarity of zoo animals to the individuals from the Sulawesi subpopulations was Principal Components Analysis (PCA). PCA analysis was used because no prior information on origin was required. This allowed visual presentation of the data to allow assessment of the distribution of zoo animals, which was not possible using the STRUCTURE analysis.

The analyses were conducted using GenStat Release 7.1 (Lawes-Agricultural-Trust, 2003). The need for the standardisation of data prior to analysis was assessed. When using the same units, as here for canonical data, there was no need to standardize units, so the variance-covariance setting was suggested. However, as there was large variation in frequency of certain alleles for some loci, normalizing of all variables was thought to be useful. This was achieved by selecting the correlation matrix setting (Everitt and Dunn, 1991). Krzanowski (1988) stated that standardization was sensible when there was a high difference in variances. As the data was in canonical form with values from 0 to 2, variance was low (0 to 0.827). The need for standardisation was assessed by analyzing data using both variance-covariance and correlation matrix settings, followed by the comparison of results. There was little difference in the percentage of variation described by each PC approach and the resulting plots showed a similar distribution of individuals. Loci with high variance did not appear to have affected the PCA result greatly. In this study there was no requirement to define patterns according to one PC; instead the aim was to gain insight into the structure of the variation. This meant that less significance might be paid to unequal representation of certain alleles. Therefore, the variance-covariance setting was employed below.

Results are shown in Figure 4.10 for the first two Principal Components (PCs). The first PC accounted for 45.15% of the variation. The second and third PCs represented 20.29% and 8.00% of the variation respectively, making a total of 73.44% variation accounted for by the first three PCs.

Before looking at the affinity of the zoo animals it should be noted in Figure 4.10 that the *in situ* individuals cluster according to their geographic origins with some introgression. The Buton and South-East individuals occupy the lower half of the plot, while the other subpopulations are found largely in the upper area (open symbols in Figure 4.10). The zoo individuals are broadly representative of the genetic variation found in the South-East, Central and North subpopulations. The ten zoo individuals were coded according to 'depressicornis' or 'quarlesi' morphotypes or from Indonesian Zoos in Figure 4.10 (closed diamonds). It should be noted that the scale on the *x* axis is reversed and there are differences in the scales of the *x* and *y* axes to give clearer representation of the data.

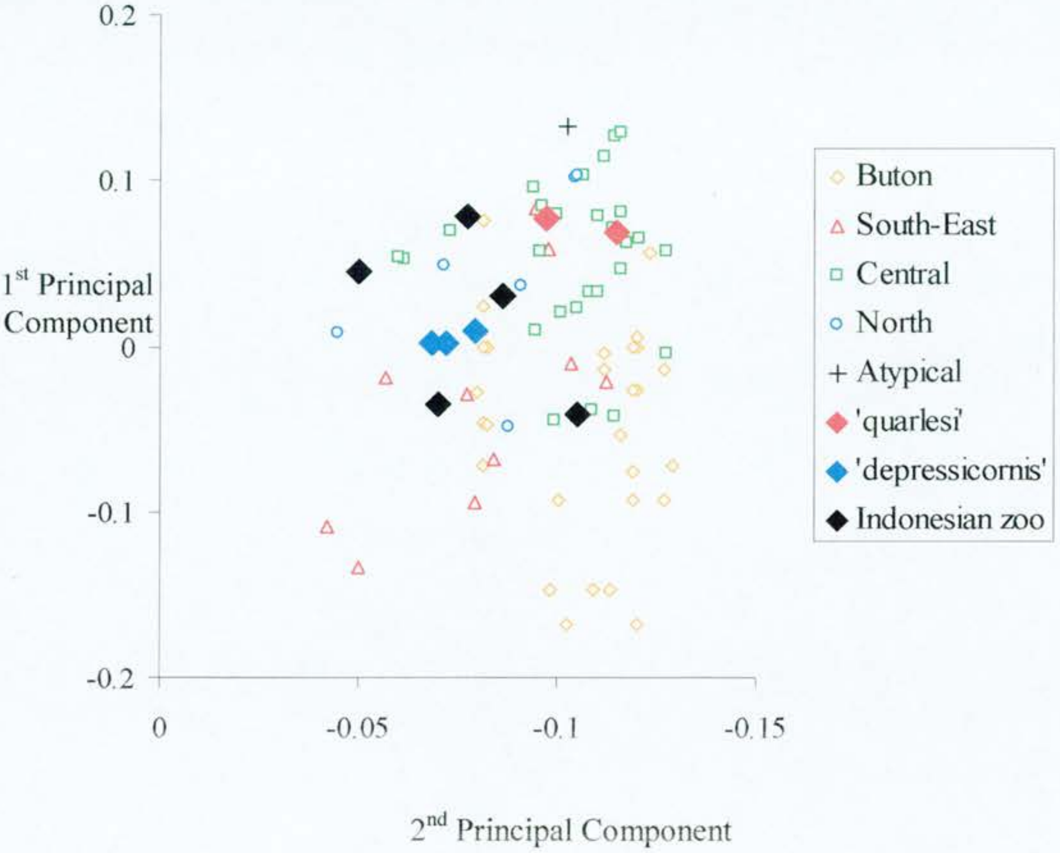


Figure 4.10. PCA plot of 89 wild anoa individuals (four geographic groups and Atypical group) and ten zoo anoa ('depressicornis', 'quarlesi' and Indonesian Zoo).

Zoo individuals of the 'depressicornis' and 'quarlesi' morphotype showed greater affinity to individuals from their morphotype, rather than the other morphotype. In the case of the 'quarlesi' individuals this is likely to be due to the fact that they are known to be closely related. The 'depressicornis' individuals showed greater affinity to wild individuals in the lower half of the PCA plot, while the 'quarlesi' individuals had greater affinity to wild individuals from the Central subpopulation. The animals sampled from Indonesian zoos showed a greater representation of the genetic variation in the wild population, than the European animals. This shows that the Indonesian zoo individuals are important for maintaining the greatest possible genetic diversity in the conservation breeding population. None of the zoo individuals showed affinity to individuals from the Buton subpopulation, which is the most genetically distinct region. This is despite both morphotypes being sampled, suggesting that further genetic variation found in the Buton individuals was absent from previous genetic studies (Schreiber et al., 1993; Schreiber et al., 1999).

4.6.7 Summary

The differentiation between the 'depressicornis' and 'quarlesi' morphotypes found in previous studies (Schreiber et al., 1993; Schreiber et al., 1999) has also been observed in the results of PCA here. However, it was not possibly to assign individuals with a high probability to their population of origin using program STRUCTURE. From the PCA results it could be cautiously suggested that the individuals from the different morphotypes may have originated from different subpopulations in the wild, but this could not be confirmed due to the level of introgression between subpopulations.

From the individuals sampled, the zoo population was shown to be partially representative of the genetic variation recorded from *in situ* anoa. Individuals present in the zoo population were thought to represent genetic variation from South-East, Central and North subpopulations.

4.7 Discussion

4.7.1 Introduction

Three aspects of the analysis of microsatellite loci will be discussed below, followed by assessment of the techniques used. Results will be reviewed to determine the existence of genetic structure in contrast to uniform gene flow throughout the population. This will determine where genetic variation exists and whether it follows the broad geographic regions or not. Next the areas of greatest genetic variation will be identified, as these are important for conservation. Any population structure will be defined by describing the degree of the division within the population. This section will also describe the level of variation in relation to that found in other species.

4.7.2 Genetic variation and the structure of the *in situ* anoa population

From Hardy-Weinberg analysis there was no evidence of a genetically homogeneous anoa population throughout Sulawesi and Buton Islands. When all individuals were pooled the results displayed significant deviation from HWE (all of the 13 loci) whereas, a sub-divided population was found to be closer to the Hardy-Weinberg equilibrium (only seven of 65 combinations deviated from HWE). This was supported by analysis using program STRUCTURE (Figure 4.3) that showed individuals clustered into four biogeographically distinct subpopulations. The lack of complete barriers to gene flow between the Sulawesi subpopulations was indicated by the truncated branches of the NJ trees (Figure 4.7), suggesting some hybridisation. An additional anomalous group (Atypical) was found within the northern peninsula that did not conform to the biogeographic pattern of Sulawesi Island, as did the other subpopulations.

Among the regional clusters of individuals, the greatest allelic variation (Mean A and Mean *Ne*) observed was in the Central population (Table 4.4). This can be defined as an area of importance for conservation efforts because this has greatest allelic variation and could act as a source population, connecting areas (Pulliam, 1988). The two sites to the north and south (South-East and North) have slightly lower allelic values, possibly due to their smaller sample sizes. The Buton cluster had the lowest genetic diversity values, probably because of its isolated nature due to the sea barrier.

Though a subdivided population was shown to be closer to HWE, there was limited evidence of any major genetic structure within the population. The analysis is comparable to results of Eizirik et al. (2001) and Culver, Jason, Pecon-Slattery et al., who found that for both puma (*Puma concolor*) and jaguar (*Panthera onca*) subspecies showed low bootstrap support for phylogenetic trees produced from Dps distances.

The degree of population structure is discussed and compared to other studies of species at varying levels of differentiation. The global F_{ST} value for anoa (0.349) was greater than those found in eight British cattle breeds (0.126). The anoa also had a higher global F_{ST} than the African buffalo (*Syncerus caffer*), 0.059 for 14 loci, from nine populations located continent wide (van Hooft et al., 2000). Comparison between three subspecies of common warthog (*Phacochoerus africanus*), gave results of F_{ST} (0.199), which were similar to the values between anoa subpopulations (Muwanika et al., 2003). The study analysed six loci for 143 individuals from 11 populations.

A study of Bighorn sheep (*Ovis canadensis canadensis*) of eight loci (Forbes et al., 1999) showed greater differentiation for Nei's distance (0.072 to 1.026) to the range observed between anoa subpopulations (0.040 to 0.302). The Bighorn sheep results showed variation at an inter-subspecific level (three Bighorn subspecies) of 1.380. At the intra- subspecies level (five Rocky mountain sheep populations) the Nei's distance was 0.171. Another subspecies comparison was possible between river type and swamp type Asian buffalo (*Bubalus bubalis*) using the $\delta\mu^2$ distance with a value of 33.63 (Ritz et al., 2000). The $\delta\mu^2$ distance value for the anoa was 8.904. In the same study of $\delta\mu^2$ distance inter-species distance varied from 14.00 for *Bos taurus* and *Bos indicus* to the largest distance of 96.88 *Bos frontalis* and *Bubalus bubalis* (river type) Therefore, the Nei's and $\delta\mu^2$ distance measures compared above suggested that the anoa has variation comparable to the intra-species level of variation. This was in agreement with the Nei's distance results from this study. However, the $\delta\mu^2$ values calculated in this study suggest the distance between at least one pair of anoa subpopulations was only slightly below that of the inter-species level of variation between Mongolian cattle and yak.

This genetic variation that is related to geographic locales of the anoa's range may be the result of recently isolated populations. Alternatively, this may be due to the effect of isolation by distance. This will be assessed in Chapter 6.

There was reasonable agreement in the clustering of individuals for most methods, which supported the grouping of individuals to allow the analysis of differentiation between subpopulations. It should be noted that an alternative to the Mongolian yak in the analysis to act as a scale of genetic distance with cattle species may have been preferable because of the possible introgression of these two species (Wiener et al., 2003).

4.7.3 Genetic variation of the *ex situ* anoa population

The aim of including anoa from the conservation breeding population was to assess the genetic representation of these individuals relative to the *in situ* population. The results showed that three breed lines from the European population were partially representative of variation in the wild population. The genetic variation found in the Indonesian zoo animals means that they are of importance in maintaining the greatest level of genetic diversity in the *ex situ* population.

4.7.4 Summary

It seems the species concept definitions employed in the assessment of microsatellite data differentiation, genotypic cluster definition and the scale of genetic distance as an indicator of distinctiveness, were unable to agree on the anoa's taxonomic status. The allelic differentiation between anoa subpopulations is less than that found between a pair of bovid species. Genetic distance measures produced contrasting results, one with comparable values to intra-species variation and one similar to inter-species variation. Defining the anoa's taxonomic status was therefore not feasible with the microsatellite data available. This was also confused by admixture between subpopulations suggested here and the lack of more detailed comparable data from other bovids. The genetic variation is related to the geographic locales of the individuals. This may be the result of a cline of genetic variation related to geographic distance or weak structure of incompletely isolated phylogenetic partitions. This analysis of genetic structure of anoa showed minimal evidence for major geographical partitions possibly due to recent admixture seen in the STRUCTURE analysis. The presence of old subdivision, or complete barriers to historical gene flow on Sulawesi Island may have been masked by this hybridisation, but an oceanic barrier clearly caused differentiation of the Buton subpopulation. Further analyses concerning the relationship of genetic and geographic distances are required to define this population structure more clearly.

Chapter 5

Results: Population Structure of Anoa Assessed using Mitochondrial DNA D-loop Sequences

5.1 Introduction

In this chapter the nucleotide variation of a section of anoa mtDNA D-loop sequence is assessed. This will allow interpretation of the matrilineal phylogenetic population structure, which may be distinct from the paternal structure defined in Chapter 4 with microsatellite alleles. Data from 35 wild anoa, one zoo animal and three outgroup individuals are presented below. Firstly, the sequence variation is presented in table format. This is followed by the construction of a neighbour joining tree and finally pairwise sequence mismatch plots are presented. By employing these methods an understanding of the evolutionary history and population structure of the anoa should be gained.

A 298 base sequence of DNA was amplified from the anoa samples to allow analysis. The results below identify three clades that broadly corresponded to three geographic regions of the anoa's range.

5.2 Variation in anoa mitochondrial DNA and identification of haplotypes

Preliminary assessment of the differences in single nucleotides in a DNA sequence can provide an indication of the genetic diversity. The nucleotide differences for 35 wild anoa D-loop sequences are displayed in Table 5.1. Individual anoa codes are in the first column with the letter representing their region of collection, ordered according to geographic origin, from south to north. (B - Buton, SE – South-East, C – Central, N – North and A – Atypical). There were 40 variable sites in a total of 298 nucleotides. A total of 13 haplotypes were observed. Ten of these were found only in individuals sampled from the Central and North subpopulations. Two haplotypes were found only in the South-East individuals. All 15 individuals sampled from Buton Island were monomorphic. Two individuals from the South-East subpopulation displayed the same sequence as the Buton haplotype and two individuals from the Atypical group were found to have the same haplotype as the majority of South-East individuals. Table 5.1 shows the difference in the level of genetic diversity between different regions of the anoa population. Individuals sampled from the Central region have a high level of sequence variation, compared to other regions, especially the monomorphic Buton individuals.

Table 5.1. Sequence variations observed in 35 anoa D-loop sequences. The first individual represents the full bases present and a dash represents the same base in following individuals or represented by a different base letter where a substitution occurred.

	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
	2	2	2	1	9	7	7	6	6	6	5	4	4	3	3	3	3	3	2	1	1	1	1	1	0	9	9	9	8	7	7	5	2	1	1	1	9	9	8	6
	6	3	0	5	7	2	0	8	3	2	8	7	2	8	7	6	5	3	9	9	7	6	1	0	1	9	7	3	1	8	7	4	2	9	7	5	1	0	9	4
B7	T	A	C	T	T	G	T	T	G	T	T	C	A	A	T	A	G	C	G	A	T	A	T	A	C	G	C	T	T	C	T	T	G	G	G	T	G	T	T	C
B8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-									
B144	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B83	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B84	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SE94	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SE95	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SE53	C	G	-	-	C	A	-	C	T	-	C	-	-	G	C	-	A	T	A	-	-	-	C	G	T	A	-	-	-	-	C	C	A	A	-	-	A	C	C	-
SE54	C	G	-	-	C	A	-	C	T	-	C	-	-	G	C	-	A	T	A	-	-	-	G	G	T	A	-	-	-	-	C	C	A	A	-	-	A	C	C	-
SE41	C	G	-	-	C	A	-	C	T	-	C	-	-	G	C	-	A	T	A	-	-	-	C	G	T	A	-	-	-	-	C	C	A	A	-	-	A	C	C	-
SE63	C	G	-	-	C	A	-	C	T	-	C	-	-	G	C	-	A	T	A	-	-	-	C	G	T	A	-	-	-	-	C	C	A	A	-	-	A	C	C	-
C173	-	G	-	-	C	A	-	C	A	C	-	-	G	G	-	-	A	T	A	G	C	-	C	-	-	A	T	-	-	-	-	C	-	-	A	C	-	C	C	-
C176					C	A	-	C	A	C	-	-	G	G	-	-	A	T	A	G	C	-	C	-	-	A	T	-	-	-	-	C	-	-	A	C	-	C	C	-
C99	C	G	T	-	-	A	-	C	A	C	-	-	-	G	C	G	A	T	A	-	C	-	C	G	-	-	-	C	-	-	C	C	-	-	A	-	-	C	C	T
C121	C	G	-	-	C	A	C	C	A	C	-	-	-	G	-	-	A	T	A	-	-	-	C	G	-	-	-	C	-	-	-	C	-	-	A	-	-	C	C	T
C189	-	G	-	C	C	A	-	C	A	C	-	-	G	G	-	-	A	T	A	G	C	-	C	-	-	A	T	-	-	-	-	C	-	-	A	C	-	-	C	C
C191	C	G	T	-	-	A	-	C	A	C	-	-	-	G	C	G	A	T	A	-	C	-	C	G	-	-	-	C	C	-	C	C	-	-	A	-	-	C	C	T
C96	C	G	-	-	-	-	C	C	A	C	-	-	-	G	C	-	A	T	A	-	C	G	C	G	-	-	-	C	-	-	C	C	-	-	A	-	-	-	C	T
C97	C	G	-	-	C	A	C	C	A	C	-	-	-	G	-	-	A	T	A	-	-	-	C	G	-	-	-	C	-	-	C	C	-	-	A	-	-	C	C	T
C180	C	G	T	C	-	A	-	C	A	C	-	-	-	G	C	G	A	T	A	-	C	-	C	G	-	-	-	C	-	-	C	C	-	-	A	-	-	C	C	T
C120	C	G	-	-	C	A	C	C	A	C	-	-	-	G	-	-	A	T	A	-	-	-	C	G	-	-	-	C	-	-	C	C	-	-	A	-	-	C	C	T
N115						-	C	C	A	C	-	A	-	G	C	-	A	T	A	-	C	G	C	G	-	-	-	C	-	-	C	C								
N116	C	G	-	-	-	-	C	C	A	C	-	-	-	G	C	-	A	T	A	-	C	G	C	G	-	-	-	C	-	-	C	C	-	-	A	-	-	-	C	T
A110	C	G	-	-	C	A	-	C	T	-	C	-	-	G	C	-	A	T	A	-	-	-	C	G	T	A	-	-	-	-	C	C	A	A	-	-	A	C	C	-
A111	C	G	-	-	C	A	-	C	T	-	C	-	-	G	C	-	A	T	A	-	-	-	C	G	T	A	-	-	-	-	C	C	A	A	-	-	A	C	C	-

5.3 Phylogenetic analysis

5.3.1 Introduction

Neighbouring joining trees are a commonly used phylogenetic analysis tool for interpreting variation in mtDNA sequence data (Avise, 1994). The aim of constructing phylogenetic trees from anoa mtDNA D-loop sequences is to identify the structure of a population. This method has been used widely in phylogeographic analysis of many species.

5.3.2 Data preparation

A suitable length of the D-loop region was sequenced for 35 wild anoa, one zoo anoa (Z236 Marwell Zoo, UK) and three outgroup individuals (O260 European cattle *Bos taurus*, O237 Mongolian cattle *Bos taurus*, O238 Mongolian yak *Bos grunniens*). The outgroup individuals were included in the analysis to act as a positive control and to allow the tree to be rooted to an individual from a distinct species. Although the more closely related Asian water buffalo would have been a more suitable outgroup, samples of this species that were included in the study, produced incomplete sequences and time did not permit re-analysis of these. Available sequences of this species should be used in future analyses. The available sequence for each of these 35 anoa individuals was used in the tree construction. Hence there is a small difference between the monomorphic Buton individuals because of the occasional missing nucleotide for some individuals.

5.3.3 Neighbour joining tree

The construction of neighbour joining trees has the advantage of being a more powerful and informative tool than genetic distance measures. This method of data presentation allows researchers to determine the cause of restricted gene flow, historical fragmentation, range expansion and genetic differentiation (Frankham et al., 2002). Phylogenetic trees should only be constructed from sequence data if an ancestral population is monomorphic, the markers are not affected by selection pressure, and that there are equal rates of evolution in different lineages. There is no evidence to suggest that these criteria have not been met by this dataset.

A neighbour joining tree was drawn using PAUP Version 4.0bi (Swofford, 2003). Individuals were classed as operational taxonomic units, rather than a tree constructed from haplotypes because of an incomplete dataset. The tree was rooted with a European cattle sample (O260). Assistance in tree construction was given by Alex Hayward (IEB, University of Edinburgh).

The neighbour joining tree grouped all anoa into a monophyletic cluster, which included three deeply divergent clades (Figure 5.1). The first clade included the haplotype sampled from individuals on Buton Island and two individuals from the nearby South-East region. The second clade included samples collected from the South-East and two individuals from the Atypical microsatellite group. The third clade consisted of individuals from the Central and North regions of Sulawesi. The root of the tree was intermediate between the first and second clade. The first clade (Buton) was the most differentiated from the other two, suggesting either a longer period of isolation or a greater degree of isolation.

The clades are largely correlated with the geographic origin of sampled individuals. The two exceptions are firstly, two individuals in the South-East region display the Buton haplotype and secondly, two individuals in the North display the haplotype of individuals in the South-East. The explanation for these exceptions is that these four individuals may have been transported prior to sample collection. This is known to occur as part of the meat trade (Clayton et al., 1999) and so the Buton individuals were classed as a monophyletic group.

The single haplotype present in the Buton Island individuals (first clade) compared to the other two clades suggests a bottleneck event may have occurred in this population's history. During colonisation of this island the bottleneck may have been casused by the oceanic barrier.

The single zoo anoa from Marwell Zoo (Z236) included in the neighbour joining tree clustered in the clade with South-East and Atypical individuals. Therefore, it could not be predicted where this animal's ancestors could have originated from, because it may have been from either of these two regions. All the outgroup individuals clustered together, with the European cattle sample used to root the tree.

