A study of the impact of population bottlenecks on the genetics and morphology of reindeer (Rangifer tarandus tarandus) on the island of South Georgia

Lovatt, Fiona Margaret

A study of the impact of population bottlenecks on the genetics and morphology of reindeer *(Rangifer tarandus tarandus)* on the island of South Georgia

Fiona Margaret Lovatt

Ph.D.

Department of Biological and Biomedical Sciences

Durham University

2007

Contents

Abstract

Numerous studies have shown that population bottlenecks result in loss of genetic diversity and as a consequence of this, it is commonly inferred that there is a loss of evolutionary potential. It is rare that circumstances are such that there should be well documented details of the founder event, such as the size and date of the bottleneck, that there should be access to subsequent demographic information and to suitable samples from both the post-bottleneck and the source populations. It is even less common for this information to be available for two separate bottlenecks that occurred in parallel in a largely unmanaged population of large mammals.

The importation by whalers of two separate groups of reindeer *(Rangifer tarandus tarandus)* onto the island of South Georgia in the early part of the twentieth century provided precisely those circumstances.

There are accurate historical records of each founder events with details of dates and numbers of reindeer. The inaccessibility and geography of the island ensures that the South Georgia reindeer herds have been isolated from immigrants and separated from each other.

The aim of this study was to test hypotheses about the impact of population bottlenecks on phenotype and genetic diversity and this was achieved by making genetic and morphometric comparisons of both post-bottleneck populations with the source population in Norway. Genetic diversity was primarily measured by allele numbers and heterozygosity based on data from thirteen microsatellite loci. Morphometric comparisons included

measures of developmental stability, notably fluctuating asymmetry (FA), as well as phenotypic variation and body size.

Each of the post-bottleneck populations showed significant genetic differentiation from the pre-bottleneck population and showed decreased levels of heterozygosity and allelic diversity. The data was used to validate commonly used 'bottleneck signatures' and considerable variability was found in the accurate detection of the known bottlenecks by the different detection methods.

Both the post-bottleneck populations showed increased FA and morphometric variation compared to the pre-bottleneck population in some of the measured traits. Both post-bottleneck populations had smaller overall skull size than the pre-bottleneck population though it is discussed that this may be due as much to a plastic environmental response as to a consequence of the genetic bottleneck.

Within each population the relationship between measures of genomic diversity and indirect measures of fitness were investigated on an individual basis. Although the results were of low significance, weak associations were found to support the hypotheses that developmental stability is correlated with measures of genomic diversity even at the level of the individual.

Acknowledgements

Funds were provided by project grants awarded by the British Ecological Society, the Shackleton Scholarship Fund and the Trans-Antarctic Association.

The original project was inspired by members of the British Schools Exploring Society Millennium Expedition – Dan Brown, Linden Davies, Gemma Ford, Chris Furse, Susie Grant, Alan Halewood, Fiona Hatchell, Richard Kendall, Peter Kenyon, Sarah Lessware, Chris Lipson, Stuart Martin, Karl McKay, Sarah Robinson and Anna Smales. Credit is due to all the individual sponsors and supporters of this expedition including parents and friends.

The BSES Millennium Expedition received support and advice from its patron Sir Martin Holdgate, from Professor David Walton and Ron Lewis-Smith of British Antarctic Survey, Bob Headland of SPRI and Professor Nigel Leader Williams of Canterbury University. The captain and crew of M/V Grigory Mikheev, Chuck Berry and Colin Monteath contributed practical and moral support to this expedition on South Georgia.

A number of individuals and expeditions contributed considerable time and effort in the collection of further samples: Mark and Fraser Carpenter, David Dyson, Keri Pashuk and the crew of the yacht Northanger, Jerome Poncet and Jean Sinclair, David Nicholls and members of the BSES Footsteps of Shackleton expedition.

Transport of the skulls would not have been possible without the help of HMS ENDURANCE and British Antarctic Survey, particularly Andy Liddell and Robert Paterson on the R.R.S. JAMES CLARK ROSS.

Generous support and practical help was given by Sarah and Pat Lurcock in South Georgia and Richard McKee from the South Georgia Government.