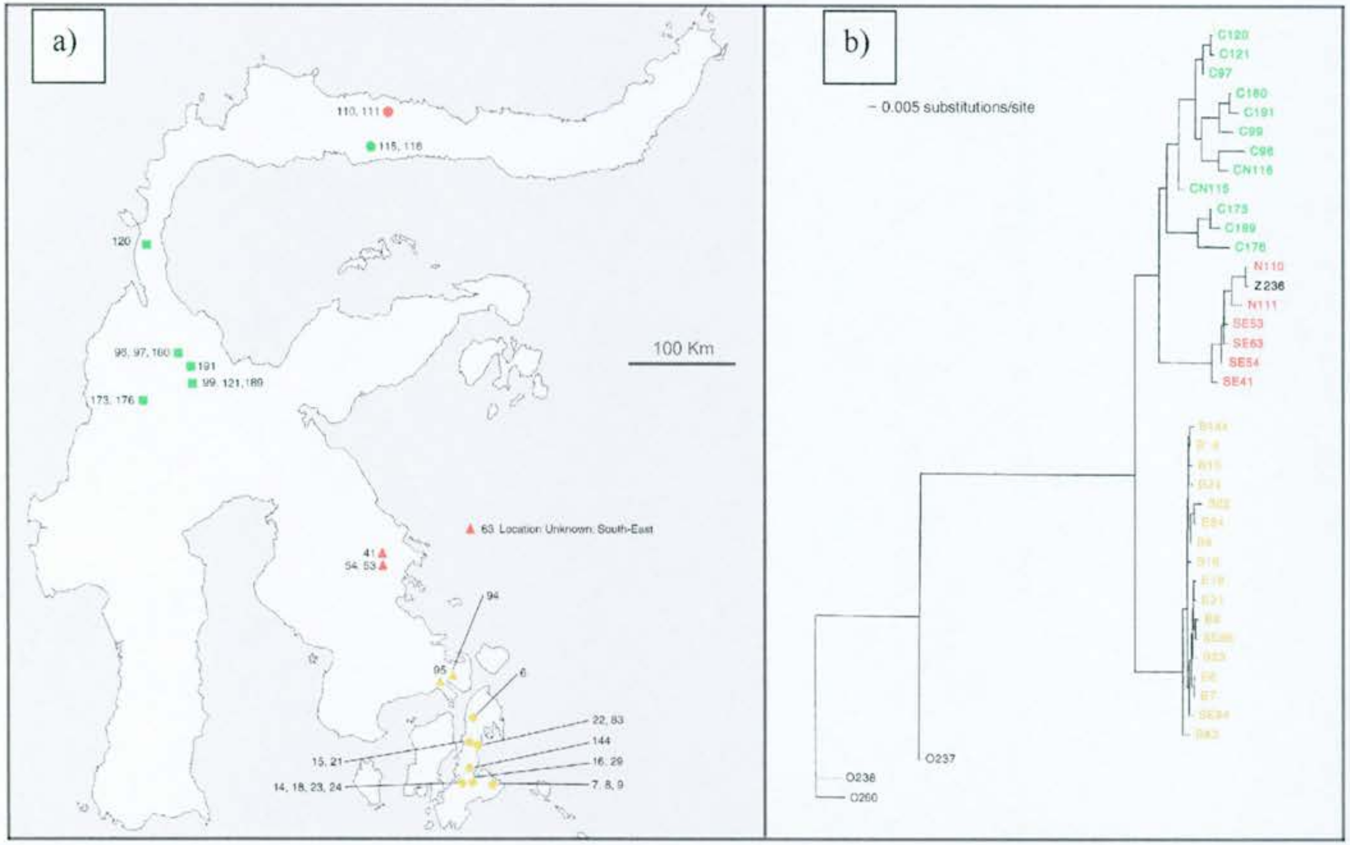


Figure 5.1. mtDNA D-loop sequence phylogeographic variation throughout Sulawesi.
a) Location of 35 mtDNA samples from Sulawesi and b) Neighbour joining tree defining three clades from a 298bp sequence. Clades are colour coded on the map and tree. Sample numbers on the map correspond with numbers on the tree, with prefixed letters to represent area of origin. The tree was constructed from number of substitutions per site and rooted to a European cattle sample (O260). The tree also includes a single European zoo individual (Z236 Marwell Zoo, UK) and two other outgroup samples (O237 Mongolian cattle, O238 Mongolian yak).

5.4 Sequence mismatch frequency plots

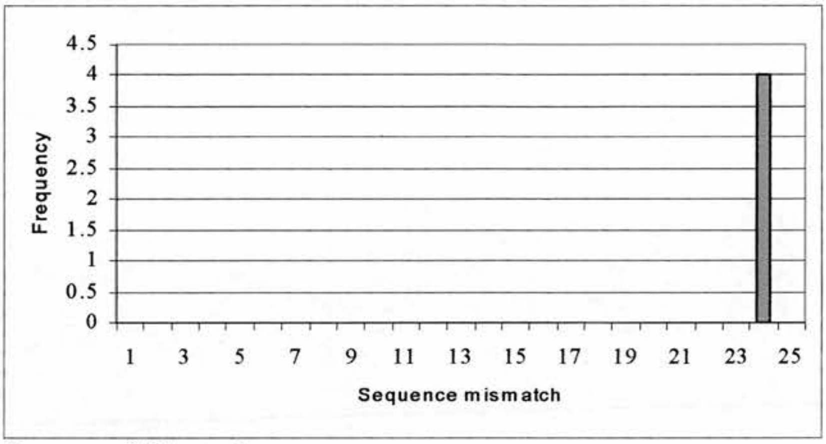
Pairwise sequence mismatch frequency plots are thought to be useful for understanding the evolutionary genetics of a population (Rogers and Harpending, 1992). Bradley et al. (1996) found a smooth peaked mismatch curve was indicative of a recent expansion event within both European and African cattle breeds, which occurred after separation of the two continental populations. These plots also display the level of sequence mismatch, represented by the location of the highest frequency on the *x* axis. This can provide information on the time since population expansion events.

The pairwise sequence mismatch method was applied to the anoa to ascertain their evolutionary genetic history. Pairwise sequence mismatches between the three clades are displayed as frequency plots in Figure 5.2a, b and c, and within each clade in Figure 5.2d, e and f. The three inter-clade plots represent the mismatch between pairs of individuals across the three clades. The sample size per clade is small, so this may not have allowed a smooth peak to be produced. So in this case it may not be possible to identify the existence of a population expansion. However, the most striking aspect of the inter-clade plots is the higher level of sequence mismatch between Buton and both the South-East and Central clades (Figure 5.2a and b), compared to between the two latter clades (Figure 5.2c). An explanation for this greater sequence differentiation of the Buton clade to the other two clades, as suggested above, is that it was due to a bottleneck effect and the time since common ancestry.

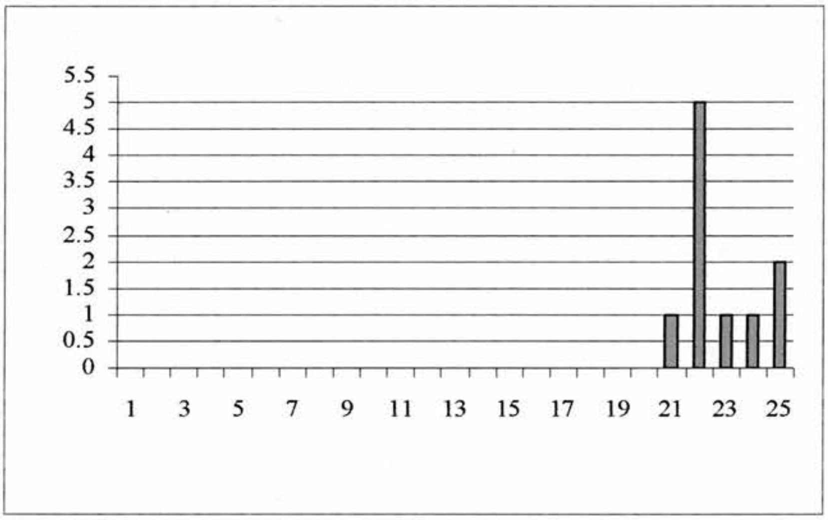
The sequence mismatch within clades showed little or no mismatch within the Buton and South-East clades (Figure 5.2d and e) and considerable mismatch in the Central clade (Figure 5.2f). This displays the high genetic diversity of individuals from the Central region. When the sequence mismatches between clades (Figure 5.2a, b and c) are compared with those found within clades (Figure 5.2d, e and f), the former has greater variation for all three pairs. This represents the greater division of these mtDNA clades in comparison to local variation within clades.

The nucleotide diversities between clades (Buton n=17, South-East n=6, and Central n=10) were calculated as a percentage of the 298 bases sequenced. Between the Buton and South-East clades the nucleotide diversity was 7.7% (23bp), between the Buton and Central clades it was 7.4% (22bp) and between the South-East and Central clades the difference was 5.4% (16bp).

a) Buton and South-East



b) Buton and Central



c) South-East and Central

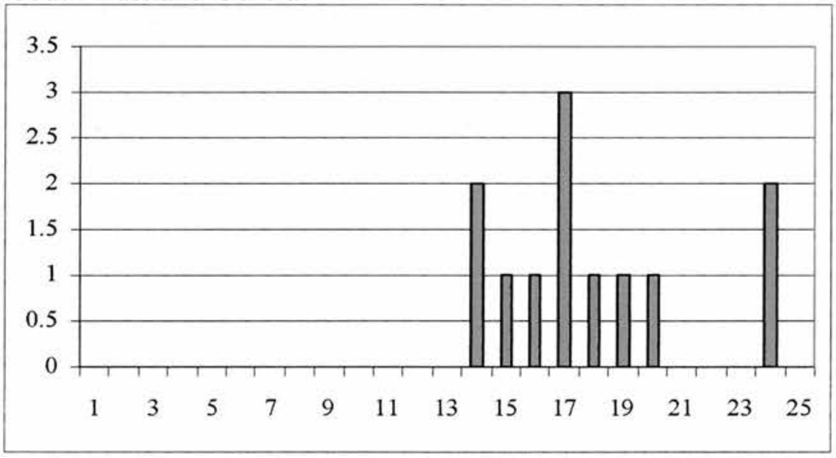
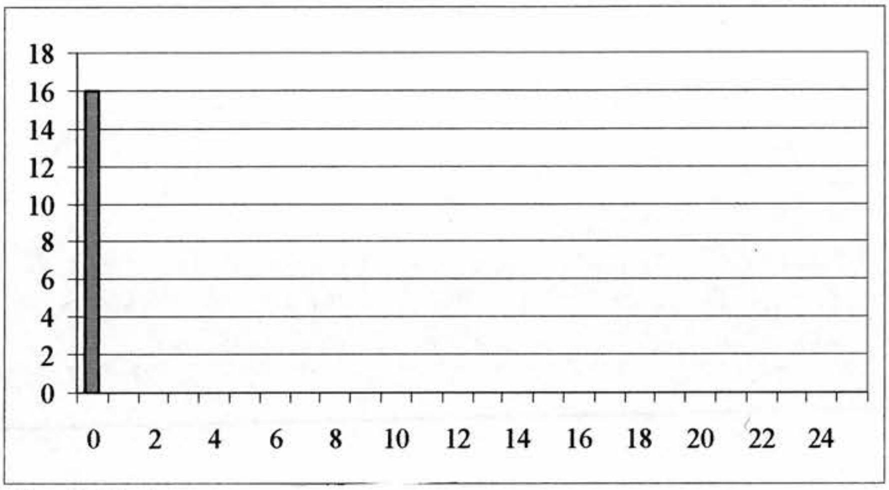
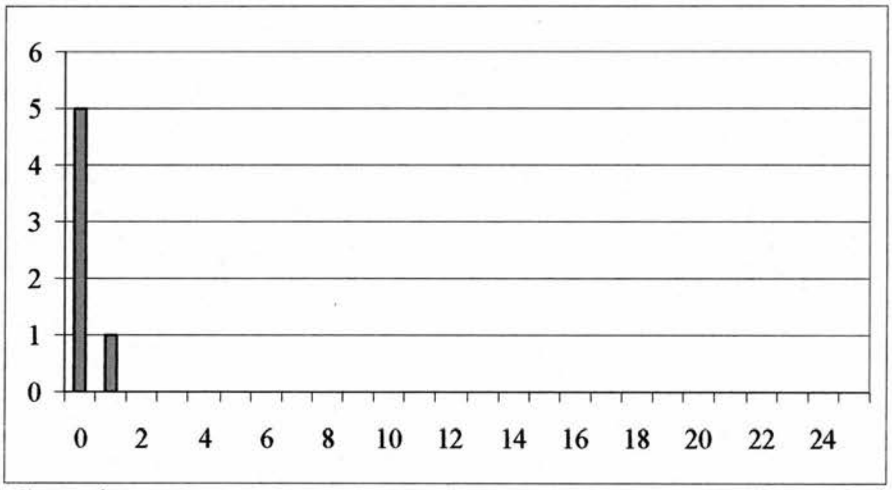


Figure 5.2. Frequency plots of pairwise sequence mismatches of D-loop sequence data (plots a, b and c) between three clades Buton, South-East, Central; and (plots d, e and f) within each clade.

d) Buton



e) South-East



f) Central

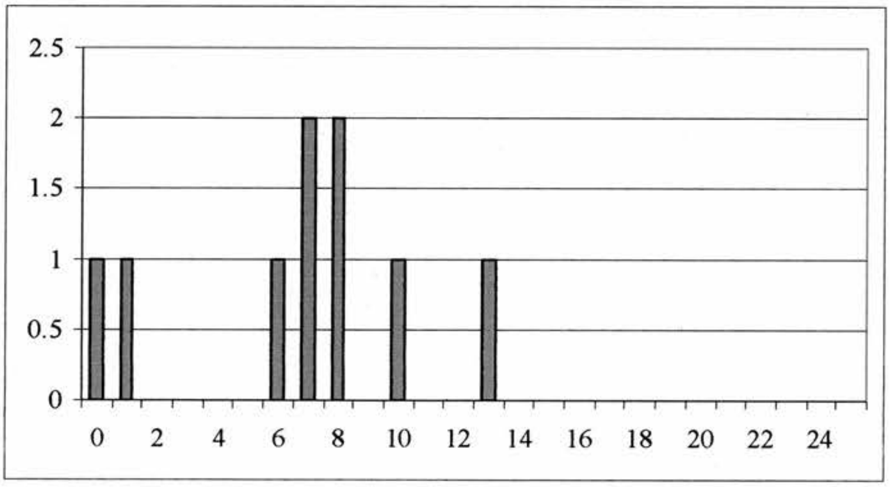


Figure 5.2. continued

5.5 Discussion

5.5.1 Matrilineal population structure

The anoa mtDNA D-loop sequences analysed here identified three clearly differentiated clades that largely represent individuals from Buton, South-East and Central/North regions. So this method successfully determined the matrilineal population structure of the anoa.

The ratio of the number of haplotypes to the sample size for the anoa is low (0.37) for D-loop sequence, compared to a study of African ungulate species, in which there was a range from the African buffalo *Syncerus caffer* (0.45) (Simonsen et al., 1998) to Grant's gazelle *Gazella granti* (0.88) (Arctander, Kat, Simonsen et al., 1996). This was because of the large number of individuals sampled from the monomorphic Buton region (n=17), compared to the other regions with 13 haplotypes (n=18). The highest genetic diversity in the anoa, represented by the greater number of ten haplotypes was present in the Central and North region (n=12). The focus of anoa conservation should be to maintain the greatest haplotype diversity. Therefore, anoa in the Central region, where most of these diverse individuals were sampled, should be identified as having high conservation status.

5.5.2 Matrilineal evolutionary history

The analyses showed that the South-East and Central clades had fewer nucleotide mutational differences than between Buton individuals and those in the other two clades. It is proposed above that the explanation for the greater differentiation of the Buton clade is due to a bottleneck event because of the oceanic barrier separating Buton and Sulawesi Islands and possibly a greater time since they shared a common ancestor. This is supported by the lack of genetic variation on Buton Island, which suggests that there has been no recent migration on a scale large enough to add genetic variability to this clade. Therefore, according to the mtDNA data, colonisation of Buton Island from Sulawesi caused a bottleneck effect and repeat colonisations are thought not to have occurred.

Comparison of the nucleotide diversity of the anoa to those of other species allowed the scale of divergence between the anoa clades to be determined. Two studies of partial D-loop sequences gave nucleotide diversity values between *Bos indicus* and *Bos taurus* the two domesticated cattle groups and also between the river and swamp subspecies of Asian water buffalo *Bubalus bubalis*. The diversity between the domesticated cattle showed 21.35% difference in sites for 90 animals (370 nucleotide sequence) (Bradley et al., 1996). A similar value of 22.78% variation for 80 animals was found between the Asian buffalo subspecies (158 nucleotide sequence) (Lau et al., 1998). These comparisons show that the nucleotide diversities between the three anoa clades (7.7%, 7.4% and 5.4%) were lower than those found between *Bos indicus* and *Bos taurus* and also between the subspecies of Asian water buffalo. The only caveat is that the smaller sample size of the anoa clades could be partially responsible for the lower value between anoa clades (Buton n=17, South-East n=6, Central n=10). This result suggests the time since common ancestry of the anoa clades was more recent than for the other species mentioned. These were estimated to have diverged 28,000 years before present for the buffalo subspecies and 117,000-275,000 years before present for the cattle, using different evolutionary rates of sequence change. Further estimations were omitted as the comparable species data was from different regions of the D-loop.

5.5.3 Comparison of microsatellite and mtDNA results

Results from the mtDNA analysis above showed a population structure related to the geographic origin of individuals, supporting the microsatellite results presented in Chapter 4. Further investigation into the correlation of geographic distance, biogeographic barriers and taxonomic criteria with genetic distance measures will be made in Chapter 6. These results do not indicate any differences in microsatellite and mtDNA population structure that could be related to paternal and matrilineal behaviour. Instead, the differences may be due to the more distant inference possible with mtDNA than nDNA (Palumbi, 1996).

The mtDNA neighbour joining tree (Figure 5.1) shows that there is a greater differentiation between the Buton individuals and both the Central and the South-East individuals, while the latter two groups are less differentiated. This is in agreement with the neighbour joining tree for subpopulations using microsatellite data (Figure 4.7). However, the neighbour joining tree for individuals as OTUs using microsatellite data shows a greater differentiation between South-East and Central individuals, than South-East and Buton individuals (Figure 4.6). The low bootstrap support for this tree indicates less reliability to this latter population structure.

The samples that were found to have identical microsatellite allele sizes (SE94 and SE95; A110 and A111) that were also analysed in this chapter showed no difference in their mtDNA sequences. The only differences were due to the occurrence of single missing nucleotides in the sequence. These were defined as either indels or sequencing errors, but it could not be determined which was responsible without repeat sequencing. Therefore, it could not be confirmed or refuted that these samples were actually duplicates or distinct individuals. The mtDNA data proved the anomalous nature of the two individuals from the Atypical group found in the microsatellite allele analysis. Instead of clustering with individuals from the neighbouring Central or North subpopulation, these individuals grouped with those from the South-East subpopulation. These results did not clarify this perturbation to the genetic structure, but suggested the likelihood of a human effect rather than that of gene flow barriers.

Further interpretation of these mtDNA sequence results with the microsatellite results (Chapter 4) will be made in Chapter 6. This will include assessment for correlation of genetic variation with biogeographic barriers, geographic distance, and characteristics currently used to define the anoa species.

Chapter 6

Co-analysis of Microsatellite and mtDNA Data with Descriptive Characteristics of Anoa Individuals

6.1 Introduction

The results presented in Chapters 4 and 5 have demonstrated that there is considerable genetic differentiation in the anoa population. The clustering of individuals into mtDNA haplotypes and tree building with individuals as OTUs for microsatellite alleles show the existence of a population structure that appears to be related to the geographic origin of individuals. In Chapter 6 I investigate further the relationship of genetic variation, as microsatellite allele and mtDNA sequence pairwise genetic distances, and geographic distance (Euclidian and shortest land route) using multivariate analyses and Mantel tests. Microsatellite allele data will be the focus analysis in Sections 6.2 and both microsatellite and mtDNA data will be used in Section 6.3 and 6.4.

The multivariate analyses investigate the spatial patterns of genetic variation, by displaying the data visually. This method presents genetic variation by a reduction in dimensionality of the data. To investigate a correlation of genetic and geographic distance the Principal Coordinates Analysis (PCoA) multivariate method was used.

The Mantel tests (Mantel, 1967; Manly, 1997) are designed to test for a significant correlation between genetic and geographic distances representing the effect of isolation by distance. This is described in Section 6.3. This analysis will determine if the population structure is more representative of a fragmented population or a cline of genetic variation across the anoa's range.

The proposed correlation of genetic and geographic distance has implications for taxonomic classification and population structure. This correlation is contrary to the previous taxonomic descriptions of the anoa (Groves, 1969). Therefore, it is of interest to test for correlations between factors previously connected with the taxonomy of the species and genetic variation found in this study, to verify if these taxonomic descriptions are valid. The two factors that Groves (1969) used to identify the two species were altitude and morphotypic descriptions. These factors and geographic distance will be tested for correlations with genetic variation. The analyses will include the multivariate statistical method Principal Coordinates Analysis (PCoA), similar to PCA previously described in Chapter 4 and simple and partial Mantel tests (Section 6.3 and 6.4), described below in Section 6.1.1. The Mantel methods will allow a more detailed assessment of the population structure of the anoa by employing hypothesis-testing techniques.

6.1.1 Mantel tests

There are two types of Mantel test, the simple Mantel test (Mantel, 1967), which tests for a correlation between two matrices, while the partial Mantel test tests three or more matrices (Smouse, Long and Sokal, 1986; Thorpe, Brown, Malhotra et al., 1991; Manly, 1997). Both of these tests have been employed in the areas of population genetics and ecology (Legendre and Legendre, 1998).

For the simple Mantel test, the null hypothesis is that distances in the two matrices are independent of each other. However, often the observed pattern of variation in matrices may be very complex, but hypotheses used to test for correlations may be limited and simpler. If several hypotheses are tested at one time this is advantageous because these hypotheses may produce intercorrelation patterns, and more than one factor may be contributing to the variation that is being tested for. For example, in studies of the common shrew (*Sorex araneus*) in the Alps the effect of glacial advances and habitat suitability were found to be more correlated to genetic variation than other proposed barriers such as moraines, rivers and a mountain pass (Lugon-Moulin, Brunner, Balloux et al., 1999; Lugon-Moulin and Hausser, 2002).

The aim of using a partial Mantel test is to test the correlation between two matrices, while at the same time controlling for the effect of a third matrix. This can remove the effect of spurious correlations caused by this third matrix. Different forms of the partial Mantel test have been tested (Anderson and Legendre, 1999). The method of permutation of raw values (Smouse et al., 1986) was found to be suitable for small sample sizes ($n < 20$), combined with highly skewed data. Whereas the permutation of the residuals of a null model (Freedman and Lane, 1983; Anderson et al., 1999) could be used in most situations. In this study the former model was used due to the small sample sizes. There has recently been some debate over the limitations of this test (Raufaste and Rousset, 2001; Castellano and Balletto, 2002), but this method has been shown to be effective in similar studies to this one (Coulon, Cosson, Angibault et al., 2004; Worley, Strobeck, Arthur et al., 2004). Therefore, partial Mantel tests that test multiple hypotheses at one time have a considerable advantage over simple Mantel tests.

6.2 Multivariate statistical analysis

Principal Coordinates Analysis (PCoA) searches for similarities between cases (Gower and Hand, 1996). This technique uses a matrix of similarity to analyse genetic distance between individuals. The resulting set of coordinates for a number of axes are plotted for visual representation. These coordinates can also be calculated from a genetic distance matrix, but here data were comprised of categorical values to avoid loss of information, described below.

6.2.1 Data preparation

Microsatellite allele size data was transformed into canonical variables. Each individual was coded with respect to particular alleles at a locus to show if it was heterozygote (1), homozygote (2) or if the given allele was absent (0). This resulted in the loss of the allele size information. However, if data had been analysed as allele frequencies then even more information would have been lost (Krzanowski, 1988).

The effect of bias in the data was reduced where possible. Individuals with a maximum of one missing value were included in the analyses (n=89). Low frequency alleles may also cause bias (Roques, Sevigny and Bernatchez, 2001). Therefore, analyses were repeated with rare alleles (present less than five times) either included or excluded. Comparison of results showed no appreciable difference between them, so the results below represent analysis containing all alleles. The analyses were conducted using GenStat Release 7.1 (Lawes-Agricultural-Trust, 2003).

6.2.2 Principal coordinates analysis plot

The results of the PCoA were defined by the coordinates on three axes. The coordinates of the first axis accounted for 38.43% of the variation, the second for 18.92% and the third for 13.70%. In relation to geographic distribution, a two dimensional PCoA plot was overlaid onto a map of the anoa's range (Figure 6.1). A rotation of 90^{o} was used to fit the PCoA plot to the map. There was slight clustering of individuals according to their geographic origin, but there was considerable introgression or overlapping between all biogeographic groups. It is also of interest to reassess the wild individuals represented by open symbols in the PCA plot in Figure 4.10. This PCA plot differs slightly from the PCoA plot here by the greater clustering of individuals from the same biogeographic region, a result of the different analysis method.

Generally there was a poor fit of the regional clusters to their geographic locales on the background Sulawesi map for the PCoA. The clustering of individuals correctly defined North and Central individuals to the north half of the plot, and Buton and the South-East to the south part of the plot, with only five and eight exceptions, respectively. Therefore, this multivariate analysis has demonstrated that genetic variation of microsatellite alleles is related to the geographic locality of individuals. The significance of this correlation will be tested below.

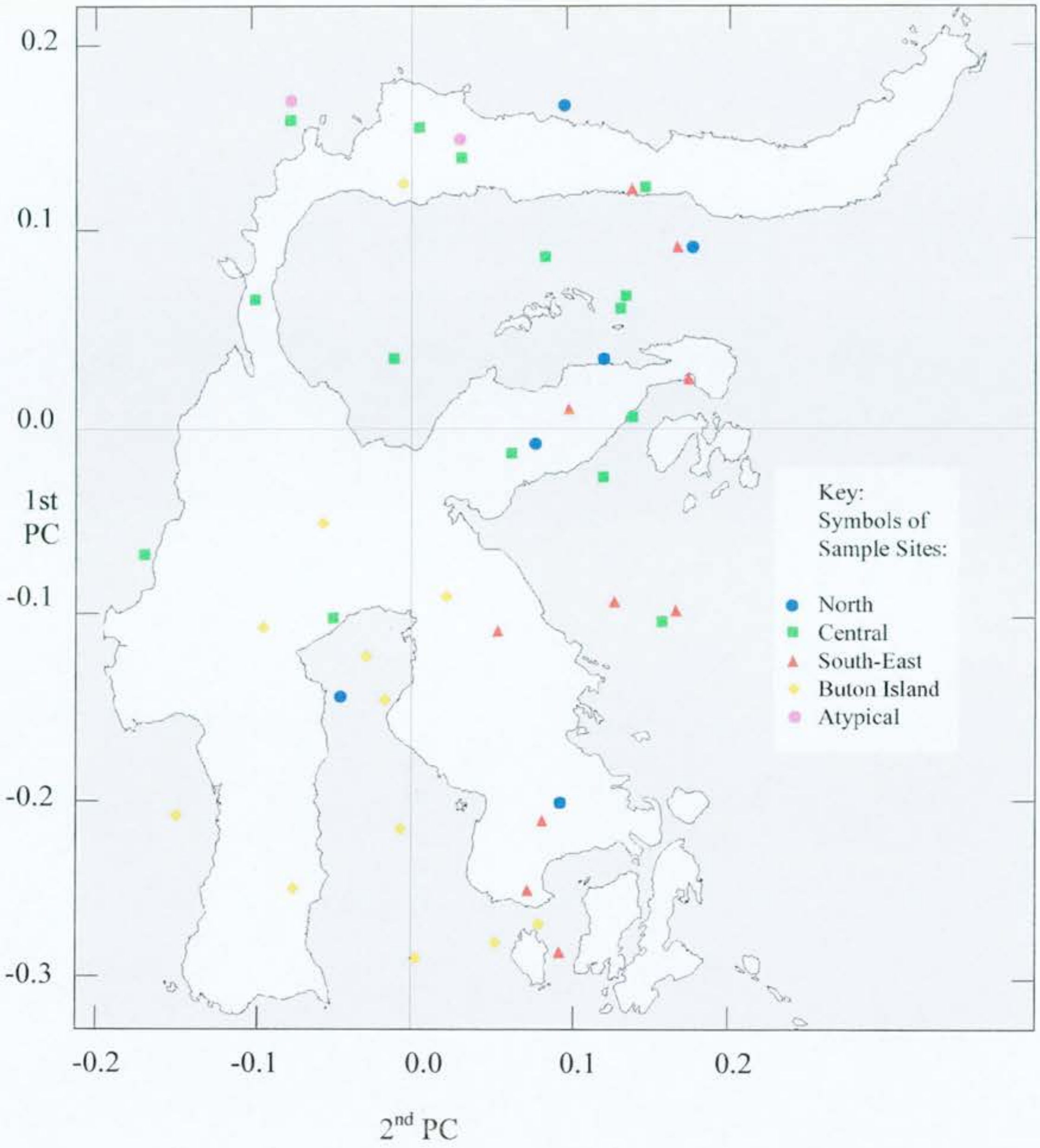


Figure 6.1. Two dimensional PCoA (rotated 90 °) of 89 anoa individuals identified into four biogeographic regions (and Atypical) overlaid onto a map of Sulawesi Island.
(PC = Principal Coordinate)

6.3 Isolation by distance analysis

6.3.1 Introduction

Biological factors mean that individual dispersal is geographically restricted for many species. Models to assess this isolation by distance were developed by Wright (1943). More recently these were designed for microsatellite data relating to sub-populations (Rousset, 1997) and for a continuous population on a local scale (Rousset, 2000). A Mantel test was employed to test for correlation between genetic (microsatellite allelic and mtDNA sequence) variation and geographic distances. The previous analyses of anoa microsatellite and mtDNA data were not able to determine if there were divisions in the anoa population. Therefore, for the microsatellite data the a_r statistic, $F_{ST}/(1- F_{ST})$, was preferred to test for the effect of isolation by distance in a continuous population.

6.3.2 Data preparation

Geographic distance was measured in two ways using a 1:100,000 scale map of Sulawesi Island. Firstly, measurements were made with a lattice (grid of coordinates) to estimate direct or Euclidean distance, and secondly, using a map wheel to measure the shortest land route (and avoiding lakes). An alternative measurement to the direct distance has been used to represent routes of movement of the European roe deer (*Capreolus capreolus*) defined by habitat preferences and called the 'Least cost distance' described by Coulon, Cosson, Angibault et al. (2004). Data on habitat fragmentation was not combined with distances measured because of the lack of availability of recent accurate forest cover maps. The $F_{ST}/(1- F_{ST})$ statistic was designed for correlation with the lattice measurements, but in the case of Sulawesi

Island the direct distance (or lattice) was not equivalent to the shortest land route from one individual to another. The shape of the island's various peninsulas mean that the direct line distance often crosses the sea, whereas gene flow would be expected to have greater effect via the land route. Therefore, both were measured and the levels of correlation to genetic variation were compared.

Sampled individuals were located on the map as accurately as records permitted. Figure 4.1 displays the location of individuals with microsatellite data and Figure 5.1a those with mtDNA data. The animals from Buton Island have been grouped for easier representation. Five animals (B27, SE63, C65, C69, and C130) were omitted from the microsatellite analysis and one from the mtDNA analysis (SE63), as their exact origin was unknown. Multiple individuals were sampled from certain localities. The distance between these localities was unclear; so all individuals were included with a minimum pairwise distance of 1km. This analysis was repeated with single individuals from each of these locations to test for the effect of bias, but all results showed the same levels of significance as the complete dataset and there was little difference in the slope and confidence limits. Isolation by distance analysis was conducted for the microsatellite data (n=84) and mtDNA distances (n=34) and on different subsets of the datasets.

6.3.3 Isolation by distance of a continuous anoa population on Sulawesi Island

The analysis of individuals in a 'continuous' population was possible using the a_r statistic for microsatellite data (Rousset, 2000), which is a generalization of the theory of $F_{ST}/(1- F_{ST})$ between pairs of individuals (Rousset, 1997). For a small geographical scale $F_{ST}/(1- F_{ST})$ has an approximately linear relationship to the logarithm of distance. The $F_{ST}/(1- F_{ST})$ model is thought to perform better than previous models, firstly, because the method makes weak assumptions about the distribution of dispersal distance. Secondly, the interpretation of F_{ST} values related to geographic distance is also more informative than a single F_{ST} value for all data (Slatkin, 1993; Rousset, 2000). The mtDNA was analysed as pairwise distances between individuals (Puorto, Da Grace Salomã, Theakston et al., 2001).

The selection of the $F_{ST}/(1- F_{ST})$ model was made following assessment of the anoa dataset under models for 'sub-population' and a 'continuous' population using the three scenarios described by Rousset (2000). The anoa dataset was similar in part to two of these scenarios, having little spatial clustering of individuals. However, the $F_{ST}/(1- F_{ST})$ method was selected because of the following factors. The current poverty of sample coverage across the whole study area meant that it was important to avoid the arbitrary setting of geographical limits, when allocating individuals into sub-populations. Also the non-random density of sampled individuals varied between regions, making standardized geographic clustering problematic.

The linear relationship between $F_{ST}/(1- F_{ST})$ genetic distance and logarithm of geographic distance is thought to hold less well at shorter distances. This is because of the effect of the dispersal distance of individuals within a population (Rousset, 1997). As the origins of individuals were only known to within a few kilometers, all individuals were included, so that unrelated individuals were not omitted. If low significance levels of the correlation had been recorded, then a minimum geographic distance would have been set to compensate for the effect of dispersal of individuals.

Analysis was conducted using GENEPOP Version 3.4 (Raymond et al., 1995) with the isolation by distance for individuals option. This was followed by a Mantel test of 1000 permutations. The regression of linear distance against $F_{ST}/(1- F_{ST})$ was plotted and confidence intervals of the regression slope were calculated.

Assessment of the effect of isolation by distance on microsatellite variation across the anoa's range was first conducted, using 84 individuals, with both the lattice and shortest land route geographic distances. Further analysis used only the lattice geographic distance. The four subpopulations (Buton, South-East, Central and North) and the Atypical cluster were analysed as subpopulations to test for the effect of IBD within each. Also analyses were conducted with the pairs of neighbouring subpopulations (Buton and South-East, South-East and Central, and Central and North) to test for isolation by distance between these biogeographic regions. The differences in *r*-value (slope) and in P values were used to define differences in the relationship of genetic and geographic distance in the results presented here.

The Mantel test of lattice geographic distance across the anoa's range showed significant correlation to microsatellite genetic variation ($r= 0.239$, $P =0.001$) with 95% confidence interval 0.34559095 to 0.15908789 (Figure 6.2). Figure 6.2 only illustrates a potential effect of isolation by distance of anoa individuals on Sulawesi and Buton Islands. Barriers to gene flow could also cause this effect. For the shortest land route distance, the values were very similar to the lattice geographic distance: $r=0.246$, $P=0.001$. This suggested that any difference between the direct line and land route geographic distance did not affect the correlation. A Mantel test of the mtDNA sequence pairwise distances and lattice geographic distances produced a significant correlation (*r*=0.032, P=0.001). The data included 34 individuals from across the anoa's range. Further Mantel tests involved microsatellite data only, due to this small sample size for mtDNA data.

When the four subpopulations (Buton, South-East, Central and North) and the Atypical group were tested independently only the Central subpopulation showed a significant correlation between geographic and genetic distance (P=0.001). Results are given in Table 6.1, with significant values shown in bold, after Bonferroni corrections. The lack of significant values in other subpopulations may be real, as seen by the *r*-value=0.0039 in Buton (n=27), or this could be explained by the low number of individuals with a sufficient pairwise geographic distance. Between the three pairs of neighbouring subpopulations (Buton and South-East, South-East and Central, and Central and North) there were significant correlations between genetic and geographic distance (P=0.001) (Table 6.1b). The highest *r*-value occurred between the Buton and South-East subpopulations (*r*=0.243), confirming the high level of differentiation between Buton and Sulawesi Islands. These islands are separated by a sea barrier of ~1km across the Muna Strait. Anoa have been reported to swim (Mustari, 1995), and it is quite possible that they might have crossed to Buton at the current sea level. Nevertheless, the sea could act as a significant barrier to gene flow. Therefore, from these analyses it remains to be clarified if geographic distance or isolation across biogeographic barriers (between subpopulations) has had the greatest effect in determining genetic distance.

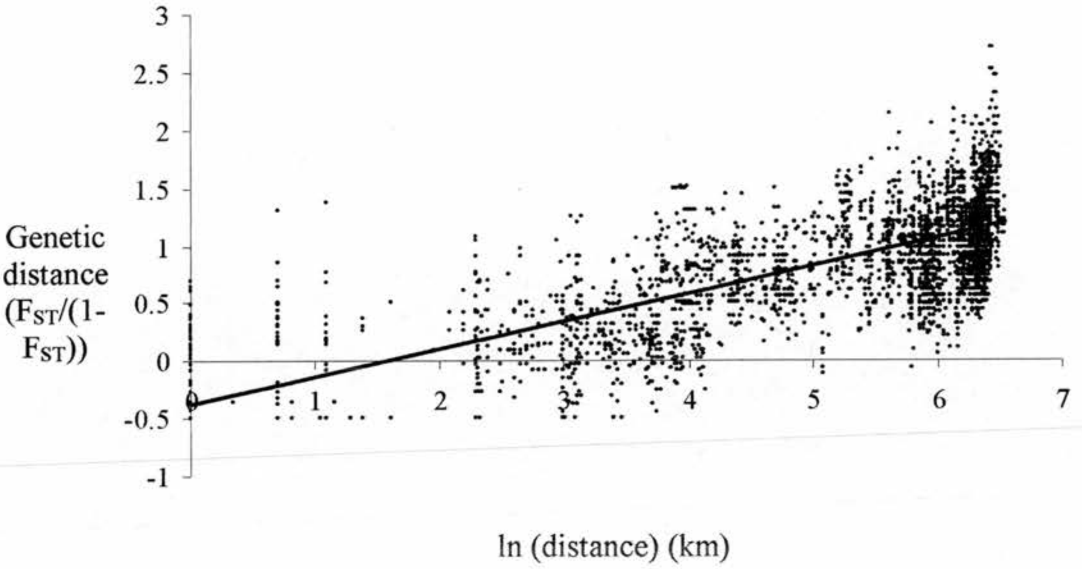


Figure 6.2. Correlation of geographic distance from the logarithm of lattice measurements and F_{ST}/(1- F_{ST}) genetic distance for 84 individuals from Sulawesi and Buton Islands.

Points show the estimates for pairs of individuals. The line represents the regression line.

Table 6.1. Mantel test results of isolation by distance correlation between logarithm of geographic distance and genetic distance (F_{ST}/(1- F_{ST})), a) within five anoa subpopulations and b) between pairs of neighbouring subpopulations.

Subpopulation/s	*r*-value	P
a)		
Buton	0.0039	0.910
South-East	0.0586	0.052
Central	**0.1096**	**0.001**
North	0.0645	0.318
Atypical	0.1213	0.089
b)		
Buton and South-East	**0.2433**	**0.001**
South-East and Central	**0.1177**	**0.001**
Central and North	**0.0758**	**0.001**

To conclude, the anoa population of Sulawesi and Buton Islands is potentially under the effect of isolation by distance, as shown by the significant correlation between geographic distance and $F_{ST}/(1- F_{ST})$ distance from microsatellite allele data. mtDNA pairwise distance also showed a significant correlation with lattice geographic distance. This supports the results of previous analyses that geographic distance is an important factor in explaining the genetic variation in the anoa population. The population structure is represented by the greatest differentiation between the Buton and South-East subpopulations. Other factors may also have affected this genetic variation, such as barriers to gene flow, which will be tested below without the effect of geographic distance.

6.4 Partial Mantel test

6.4.1 Introduction

The aim of this partial Mantel analysis was to test for a correlation between genetic distance (both microsatellite and mtDNA) and multiple factors that could have singly, or in combination, affected genetic variation within the anoa population. Two sets of partial Mantel tests were conducted to answer two sets of hypotheses. The first was a continuation of the previous section assessing the level of correlation of geographic distance and biogeographic barriers with genetic distance. Secondly, the relationships of elevation and species morphotype to genetic distance were analysed to determine if there was any relevance in the current taxonomic classification of the anoa (Groves, 1969). The partial Mantel method tested the correlation between two matrices, while at the same time controlling the effect of a third matrix. It has been employed in recent studies in a similar way to identify factors that are associated with genetic differentiation (Thorpe, Malhotra, Black et al., 1995; Lugon-Moulin et al., 1999; Riginos and Nachman, 2001). The advantage of this method is that although this study found geographic distance to be significantly correlated with genetic distance in the Mantel test above, other factors may impact on the genetic variation. Therefore, a correlation between genetic distance and elevation or morphotype can be tested for, while removing the effect of geographic distance.

The combination of data used for testing correlations of genetic distance and geographic distance or biogeographic barriers is summarized by the null hypotheses:

H_o= There is no correlation between each of three potential biogeographic barriers tested independently and genetic variation within the two neighbouring subpopulations (remove effect of geographic distance).

H_o= There is no correlation between any of three potential biogeographic barrier and genetic variation across the anoa's range (remove effect of geographic distance).

H_o= There is no correlation between the Buton and South-East potential biogeographic barrier and genetic variation across the anoa's range (remove effect of geographic distance).

These null hypotheses were tested as stated above and also for the reverse, by removing the effect of the barriers and testing for a correlation with geographic distance. Each test was repeated with microsatellite data and the latter two null hypotheses were repeated with mtDNA genetic distance data, so a total of 14 tests were conducted.

Secondly, the combination of data used for testing correlations of genetic distance and elevation or species morphotype is summarized by the null hypotheses:

H_o= There is no correlation between elevation and genetic variation (remove effect of geographic distance).

H_o= There is no correlation between species morphotype and genetic variation (remove effect of geographic distance).

H_o= There is no correlation between geographic distance and genetic variation (remove effect of altitude).

Again, the null hypotheses were also tested by removing the effect of elevation and species morphotype to test for a correlation with geographic distance. All tests were repeated for microsatellite and mtDNA data, totaling 12 tests.

6.4.2 Data preparation and methodology

Data matrices were produced for all the datasets mentioned above. The genetic distance data for microsatellite alleles and mtDNA sequences were prepared as described in Section 6.3. The lattice geographic distance data was also formatted as described in Section 6.3. Matrices were created to account for the effect of each of the biogeographic barriers (between Buton and South-East, South-East and Central, and Central and North). Those individuals in the same subpopulation (0) were differentiated from those separated by one of the barriers (1). The dataset consisted of 72 anoa, following the removal of individuals in the Atypical group that did not fit into the biogeographic pattern of the Wallacea region.

The elevation or altitude of the sampling location of individuals was recorded using 1:25,000 scale maps. This was converted to variation in elevation between individuals, as was calculated by Knoll and Rowell-Rahier (1998). These values were then transformed into two categories for pairs of individuals, greater than 1000m difference in elevation (1) or less than 1000m difference in elevation (0) between two individuals. The species morphotype identifications were taken from descriptive notes collected with each sample, where available. Where these descriptions were absent or insufficient individuals were removed from the analysis, leaving a sample size of 60 and 25 anoa for the microsatellite and mtDNA datasets, respectively. The criteria for classification of the two morphotypes have been widely used in anoa taxonomic studies (Groves, 1969; Burton et al., 2005). The morphological descriptions are given in Appendix 2 and the data used in this analysis is described in Appendix 3. Two individuals with morphological characteristics representing the same species morphotype (0) were differentiated from those displaying characteristics of different morphotypes (1).

The method used followed the procedures of Manly (1997) of a standard multiple regression. The FSTAT program version 2.9.3 was used for the analysis (Goudet, 1995). All partial Mantel tests were run using 10 000 randomizations of the matrix of observed distances. Combinations of the matrices described above were analysed to test the H_0 stated above. A Bonferroni correction (Sokal and Rohlf, 1995) was used to reduce the likelihood of type I errors among the partial Mantel tests performed.

6.4.3 Partial Mantel test results

All the results of partial Mantel tests are presented in Table 6.2, with the assessment of biogeographic barriers and geographic distance given in a), and the elevation and species morphotype results shown in b). The values are the genetic variance explained by the second matrix, once the third matrix is taken into account. Significant results show probability that the variance is different from zero. These are shown in bold, after Bonferroni corrections.

The results relating to biogeographic barriers and geographic distance showed variation in the effect of these two factors across the anoa's range. The test for a correlation of genetic distance and the presence of a barrier dividing Buton and

South-East subpopulations, with removal of the effect geographic distance, showed a considerable amount of the genetic variation was explained by the barrier (0.3132), with a significant correlation (P=0.0075). None of the other tests involving one barrier and two subpopulations were significant. The lack of a significant correlation might have been due to the small sample size or the small genetic variation caused by geographic distance or gene flow barriers, in contrast to greater variation caused by other factors such as habitat, river barriers or the anoa's behaviour. For example, this was shown by the fact that only one of five subpopulations (Central) gave a significant correlation between genetic and geographic distance in the Mantel test (Table 6.1).

By testing all three biogeographic barriers together across the anoa's range, a significant correlation was observed with genetic distance (0.3051, P=0.0104), when the geographic distance effect was removed. Although geographic distance was not significantly correlated to genetic distance, once the effect of these barriers had been removed, some variance was explained by this factor (0.2407). Therefore, the significant correlation between genetic and geographic distance observed in the Mantel test above (Figure 6.2) may be in part due to the effect of barriers to gene flow across the anoa's range, rather than solely due to geographic distance. This may be due to one rather than multiple barriers, as the first three partial Mantel tests suggest. The oceanic barrier between Buton and South-East subpopulations was the only result of the three tests that was significantly correlated. The effect of the sea between Buton and Sulawesi Islands may have acted as a stronger dispersal barrier than predicted by geographic distance alone. Therefore, a combination of geographic distance and barriers to gene flow appear to be the most likely explanation for the genetic variation across the anoa's range of Buton and Sulawesi Islands.

Partial Mantel test results in Table 6.2b show that neither elevation nor species morphotype descriptions have a significant correlation with gene flow once isolation by distance has been accounted for. The extra variance explained by the elevation matrix was 5.4%, P=0.0745, and for the morphological species matrix the extra variance explained was only 2.6%, P = 0.2256.

The results of partial Mantel tests of the mtDNA data (Table 6.2a) did not show a significant correlation between pairwise genetic distance and geographic distance or any barriers, where data was sufficient for these tests. Significant correlations were observed between mtDNA genetic distance and elevation, with the removal of geographic distance; and between mtDNA genetic distance and geographic distance, with the removal of elevation and species morphotype (Table 6.2b). The correlation of elevation with genetic distance may be due to all the individuals in the higher altitude range (n=9) comprising all except one of the Central subpopulation. It cannot be ruled out that this significant value is actually representative of the genetic variation of this subpopulation, rather than an altitudinal effect. As the microsatellite data with a greater sample size does not support this correlation, there cannot be presumed to be a significant correlation between mtDNA genetic distance and altitude from this data.

Table 6.2 Results of partial Mantel tests for microsatellite and mtDNA data, a) between combinations of subpopulations for lattice geographic distance (km), genetic distance $F_{ST}/(1- F_{ST})$ or mtDNA pairwise distance and barriers between neighbouring populations and b) for elevation matrix, species morphological classification matrix, and geographic distance matrix.