Considerable help in the research to trace the source Norwegian herds was provided by Erik Gossner (Tonsberg library), Jane Cameron (FIG archives), Anna Smales, Chris Furse and Chris Lipson (BSES Millennium expedition).

Asgrim Opdal and Olaf Ødigaard from Filefjell Reinlag gave very generous help and hospitality in Norway.

Cameron Bell and Steve Pointing, veterinary surgeons in the Falklands Islands and Morgens Østenvig, veterinary surgeon in Hemsedal were very helpful in giving practical help and in certifying the health status of samples for import into the UK.

Practical assistance, advice and encouragement was provided by my contempories in the laboratory: Pia Anderwald, Eulalia Banguera-Hinestroza, Dilletta Castellani, Laura Corrigan, Vittoria Elliott, David Faini, Andy Foote, Carlos de-Luna-Lopez, Courtney Nichols, Colin Nicholson, Ana Topf, Demetrios Vassilakos and my supervisor, Rus Hoelzel.

Finally, I am very grateful to Pete for his loving support and encouragement and to Jessica for her patience.

Chapter One - Introduction

1.1 Introduction

South Georgia, a sub-antarctic island which lies between latitudes 54° and 55°S and longitudes 35° and 38°W, was home to shore-based whaling stations from 1904. Small groups of reindeer *(Rangifer tarandus tarandus)* from Norway were introduced onto the island between 1911 and 1928 by the managers of the whaling stations. The reindeer were intended to provide sport and an alternative source of fresh meat for the whalers (Leader-Williams 1988).

Figure 1.1 Map of South Georgia (after Leader-Williams 1988) showing the areas currently occupied by reindeer herds. Scale: |_____| 30 km

There are two herds of reindeer on South Georgia today, one herd on the Barff peninsula (in red in figure 1.1) and one on Busen Point (in green in figure 1.1). The founder group of the Barff herd comprised three males and

seven females introduced to Ocean Harbour in 1911. The Busen (hereafter referred to as Husvik) herd was founded by the introduction of three males and four females into Husvik Harbour in 1925. Records suggest that both founder groups of reindeer came from Filefjell Reinlag, a reindeer herd based at Hemsedal, Buskerud, southern Norway (60°N, 8°E). A single newspaper report from 1928 suggested that a further four females were sent from Norway to the Husvik herd in that year though there are no further references to these extra reindeer and no indication of their exact origin. Full details of the historical evidence for the separate introductions are included in Chapter Two.

The source Norwegian herd, owned by the Opdal family, have been kept extensively for commercial meat production over the whole of the last century. They range freely over a mountainous area of 2000 square kilometres and forage supplies are such that it is never considered necessary to give supplementary feeding. The herd is gathered once or twice a year for the slaughter of young males (at just under 2 years old) and the culling of old females (at 11 to 12 years old) to maintain numbers at 3000 over winter to 5000 in summer. The owner reports that the level of deaths due to natural causes is insignificant and that due to extensive grazing and mild conditions the herd thrives in comparison to more northerly herds. He has no concerns about predation of the herd and considers routine treatment against parasites unnecessary (Asgrim Opdal – personal communication).

The reindeer on South Georgia are unmanaged and have been protected by legislation since 1912 though licences were issued for hunting

the males. At the peak of hunting between the 1930s and 1950s, up to 100 a year were shot from the Barff herd and it is thought that poaching of the Husvik herd took place during this period (Leader-Williams 1988). Approximately 400 reindeer from the Barff herd and 100 from the Husvik herd were shot for research purposes in the seventies (Leader-Williams 1988) but since then reindeer have been shot only infrequently for food (McIntosh and Walton 2000). Visitors to South Georgia are strictly controlled and the Gritviken area where permanent residents (mainly garrison soldiers and scientists) have been accommodated does not coincide with either of the reindeer areas. A couple of hundred Husvik reindeer were gathered in 2001 to capture approximately 60 calves for translocation to a commercial operation on the Falkland Islands.

The South Georgia reindeer have no predators or competitors and a low burden of internal parasites (Leader-Williams 1988). Soon after arrival, the founder groups adapted their winter diet and came to depend almost entirely on the coastal tussock grass *(Poa flabellata)* for winter forage rather than the lichen on which continental European reindeer depend. The rapid recovery of overgrazed tussock in comparison to lichen was probably a major contributory factor in the continued success of the reindeer following their initial introduction to the island (Leader-Williams 1988). Both Barff and Husvik populations have thrived though there have been fluctuations in population numbers over the past ninety years (Headland 1984; Leader-Williams 1988). The two herds are geographically separated by mountains and glaciers and there is no possibility of mixing.

The South Georgia calves generally either die perinatally or in their first winter but due to the lack of predators, rates of calf mortality are low compared to most reindeer populations (Leader-Williams 1988). Causes of mortality for both yearlings and adults follow a typical pattern with reindeer mainly dying at the end of the period of most reproductive stress: after the rut for males and at the end of winter for females. However, falling off cliffs was the most common cause of death with disease very uncommon except for an unusually high prevalence of mandibular pathology (Leader-Williams 1988).

The future of the South Georgia reindeer has been the subject of much debate over recent years as an environmental management plan has been drawn up for South Georgia (McIntosh and Walton 2000). The concerns are that the reindeer, which occupy the most extensive and species-rich vegetated areas of South Georgia, have damaged native vegetation by over-grazing. This overgrazing has lead to soil erosion at some sites, changes in plant community structures and widespread distribution of introduced plant species (McIntosh and Walton 2000). Further concerns are that the retreat of glaciers will open up new areas to the reindeer, leading to further degradation of plant communities (Moen and MacAlister 1994). As a long-term policy, the Government of South Georgia seeks to remove all non-indigenous flora and fauna from South Georgia and the reindeer are the first priority in this regard (McIntosh and Walton 2000).

Field studies of the impact of genetic bottlenecks and the foundation of island populations have significant implications for the conservation of

endangered species. Following a bottleneck, there is a reduced evolutionary potential of the population to respond to environmental change (Lacy 1997). This is due not only to the consequences of the founder event but also to subsequent inbreeding that occurs in a small isolated population. The consequences of demographic and environmental stochasticity, inbreeding depression, loss of genetic variation and the fixation of mildly deleterious alleles all contribute to the increased probability that the population will become extinct (Caughley 1994; Frankham 1995; Frankham 1997; Frankham 1998). For these reasons, population bottlenecks have attracted considerable research, both in experimental populations and more recently in natural populations. However, few studies have been undertaken on naturally occurring populations where the exact history of the bottleneck is known (Ardern and Lambert 1997; Le Page *et al.* 2000).

This study has the benefit of considerable information on the dates and numbers in the founder populations and access to samples from both the source and post-bottleneck populations. Thus there is the opportunity to investigate in the field the genetic and morphological impact of a bottleneck on each of two populations in parallel.

1.2 Genetic comparison of pre- and post-bottleneck populations

Microsatellites are DNA markers that are selectively neutral, show nuclear co-dominant inheritance and are characterised by short tandem repeat sequences. The relevant segment of DNA is flanked by unique conserved sequences and amplification occurs by the polymerase chain

reaction (PCR), which is a routine procedure (eg see Avise 1994). Microsatellite loci can have as many as ten or more alleles per locus and heterozygosity levels of between 50% and 90%, which makes them highly sensitive measures of genetic variation in wild species (Dietrich *et al.* 1992; Taylor *et al.* 1994; Houlden *et al.* 1996).

A study of microsatellite variation in Scandinavian cervids concluded that reindeer have a high degree of polymorphism compared to other cervids and that they have not been exposed to severe population bottlenecks or genetic drift in recent times (Roed 1998b).

Genetic variation within a population is most commonly quantified by gene diversity (the heterozygosity observed or expected under the Hardy Weinberg equilibrium), the average numbers of distinct alleles per locus or the proportion of polymorphic loci (Lacy 1997; England and Osler 2001).

Computer simulation models (Hoelzel *et al.* 1993; England and Osler 2001) are available to predict the theoretical consequences of a population bottleneck on genetic diversity. This enables comparisons to be made between the average outcome of many iterations of a hypothetical populations and the genetic diversity of actual post-bottleneck populations.