	Microsatellite		**mtDNA**	
Matrices compared	Extra variance explained	P	Extra variance explained	P
a)				
Buton and South-East barrier:				
Gendist.BioBSE.Geog	**0.31320**	**0.0075**	-	-
Gendist.Geog.BioBSE	0.19205	0.1046	-	-
South-East and Central barrier:				
Gendist.BioSEC.Geog	-0.25634	0.1282	-	-
Gendist.Geog.BioSEC	0.28787	0.0795	-	-
Central and North barrier:				
Gendist.BioCN.Geog	-0.09207	0.6100	-	-
Gendist.Geog.BioCN	0.08067	0.6581	-	-
Four subpopulations, three barriers:				
Gendist.BioALL.Geog	**0.30511**	**0.0104**	0.25027	0.1520
Gendist.Geog.BioALL	0.24071	0.0438	0.22832	0.1949
Four subpopulations, Buton and South-East barrier:				
Gendist.BioISL.Geog	**0.29959**	**0.0107**	0.25027	0.1541
Gendist.Geog.BioISL	**0.89198**	**0.0001**	0.22832	0.1919
b)				
Gendist.Elev.Geog	-0.01596	0.9014	0.19779	0.3323
Gendist.Elev.Morph	0.23069	0.0730	**0.55060**	**0.0049**
Gendist.Morph.Geog	-0.07553	0.5694	-0.11907	0.5658
Gendist.Morph.Elev	-0.15966	0.2248	-0.28221	0.1521
Gendist.Geog.Elev	**0.86444**	**0.0001**	**0.50829**	**0.0042**
Gendist.Geog.Morph	**0.88342**	**0.0001**	**0.67597**	**0.0006**

X.Y.Z is the amount of variance of X explained by Y after Z is taken into account. Gendist = genetic distance; Geog = lattice geographic distance; BioBSE = biogeographic barrier between Buton and South-East subpopulations; BioSEC = barrier between South-East and Central; BioCN = barrier between Central and North; BioALL = all three barriers; BioISL = barrier between Buton and South-East including data from all Sulawesi anoa; Elev = elevation; Morph = species morphotype (see text for explanation of matrices)

6.5 Discussion

Population genetic analyses of both microsatellite and mtDNA variation refute the hypothesis of the anoa population structure being related to altitude or morphological variation. These results provide clear evidence of a population structure related to a combination of geographic distance and known biogeographic barriers. The isolation by distance analysis and PCoA suggested there was a correlation between geographic distance and genetic variation. Both these methods have been successfully used in previous studies of ungulates to show the effect of isolation by distance on genetic variation (MacHugh, 1996; Forbes et al., 1999; Worley et al., 2004). Hypothesis testing was applied to determine how microsatellite and mtDNA variation is correlated to altitudinal separation, species morphotype (previous taxonomic criteria), biogeographic barriers and geographic distance.

Partial Mantel tests confirmed that a contribution of gene flow barriers and geographic distance were correlated to the genetic variation. The oceanic barrier between Sulawesi and Buton Islands acted as a stronger dispersal barrier than predicted by geographic distance alone. The effect of a barrier to gene flow has been shown in a study of túngara frog (*Physalaemus pustulosus*) (Lampert, Rand, Mueller et al., 2003) and the common shrew (*Sorex araneus*) (Lugon-Moulin et al., 1999; Lugon-Moulin et al., 2002) using Mantel and partial Mantel tests. On Sulawesi range fragmentation was less apparent. The partial Mantel tests performed could not determine if geographic distance or gene flow barriers were the more closely correlated to genetic variation. However, isolation by distance was shown be correlated to genetic variation within the Central subpopulation by a simple Mantel test. So the genetic variation of Sulawesi's anoa is partially due to the effect of isolation by distance.

These results are clearly in contradiction to the previous taxonomic status of the anoa, two species defined by morphological characteristics, inhabiting distinct altitudinal ranges (Groves, 1969). Results of hypothesis testing found no significant correlation of elevation and species morphotype with microsatellite variation, and species morphotype with mtDNA variation. The only significant correlation (mtDNA variation and elevation) could not be accepted as a true result because of the lack of sufficient data.

In summary, elevation and species morphotype descriptions, which previously were used to describe the two anoa species, were not correlated to genetic variation found in microsatellite alleles or conclusively to that found in mtDNA sequence data. Instead, microsatellite genetic variation was significantly correlated to at least one biogeographic barrier (between Buton and South-East subpopulations), and a proportion of the variation was explained by geographic distance. Therefore, reassessment of the anoa's taxonomic status is required and this will be discussed in the following chapter.

Chapter 7

Discussion and Conclusions

7.1 Introduction

The first phase of management of a species is clarification of taxonomic uncertainties and population structure. This has been the overall focus of this study, with the results presented in Chapters 4, 5 and 6. These results and those of previous analyses will be summarized below. The complexities of the population colonisation and structure will be discussed first in order to define the historical and recent genetic variations to give some background for the following sections. Comparisons are made between the anoa and species with similarly complex phylogeographic distributions in Sulawesi. These distributions are contrasted with the evolutionary biogeography of the region using 'areas of endemism' (AOEs) (Evans et al., 2003c). In the following section (6.3) the genetic variation between geographic regions in relation to the anoa's species status are discussed. Then interpretation of these results is made in relation to conservation and management strategies required, resulting from the proposed taxonomic status. The final sections assess the applicability of techniques used here as conservation tools and suggest further analysis that could be conducted.

7.2 The Anoa and Sulawesi's biogeography

7.2.1 Anoa's colonisation of Sulawesi

The location on Sulawesi of the colonisation by anoa could not be determined from these results. Certain results suggested that the Central region could have been a site of anoa colonisation. These results included the presence of the greatest microsatellite allelic variation and the highest number of private alleles from this area. Also the greatest number of mtDNA haplotypes were found in the same region. However, these results are inconclusive and could have been caused by hybridisation of two previously isolated populations, as reported by Kumar et al. between *Bos taurus* and *B. indicus* cattle (2003). Other species, such as the macaques colonised Sulawesi twice; to either the central west, northwest or southwest regions (Fooden, 1969; Albrecht, 1978; Evans et al., 1999). The possibility of the anoa arriving in the southwest cannot be tested because sample collection was not feasible in this area due to the small population size. The macaques and the anoa were thought by Groves (2001) to be members of the 'old endemics' or early arrivals on Sulawesi. Therefore, it is possible that the anoa colonised Sulawesi in the southwest following the same route as the macaques. Kakio et al. (1994) proposed a second colonisation by the anoa based on variation in amino acid sequences of hemoglobin beta chains. From the evidence mentioned above and that given in the following paragraph the colonisation route of the anoa remains unconfirmed.

If the anoa did arrive on the Central region of Sulawesi, the route of colonisation may have been across the Makassar Strait (Musser, 1987; Moss et al., 1998), which would have been assisted by the much narrowed sea barrier of an estimated 40km between Borneo and Sulawesi in the Pleistocene era (Voris, 2000). To date it has not been suggested that the anoa might have arrived from the north and that it could be most closely related to the endemic tamaraw (*Bubalus mindorensis*) of the Philippines. Other authors found evidence of the anoa being close relatives of the Asian water buffalo (*Bubalus arnee*) (Groves, 1969; Geraads, 1992; Tanaka et al., 1996; Schreiber et al., 1999). This species' range historically only covered mainland Asia as far south as Malaysia and there are no records of their existence on Borneo (Hedges, 1998). The only bovid native to Borneo is the distantly related banteng (*Bos javanicus*) (MacKinnon et al., 1996). Therefore, the Asian water buffalo could have been an early ancestor of the anoa.

Van den Berg et al. (2001) hypothesized that the anoa arrived in the Late Pleistocene or Holocene. They thought that the anoa dispersed to the South, crossing the Tempe Depression (at the northern end of the south peninsula, near its meeting point with the central region) during the sea level lowering of the Mid-Pleistocene, and were therefore present in the fossil records of that period in South Sulawesi. The results of Tanaka et al. (1996) are also in agreement with the proposed expansion of the anoa from Central Sulawesi to the South (Van den Berg et al., 2001) in this period. Unfortunatly, results presented here cannot confirm this.

7.2.2 Anoa population structure

The identification of the level of population structure and the location of the barriers to gene flow are addressed below. Genetic variation of anoa across Sulawesi was found to be significantly correlated with geographic distance using Mantel tests, showing the effect of isolation by distance. This is likely to have been reinforced by the territorial behaviour of the anoa and the geographical structure of the island. The existence of geographic barriers across the anoa's range (Fooden, 1969) may have caused isolation of anoa in geographic locales. This was supported by partial Mantel tests, which defined that genetic variation was significantly correlated to a barrier to gene flow between Buton and Sulawesi Islands. The microsatellite allele Dps distance, mtDNA neighbour joining trees, STRUCTURE analysis and Hardy-Weinberg equilibrium tests all support structuring of the population caused by isolation by distance or barriers to gene flow. Also distinctive to each geographic group were certain private alleles and mtDNA haplotypes, indicative of limited gene flow and a structured population.

The current sea barrier dividing Buton from Sulawesi was manifest in the clear discontinuity in genetic composition between individuals from Buton and South-East Sulawesi seen in the mtDNA sequence mismatch plots, the presence of private alleles in Buton individuals, and partial Mantel tests. This was defined as range fragmentation. These features of genetic variation are indicative of historical structuring of the anoa population. Other gene flow barriers on Sulawesi may have been influential in determining the population structure of the anoa. However, only isolation by distance was shown to be significantly correlated to genetic variation on the mainland (in the Central subpopulation). The truncated branch lengths of microsatellite allele NJ trees also supported the presence of gene flow between the Sulawesi subpopulations. Additionally, the partial Mantel test results were not able to determine a significant correlation of genetic variation and other barriers, once the effect of geographic distance had been removed. Sulawesi's geological history and other Sulawesi species' population structure are compared with the results from the anoa below.

7.2.3 Biogeographic regions: Comparison of anoa and other Sulawesi species' gene flow barriers and the Island's history

In general, the anoa microsatellite and mtDNA data presented in this thesis support the 'areas of endemism' (AOEs) of Sulawesi proposed by Evans et al. (2003c). This structuring was defined by the large variation in allele frequencies and mtDNA sequences between individuals from different localities. Defining the exact geographical location of barriers between anoa clusters (groups of individuals) was more difficult because this genetic variation may have been an effect of isolation by distance and also due to the lack of samples in certain areas. Instead, areas were defined that represented the genetic subpopulations of anoa (Figure 7.1).

The effect of an oceanic barrier between Buton and Sulawesi Islands found in microsatellite and mtDNA results of the anoa here was also defined by mtDNA and aDNA in the macaque population (Evans et al., 2003c), but absent in the Sulawesi toad's (*Bufo celebensis*) population structure. It is proposed that a bottleneck event occurred when anoa arrived on Buton, followed by a lack of recent migration on a scale large enough to add genetic variability to this subpopulation. Firstly, this is shown by the lower genetic diversity in this subpopulation compared to others. The considerable divergence of the Buton subpopulation compared to the South-East and Central mtDNA clades (7.4% and 7.7%, respectively) supports this bottleneck. These divergence values suggest a considerable time since common ancestry. Although they were less than between *Bos taurus* and *B. indicus* cattle (Bradley et al., 1996) and between the subspecies of Asian water buffalo (Lau et al., 1998). This is further supported by STRUCTURE analysis and the greater $(\delta\mu)^2$ genetic distance differentiating the Buton individuals from those in other subpopulations. Considering this clear differentiation of the Buton Island subpopulation, and the existence of a distinct macaque species, Buton should be classed as a distinct AOE, which it currently is not. Between the South-East and Central anoa subpopulations the effect of isolation by distance was found, this coincides with the boundary between two AOEs. Four anoa individuals from the east peninsula of Sulawesi showed similarity in microsatellite allele sizes to animals in the Central and North subpopulations in STRUCTURE analysis, but formed a distinct cluster in the Dps distance neighbour joining tree. The genetic differentiation of individuals in this region is in agreement with the variation in mtDNA and aDNA of the macaques and mtDNA of the toad, which suggest that this region forms a distinct AOE.

Discrepancies were found in the distribution of variation in autosomal DNA (aDNA) of macaque species compared to mtDNA (Evans et al., 2003b). This was also the case for the population structure of the anoa, with the North and Central microsatellite subpopulations forming one haplogroup or clade in the mtDNA NJ tree.

These differences in the location of barriers for different species is illustrated in the partition in the narrow neck of land to the north of the central area, where the location of the barrier between macaque species (*M. hecki* and *M. tonkkeana*) was discordant with that of the *Bufo celebensis* toads (Evans et al., 2003c). There was also discordance between the macaques and the anoa, observed from STRUCTURE analysis and Dps neighbour-joining tree results. Another interpretation of this is that the second colonisation by macaques caused this partition, which did not occur for the toad or anoa. The partitions of the toad and the anoa subpopulations in this region are similar, while the partitions of macaques are different from these two species.

In the northern peninsula the individuals from the North subpopulation were distributed across two AOEs. The geographic distribution of the Atypical group of individuals, that showed low allelic variation, did not conform to the distribution of AOEs. The factors that originally caused the barriers to gene flow might not remain, so these partitions may have become more fluid. The loss of a geological barrier could affect different species in different ways (Evans et al., 2003c). Therefore, it is feasible that the Atypical group may represent the genetic diversity from an AOE to the east.

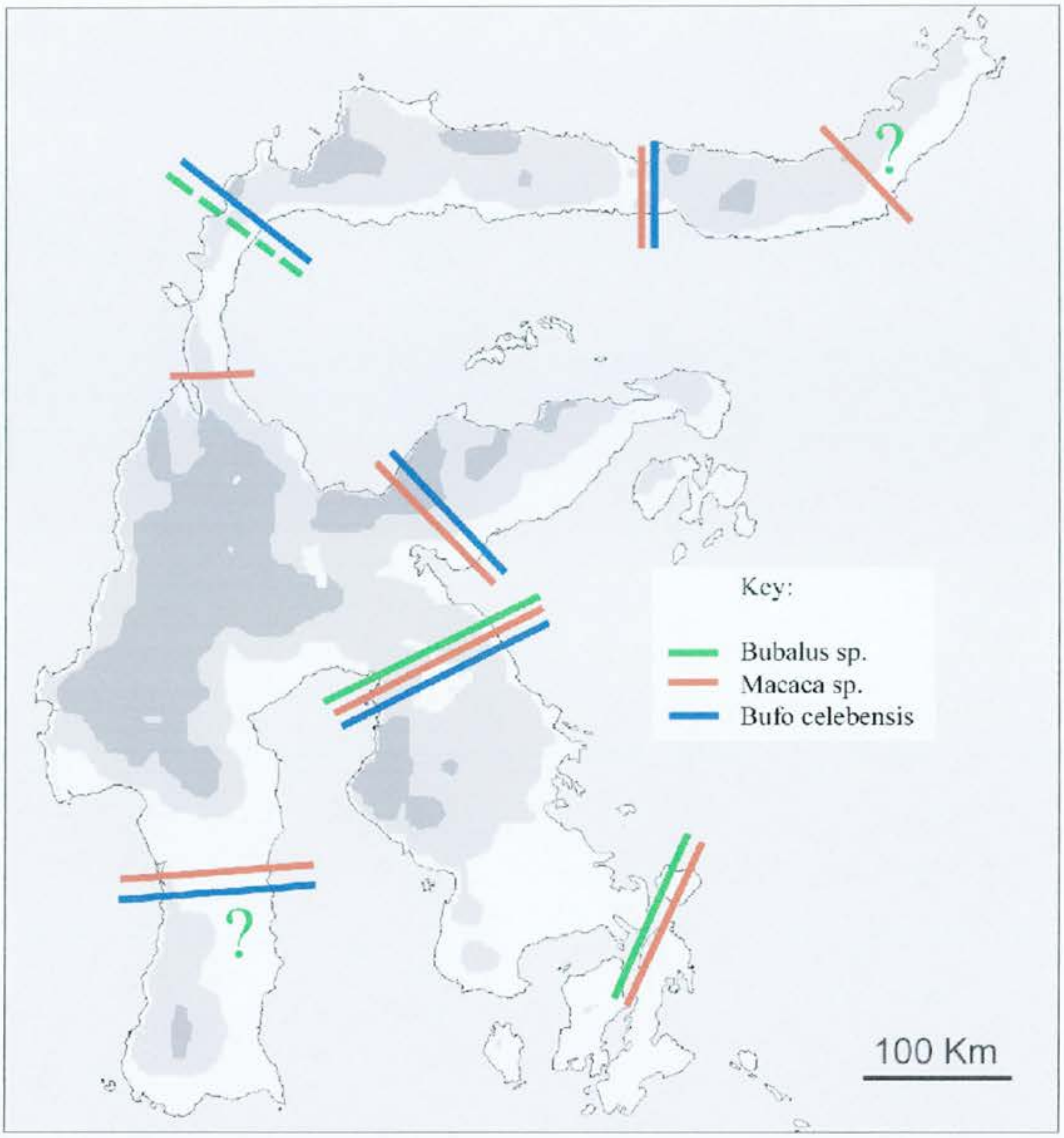


Figure 7.1. Biogeographic regions or 'Areas of Endemism' in Sulawesi and off shore islands for the macaque species and the toad *Bufo celebensis*, and population structure of the anoa according to microsatellite (only= dashed line) and mtDNA variation (both= solid line).
Shaded areas represent altitude ranges, light grey above 1000m asl, dark grey above 2000m asl. The Bubalus species data is from this study and the Macaca species and *Bufo celebensis* data are from Evans et al. (2003c). Question marks represent the ends of the island's peninsulas where anoa were very rare, making sampling not possible.

The biogeographic regions described above broadly correspond to the distribution of other species. This includes populations of grasshoppers (*Chitaura*) that were described as 'behaving like genetically isolated species' (Butlin et al., 1998). Also, the Sulawesi tarsiers (*Tarsius* sp.) have a pre-mating isolation mechanism in that individuals from the different regions have distinct duet calls (Nietsch, 1999), which vary throughout the island (Shekelle et al., 1997; Nietsch et al., in prep.). Where data was available these two groups of species have been shown to have a broad geographical separation similar to the anoa and macaques. Additionally, these morphologically smaller species were reportedly partitioned according to other phylogenetic barriers, such as mountaintop habitats or rivers (Shekelle et al., 1997; Butlin et al., 1998).

7.2.4 Sulawesi's biogeography and anoa evolution

The previously reported divergence in mtDNA and chromosome numbers suggested that for the anoa there were strong barriers to gene flow in the past which affected the population structure (Schreiber et al., 1993; Schreiber et al., 1999). Microsatellite data agrees with the existence of at least one strong barrier to gene flow, as described above. The variation of anoa mtDNA sequences defined a population structure similar to the microsatellite data. Isolation by distance was identified as affecting gene flow between anoa individuals for both anoa microsatellite and mtDNA data. The locations of one barrier to gene flow between anoa have been shown to coincide with certain boundaries of AOEs. The partitions that caused the AOEs have either been attributed to sea level variation (Fooden, 1969), or their cause is unknown (Evans et al., 2003c). They were not attributed to the sutures of Sulawesi's precursor islands (Evans et al., 2003c), one of the scenarios proposed by Schreiber et al. (1999). Habitat variation is less well understood, but temperature and rainfall variation in the Pleistocene era is likely to have replaced forest with savannah or seasonal forests in Sulawesi, as observed in predictions for south-east Asia (Whitmore, 1987; Whitten et al., 2002). It is uncertain if this would have had a role to play in anoa population fragmentation. The effect of the historical isolation might have resulted in the karyotypic variation reported in the zoo population (Schreiber et al., 1993). The patterns of variation in anoa nuclear DNA were interpreted as having been produced by multiple demographic phenomena.

In summary, genetic variation between geographical clusters of anoa is likely to have been caused by historical isolation, initially caused by oceanic barriers to gene flow (at least between Buton and South-East subpopulations), and combined with the effect of isolation by distance. The population as a whole can be described as having high microsatellite allele variation and high microsatellite differentiation between geographical areas.

7.3 Anoa taxonomic status

7.3.1 Relevance of the current putative two species status

Hypothesis testing methods confirmed that variation in microsatellite allele size and mtDNA sequences were correlated with geographic distance and biogeographic barriers. The partial Mantel tests found no evidence of a correlation between genetic variation and species morphological features or altitude of wild anoa. Therefore, the two species theory that the taxonomic status of the anoa was related to altitude and gross morphology (Groves, 1969) appears erroneous.

In this study the individual morphological characteristics (such as coat colour) that were used for defining the two species morphotypes were not analysed independently of each other. This was due to the small number of individuals for which these characteristics were recorded. However, certain characteristics may be representative of the subpopulations defined here, while others may not. For example, the white coat marking variation might be genetically correlated to the anoa subpopulations as has been observed in cattle breeds (Berryere, Schmutz, Schimpf et al., 2003). Coat colour has also been suggested as being related to age (Burton et al., 2005) and is known to vary widely within domestic cattle breeds (Casellas, Jimenez, Fina et al., 2004). It is feasible that the 'wooliness' or length of the anoa's coat hairs, which has been described as a distinctive feature for identification of the two putative anoa species (Groves, 1969) is related to the temperature gradient due to altitude. It seems unlikely that this is an ecological adaptation that has caused genetic variation because there is no known barrier to gene flow between different altitudinal ranges. Rather, it is more likely that hair length occurs on a cline through the altitudinal range. Therefore, this can be described as intra-species polymorphism and not a criterion for defining separate species. However, because of recent habitat disturbance any polymorphism caused by altitudinal variation is unlikely to remain in certain locations, as suggested by observation of both morphotypes at the opposite extremes of their altitude range (Burton et al., 2005). Other features such as the ridges or rings on horns, triangular cross-section of horns, and horn length might be related to age.

7.3.2 Interpretation of anoa taxonomic status

This section will relate the differences in microsatellite and mtDNA results of population structure (Section 6.2) and previous studies to species concepts with the aim of clarifying the taxonomic status of the anoa.

The widely employed biological species concept defines species as reproductively isolated units (Mayr, 1963) that can be defined using indirect data, such as differentiation of morphological features between populations. The early definitions of anoa taxonomy used this definition (Groves, 1969; Weise, 1979). They attempted to assess if intermediate morphological characteristics were present or absent between individuals from two morphotypic extremes, suggesting one or two distinct species of anoa, respectively. Often different criteria have been used to investigate this concept. Two studies relating to skull morphology of the anoa (Groves, 1969; Weise, 1979), produced different results that defined the existence of two and one species, respectively. Therefore, this skull morphology was not used in the current study. Additional reasons for omitting this approach include the circular argument within this concept described by Wallace (1865). His argument is that because theories of speciation involve the reduction in ability or tendency to interbreed, species cannot themselves be defined by interbreeding without confusing cause and effect. So this concept is essentially flawed. Another reason against using this concept is that hybridization compatibility has since been reported in wild anoa (Schreiber et al., 1999; Burton et al., 2005). Alternative species concepts were employed to test the hypothesis that multiple anoa species were geographically distributed.

A more appropriate concept for the analysis of the anoa dataset was the genotypic cluster definition (Mallet, 1995). The null hypothesis is that there are multiple independent clusters of alleles defining each species for single and multiple loci. However, the allele frequency distribution plots for anoa showed that genotypic clustering did not occur. No independent clusters in allele frequency plots were observed, except for the clusters of individuals at the opposite extremes of the anoa's range. This suggested limited differentiation between populations, either due to some admixture between previously distinct species, or between two populations of the same species. The division between anoa clusters (Figure 4.4) were less defined than that observed between two distinct bovid species (Mongolian yak and cattle in Figure 4.5). Although under the genotypic cluster concept the evidence supports the one species hypothesis for the anoa population, it was felt that multiple species comparisons should be conducted to verify this result.