1.3 Bottleneck signatures

For loci which are neutral to the effects of selection in a natural population, the allele numbers and frequencies are affected by mutation and genetic drift.

Mutational models can be used to predict the outcome of a population at mutation-drift equilibrium. Populations which have experienced a recent bottleneck show characteristic patterns in the reduction in allele numbers, reduction in the range of allele size, reduction in heterozygosity and alterations to allele frequency distributions (Nei *et al.* 1975; Chakraborty and Nei 1977; Chakraborty *et al.* 1980; Cornuet and Luikart 1996; Luikart *et al.* 1998a; Garza and Williamson 2001). These patterns are termed 'bottleneck signatures' and their validity is important to conservationists as they are commonly used to assess whether a population has experienced an historical bottleneck. Recent studies that used these signatures to suggest whether an historical bottleneck had occurred include that of the endangered black rhinoceros (Harley *et al.* 2005), a population of Bowhead whales (Rooney *et al.* 1999) and a study into whether a severe viral epizootic caused a bottleneck in wild European rabbits (Queney *et al.* 2000).

1.4 Developmental stability

Developmental homeostasis results from the combination of two principal factors: canalization, in which one definite end-result occurs despite minor variations in genetic and environmental conditions (Waddington 1942), and developmental stability, which relates to processes that reduce the effect of developmental accidents on phenotypic variation (Clarke 1992).

Since Waddington (1942) described how the phenotypic constancy of the wild type demonstrated a buffering of the genotype, levels of phenotypic variability in the population have been used as a measure of developmental

stability. Investigations of populations of wild strains and inbred *Drosophila* showed higher morphological variance in inbred lines (Robertson and Reeve 1952) and Lerner (1954) described how heterozygotes are better canalized or buffered in their developmental processes than homozygotes. A number of further studies have demonstrated an inverse relationship between heterozygosity and morphological variance; examples include populations of killifish (Mitton 1978), Monarch butterflies (Eanes 1981) and house sparrows (Fleischer 1983).

As the development of each side of a bilaterally symmetrical organism is influenced by precisely the same genetic and environmental conditions, any disruption to this symmetry presumably results from a developmental accident. For this reason fluctuating asymmetry (FA) is commonly used as a measure for developmental stability (Van Valen 1962). Indeed FA may be a more reliable indicator of developmental stability than phenotypic variability due to the very fact that there is no difference in either the genetic or environmental conditions affecting the development of the left and the right side (Pertoldi *et al.* 2006).

The symmetry of sternopleural chaetae of *Drosophila* has often been used as an experimental measure of developmental stability and a number of studies have found more asymmetry in the inbred lines compared to the crosses (Mather 1953; Beardmore 1960). Further studies of natural populations have also found a negative correlation between asymmetry and heterozygosity on a population level; examples include island populations of

lizards (Soule 1979), Poecilid fish (Vrijenhoek and Lerman 1982) and 33 hatchery and natural populations of trout (Leary and Allendorf 1989).

Measures of developmental stability, such as fluctuating asymmetry, are useful to identify populations that are subject to genetic or environmental stress before there is an effect on the fitness or viability of the population (Clarke 1995). In a study of the breeding records of small captive populations of ungulates (including reindeer), the juvenile mortality of inbred animals was found to be higher than that of non-inbred young in 15 out of 16 populations (Ballou 1997). Inbreeding depression was indicated by lower levels of larval survival, adult longevity and egg-hatching rate in natural populations of the Glanville fritillary butterfly (Saccheri 1998) and smaller litter size and a higher incidence of stillborn or deformed offspring in an isolated population of adders with a small effective population size (Madsen *et al.* 1996).

1.5 Pleiotropic interactions

Subtle phenotypic consequences of a population bottleneck are difficult to predict due to the stochastic nature of the genetic effect and further complications due to pleiotropic interactions between genes. Pleiotropy describes the situation where an allele at one locus might have multiple phenotypic effects. Pleiotropic interactions include those that are interallelic such as dominance or overdominance (both described below) and those that are interlocus such as linkage, where the association of alleles at loci is not random, or epistasis where one gene has a controlling effect on other genes.