The cohesion species concept employs lineage associations of mtDNA data to determine species status (Templeton, 1989). The mtDNA showed a single haplotype present on Buton Island (Table 5.1) and two haplogroups present on Sulawesi, all three groups were clustered phylogenetically and geographically. The only exceptions were caused by introgressions into other subpopulations. So there are three monophyletic groups shown in the NJ tree (Figure 5.1). This suggested the existence of multiple anoa species from mtDNA data, according to the cohesion species concept (Templeton, 1989).

Stasipatric speciation (White, 1978), or speciation due to chromosomal evolution is not always a reliable indicator of distinct species, especially in large bovid species. For example the river and swamp buffalo populations (*Bubalus arnee*) have distinct karyotypes but interbreed (Fischer, 1974). The Cape and red buffalo (*Synceros caffer caffer* and *S. c. nanus*) have different karyotypes and have heteromorphic karyotypes, suggesting crossbreeding at a secondary contact zone (Buckland and Evans, 1978). As pointed out by Schreiber et al. (1993) domestic species have impaired fertility, such as in taurine cattle, caused by a centric fusion (Long, 1988).

With the wider use of microsatellite markers further investigation into chromosome races has recently been carried out (Morgan-Richards, Trewick and Wallis, 2000; Andersson, Narain, Tegelstrom et al., 2004), but there remains a lack of understanding of the degree of reduction in fertility between races. Karyotypic variation should not be used as a criterion for defining species but may form an important part of the management design for the anoa.

In the absence of one clear concept for defining species many studies have resorted to the use of an indicator of distinctiveness. This is used to make comparisons of genetic distance between study populations and between clearly defined closely related species. The method is employed to delineate allopatric species using genetic markers. Three measures of microsatellite allele size variation were used here for comparison between other related species (Nei's, $\delta\mu^2$ genetic distance and F_{ST}). The resulting Nei's and F_{ST} values within anoa were found to be less than those observed between two different species (Mongolian yak and cattle). In contrast, certain $\delta\mu^2$ measures between pairs of anoa subpopulations were similar to the values between the two species (Mongolian yak and cattle). The three measures within anoa were less than those values between species in previously reported studies, as detailed in the Chapter 4 Discussion. For example, the values of Nei's genetic distance for the anoa were similar to those reported for intra-specific variation (Forbes et al., 1999). However, the contrast in results between the genetic distance measures presented in this thesis does not conclusively reject the null hypothesis of two species or confirm the single species hypothesis, according to the scale of genetic distance as an indicator of distinctiveness.

Previous studies found that differentiation of mtDNA between anoa individuals was similar to that found between individuals from different species (Tanaka et al., 1996; Kikkawa et al., 1997; Schreiber et al., 1999). However, analysis of allozyme data showed that the variation was comparable to that found within other bovid species (Schreiber et al., 1993). The geographic origins of the individuals used in the study were not known, and it is possible that they showed the variation at the extremes of the anoa's range, but did not include the linking populations between them. Schreiber et al. (1999) proposed that the difference between mtDNA and nDNA was related to the more philopatric behaviour of females than males (Schreiber, Klein and Lang, 1994; Tiedemann, Hardy, Vekemans et al., 2000), which had been observed in the anoa (Freer, 1994; Mustari, 1995). So there remains some uncertainty as to the taxonomic relationship between anoa populations, partly due to the differences in classification from the species concepts employed.

The management of the anoa population is a major reason for conducting the current study. There is an urgent requirement for management recommendations for both *in situ* and *ex situ* populations. The criteria of ESUs developed by Crandall et al. (2000) was employed to define units within the anoa population. They defined ESUs according to the degree of ecological and genetic exchangeability in recent or historical time, rather than a dichotomous definition. An ESU is a group of individuals that do not show ecological or genetic exchangeability. The data required for assessment under this set of criteria were not fully available from this study. Additional data was included from other studies and though there remained missing data, some interpretations were possible. In terms of genetic exchangeability it was observed from the existence of private alleles that recent genetic exchangeability was limited between the five clusters of individuals, either by isolation by distance or barriers to gene flow. This was shown by the microsatellite data, which has a rapid rate of change over time, making it most suitable for identifying differentiation over short time frames, compared to other techniques such as mtDNA. The mtDNA NJ tree showed three clades that were related to the geographic origin of individuals. This phylogenetic divergence was concordant with proposed historical geographic barriers as reported for the macaque species (Evans et al., 2003c). Recent genetic exchange clearly has not produced a genetically homogeneous population (PCA, STRUCTURE analysis, genetic distance measures, and a greater value for F_{ST} than F_{IS} values). Instead the genetic variation appears to have been affected by isolation by distance and barriers to gene flow. Visible physical features do not seem to be the barriers to gene flow, but this population structure may be a relic of isolation in the past. The relationship of karyotypic variation to the subpopulations is still unclear. Assignment analysis of zoo animals was unable to determine if this was related to geographical separation. This should be kept in mind when devising management strategies.

The ecological exchangeability can be assessed with the data available in the current study and information previously reported. Partial Mantel tests conducted in Chapter 6 did not find significant evidence of a correlation between genetic variation and species morphotype for both microsatellite and mtDNA data, when the effect of geographic distance was removed. Therefore, this contradicts the previous taxonomic status, which proposed two species of anoa related to skull morphometrics and gross body morphology (Groves, 1969). However, Weise (1979) found this to be simply related to a cline in size. His results should be interpreted cautiously, because *a priori* skull size and horn characteristics were used to identify the species prior to analysis meaning the methods were flawed (Burton et al., 2005). Though the anoa is polymorphic with respect to morphology, there has been no reported ecological adaptation by the different subpopulations on Sulawesi (Burton et al., 2005). This could be due to the lack of data. Additionally, there appears to be no habitat variation concordant with the barriers between the subpopulations that could have caused ecological adaptation by the anoa (Whitten et al., 2002).

The lack of a correlation between genetic variation and species morphotype suggests that there is no evidence to reject the null hypothesis of ecological exchangeability.

Therefore, even though the genetic exchangeability has been limited both historically and recently, there appears to be no regional ecological adaptation. So under the ESU criteria of Crandall et al. (2000) there is no evidence that would provide reason for separate management of these subpopulations. Instead, gene flow should be managed at the current level (before human induced habitat fragmentation occurred) to maintain the genetic diversity between the subpopulations. The results above should be supported by further work to assess ecological exchangeability before management actions are taken.

In summary, the interpretation of results in relation to the anoa's taxonomic status suggest a complex population structure of at least three mtDNA lineages, possibly with multiple cryptic chromosome races. These results also refute the existence of a correlation between genetic variation and morphological characteristics distributed according to altitude, which were central to the two species definition. The management of the different populations will be discussed below in Section 7.5. The species should remain *Bubalus depressicornis* and *B. quarlesi*, however it is expected further analyses may require this to be altered.

7.4 Significance of the *ex situ* anoa population

The individuals from the *ex situ* population of anoa were included in the study to determine the genetic representation in the zoo population in relation to the wild. The results of PCA found that the three European breed lines sampled (two 'depressicornis' and one 'quarlesi') were partially representative of the genetic variation of *in situ* animals. Anoa from two Indonesian zoos, Ragunan and Surabaya showed more genetic variation than in the European breed lines. The Indonesian conservation breeding population is therefore important for maintaining diversity in the *ex situ* population.

The assignment analysis with program STRUCTURE was less successful. The low level of assignment for all zoo individuals meant cautious interpretation was necessary (Pritchard et al., 2000). One possible explanation for these low values was that these individuals were of mixed origin, due to parental introgression between two subpopulations, in a zoo situation. Also these zoo individuals may have originated from localities that were not sampled during this study. The most likely explanation however, is the effect of inbreeding in the zoo population.

Management strategies of the *ex situ* population are in need of reassessment. As it was predicted that the three European zoo breed lines surveyed were only partially representative of the genetic variation of the *in situ* anoa population, it is important to include individuals from the Indonesian institutions in future breeding plans.

7.5 Conservation management implications

In this thesis the results I have presented require that changes be made to the current conservation plans for anoa on Sulawesi and Buton Islands. The affect on the current conservation measures of the finding that the anoa population structure is related to biogeographic areas is assessed below.

7.5.1 Population viability and maintenance of genetic diversity

A measure that is gaining more support in species conservation is the minimum viable population (MVP) of a species (Schaffer, 1981), defined as 'the smallest size required for a population or species to have a predetermined probability of persistence for a given length of time'. Reed, O'Grady, Brook et al. (2003) proposed that approximately 7000 adult vertebrate individuals should be preserved in order to ensure 99% probability of persistence for the 'long-term' or 40 generations. Numerous other multi-species estimates have been made (Franklin et al., 1998; Lynch et al., 1998) and all should only be used as a guide. The importance of this single measurement in estimating population viability has recently been proved; population size was found to be the best surrogate indicator (along with population trend) of extinction risk of a species (O'Grady et al., 2004).

The *in situ* anoa population is in rapid decline due to a combination of habitat degradation and hunting / human activity within the forested areas of Sulawesi and Buton Islands. Using program VORTEX it was predicted that all anoa populations would be extinct within 100 years with a 15% off-take hunting level per year as suggested by O'Brien et al. (1996); from the largest estimated population of 2860 individuals only 232 animals would remain with a 4% off-take (Manansang et al., 1996). Sulawesi's remaining protected forest covers an area of 1,823 km^2. This is 82% of the area under protection, which itself is 11.9% of the region (Aden et al., 2001). Considerable areas of unprotected forest are also being destroyed annually, as are areas within certain protected forests.

The geographic distribution of the four anoa subpopulations and the Atypical group are presented in Figure 7.2 together with recent data on the protected area status (UNDP/FAO, 1982; Anon., 1991; McCarthy, 1991) and forest cover of Sulawesi and Buton Islands (MacKinnon, 1997). There are considerable differences in the total land area covered by protected status within the areas of distribution of each of the anoa subpopulations. This imbalance was also reported for the proposed AOEs (Evans et al., 2003c). The Buton, Central and North subpopulations have the largest proportion of their range protected. A considerable number of the protected areas in the South-East are either of unsuitable habitat type for the anoa or are heavily hunted. Figure 7.2 also demonstrates the fragmented nature of the protected forest areas that could sustain anoa in the future. With limited natural gene flow between these protected areas they will become genetically distinct populations.

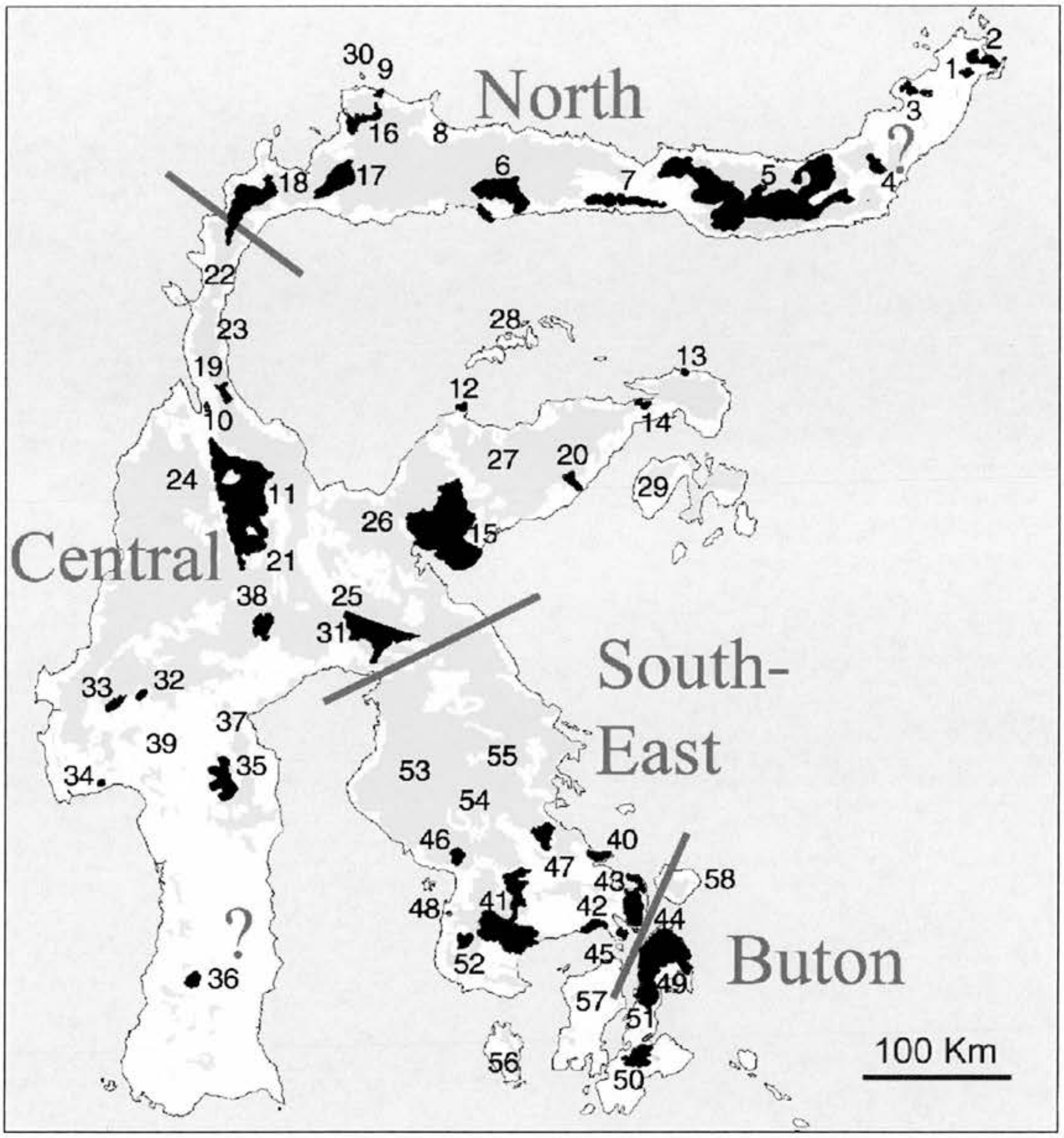


Figure 7.2. The location of the four anoa subpopulations and forest cover in Sulawesi and offshore islands. Protected areas are shaded in black. Forested areas are shaded in light grey according to surveys between 1989-1995 (MacKinnon, 1997).
Numbers relate to the forested regions of Sulawesi. See Appendix 1, Table 1 for the names of areas. Light shading indicates remaining forest.

The largest protected areas found within the range for each anoa subpopulation are listed in Table 7.1. This shows that the largest total population size (actual population size) in each subpopulation is smaller than the proposed MVP of 7000 animals (Reed et al., 2003). While the total population of the larger protected areas in each region is closer to the MVP. The largest estimated population in a protected area (in Bogani Nani-Wartabone National Park) is either 1608 or 3158 animals at densities of 0.56 or 1.1 animals / km^2. A single area of between 7700km^2 and 15400km^2 (current size 2871km^2) would be required to accommodate an MVP of 7000 animals at densities of 1.1 to 0.56 animals / km^2. These estimated densities represent the density under limited hunting threat and that of a population under hunting pressure (Mustari, 1995; O'Brien et al., 1996). Therefore, the protection of an MVP for anoa under the current area management structure is clearly not met for the population as one ESU defined in this study. To achieve this level of protection for the anoa, either larger protected areas are required or the movement of individuals between protected areas, to maintain gene flow is needed. This managed migration will introduce variation to maintain genetic variability and reduce possible inbreeding (Hedrick and Miller, 1992).

Table 7.1. Major Protected Areas in the five anoa subpopulations, with area (km^2) and estimated anoa population size (total population size).

Sub-population	Protected Areas (map location code on Figure 7.2)	Area (km^2)	Estimated population size	Notes
Buton	Buton Utara WR (49)	820	459 / 902	
	Lambusango WR (50)	250	140 / 275	
South-East	Tanjung Peropa NR (43)	389	218 / 427	
	Rawa Aopa-Watumohai NP (41)	968	-	Poor habitat
Central (to the east)	Lore Lindu NP (11)	2310	1294 / 2541	
	Morowali NR (15)	2250	1260 / 2475	
North	Panua NR (6)	940	526 / 1034	
	Bogani Nani-Wartabone NP (5)	2871	1608 / 3158	
Atypical				No protected areas defined as range uncertain
TOTAL (for largest protected areas)		9830	5505 / 10813	

Protected areas: NP - National Park; NR - Nature Reserve; WR - Wildlife Reserve.

The basic goal of conservation is to maintain genetic diversity in order to prevent species extinction (Frankham, 1995). In the anoa subpopulations defined here, the greatest allelic variation (Mean A and Mean *Ne*) was observed in the Central subpopulation (Table 4.4). A high level of diversity was also found in the mtDNA, with ten haplotypes in this subpopulation (n=10). Therefore, individuals in this area should be a high priority for conservation efforts, and also because they could act as a source population, connecting other areas (Pulliam, 1988).

7.5.2 Sulawesi phylogeography and regional conservation priority setting

The comparative analysis of co-distributed species can lead to more powerful interpretation of an area's biogeographic history (Avise, 1994). It can also allow assessment of genetic partitioning in relation to geographic separation, which is important for the management of regions (Moritz and Faith, 1998). The distribution of anoa subpopulations broadly confirms the Areas of Endemism (AOE) proposed to represent genetic endemism of the Sulawesi macaques and toads (Evans et al., 2003c) (Section 6.2.3). In the case of the anoa this may be due to a combination of isolation by distance and biogeograpic barriers. This agreement is a significant point, as it suggests that the divisions of AOEs may hold for other species, advocating the protection of a part of each of these areas as reserves for protection of endemism. Therefore, as stated by Evans et al. (2003c) a greater percentage of the total area of certain AOEs with a low proportion of protected habitat should be gazetted for wildlife preservation.

The distribution of anoa subpopulations and the AOEs of Sulawesi can be viewed in comparison with priority-setting analyses for conservation. The biogeographic provinces and biounits of the Dasmann-Udvardy system (MacKinnon et al., 1981; MacKinnon, 1997) represent the genetic and karyotypic differentiation far better than the ecoregions system (Wikramanayake, Dinerstein, Loucks et al., 2001). As observed by Jepson and Whittaker (2002) defining Sulawesi's lowland forest as one uniform region does not define the significance to the variation in biogeography and threat level of each peninsula. Therefore, the biounits of the Dasmann-Udvardy system are a better basis for conservation planning in this region and are likely to be beneficial to the anoa.

7.6 Genetic techniques as a conservation tool

Microsatellite markers have been employed in many ways to test conservation related hypotheses, from taxonomic uncertainties and population structure of wild populations, to implications for management of *ex situ* individuals (Goodman et al., 1999; Culver et al., 2000; Eizirik et al., 2001). The availability of a broad database of bovine microsatellite markers from closely related domestic species (European Cattle Diversity Database) greatly reduced the laboratory preparation time before their inclusion in endangered species studies (http://www.projects.roslin.ac.uk/cdiv/markers.html, January 2004).

The additional use of mtDNA added support to the conclusions drawn from the microsatellite data. For example, in determining the population structure both methods were broadly in agreement, with a discrepancy of microsatellite differentiating between North and Central individuals when the mtDNA did not.

The recent perturbations of the population caused by habitat fragmentation might have added to the difficulties of defining population structure (Goodman et al., 1999). Greater definition might have been produced using a larger set of variable microsatellite markers and larger sample size (Cornuet et al., 1999). The target at the outset of this study was 30 markers and 30 individuals per location. However, not surprisingly many of the bovine microsatellite markers failed when used in the anoa. The logistics of sample collection in Sulawesi made achieving the target of 30 samples, particularly in some regions, unrealistic given the resources available. Though these outcomes were planned for, a larger number of non-polymorphic markers than expected were found.

The use of microsatellite markers as a technique for resolving taxonomic uncertainties has proved effective in the case of the anoa. Analytical methods for hypothesis testing of population structure are currently a field of great development (Goodman et al., 1999; Beaumont et al., 2004). The use of Bayesian methodologies allowed greater insight into the levels of structure in the wild anoa population compared to traditional techniques.

7.7 Future anoa genetic research

There are five areas where future genetic surveys would assist in conservation of the anoa. The main goal is to determine the number of species of anoa, by the addition of further comparative analyses with closely related bovid species.

Having demonstrated that the use of microsatellite markers is a valuable tool for the identification of variation in the anoa, there is the need to complete a full assessment of the relative representation of the variation retained in *ex situ* locations and individuals' relationships to the *in situ* population. This is especially important for Antwerp / Plankendael Zoo and Namegawa Island Zoo (Chiba, Japan) whose breed-lines are not represented here. As the zoo populations will become more important as a conservation resource in future years, it is vital to identify the *in situ* genetic diversity that is not well represented in the world conservation breeding population. This should include all individuals kept in Indonesian collections, as these are unlikely to be represented outside Indonesia.

The difficulty of acquiring detailed morphological data related to DNA samples limited the analyses of these in relation to genetic variability. The zoo population of anoa could act as a suitable source of easily accessible DNA material and detailed morphological descriptions. Some areas of interest include the variation in overall coat colour and specifically the white pelage markings that have been observed (Dolan, 1965). The current lack of skeletal and DNA material from the same individuals has hampered the assessment of variation of skull morphology. The more systematic collection of skeletal and DNA material would be beneficial, as is currently being undertaken for the babirusa (*Babyrousa babyrussa*) (A. A. Macdonald, pers. comm.).

The phylogenetics of the genus Bubalus remains of interest as there are currently alternative hypotheses as to the phylogenetic position of the anoa (Groves, 1981; Pitra et al., 1997). This type of study is frequently conducted using mtDNA phylogenies, with microsatellites also having been used in the past (Ritz et al., 2000).

Site-specific sampling, relying on non-invasive, or non-capture techniques, such as dung or hair samples from the wild, would greatly increase the detail and accuracy of defining gene flow throughout Sulawesi and the location of introgression zones that exist. This would allow comparison to the AOEs described for other species above (Evans et al., 2003c), so assisting the planning of conservation priorities for the anoa. This could also encompass ecological studies, some areas of which have been paid recent attention (Mustari, 2004).

In conclusion, the use of microsatellite markers and mtDNA sequencing has proved to be a valuable conservation tool in this example. The genetic insight that has been reveale has clarified our understanding of the population structure of the anoa and identified *in situ* requirements for effective conservation.

7.8 Summary

As the largest mammal found in the Wallacea region, one of the 25 hotspots of high species endemism and high degree of threat (Myers et al., 2000), the anoa is an important species to use as a model for conservation of the region. The anoa should be managed on the islands of Sulawesi and Buton as a single ESU, maintaining the current level of gene flow. There is a requirement for active management to form a network of protected areas within each of the four subpopulations to maintain gene flow between the different fragmented groups. Further assessment of gene flow and the distribution of chromosome races are proposed to clarify the degree of fragmentation within and introgression between subpopulations.

The results of microsatellite and mtDNA analysis have provided a vital insight into the management requirements of this species. As pressure on the habitats of Sulawesi and the remaining wild anoa become more destructive (Burton et al., 2005), and the inbreeding of the anoa in zoological collections continues there is an urgent need to take informed management actions to maintain genetically representative and viable populations of the anoa based on applied research.