Inbreeding depression and heterosis have been explained by the two phenomenons of overdominance and partial dominance. Overdominance or heterozygote advantage describes the superiority of heterozygotes over homozygotes at individual loci for a given trait. One explanation of this superiority is that it may be due to an improved ability to buffer biochemical pathways against negative genetic effects on account of the diversity of biochemical products that result from heterozygous genotypes (Pertoldi *et al.* 2006). The partial dominance theory suggests that inbreeding depression occurs when inbred lines have become fixed for recessive or partially recessive deleterious alleles. Once these inbred lines are crossed, the next generation is heterozygous and thus superior as the recessive deleterious allele is not expressed (Wright 1921; Charlesworth and Charlesworth 1987). According to the overdominance theory, biochemical efficiency, fitness and developmental stability will always decrease with increasing homozygosity. However, according to the partial dominance theory, the direct association of developmental stability and heterozygosity is complicated by the purging of deleterious alleles from the population (Pertoldi *et al.* 2006).

Studies in flies (Tebb and Thoday 1954), honey bees (Clarke 1992) and rats (Borisov *et al.* 1997) have shown that there is not necessarily a simple relationship between developmental stability and heterozygosity but it is the balance of genes within the chromosome that is important.

A reduction in simple additive genetic variance might be expected to reduce quantitative variance whilst a reduction in non-additive interactions, such as linkage, dominance or epistasis might be expected to increase

quantitative variance due to the removal of buffering (Templeton 1980; Carson and Templeton 1984; Goodnight 1987). Experimental studies of bottlenecks in populations of houseflies showed an increase in morphological variance (Bryant 1986; Bryant *et al.* 1986).

In a review of the heritability of FA, Leamy and Klingenberg (2005) conclude that FA has a predominantly nonadditive genetic basis with substantial dominance and especially substantial epistasis.

There is general agreement in the literature that island populations that have experienced bottlenecks or been isolated at a small effective population size have lower genetic variation than large outbred mainland populations but there is an ongoing controversy as to how this impacts on their developmental stability.

Much debate has surrounded studies of wild populations of cheetah which show dramatically low genetic variation, purportedly due to an historical bottleneck. Early work showed significantly more fluctuating asymmetry compared to other felidae but this has been challenged by further studies (Wayne 1986; Modi *et al.* 1987; Kieser and Groeneveld 1991; Merola 1994).

More recent studies of wild populations that had experienced a genetic bottleneck, notably island populations of moose (Broders *et al.* 1999), black robins (Ardern and Lambert 1997) and brown hares (Hartl 1995; Suchentrunk 1998) show a reduction in genetic variation with apparently no adverse phenotypic consequences.

However, the Northern Elephant Seal, which shows low levels of molecular genetic variation following a bottleneck due to commercial hunting, does have increased levels of fluctuating asymmetry and morphological variation in some quantitative characters (Bonnell and Selander 1974; Hoelzel *et al.* 1993; Hoelzel 1999; Weber *et al.* 2000; Hoelzel 2002). An island population of black-footed rock wallabies showed very low levels of genetic variation and showed indications of inbreeding depression manifested in reduced female fecundity, skewed sex ratios and increased fluctuating asymmetry (Eldridge *et al.* 1999).

1.6 Relationship between developmental stability, fitness and genomic diversity on an individual basis

It was shown from a study of song sparrows that following a natural population bottleneck due to environmental factors, natural selection favoured the survival of outbred individuals. This suggested that even in a bottleneck that was apparently due entirely to environmental factors, there was still a significant genetic effect on survival (Keller *et al.* 1994).

In a study of individuals within a population of rainbow trout, a significant correlation was found between the proportion of heterozygous loci and proportion of asymmetric characters (Leary 1983). However, this direct association between individuals has not been found to be universal and examples of studies that have failed to find a correlation between individual heterozygosity and assymetry include a natural population of feral house mice (Wooten 1986), forked fungus beetles (Whitlock 1993) and a large sample of

Drosophila (Fowler and Whitlock 1994). Indeed when the FA of several traits are examined together there does not seem to be consistency between the rank order of individuals (Rasmuson 2002) which may suggest that FA is an imperfect tool to measure developmental stability on an individual level.