References

Aden, J., G. Dore, J. Vincent and T. Walton (2001) Indonesia: Environmental and the natural resource management in a time of transition, World Bank Report.

Agapow, P. M., O. R. P. Bininda-Emonds, K. A. Crandall, J. L. Gittleman, G. M. Mace, J. C. Marshall and A. Purvis (2004) The impact of species concept on biodiversity studies. *The Quarterly Review of Biology* **79**(2): 161-179.

Albrecht, G. H. (1978) The craniofacial morphology of the Sulawesi Macaques. *Contribution to Primatology (Basel Karger)* **13**: 1-152.

Amano, T. and H. Martojo (1983) Karyotypes of water buffaloes and anoas. *Report for Society of Research on Native Livestock* **10**: 98-110.

Amos, W. and A. R. Hoelzel (1991) Long-term preservation of whale skin for DNA analysis. In *Genetic ecology of whales and dolphins*. A. R. Hoelzel. Report for the International Whaling Commission (Special Issue, 13)**:** pp.99-104.

Anderson, M. J. and P. Legendre (1999) An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistics Computation and Simulation* **62**: 271-303.

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

Andersson, A. C., Y. Narain, H. Tegelstrom and K. Fredga (2004) No apparent reduction of gene flow in a hybrid zone between the West and North European karyotypic groups of the common shrew, *Sorex araneus*. *Molecular Ecology* **13**(5): 1205-1215.

Anon. (1991) Biodiversity Action Plan For Indonesia, Final Draft August, 1991, WWF-IP/PHPA, Jakarta, Indonesia.

Applied-Biosystems (1996a) Genescan 3.0.

Applied-Biosystems (1996b) Genotyper 2.0.

Applied-Biosystems (1999) ABI 3700.

Aranguren-Mendez, J., J. Jordana and M. Gomez (2002) Genetic conservation of five endangered Spanish donkey breeds. *Journal of Animal Breeding and Genetics* **119**(4): 256-263.

Arctander, P., C. Johansen and A.-M. Coutellec-Vreto (1999) Phylogeography of three closely related African bovids (Tribe Alcelaphini). *Molecular Biology and Evolution* **16**(12): 1724–1739.

Arctander, P., P. W. Kat, B. T. Simonsen and H. R. Siegismund (1996) Population genetics of Kenyan impalas: consequences for conservation. In *Molecular genetic approaches in conservation.* Eds: B. T. Smith and R. K. Wayne. Oxford, UK, Oxford University Press: pp.399-412.

Attardi, G. (1985) Animal mitochondrial DNA: an extreme example of genetic economy. *International Review of Cytology* **93**: 93-145.

Avalard, M. (1999) The impact of traditional subsistence hunting and trapping on prey populations: Data from Wana horticulturalists of upland Central Sulawesi, Indonesia. In *Hunting for Sustainability in Tropical Forests*. Eds: J. G. Robinson and E. L. Bennett. Columbia University Press, New York, USA: pp.214-230.

Avise, J. C. (1989) A Role for Molecular Genetics in the Recognition and Conservation of Endangered Species. *Trends in Ecology and Evolution* **4**(9): 279-281.

Avise, J. C. (1994) *Molecular markers, natural history and evolution*, Chapman & Hall.

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

Avise, J. C. and R. M. Ball (1990) Principles of geanological concordance in species concepts and biological taxonomy. *Oxford surveys in evolutionary biology* **7**: 45-67.

Avise, J. C. and G. C. Johns (1999) Proposal for a standardized temporal scheme of biological classification for extant species. *Proceedings of the National Academy of Sciences (USA)* **96**: 7358-7363.

Avise, J. C. and W. S. Nelson (1989) Molecular genetic relationships of the extinct dusky seaside sparrow. *Science* **243**: 646-648.

Avise, J. C. and K. Wollenberg (1997) Phylogenetics and the origin of species. *Proceedings of the National Academy of Sciences (USA)* **94**: 7748-7755.

Backmann, J. S. and J. L. Weber (1992) Survey of human and rat microsatellites. *Genomics* **12**(4): 627-631.

Baker, A. J. and S. R. Palumbi (1996) Population structure, molecular systematics and forensic identification of whales and dolphins. In *Conservation genetics: Case histories from nature*. Eds: J. C. Avise and J. L. Hamrick. Chapman and Hall, New York: pp.10-49.

Ballou, J. D. and R. C. Lacy (1995) Identifying genetically important individuals for management of genetic variation in pedigreed populations. In *Population management for survival and recovery: analytical methods and strategies in small population conservation*. Eds: J. D. Ballou, M. Gilpin and T. J. Foose. Columbia University Press.

Balloux, F., H. Brünner, N. Lugon-Moulin, J. Hausser and J. Goudet (2000) Microsatellites can be misleading: an empirical and simulation study. *Evolution* **54**: 1414-1422.

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

Barendse, W., D. Vaiman, S. Kemp, Y. Sugimoto, S. Armitage, J. Williams, S. Sun, A. Eggen, M. Agaba, A. Aleyasin, M. Band, M. Bishop, J. Buitkamp, K. Byrne, F. Collins, L. Cooper, W. Coupettiers, B. Denis, R. Drinkwater, K. Easterday, S. Ennis, G. Erhardt, L. Ferretti, Q. Gao, M. Georges, R. Gurung, Harlizius, G. Hawkins, J. Hetzel, T. Hirano, C. Joergensen, M. Kessler, B. Kirkpatrick, B. Konfortov, C. Kuhn, H. Lenstra, H. Leveziel, H. Lewin, B. Leyhe, L. Li, I. Martin Burriel, R. McGraw, R. Miller, D. Moody, S. Moore, S. Nakane, I. Nijman, I. Olsaker, D. Pomp, A. Rando, M. Ron, A. Shalom, M. Soller, A. Teale, U. Thieven, D. Vage, S. Varvio, R. Velmalla, J. Vilkki, R. Weikard, C. Woodside, J. Womack, M. Zanotti and Zaragoza (1997) A medium-density genetic-linkage map of the Bovine Genome. *Mammalian Genome* **8**: 21-28.

Barker, J. S. F., S. S. Moore, D. J. S. Hetzel, D. Evans, S. G. Tan and K. Byrne (1997) Genetic diversity of Asian water buffalo (*Bubalus bubalus*): Microsatellite variation and a comparison with protein-coding loci. *Animal Genetics* **28**: 103-115.

Barton, N. H. and G. M. Hewitt (1989) Adaptation, speciation and hybrid zones. *Nature* **341**: 497-503.

Baverstock, P. R. and C. Moritz (1996) Project design. In *Molecular Systematics*. Eds: D. M. Hillis, C. Moritz and B. K. Mable. Sunderland, Massachusetts, Sinauer Associates Inc.

Beaumont, M. and B. Rannala (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics* **5**: 251-261.

Benirschke, K. and A. T. Kumamoto (1991) Mammalian cytogenetics and conservation of species. *Journal of Hereditary* **82**: 187-191.

Berryere, T. G., S. M. Schmutz, R. J. Schimpf, C. M. Cowan and J. Potter (2003) TYRP1 is associated with dun coat colour in Dexter cattle or how now brown cow? *Animal Genetics* **34**(3): 169-175.

Blott, S. C., J. L. Williams and C. S. Haley (1998) Genetic relationships among European cattle breeds. *Animal Genetics* **29**: 273-282.

Bohlken, H. (1958) Vergleichende untersuchengen an wildridern (Tribus Bovini Simpson, 1945). *Zoologische Jahrbuecher (Phys.)* **68**: 113-220.

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

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

Buckland, R. A. and H. J. Evans (1978) Cytogenetic aspects of phylogeny in the Bovidae I. G-banding. II. C-banding. *Cytogenetic Cellular Genetics* **21**: 42-71.

Burton, J. A., S. Hedges and A. H. Mustari (2005) The taxonomic status, distribution and conservation needs of the lowland anoa *Bubalus depressicornis* and mountain anoa *Bubalus quarlesi*. *Mammal Review* **35**(1): 25-50.

Butlin, R. K., C. Walton, K. A. Monk and J. R. Bridle (1998) Biogeography of Sulawesi grasshoppers, genus Chitaura, using DNA sequence data. In *Biogeography and Geological Evolution of SE Asia*. Eds: R. Hall and J. D. Holloway. Backhuys, Leiden, Netherlands: pp.355-359.

Bynum, E. L. (1999) Biogeography and evolution of Sulawesi macaques. *Tropical Biodiversity* **6**(1&2): 19-36.

Canon, J., P. Alexandrino, I. Bessa, C. Carloes, Y. Carretero, S. Dunner, N. Ferran, D. Garcia, J. Jordana, D. Laloa, A. Pereira, A. Sanchez and K. Moazami-Goudarzi (2001) Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genetics, Selection, Evolution* **33**: 311-332.

Casellas, J., N. Jimenez, M. Fina, J. Tarres, A. Sanchez and J. Piedrafita (2004) Genetic diversity measures of the bovine Alberes breed using microsatellites: variability among herds and types of coat colour. *Journal of Animal Breeding and Genetics* **121**(2): 101-110.

Castellano, S. and E. Balletto (2002) Is the partial Mantel test inadequate? *Evolution* **56**(9): 1871-1873.

Cavalli-Sforza, L. L., P. Menozzi and A. Piazza (1994) *The history and geography of human genes*, Princeton University Press.

Cegelski, C. C., L. P. Waits and N. J. Anderson (2003) Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. *Molecular Ecology* **12**(11): 2907-2918.

Chakraborty, R. (1986) Gene admixture in human populations: models and predictions. *Yearbook of Physical Anthropology* **29**: 1-43.

Charlesworth, B. (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* **15**: 538-543.

Ciani, A. C., R. Stanyon, W. Scheffrahn and B. Sampurno (1989) Evidence of gene flow between Sulawesi macaques. *American Journal of Primatology* **17**(4): 257-270.

Ciofi, C., M. A. Beaumont, I. R. Swingland and M. W. Bruford (1999a) Genetic divergence and units for conservation in the Komodo dragon *Varanus komodoensis*. *Proceedings of the Royal Society of London, Series B* **266**: 2269-2274.

Ciofi, C. and M. W. Bruford (1999b) Genetic structure and gene flow among Komodo dragon populations inferred by microsatellite loci analysis. *Molecular Ecology* **8**(12 Suppl. 1): S17-S30.

Claridge, M. F., H. A. Dawah and M. R. Wilson (1997) Practical approaches to species concepts for living organisms. In *Species: The units of biodiversity*. Eds: M. F. Claridge, H. A. Dawah and M. R. Wilson. London, Chapman and Hall.

Clayton, L. and E. J. Milner-Gulland (1999) The Trade in Wildlife in North Sulawesi, Indonesia. In *Hunting for Sustainability in Tropical Rainforests*. E. L. Bennett. Columbia University Press, New York, USA**:** pp.473-496.

Clayton, L. M., M. J. Keeling and E. J. Milner-Gulland (1997) Bringing home the bacon: A spatial model of the wild pig hunting in Sulawesi, Indonesia. *Ecological Applications* **7**(2): 642-652.

Cockerham, C. C. and B. S. Weir (1984) Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics* **40**: 157-164.

Cohan, F. M. (1994) Genetic exchange and evolutionary divergence in prokaryotes. *Trends in Ecology and Evolution* **9**: 175-180.

Coote, T. and M. W. Bruford (1996) A set of human microsatellites amplify polymorphic markers in Old World apes and monkeys. *Journal of Hereditary* **87**: 406-410.

Corbert, G. B. (1997) The species in mammals. In *Species: The units of biodiversity*. Eds: M. F. Claridge, H. A. Dawah and M. R. Wilson. Chapman & Hall, UK**:** pp.341-356.

Corbet, G. B. and J. E. Hill (1992) *The mammals of the Indomalayan Region: a systematic review*, Natural History Museum Publications and Oxford University Press, Oxford, UK.

Cornuet, J.-M., S. Piry, G. Luikart, A. Estoup and M. Solignac (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**: 1989-2000.

Coulon, A., J. F. Cosson, J. M. Angibault, B. Cargnelutti, M. Galan, N. Morellet, E. Petit, S. Aulagnier and A. J. M. Hewison (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. *Molecular Ecology* **13**(9): 2841-2850.

Cracraft, J. (1983) Species concepts and speciation analysis. *Current ornithology* **1**: 159-187.

Cracraft, J. (1989) Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In *Speciation and its consequences*. Eds: D. Otte and J. A. Endler. Sinauer Associates, MA, USA**:** pp.28-59.

Crandall, K. A., O. R. P. Bininda-Emonds, G. M. Mace and R. K. Wayne (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* **15**: 290-295.

Culver, M., W. E. Johnson, J. Pecon-Slattery and S. J. O'Brien (2000) Genetic ancestry of the American Puma (*Puma concolor*). *Journal of Heredity* **91**: 186-197.

Cunningham, E. P., J. J. Dooley, R. K. Splan and D. G. Bradley (2001) Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Animal Genetics* **32**(6): 360-364.

Custodio, C. C., M. V. Lepiten and L. R. Heaney (1996) *Bubalus mindorensis*. *Mammalian Species* **520**: 1-5.

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

De Siquera, M. F., A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. Van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips and S. E. Williams (2004) Extinction risk from climate change. *Nature* **427**: 145-148.

Di Rienzo, A., A. C. Peterson, A. C. Garza, A. M. Valdes, M. Slatkin and F. N. B. (1994) Mutational processes of simple-sequences repeat loci in human populations. *Proceedings of the National Academy of Sciences (USA)* **91**: 3166-70.

Diez-Tascon, C., R. P. Littlejohn, P. A. R. Almeida and A. M. Crawford (2000) Genetic variation within the Merino sheep breed: analysis of closely related populations using microsatellites. *Animal Genetics* **31**: 243-251.

Dizon, A. E., C. Lockyer, W. F. Perrin, D. M. Demaster and J. Sisson (1992) Rethinking the stock concept: A phylogenetic approach. *Conservation Biology* **6**(24-36).

Dolan, J. M. (1965) Breeding of the Lowland Anoa, *Bubalus (Anoa) d. depressicornis* (H. Smith, 1927 in the San Diego Zoological Garden. *Zeitschrift Zoologische Saugetierkunde* **30**(4): 241-248.

Dowling, T. E., C. Moritz, J. D. Palmer and L. H. Rieseberg (1996) Nuclei acids III: analysis of fragments and restriction sites. In *Molecular Systematics*. Eds: D. M. Hillis, C. Moritz and B. K. Mable. Sinauer Associates, Sutherland, USA: pp.249-320.

Eizirik, E., J. H. Kim, M. Menotti-Raymond, P. G. Crawshaw, S. J. O'Brien and W. E. Johnson (2001) Phylogeography, population history and conservation genetics of jaguars (*Panthera onca*, Mammalia, Felidae). *Molecular Ecology* **10**: 65-79.

Eldridge, M. D. B., J. M. King, A. K. Loupis, P. B. S. Spencer, A. C. Taylor and L. C. Pope (1999) Unprecedented low levels of genetic variation and inbreeding depression in an island population of the black-footed rock-wallaby. *Conservation Biology* **13**: 531-541.

Engel, S. R., R. A. Linn, J. F. Taylor and S. K. Davis (1996) Conservation of microsatellite loci across species of artiodactyls: Implications for population studies. *Journal of Mammalogy* **77**(2): 504-518.

Estoup, A., L. Garnery, M. Solignac and J.-M. Cornuett (1995) Microsatellite variation in honey bee (*Apis mellifera*) populations: Hierarchical genetic structure and test of the Infinite Allele and Stepwise Mutation Model. *Genetics* **140**: 679-695.

Evans, B. J., R. M. Brown, J. A. Mcguire, J. Supriatna, N. Andayani, A. Diesmos, D. Iskandar, D. J. Melnick and D. C. Cannatella (2003a) Phylogenetics of Fanged Frogs: Testing Biogeographical Hypotheses at the Interface of the Asian and Australian Faunal Zones. *Systematic Biology* **52**(6): 794–819.

Evans, B. J., J. C. Morales, J. Supriatna and D. J. Melnick (1999) Origin of the Sulawesi macaques (Cercopithecidae: Macaca) as suggested by mitochondrial DNA phylogeny. *Biological Journal of the Linnean Society* **66**(4): 539-560.

Evans, B. J., J. Supriatna, N. Andayani and D. J. Melnick (2003b) Diversification of Sulawesi macaque monkeys: Decoupled evolution of mitochondrial and autosomal DNA. *Evolution* **57**: 1931–1946.

Evans, B. J., J. Supriatna, N. Andayani, M. I. Setiadi, D. J. Melnick and D. C. Cannatella (2003c) Monkeys and toads define areas of endemism on Sulawesi. *Evolution* **57**: 1436–1443.

Evans, B. J., J. Supriatna and D. J. Melnick (2001) Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. *Evolution* **55**: 1685–1702.

Everitt, B. S. and G. Dunn (1991) *Applied Multivariate Data Analysis*, Arnold.

Felsenstein, J. (1993) PHYLIP 3.5. Seattle, Washington University.

Fischer, H. (1974) Cytogenetic observation on crossbreeds between swamp and murrah buffaloes. *Zuchthygiene* **9**: 105-110.

Fooden, J. (1969) *Taxonomy and evolution of the monkeys of Celebes. (Primates: Cercopithecidae)*, Basel: S. Karger (Bibliotheca Primatologica, No. 10).

Forbes, S. H. and J. T. Hogg (1999) Assessing population structure at high levels of differentiation: microsatellite comparisons of bighorn sheep and large carnivores. *Animal Conservation* **2**: 223-233.

Frankham, R. (1995) Inbreeding and extinction: A threshold effect. *Conservation Biology* **12**: 665-675.

Frankham, R., J. D. Ballou and D. A. Briscoe (2002) *Conservation Genetics*, Cambridge University Press.

Franklin, I. R. and R. Frankham (1998) How large must populations be to retain evolutionary potential. *Animal Conservation* **1**: 69-73.

Freedman, D. and D. Lane (1983) A nonstochastic interpretation of reported significance
levels. *Journail of Business and Economics Statistics* **1**: 292-298.

Freer, F. (1994) Observations ethologigues sur *Bubalus [Anoa] quarlesi* Owens, 1910 (Ruminantia, Bovidae) en captivite. *Zeitschrift fur Saugetierkunde* **59**: 139-154.

Gaggiotti, O. E., O. Lange, K. Rassmann and C. Gliddons (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology* **8**: 1513–1520.

Gallagher, D. S. and J. E. Womack (1992) Chromosome conservation in the Bovidae. *Journal of Hereditary* **83**: 287-298.

Gatesy, J., D. Yelon, R. Desale and E. Vrba (1992) Phylogeny of the Bovidae (Artiodactyla, Mammalia), based on Mitochondrial Ribosomal DNA sequences. *Molecular Biological Evolution* **9**(3): 433-446.

Geraads, D. (1992) Phylogenetic analysis of the tribe Bovini (Mammalia: Artiodactyla). *Zoological Journal of the Linnaean Society* **104**: 193-207.

Gilpin, M. (1991) The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society* **42**: 165-176.

Gilpin, M. and M. E. Soulé (1986) Minimum viable populations: processes of species extinctions. In *Conservation biology: the science of scarcity and diversity*. M. E. Soulé. Sinauer, Sutherland, MA**:** pp.19-34.

Godfray, H. C. J. (2002) Challenges for taxonomy. *Nature* **417**: 17-19.

Goldstein, D. B., R. Desalle, G. Amato and A. P. Vogler (2000) Conservation genetics at the species boundary. *Conservation Biology* **14**(1): 120-131.

Goldstein, D. B. and D. D. Pollock (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. *Journal of Heredity* **88**: 335-342.

Goldstein, D. B., A. Ruiz-Linares and L. L. Cavalli-Sforza (1995a) An evaluation of genetic distance for use with microsatellite loci. *Genetics* **139**: 463-471.

Goldstein, D. B., A. Ruiz-Linares, L. L. Cavalli-Sforza and M. W. Feldman (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Science of the USA* **92**: 6723-6727.

Goldstein, D. B. and C. Schlötterer (1999) *Microsatellites: evolution and applications*, Oxford University Press, New York.

Goodman, S. J. (1997) RST Calc: A collection of computer-programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* **6**: 881-885.

Goodman, S. J., N. H. Barton, G. Swanson, K. Abernethy and J. M. Pemberton (1999) Introgression through rare hybridization: A genetic study of a hybrid zone between red and sika deer (genus Cervus), in Argyll, Scotland. *Genetics* **152**: 355-371.

Goodman, S. J., H. B. Tamate, R. Wilson, J. Nagata, S. Tatsuzawa, G. Swanson, J. M. Pemberton and D. R. McCullough (2001) Bottlenecks, drift and differentiation: The population structure and demographic history of sika deer (*Cervus nippon*) in the Japanese archipelago. *Molecular Ecology* **10**: 1357–1370.

Goudet, J. (1995) FSTAT Version 1.2: a computer program to calculate F-statistics. *Journal of Heredity* **86**: 485–486.

Gower, J. C. and D. J. Hand (1996) *Biplots*, Chapman & Hall/CRC.

Groombridge, J. J., G. C. Jones, M. W. Bruford and R. A. Nichols (2000) 'Ghost' alleles of the Mauritius kestrel. *Nature* **403**: 616.

Groves, C. P. (1969) Systematics of the anoa (Mammalia, Bovidae). *Beaufortia* **17**: 1-12.

Groves, C. P. (1981) Systematic relationships in the Bovini (Artiodactyla, Bovidae). *Zeitschrift Zoologische Systematik Evolution-forschung* **19**: 264-278.

Groves, C. P. (1998) Systematics of the tarsier and lorises. *Primates* **39**: 13-27.

Groves, C. P. (2001) Mammals in Sulawesi: where did they come from and when, and what happened to them when they got there? In *Faunal migrations and evolution in SE Asia-Australasia*. I. S. Metcalfe, J. M. B., Morwood, M. & Davidson, I., A.A. Balkema / Lisse / Abingdon / Exton (PA) / Tokyo: pp.333-342.

Guo, S. W. and E. A. Thompson (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* **48**: 361-372.

Hall, R. (1996) Reconstructing Cenozoic SE Asia. In *Tectonic Evolution of Southeast Asia, Special Publication 106*. R. Hall, Blundell, D., London: Geological Society: pp.153-184.

Hall, R. (1998) The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. In *Biogeography and geological evolution of SE Asia*. Eds: R. Hall and J. D. Holloway. Backhuys: Leiden: pp.99-131.

Hamilton Smith, C. (1827) Arranged in Conformity with its Organization, by the Baron Cuvier, with additional descriptions by Edward Griffith and others, Vol. 4. In *The Animal Kingdom*. G. B. Whittaker. London, UK.

Harper, F. (1945) Extinct and vanishing mammals of the Old World. *American Committee for International Wild Life Protection, Special Publication*, New York Zoological Park, New York, USA. **12**.

Hartl, D. L. and A. G. Clark (1997) *Principles of Population Genetics*, Sinauer Associates, Inc, Sunderland, MA.

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

Hedges, S. (1998) Asian water buffalo *Bubalus arnee*: Species 30, Asian Wild Cattle Specialist Group.

Hedin, M. and D. A. Wood (2002) Genealogical exclusivity in geographically proximate populations of *Hyplochilus thorelli Marx* (Araneae, Hypochilidae) on the Cumberland Plateau of North America. *Molecular Ecology* **11**: 1975-1988.

Hedrick, P. W. (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**(2): 313-318.

Hedrick, P. W. and P. S. Miller (1992) Conservation genetics: Techniques and fundamentals. *Ecological Applications* **2**(1): 30-46.

Heller, K. M. (1889) Der Urbuffel von Celebes, *Anoa depressicornis* (H. Smith). *Versuch einer Monographie*. Dresden: H. Grunberg.

Henderson, S. T. and T. D. Petes (1992) Instability in simple sequence DNA in *Saccharomyces cerevisiae*. *Molecular Cell Biology* **12**: 2749-2757.

Higuchi, R., C. H. von Beroldingen, G. F. Sensabaugh and H. A. Erlich (1988) DNA typing from single hairs. *Nature* **332**: 543-546.

Holder, M. and P. O. Lewis (2003) Phylogeny estimation: Traditional and Bayesian approaches. *Nature Review Genetics* **4**: 275-284.

Holloway, D. J. (1998) Geological signal and dispersal noise in two contrasting insect groups in the Indo-Australian tropics: R-mode analysis of pattern in Lepidoptera and cicadas. In *Biogeography and geological evolution of SE Asia*. Eds: R. Hall and J. D. Holloway. Backhuys: Leiden**:** pp.291-313.

Holmes, P. R. and H. J. Holmes (1985) Notes on Zostrops spp. From the lake Matano area of South-East Sulawesi, Indonesia. *Bulletin of the British Ornithological Club* **105**: 136-140.

Hooijer, D. A. (1950) Man and other mammals from Toalian sites in southwestern Celebes. *Verhandelingen der Koninklijke Nederlandsche Akademie Wetenschappen: Afdeeling Natuurkunde* **46**(2): 131-144.

Hoss, M., M. Kohn, S. Pääbo, F. Knauer and W. Schröder (1992) Excrement analysis by PCR. *Nature* **359**: 199.

http://dad.fao.org/en/refer/library/guidelin/marker.pdf (August 2004) Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans. Measurement of Domestic Animal Diversity (MoDAD): Recommended Microsatellite Markers.

http://www.iucn.org/redlist/2000/news.html (September 2004) Confirming the Global Extinction Crisis: A call for international action as the most authoritative global assessment of species loss is released, The World Conservation Union.

http://www.projects.roslin.ac.uk/cdiv/markers.html (January 2004) European Cattle Diversity Database.

http://www.redlist.org/ (August 2004) The 2003 IUCN Red List of Threatened Species: *Bubalus quarlesi* and *Bubalus depressicornis*, The World Conservation Union.

Hudson, R. (1990) Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology* **7**: 1-44.

Humphries, C. J., P. H. Williams and R. I. Vane-Wright (1995) Measuring biodiversity value for conservation. *Annual Review of Ecology and Systematics* **26**: 93-111.

Ikeobi, C. O. N., J. A. Woolliams, D. R. Morrice, A. Law, D. Windsor, D. W. Burt and P. M. Hocking (2002) Quantitative trait loci affecting fatness in the chicken. *Animal Genetics* **33**(6): 428-435.

Jahja, M. M. (1987) The possibility of breeding anoa in captivity: an alternative for conservation of the species. *BIOTROP, Special Publication* **30**: 101-108.

Jarman, S. N. and N. G. Elliott (2000) DNA evidence for morphological and cryptic Cenozoic speciations in the Anaspididae, 'living fossils' from the Triassic. *Journal of Evolutionary Biology* **13**: 624-633.

Jarne, P. and P. J. L. Lagoda (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**(10): 424-429.

Jeffreys, A. J., V. Wilson and S. L. Thein (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**: 67-73.

Jenkins, M. (2001) Prospects for biodiversity. *Science* **302**: 1175-1177.

Jepson, P., J. K. Jarvie, J. MacKinnon and K. A. Monk (2001) The end of Indonesia's lowland rainforest. *Science* **292**: 859-861.

Jepson, P. and R. J. Whittaker (2002) Ecoregions in context: a critique with special reference to Indonesia. *Conservation Biology* **16**(1): 42-57.

Jones, K. L., T. C. Glenn, R. C. Lacy, J. R. Pierce, N. Unruh, C. M. Mirande and F. Chavez-Ramirez (2001) Refining the Whooping Crane Studbook by incorporating microsatellite DNA and leg-banding analysis. *Conservation Biology* **16**(3): 789-799.

Kakoi, H., T. Namikawa, O. Takenaka, A. Takenaka, T. Amano and H. Martojo (1994) Divergence between the anoas of Sulawesi and the Asiatic water buffaloes, inferred from their complete amino acid sequences of hemoglobin ß chains. *Zeitschrift für Zoologische Systematik und Evolutionsforschung* **32**(1): 1-10.

Kikkawa, Y., H. Yonekawa, H. Suzuki and T. Amano (1997) Analysis of genetic diversity of domestic water buffaloes and anoas based on variations in the mitochondrial gene for cytochrome b. *Animal Genetics* **56**: 21-24.

Kimura, M. and T. Otha (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences (USA)* **75**: 2868-2872.

Kimura, Y. and J. F. Crow (1964) The number of alleles that can be maintained in a finite population. *Genetics* **143**: 725-738.

Knoll, S. and M. Rowell-Rahier (1998) Distribution of genetic variance and isolation by distance in two leaf beetle species: *Oreina cacaliae* and *Oreina speciosissima*. *Heredity* **81**: 412-421.

Knowles, L. L. (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology* **17**(1): 1-10.

Knowles, L. L. and W. P. Maddison (2002) Statistical phylogeography. *Molecular Ecology* **11**: 2623–2635.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca and A. C. Wilson (1989) Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Science (USA)* **86**: 6196-6200.

Koulischer, L., J. Tyskens and J. Mortelsmanns (1972) Mammalian cytogenetics 6. The chromosomes of a male specimen of Anoa depressicornis quarlesi. *Acta Zoologica et Pathologica, Antverpiensia No.* **56**: 21-24.

Krzanowski, W. J. (1988) *Principles of Multivariate Analysis*, Clarendon Press.

Kumar, P., A. R. Freeman, R. T. Loftus, C. Gaillard, D. Q. Fuller and D. G. Bradley (2003) Admixture analysis of South Asian cattle. *Heredity* **91**: 43-50.

Lampert, K. P., A. S. Rand, U. G. Mueller and M. Ryan (2003) Fine-scale genetic pattern and evidence for sex-biased dispersal in the túngara frog, *Physalaemus pustulosus*. *Molecular Ecology* **12**: 3325–3334.

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

Laval, G., N. Iannuccelli, C. Legault, D. Milan, M. Groenen, E. Guiffra, L. Anderson, P. Nissen, C. Jorgensen, P. Beeckmann, H. Geldermann, J. L. Foulley, C. Chevalet and L. Olliver (2000) Genetic diversity of eleven European pig breeds. *Genetics Selection Evolution*: 187-203.

Lawes-Agricultural-Trust (2003) Program GENSTAT Version 7.1.

Leakey, R. and R. Lewin (1995) *The sixth extinction: Biodiversity and its survival*. London, Phoenix.

Legendre, P. and L. Legendre (1998) *Numerical Ecology*, Elsevier.

Levison, G. and G. A. Gutman (1987) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 *Escherichia coli* K-12. *Nucleaic Acid Research* **15**: 5323-5338.

Linnaeus, C. (1753) *Species Plantarum Laur*, Salvius.

Linnaeus, C. (1758) *Systema Naturae, Editio decima Laur*, Salvius.

Litt, M. and J. A. Luty (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* **44**(3): 397-401.

Long, S. E. (1988) Segregation patterns and fertility of domestic mammals with chromosome translocations. In *The cytogenetic of mammalian autosomal rearrangements*. A. Daniel. Alan Liss, New York**:** pp.383-396.

Lugon-Moulin, N., H. Brunner, F. Balloux, J. Hausser and J. Goudet (1999) Do riverine barriers, history or introgression shape the genetic structuring of a common shrew (*Sorex araneus*) population? *Heredity* **83**: 155-161.

Lugon-Moulin, N. and J. Hausser (2002) Phylogeographical structure, postglacial recolonization and barriers to gene flow in the distinctive Valais chromosome race of the common shrew (*Sorex araneus*). *Molecular Ecology* **11**(4): 785-794.

Luikart, G. and P. R. England (1999) Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution* **14**: 253-255.

Lynch, M. and R. Lande (1998) The critical effective size of a genetically secure population. *Animal Conservation* **1**: 70-71.

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

MacHugh, D. E., R. T. Loftus, D. G. Bradley, P. M. Sharp and P. Cunningham (1994) Microsatellite DNA variation within and among European cattle breeds. *Proceedings of the Royal Society of London Series B-Biological Sciences* **256**(1345): 25-31.

MacHugh, D. E., R. T. Loftus, P. Cunningham and D. G. Bradley (1998) Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Animal Genetics* **29**(5): 333-340.

MacHugh, D. E., M. D. Shriver, R. T. Loftus, P. Cunningham and D. G. Bradley (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146**(3): 1071-86.

MacKinnon, J. R. (1997) *Protected Areas System Review of the Indo-Malayan Realm*, The Asian Bureau for Conservation Limited, Canterbury, UK.

MacKinnon, J. R. and J. Wind (1981) *Birds of Indonesia. Special report FO/INS/78/061*, Food and Agriculture Organisation of the United Nations, Rome.

MacKinnon, K., G. Hatta, H. Halim and A. Mangalik (1996) *The ecology of Kalimantan*, Periplus Editions, Hong Kong.

Mallet, J. (1995) A Species Definition for the Modern Synthesis. *Trends in Ecology and Evolution* **10**(7): 294-299.

Manansang, J., S. Hedges, S. Dwiatmo, P. Miller and U. S. Seal (1996) *Population and habitat viability assessment workshop for the anoa (Bubalus depressicornis and Bubalus quarlesi) report*. Population and Habitat Viability Assessment Workshop for the Anoa (*Bubalus depressicornis* and *Bubalus quarlesi*), Bogor Indonesia, IUCN/SSC Conservation Breeding Specialist Group, Apple Valley, MN., USA.

Manel, S., P. Berthier and G. Luikart (2002) Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Conservation Biology* **16**: 650–659.

Manly, B. F. J. (1994) *Multivariate Statistical Methods: A Primer*, Chapman and Hall.

Manly, B. F. J. (1997) *Randomization, bootstrap and Monte Carlo methods in biology*. London, Chapman and Hall.

Mannen, H., S. Tsuji, R. T. Loftus and D. G. Bradley (1998) Mitochondrial DNA Variation and Evolution of Japanese Black Cattle (Bos taurus). *Genetics* **150**: 1169–1175.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209–220.

Marsono, N., M. Utami, W. E. Setiowati, T. S. Puntoprini, D. P. Astuti, H. Andayani and Vega (2001) Genetic management of lowland anoa and mountain anoa at Ragunan Zoo, South-East Asian Zoo Association.

Maudet, C., G. Luikart and P. Taberlet (2001) Development of microsatellite multiplexes for wild goats using primers designed from domestic Bovidae. *Genetics Selection and Evolution* **33**: 5193-5204.

Maudet, C., G. Luikart and P. Taberlet (2002) Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. *Journal of Animal Science* **80**(4): 942-950.

Mayden, R. L. (1997) A hierarchy of species concepts: the denouncement in the saga of the species problem. In *Species: The units of biodiversity*. Eds: M. F. Claridge, H. A. Dawah and M. R. Wilson. Chapman & Hall, UK**:** pp.381-424.

Mayr, E. (1963) *Animal species and evolution*, Harvard University Press, Cambridge.

Mayr, E. (1969) *Principles of systematic zoology*, McGraw Hill, New York.

Mayr, E. (1991) *Principles of systematic zoology*, McGraw-Hill, New York.

McCarthy, A. J. (1991) Conservation Areas of Indonesia. [Final draft, September 1991], World Conservation Monitoring Centre, Cambridge, UK in collaboration with the Directorate General of Forest Protection and Nature Conservation (PHPA), Ministry of Forestry, Republic of Indonesia, and IUCN-Commission on National Parks and Protected Areas.

McDougall, I., F. H. Brown, T. E. Cerling and J. W. Hillhouse (1992) A reappraisal of the geomagnetic polarity time scale to 4 Ma using data from the Turkana Basin, East Africa. *Geophysical Research Letters* **19**: 2349-2532.

Meijaard, E. (2001) Successful Sea-crossing by Land Mammals; A Matter of Luck, and A Big Body: A preliminary and simplified model. *Geological Research and Development Centre's Special Publication* **27**: 87-92.

Milner-Gulland, E. J., E. Bennett, L.,, K. Abernethy, M. Bakarr, R. Bodmer, J. Brashares, G. Cowlishaw, P. Elkan, H. Eves, J. Fa, C. Peres, C. Roberts, J. Robinson, RowcliffeM. and Wilkie.D. (2002) Wild meat: the bigger picture. *Trends in Ecology and Evolution* **18:.** 351-357.

Minch, E., A. Ruiz-Linares, D. B. Goldstein, M. W. Feldman and L. L. Cavalli-Sforza (1995) Microsat (version 1.4): A computer program for calculating various statistics on microsatellite allele data. www:http://lotka.stanford.edu/research.

Mohr, E. (1921) Die Geographische Verbreitung der Anoa-Arten auf Celebes. *Arch. Naturgeschichte* **87**(6): 208-214.

Moore, S. S., L. L. Sargeant, T. J. King, J. S. Mattick, M. Georges and D. J. S. Hetzel (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* **10**: 654-660.

Morgan-Richards, M., S. A. Trewick and G. P. Wallis (2000) Characterization of a hybrid zone between two chromosomal races of the weta *Hemideina thoracica* following a geologically recent volcanic eruption. *Heredity* **85**(6): 586-592.

Moritz, C. (1994) Defining 'Evolutionary Significant Units' for conservation. *Trends in Ecology and Evolution* **9**(10): 373-375.

Moritz, C. (1995) Uses of molecular phylogenies for conservation. *Philosophical transactions of the Royal Society of London. Series B* **349**: 113-118.

Moritz, C. and D. P. Faith (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology* **7**: 419-429.

Moss, S. J. and M. E. J. Wilson (1998) Biogeographic implications of the Tertiary palaeogeographic evolution of Sulawesi and Borneo. In *Biogeography and Geological Evolution of Southeast Asia*. Eds: R. Hall and D. J. Holloway. Backhuys, Leiden, The Netherlands: 133–163.

Murphy, R. W., J. W. Sites, D. G. Buth and C. H. Haufler (1996) Proteins: isozyme electrophoresis. In *Molecular Systematics*. Eds: D. M. Hillis, C. Moritz and B. K. Mable. Sinauer Associates, Sutherland, USA: pp.51-120.

Musser, G. G. (1987) The mammals of Sulawesi. In *Biogeographical Evolution of the Malay Archipelago*. T. C. Whitmore. Oxford: Clarendon Press: pp.73-93.

Mustari, A. H. (1995) Population and Behaviour of Lowland Anoa (*Bubalus depressicornis* Smith) in Tanjung Amolengu Wildlife Reserve South-East Sulawesi, Indonesia. [MSc Thesis]. Faculty of Forestry Science, Georg-August University, Göttingen, Germany.

Mustari, A. H. (2004) Ecology and Conservation of Lowland Anoa in Southeast Sulawesi, Indonesia. [PhD thesis], University of New England, Australia.

Muwanika, V. B., S. Nyakaana, H. R. Siegismund and P. Arctander (2003) Phylogeography and population structure of the warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity* **91**: 361-372.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca and J. Kent (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**: 853-858.

Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* **41**: 225-233.

Nei, M. (1987) *Molecular Evolutionary Genetics*, Columbia University Press, NY, USA.

Nersting, L. G. and P. Arctander (2001) Phylogeography and conservation of impala and greater kudu. *Molecular Ecology* **10**: 711-719.

Newmark, W. D. (1995) Extinction of mammal populations in western North American national parks. *Conservation Biology* **9**: 512-526.

Niemitz, C., A. Neitsch, S. Warter and Y. Rumpler (1991) *Tarsius dianae*: A New Primates Species from Central Sulawesi (Indonesia). *Folia Primatologia* **56**: 105-116.

Nietsch, A. (1999) Duet vocalizations among different populations of Sulawesi Tarsiers. *International Journal of Primatology* **20**(4): 567-583.

Nietsch, A. and J. A. Burton (in prep.) Tarsiers in South and Southeast Sulawesi - evidence for new taxa based on vocalizations.

Nijman, I. J., M. Otsen, E. L. C. Verkaar, C. de Ruijter, E. Hanekamp, J. W. Ochieng, S. Shamshad, J. E. O. Reges, O. Hanotte, M. W. Barwegen, T. Sulawati and J. A. Lenstra (2003) Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity* **90**(1): 10-16.

Nijman, I. J., J. B. L. Vastenburg, J. L. Williams, I. Olsaker, U. Thieven and J. A. Lenstra (1998) Thirteen bovine microsatellite markers that are polymorphic in sheep. *Animal Genetics* **29**(6): 474-475.

Nixon, F. C. and Q. D. Wheeler (1990) An amplification of the phylogenetic concept. *Cladistics* **6**: 211-223.

Nötzold, G. (1999) International Studbook: Anoa, (*Bubalus [Anoa] depressicornis* Smith, 1827), Leipzig Zoo, Germany.

O'Brien, T. G. and M. F. Kinnaird (1996) Changing populations of birds and mammals in North Sulawesi. *Oryx* **30**(2): 150-156.

O'Grady, J. J., D. H. Reed, B. W. Brook and R. Frankham (2004) What are the best correlates of predicted extinction risk? *Biological Conservation* **118**: 513-520.

O'Ryan, C., E. H. Harley, M. W. Bruford, M. Beaumont, R. K. Wayne and M. I. Cherry (1998) Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Animal Conservation* **1**: 85-94.

Page, R. D. M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358.

Palumbi, S. R. (1996) Nucleic acids II: the polymerase chain reaction. In *Molecular Systematics*. Eds: D. M. Hillis, C. Moritz and B. K. Mable. Sutherland, USA, Sinauer Associates: pp.205-248.

Pardue, M. L., K. Lowenhaupt, A. Rich and A. Nordheim (1987) (dC-dA)n.(dG-dT)n sequences have evolutionary conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. *European Molecular Biology Organization Journal* **6**: 1781-1789.

Parker, H. G., L. V. Kim, N. B. Sutter, S. Carlson, T. D. Lorentzen, T. B. Malek, G. S. Johnson, H. B. DeFrance, E. A. Ostrander and L. Kruglyak (2004) Genetic Structure of the Purebred Domestic Dog. *Science* **304**: 1160-1164.

Pemberton, J. M., J. L. Slate, D. R. Bancroft and J. A. Barrett (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology* **4**: 249-252.

Pitra, C., R. Furbass and H. M. Seyfert (1997) Molecular phylogeny of the tribe Bovini (Mammalia: Artiodactyla): alternative placement of the Anoa. *Journal of Evolutionary Biology* **10**(4): 589-600.

Polhemus, D. and J. T. Polhemus (1986) The zoogeography and phylogeny of the genus *Ptilomen* Amyet and Serville (Hemiptera: Gerridae). *Abstract Bulletin of the American Benthological Society* **3**: 67.

Pranadewi, V. (1998) Analasis kromosom pada kelompok anoa (Anoa Smith 1827) di Kebun Binatang Ragunan Jakarta. Thesis. Fakultas Matematika dan Ilmu Pengetahuan Alam, Jurusan Biologi, Universitas Indonesia.

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

Pritchard, J. K. and W. Wen (2002) Documentation for STRUCTURE 2.0. Available at: http://pritch.bsd.uchicago.edu (May 2004).

Pulliam, H. R. (1988) Sources, sinks, and population regulation. *American Naturalist* **132**: 652-661.

Puorto, G., M. Da Grace Salomã, R. D. G. Theakston, R. S. Thorpe, D. A. Warrell and W. Wuster (2001) Combining mitochondrial DNA sequences and morphological data to infer species boundaries: phylogeography of lanceheaded pitvipers in the Brazilian Atlantic forest, and the status of Bothrops pradoi (Squamata: Serpentes: Viperidae). *Journal of Evolutionary Biology* **14**: 527-538.

Rannala, B. and J. L. Mountain (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Science (USA)* **94**: 9197-9221.

Raufaste, N. and F. Rousset (2001) Are partial Mantel tests adequate? *Evolution* **55**(8): 1703–1705.

Raymond, M. and F. Rousset (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248-249.

Reed, D. H., J. J. O. O'Grady, B. W. Brook, J. D. Ballou and R. Frankham (2003) Estimates of minimum viable population size for vertebrates and factors influencing those estimates. *Biological Conservation* **113**(1): 23-34.

Reich, D. E. and D. B. Goldstein (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proceedings of the National Academy of Sciences (USA)* **95**: 8119-8123.

Rice, W. R. (1989) Analysing tables of statistical tests. *Evolution* **43**: 223-225.

Riginos, C. and M. W. Nachman (2001) Population subdivision in marine environments: the contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, *Axoclinus nigricaudus*. *Molecular Ecology* **10**(6): 1439-1453.

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

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

Rojas, M. (1992) The species problem and conservation: What are we protecting? *Conservation Biology* **6**(2): 170-178.

Roques, S., J. M. Sevigny and L. Bernatchez (2001) Evidence for broadscale introgressive hybridization between two redfish (genus Sebastes) in the North-west Atlantic: A rare marine example. *Molecular Ecology* **10**(1): 149-165.

Rossiter, S. J., G. Jones, R. D. Ransome and E. M. Barratt (2000) Genetic variation and population structure in the endangered greater horseshoe bat *Rhinolophus ferrumequinum*. *Molecular Ecology* **9**: 1131-1135.

Rousset, F. (1997) Genetic differentiation and estimation of gene flow from *F*-Statistics under isolation by distance. *Genetics* **145**: 1219-1228.

Rousset, F. (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**: 58-62.

Ruedi, M., M. Auberson and V. Savolainen (1998) Biogeography of Sulawesian shrews: testing for their origin with a parametric bootstrap on molecular data. *Molecular Biological Evolution* **9**: 567-571.

Ryder, O. A. (1986) Species conservation and systematics: dilemma of subspecies. *Trends in Ecology & Evolution* **1**: 9-10.

Ryder, O. A., A. T. Kumamoto, B. S. Durrant and K. Benirschke (1989) Chromosomal divergence and reproductive isolation in Dik-Diks. In *Speciation and its consequences*. Eds: D. Otte and J. A. Endler. Sinauer Associates, MA, USA: pp.208-228.

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

Sambrook, J., E. F. Fritsch and T. Maniatis (1989) *Molecular Cloning – A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

Sarasin, P. and F. Sarasin (1901) Entwurf einer Geographisch-Geologischen Beschreibung der Insel Celebes. *Materialien zur Naturgeschichte der Insel Celebes*, Weisbaden: C.W. Kreidel.

Schaffer, M. K. (1981) Minimum viable population size for species conservation. *Bioscience* **21**: 131-134.

Schneier, S., D. Roessli and L. Excoffier (2000) Arlequin: A software for population genetic data analysis. Ver. 2.000. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Schreiber, A. and R. Goltenboth (1990) The haemoglobins of wild cattle (Bovini Simpson, 1945). *Zeitschrift fur Saugetierkunde* **55**: 276-283.

Schreiber, A., F. Klein and G. Lang (1994) Transferrin polymorphism of red deer in France: Evidence for spatial genetic microstructure of an autochthonous herd. *Genetics Selection Evolution* **26**: 187-203.