Microsatellite loci are ideal to investigate fitness consequences of short-term inbreeding due to the increase in heterozygosity-fitness correlation with increased marker diversity and high marker mutation rates (Tsitrone *et al.* 2001). Modern molecular techniques are able to provide reliable estimates of genome-wide genetic variability (Vollestad *et al.* 1999) and microsatellites allow the use of different measures of the genomic diversity of the individual, such as heterozygosity, mean d^2 and internal relatedness. Heterozygosity is the proportion of typed loci for which the individual is heterozygous . Mean d^2 is a measure of the genetic distance between the gametes that formed the individual (Coulson *et al.* 1998) and is thus argued to be based on the time since coalescence. Internal relatedness is a measure of inbreeding that considers the frequency of each allele and weights the sharing of rare alleles more than the sharing of common alleles (Queller and Goodnight 1989; Amos *et al.* 2001).

A number of studies in recent years have used these measures of genomic diversity to investigate correlations between fitness and levels of inbreeding at the level of the individual (Coltman *et al.* 1998; Coulson *et al.* 1998; Coltman *et al.* 1999; Coulson *et al.* 1999; Slate *et al.* 2000; Amos *et al.* 2001; Hedrick *et al.* 2001; Slate and Pemberton 2002). There has been much

debate as to their validity as measures of inbreeding and whether they are consistently correlated with indicators of fitness or developmental stability.

Heterozygosity-fitness correlations have been widely reported and their existence seems to have been widely accepted (Mitton and Grant 1984; David 1998)) despite the fact that a number of studies have published null results e.g. Wooten (1986), Whitlock (1993) and Fowler and Whitlock (1994). By meta-analysis of published studies, Britten (1996) concluded that there was a weakly significant positive correlation between allozyme heterozygosity and growth rate and a weakly significant negative correlation between heterozygosity and fluctuating asymmetry. Another meta-analysis of a number of studies which considered fluctuating asymmetry in relation to heterozygosity concluded that there was only a weak negative association with greater effects seen in the among-population rather than within-population studies (Vollestad *et al.* 1999).

Meta-analyses of both published and unpublished studies of microsatellites indicated that there has indeed been a bias towards the publication of significant results. The mean effect size of unpublished results was small and not significantly different from zero despite, on average, being based on greater numbers of sampled individuals than those in the published results (Coltman and Slate 2003). The conclusions of this meta-analysis indicated that the effects on life-history trait variation were significantly greater than zero for both heterozygosity and mean d^2 but that the effects on morphometric traits were small and not significant for either index (Coltman and Slate 2003).

What a number of authors agree on is that effects are taxon-, population-, trait- and environment specific (David 1998; Vollestad *et al.* 1999; Keller and Waller 2002). However, there has been a huge bias in the taxonomic range of the studies with most studies on ectotherms, specifically salmonid fish, bivalves and pine trees (David 1998; Vollestad *et al.* 1999).

This study is an important contribution of data from three semi-wild and largely unmanaged populations of reindeer that share the same origin. Due to the collection of corresponding genetic and morphometric data from different individuals it has been possible to investigate the relationship between genomic diversity and developmental stability on both an among-population and within-population basis.

Although considerable ecological data exists from extensive studies in the 1970s (Leader-Williams 1988), there have been no previous genetic studies of the reindeer on South Georgia. The apparent intentions of the Government to eradicate the reindeer from South Georgia as a matter of priority (McIntosh and Walton 2000) have given a sense of urgency to this work.

1.7 Hypotheses

This study aims to test the following hypotheses:

1. There will be reduced genetic diversity in the post-bottleneck populations compared to the pre-bottleneck population.

2. The genetic composition of the post-bottleneck populations matches that predicted by the mutation-drift models and the post-bottleneck populations show typical 'bottleneck signatures'.
3. Indicators of developmental instability, such as FA and morphological variability, will show increased values in the post-bottleneck populations compared to the pre-bottleneck population.
4. On an individual level, there will be a relationship between levels of genetic diversity and indirect measures of fitness within each of the populations with a stronger effect expected in the bottlenecked populations where it is expected that there will have been more intensive inbreeding.