Schreiber, A. and G. Nötzold (1995) One EEP, but how many anoas? In *EEP Yearbook 1994/95*. F. Rietkerk, Brouwer, K., and Smits, S., EAZA/EEP Executive Office, Amsterdam, The Netherlands: pp.419-424.

Schreiber, A., G. Nötzold and M. Held (1993) Molecular and chromosomal evolution in anoas (Bovidae: *Bubalus* spec.). *Zeitschrift Zoologische Systematik Evolutionforschung* **31**(1): 64-79.

Schreiber, A., I. Seibold, G. Nötzold and M. Wink (1999) Cytochrome b gene haplotypes characterize chromosomal lineages of anoa, the Sulawesi dwarf buffalo (Bovidae: *Bubalus* sp.). *The Journal of Heredity* **90**: 165-176.

Shekelle, M., S. M. Leksono, L. L. S. Ichwan and Y. Masala (1997) The natural history of the tarsiers of North and Central Sulawesi. *Sulawesi Primate Newsletter* **4**(2): 4-8.

Simonsen, B. T., H. R. Siegismund and P. Arctander (1998) Population structure of African buffalo inferred from mtDNA sequences and microsatellite loci: High variation but low differentiation. *Molecular Ecology* **7**: 225-237.

Sites, J. W. and K. A. Crandall (1997) Testing species boundaries in biodiversity studies. *Conservation Biology* **11**(6): 1289-1297.

Slate, J., D. W. Coltman, S. J. Goodman, I. Maclean, J. M. Pemberton and J. L. Williams (1998) Bovine microsatellite loci are highly conserved in red deer (*Cervus elaphus*), sika deer (*Cervus nippin*) and Soay sheep (*Ovis aries*). *Animal Genetics* **29**: 307-315.

Slatkin, M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264-279.

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

Slatkin, M. and W. P. Maddison (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**: 603-613.

Smouse, P. E., J. C. Long and R. R. Sokal (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**: 627-632.

Sokal, R. R. and F. J. Rohlf (1995) *Biometry: the principles and practice of statistics in biological research*. New York, W. H. Freeman and Co.

Soulé, M. E. and L. S. Mills (1998) No need to isolate genetics. *Science* **282**.

Soulé, M. E. and D. S. Simberloff (1986) What do genetics and ecology tell us about the design of nature reserves? *Biological Conservation* **35**: 19-40.

Sugiri, N. and N. Hidayat (1996) The diversity and hematology of anoa from Sulawesi. In *Population and Habitat Viability Analysis Workshop Report: Anoa Species Bubalus quarlesi and Bubalus depressicornis*. Eds: J. Manansang, S. Hedges, S. Dwiatmo, P. Miller and U. S. Seal. Bogor, Indonesia: pp.138-151.

Swofford, D. L. (2003) PAUP - Phylogenetic Analysis Using Parsimony, Version 4bi, Sinauer Associates, Sunderland, Massachusetts.

Swofford, D. L., G. J. Olsen, P. J. Waddell and D. M. Hillis (1996) Phylogenetic inference. In *Molecular Systematics*. C. M. D. M. Hillis, and B. K. Mable., Sinauer Associates, Sunderland, Massachusetts.: pp.407-514.

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

Tanaka, T., C. D. Solis, J. S. Masangkay, K. Maeda, Y. Kawamoto and T. Namikawa (1996) Phylogenetic relationship among all living species of the genus *Bubalus* based on DNA sequences of cytochrome b gene. *Biochemical Genetics* **34**(11-12): 443-452.

Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463-6471.

Tautz, D. and M. Renz (1984) Simple DNA sequences of *Drosophila virilis* isolated by screening with RNA. *Journal of Molecular Biology* **172**(2): 229-235.

Templeton, A. R. (1989) The meaning of species and speciation: A genetic perspective. In *Speciation and its Consequences*. Eds: D. Otte and J. A. Endler. Sinauer Associates, MA, USA: pp.3-27.

ter Braak, C. J. F. (1987) Ordination. In *Data Analysis in Community Ecology*. Eds: R. H. Jongman, C. J. F. ter Braak and O. F. R. van Tongeren. Pudoc, Wageningen, The Netherlands: pp.91-173.

Thorpe, R. S., R. P. Brown, A. Malhotra and Wüster.W. (1991) Geographic variation and population systematics: distinguishing between ecogenetics and phylogenetics. *Bollettino di Zoologia* **58**: 329–335.

Thorpe, R. S., A. Malhotra, H. Black, J. C. Dalrty and W. Wuster (1995) Relating biogeographic patters to phylogenetic process. *Philosophical transactions of the Royal Society of London. Series B* **349**(1327): 61-68.

Tiedemann, R., O. Hardy, X. Vekemans and M. C. Milinkovitch (2000) Higher impact of female than male migration on population structure in large mammals. *Molecular Ecology* **9**: 1159–1163.

Troy, C. S., D. E. MacHugh, J. F. Bailey, D. A. Magee, R. T. Loftus, P. Cunningham, A. T. Chamberlain, B. C. Sykes and D. G. Bradley (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**(6832): 1088-1091.

Troyanskaya, O., M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein and R. B. Altman (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**(6): 520-525.

UN (1992) Long Range World Population Projections Two Centuries of Population Growth 1950-2150, United Nations Department of International Economic and Social Affairs (UNDIESA), United Nations, New York.

UNDP/FAO (1982) National conservation plan for Indonesia, Volume VI: Sulawesi, UNDP/FAO.

UNEP-WCMC (2000) UNEP-WCMC Species Database: CITES-Listed Species, UNEP-WCMC, http://www.cites.org/eng/resources/species.html (May 2004).

USFWS (1994) *Endangered and threatened wildlife and plants. Document 1994-380-789/20165*, Unites States Fisheries and Wildlife Service, Washington (DC): U.S. Government Printing Office.

Valdderama, X., W. B. Karesh, D. E. Wildman and D. J. Melnick (1999) Noninvasive methods for collecting fresh hair tissue. *Molecular Ecology* **8**: 1749-1752.

Van den Berg, G. D. (1999) The Late Neogene elephantoid-bearing faunas of Indonesia and their paleozoogeographic. A study of the terrestrial faunal succession of Sulawesi, Flores and Java, including evidence for the early hominid dispersal east of Wallace's Line implications. *Scripta Geology* **117**: 1-419.

Van den Berg, G. D., J. de Vos and P. Y. Sondaar (2001) The late Quaternary palaeogeography of mammal evolution in Indonesian Archipelago. *Palaeontology, Palaeoclimatology, Palaeoecology* **171**: 385-408.

van der Vecht, J. (1953) The carpenter bees (*Xylocopta* Latr.) of Celebes. *Idea* **9**: 57-59.

van Hooft, W. F., A. F. Groen and H. H. T. Prins (2000) Microsatellite analysis of genetic diversity in African buffalo (*Syncerus caffer*) populations throughout Africa. *Molecular Ecology* **9**: 2017-2025.

Visscher, P. M., D. Smith, S. J. G. Hall and J. L. Williams (2001) A viable herd of genetically uniform cattle. *Nature* **409**: 303.

Vogler, A. P. and R. Desalle (1994) Diagnising units for conservation management. *Conservation Biology* **8**(2): 354-363.

Voris, H. (2000) Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. *Journal of Biogeography* **27**: 1153-1167.

Wallace, A. R. (1865) On the phenomena of variation and geographical distribution as illustrated by the Papilionidae of the Malayan region. *Transactions of the Linnaean Society of London* **25**: 1-71.

Walling, G. A., A. L. Archibald, J. A. Cattermole, A. C. Downing, H. A. Finlayson, D. Nicholson, P. M. Visscher, C. A. Walker and C. S. Haley (1998) Mapping of quantitative trait loci on porcine chromosome 4. *Animal Genetics* **29**: 415-424.

Wang, J. and A. Caballero (1999) Developments in predicting the effective size of subdivided populations. *Heredity* **82**: 212-226.

Waples, R. S. (1991) Pacific salmon, *Oncorhynchus* spp., and the definition of 'species' under the Endangered Species Act. *Marine Fisheries Review* **53**: 11-22.

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

Watanabe, K. and S. Matsumura (1991) The borderlands and possible hybrids between three species of macaques, *M. nigra*, *M. nigrescens*, and *M. hecki*, in the northern peninsula of Sulawesi. *Primates* **32**(3): 365-369.

Wayne, R. K. (1996) Conservation genetics in the Canidae. In *Conservation Genetics: Case Histories From Nature*. Eds: J. C. Avise and J. L. Hamrick. New York, Chapman & Hall**:** pp.75-118.

Weber, J. L. and P. E. May (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388-396.

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

Weber, M. (1890) Mammalia from the Malay Archipelago. I. Primates, Prosimiae, Galeopithecidae, Carnivora, Artiodactyla, Edentata, Marsupiala. *Zoologische Ergebnisse einer Reise in Niederlandisch-Ostindien* **1**: 93-114.

Weir, B. S. (1996a) *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers: Sunderland, Massachusetts.

Weir, B. S. (1996b) Intraspecific differentiation. In *Molecular Systematics*. Eds: D. M. Hillis, C. Moritz and B. K. Mable. Sinauer Associates, Sutherland, USA**:** pp.385-406.

Weir, B. S. and C. C. Cockerham (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.

Weise, R. (1979) Untersuchungen zur innerartlichen Variabilität beim Anoa (*Anoa depressicornis*). Staatsexamenarbeit, Universität Kiel, Germany.

Wheeler, Q. D. and R. Meier (2000) *Species concepts and phylogenetic theory*, New York: Columbia University Press.

White, M. J. D. (1978) *Modes of speciation*, W. H. Freeman, San Francisco, California.

Whitmore, T. C. (1987) *Biogeographical Evolution of the Malay Archipelago*, Clarendon Press, Oxford.

Whitten, A. J., G. S. Henderson and M. Mustafa (2002) *The Ecology of Sulawesi*, Periplus, Hong Kong.

Wiener, G., H. Jianlin and L. Ruijun (2003) The yak, FAO of the United Nations, Bangkok: Regional Office for Asia and the Pacific.

Wiener, P., D. Burton and J. L. Williams (2004) Breed relationships and definition in British cattle: a genetic analysis. *Heredity* **93**: 597-602.

Wiener, P., I. Maclean, J. L. Williams and J. A. Woolliams (2000) Testing for the presence of previously identified QTL for milk production traits in new populations. *Animal Genetics* **31**(6): 385-395.

Wikramanayake, E., E. Dinerstein, C. J. Loucks, D. M. Olson, J. Morrison, J. Lamoreux, M. McKnight and P. Hedao (2001) *Terrestrial Ecoregions of the Indo-Pacific: A Conservation Assessment*, Island Press, Washington D.C.

Wilson, D. E. and D. M. Reeder (1993) *Mammal Species of the World: A Taxonomic and Geographic Reference*, Smithsonian Institution Press, Washington, USA and London, UK.

Wilson, G. A. and C. Strobeck (1999) Genetic variation within and relatedness among wood and plains bison populations. *Gemone* **42**: 483-496.

Wilson, P. J., S. Grewal, I. D. Lawford, J. N. M. Heal, A. G. Granacki, D. Pennick, M. T. Theberges, D. R. Voigt, W. Waddell, R. E. Chambers, P. C. Paquet, G. Goulet, D. Cluff and B. N. White (2000) DNA profiles of the eastern Canadian wolf and the red wolf provide evidence for a common evolutionary history independent of the grey wolf. *Canadian Journal of Zoology* **78**: 2156-2166.

Worley, K., C. Strobeck, S. Arthur, J. Carey, H. Schwantje, A. Veitch and D. W. Coltman (2004) Population genetic structure of North American thinhorn sheep (*Ovis dalli*). *Molecular Ecology* **13**: 2545–2556.

Wright, S. (1943) Isolation by distance. *Genetics* **28**: 139-156.

Wright, S. (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**(3): 395-420.

Wright, S. (1978) *Evolution and the Genetics of Population, Variability Within and Among Natural Populations*, University of Chicago Press, Chicago, IL.

Xu, X. and U. Arnason (1996a) The mitochondrial DNA molecule of Sumatran Orangutan and a molecular proposal for two (Bornean and Sumatran) species of Orangutan. *Journal of Molecular Evolution* **43**: 431–437.

Xu, X., A. Gullberg and U. Arnason (1996b) The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *Journal of Molecular Evolution* **43**: 438-446.

Yang, Z. H. and A. D. Yoder (2003) Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Systematic Biology* **52**(5): 705-716.

Appendix

Appendix 1: Protected areas of Sulawesi and biogeographic divisions of the Wallacea region.

Relevant to Section 2.3.3 (Figure 2.2) and Section 7.5.1 (Figure 7.2).

Table 1. Details of Major Forested Areas of Sulawesi and Buton Islands, including protection status and area size.

Map code	Area name	Status [1]	Area (km^2)
1	Gunung Klabat	P. NR	53
2	Tangkoko Batuangus – Dua Saudara	NR	75
3	Gunung Manembo-nembo	WR	65
4	Gunung Ambang	NR	250
5	Bogani Nani-Wartabone	NP	2871
6	Panua (in Marisa Complex)	NR	940
7	Upper Paguyaman / Nantu area.	none	315
8	Buol / Toli-Toli	P. NR	225
9	Pinjam/Tanjung Matop	WR	16
10	Poboya	NR	1000
11	Lore Lindu	NP	2310
12	Tanjung Api	NR	42
13	Pati-Pati	P. WR	
14	Lombuyan / Pagimanan	WR	36
15	Morowali	NR	2250
16	Gunung Dako	P. NR	
17	Gunung Tinombala	none	
18	Gunung Sojol	NR	500
19	Pangi Binangga	P. NR	60
20	Bakiriang	P. WR	10
21	Bada Valley	none	
22	Pantai Timur	none	
23	Pantai Barat	none	
24	Pengunungan Takolekaju	none	
25	Mayoa area	none	
26	Gunung Pompangeo	none	
27	Gunung Lumut, Gunung Tempu, Gunung Hohoban	none	
28	Togians Islands	(MR)	
29	Peleng / Banggai Islands	none	
30	Dolongan Island	WR	
31	Pengunungan Faruhumpenai	NR	900
32	Masupu	WR	25
33	S. Katena / Rompi	PF	100
34	Lampoko Mampie	WR	20
35	Pengunungan Latimojong	P. NR	580

continued overleaf

Table 1 continued

Map code	Province & area name	Status [1]	Area (km^2)
36	Bulu Saraung	NR	57
37	Makale area.	none	
38	Kambuno Katena	PF	
39	Sumarorang	P. WR	
40	Lasolo-Sampara	P. NR	450
41	Rawa Aopa-Watumohai	NP	968
42	Polewali	P. WR	80
43	Tanjung Peropa	NR	389
44	Tanjung Amolengu	WR	8.5
45	Tanjung Batikolo	WR	55
46	Kolaka Utara	NR	
47	Toronipa.	NR	20?
48	Lamedia	NR	5
49	Buton Utara	WR	820
50	Lambu Sango	WR	250
51	Kakinawe	NR	50
52	Padang Mata Osu	WR	
53	Pengunungan Tangkeleboke	none	
54	Pengunungan Abuki	none	
55	Pengunungan Matarombea	none	
56	Kabaena Island	none	
57	Muna Island	none	
58	Wowoni Island	none	

[1] **Abbreviations:** 1. Protected areas: NP - National Park; NR - Nature Reserve; WR - Wildlife Reserve; PF - Protection Forest; P. - Proposed status; MR – Marine Reserve.

Data taken from (Burton, Hedges and Mustari, 2005).

Appendix 2: Definition of morphological features used to describe the two putative anoa species

Relevant to Section 2.2.2

Table 1 summarizes the most important morphological features from previous descriptions of anoa (Heller, 1889; Dolan, 1965; Groves, 1969; Walker, Warnick, Hamlet et al., 1975; Groves, 1982; NRC, 1983; Grzimek, 1990). These closely follow Groves (1969) criteria for the debated two species classification. Those criteria marked in **bold face** are those used to provisionally identify the species. Anoa were categorised if an individual showed morphological characteristics from both species, or if no data relating to these characteristics were available. Juvenile *Bubalus depressicornis* (lowland) are similar to adult *B. quarlesi* (mountain) in some features. In order to identify a juvenile *B. depressicornis*, either white leg markings had to be present or there had to be clear proof that the animal was indeed a juvenile (indicated by the lack of 3rd molar teeth). As teeth were often not observed, but age was estimated from observation of the individual an age criteria was given. From observation of captive animals it is thought that sexual maturity is reached at two to three years of age (NRC, 1983; Jahja, 1987). It is therefore estimated that M3 may occur in the third or forth year. So animals estimated to be over four years and supporting characteristics of *B. quarlesi* will be identified as this species. Under this age they will not be identified to species, as they could possibly be either species. In Figure 2.4 data was classed as: *B. depressicornis* and *B. quarlesi or* Unknown *Bubalus* species - If an individual showed a morphological characteristic from both species or there was no data on characteristics, then the individual was included in the data set as "Unknown *Bubalus* species".

Details relating to Table 1:
The features in bold face are those used by the present authors to provisionally identify the species.

[1] Groves (1969) Species Identification Characteristics for the two species

[2] Hedges (in prep.). Species Identification Characteristics. Taken from: (Harper, 1945; Dolan, 1965; Groves, 1969; Walker et al., 1975; Groves, 1982; NRC, 1983; Grzimek, 1990);

[3] Schreiber *et al.* (1993) Species Identification Characteristics

^ National Research Council (1983) states that the entire lower limbs of mountain anoa are creamy white in colour, but all other published descriptions have emphasized the generally dark-coloured legs of this species.

Table 1. Morphological characteristics of two anoa species: *Bubalus depressicornis* and *Bubalus quarlesi.* (from Burton et al., 2005).

Species	**Lowland anoa *Bubalus depressicornis***	**Mountain anoa *Bubalus quarlesi***
Body size	[3] Shoulder height: 60–100cm; Mass <300kg Body length- 170–188cm	Shoulder height: <75cm; Mass <150kg; Body length- 122–153cm
Horns [1, 3]	[1, 2] **Triangular cross-section flattened, marked transverse ridges & marked external keel; Horn** length: Lowland male 271–373mm; female 183–260mm	[1] **Short, conical, and rounded cross section, no marked ridges or external keel in juvenile & adult**; Both sexes: horn length 146–199mm
Overall body pelage colour	[1] **Black & sparsely** (woolly brown in juvenile [M3]); sparse & straight, often rubbed off with age	[1] **Dark brown to black adult & thick & woolly haired into adulthood**; female coat especially woolly
White facial / neck markings	[1] White facial markings present [1, 2, 3] Often white throat markings	White facial markings sometimes present [1] White throat markings never present
White leg markings (see age dimorphism) [2, ^]	[3] **Forelegs: always white to yellowish-white from knee to hoof (black strip down front & across pasterns).** Hind legs: conspicuous white spots above hooves. Yellow in juvenile, to white in adult	^ Foreleg: only whitish-yellow spots above hooves, sometimes absent
Tail length (as % of total body length) [1, 2]	[1] 19.8–25.8% (9 skins)	[1] 14.6–17.8% (5 skins)
Groin	[1] Light coloured to white	[1] Light coloured but not white
Age related dimorphism	[1] Juvenile form resembles adult mountain anoa (in following characteristics: pelage colour, texture, horn shape, body size), **BUT** lowland anoa always has white forelegs, conspicuous white spots on hind legs above hooves (See Dolan 1965)	[1] Horn index of least -greatest basal horn breadth(anterio-posterior to bilateral)
Skull length	[1] Lowland anoa male 293–322mm; female 290–300mm	[1] Mountain anoa: both sexes skull length 244–290mm

See notes related to this table in the text above.

Appendix 3: Additional descriptive data for anoa individuals used in Mantel and partial Mantel tests

Table 1. Anoa individual's characteristics and genetic data available for wild and zoo individuals.

		Individuals Characteristics		Genetic data	
Sample Code	**Altitude (m asl)**	**Sex**	**Species identification**	**mtDNA data**	**Microsatellite data**
B7	50	F	D	1	1
B8	50	F	D	1	
B9	50	F	D	1	1
B141	200	F	D		1
B145	200	F	D		1
B5	50	M	D		1
B16	50		Q	1	1
B29	50	F			1
B14	200	M	D	1	1
B18	50	F		1	1
B23	50	M		1	1
B24	200			1	1
B36	200	M	D		1
B144	200	M	D	1	1
B22	50	M	D	1	1
B28	50	M	D		1
B83	200	M	D	1	1
B86	50	M			1
B12	50	F	D		1
B15	50	F		1	1
B21	50	M	Q	1	1
B25	50	M	D		1
B85	50	M	D		1
B82	50	M	D		1
B6	200		Q	1	
B13	200	M	D		1
B84	200	F		1	1
B87	200	F	D		1
B147	200	M	D		1
B27	-	M	D		1

M= male, F= female. D= depressicornis morphotype, Q= quarlesi morphotype.
1= genetic data included in analyses.

Continued overleaf

Table 1. continued

		Individuals Characteristics		**Genetic data**	
Sample Code	**Altitude (m asl)**	**Sex**	**Species ident-ification**	**mtDNA data**	**Microsat-ellite data**
SE45	50	M			1
SE94	50	F		1	1
SE95	50	F		1	1
SE60	200	F			1
SE209	200	F	D		1
SE210	200	M	D		1
SE211	200	F	D		1
SE54	200	F	D	1	1
SE53	200	F	D	1	1
SE41	200	F	D	1	1
SE63	-			1	1
SE212	1250	F			1
E163	200				1
E167	200				1
E170	200				1
E171	200				1
C100	1250	M	Q		1
C187	1250	F			1
C173	1250		D	1	1
C176	1250		D	1	1
C99	1250	M	D	1	1
C121	1250	F	D	1	
C189	1250	M	D	1	1
C190	1250	F	D		1
C184	1250	F			1
C191	1250	M	D	1	1
C192	1250	M	D		1
C194	1250	F			1
C96	1250	M	D	1	1
C97	1250	M	D	1	1
C119	1250	F	D		1
C178	1250	F	D		1
C180	1250	F	D	1	1
C182	1250	F			1

Continued overleaf

Table 1. continued

		Individuals Characteristics		**Genetic data**	
Sample Code	**Altitude (m asl)**	**Sex**	**Species identification**	**mtDNA data**	**Microsatellite data**
C65	-	F			1
C69	-	M			1
C130	-	M	D		1
C105	200				1
C66	200	M			1
C104	200				1
C107	500	M	D		1
C120	200	M	D	1	1
N122	200	F			1
N114	50		Q		1
N115	50		D	1	1
N116	50		D	1	
N128	50		D		1
N140	50		D		1
A110	50				1
A111	50		D		1
A124	50				1
A125	50		D		1
A126	50		D		1
A127	50		D		1
A134	50		D		1
A138	50		D		1
A112	50		D		1
A135	200		D		1
A136	200		D		1
A133	50				1
N139	50		D		1
N113	200		D		1
N123	500		D		1

Table 1. continued

		Individuals Characteristics		**Genetic data**	
Sample Code	**Altitude (m asl)**	**Sex**	**Species ident-ification**	**mtDNA data**	**Microsat-ellite data**
Z222	-	F	Q		1
Z221	-	M	Q		1
Z225	-	M	D		1
Z235	-	F	D		1
Z236	-	M	D	1	
Z226	-	F	D		1
Z70	-	F			1
Z71	-	F			1
Z77	-				1
Z78	-				1
Z229	-				1