1.8 Scope of this study

Chapter two describes the genetics on a population level and considers hypotheses one and two. Amplification of microsatellites from the pre-bottleneck (Norwegian) and post-bottleneck (Barff and Husvik) populations allowed direct genetic comparison. Due to reliable information on the founder numbers and exact dates, this was useful in testing the validity of bottleneck signatures and the accuracy of modeling a naturally occurring wild population.

Chapter three considers hypothesis three. Direct morphological comparison of the populations was undertaken following bilateral measurements of different skull traits. In the comparison of populations three aspects were considered: FA, as a measure of variation within the

individual, phenotypic variability among individuals and overall skull size as a measure of absolute morphological difference.

Chapter four considers hypothesis four. Indices of genomic diversity were compared with indirect measures of fitness at the level of the individual within each of the populations.

Chapter five brings together the results and conclusions, comments on the relevance of this study within the context of environmental management decisions for the island of South Georgia and discusses possible areas of future work.

Chapter Two

Direct comparison of genetic diversity between pre-bottleneck and post-bottleneck populations.

2.1 Introduction

The present day reindeer on South Georgia are composed of distinct populations, the Barff herd, founded in 1911 and the Husvik herd, founded in 1925. The Barff herd spread across in front of the Cook glacier in the early 1960s to form a further herd in the Royal Bay area (Leader-Williams 1988) but all samples collected for this study were from reindeer in one of the original two areas.

There are a number of sources of information on the first introduction of reindeer onto the Barff peninsula. Although there was some contradiction between original reports, Leader Williams (1988) distinguished between letters and articles written at the time of the introduction (Norwegian newspapers Tidens Tegn 17/10/1911 and Sandefjords Blad 18/10/1911) with later reports that relied on verbal consultation (Olstad 1930). He concluded from the more reliable sources that there were ten reindeer (seven females and three males) that founded the Barff population having been introduced into New Fortuna Bay (renamed Ocean Harbour) in November 1911 (Leader-Williams 1978). Eleven reindeer were sold by Ivar Opdal of Filefjell Reinlag, Hemsedal, southern Norway but one animal died on the journey. Direct descendants of these reindeer are still found at Filefjell Reinlag, Hemsedal

where there is currently a herd of 3000-5000 reindeer (Asgrim Opdal - personal communication).

In the same 1911/1912 season, there was a second introduction of two males and three females into Leith Harbour, Stromness Bay. However, "having increased to a number of about 20, they all perished in a snow slide that swept them into the sea" (Olstad 1930). This Leith population would only be relevant to this study if there were any reindeer that did not die in the snow slide or if there were any carcasses remaining that were still present in 2003 when samples were collected for this study. Apart from Olsen's account, the last recorded sighting of this herd was noted in a letter dated 30th November 1917, written by Edward Binnie, Magistrate in South Georgia. He stated that the Leith herd was last seen in August 1917 when there were 17 reindeer that were often seen moving between Cape Saunders and Fortuna Bay. It seems unlikely that any reindeer would have survived the snow slide without there being a mention in any of the reports. It also seems unlikely that any carcase remnants from the original herd would have remained undisturbed for 85 years as Leith Harbour was the site of considerable human activity with an resident human population throughout each year from 1909 to 1966 (Headland 1984).

There has been confusion as to the details of the third introduction of reindeer. It had been proposed that the present day Barff and Busen herds were from different unrelated stock on account of different behaviour patterns. The Barff herd has been described as wild and untamed whereas

the Busen herd were friendlier, quieter and of more sedentary nature (Bonner 1958; Leader-Williams 1978).

Olstad received verbal information from whaling managers in 1928 and reported that there were three males and four females landed on Husvik Harbour in 1925 . However a short article in the Tønsberg Blåd newspaper (27/8/1924) stated that 'the Tønsberg Hvalfangeri (a whaling company) were sending ten reindeer from Filefjeld down to South Georgia on board the ship 'BUSEN". It is possible that ten reindeer left Norway but that due to deaths on the journey only seven arrived in South Georgia. The discovery of this newspaper article in Tønsberg library archives in 2001, was the first suggestion that the present day herds were both sourced from the same herd, Filefjeld. All literature prior to 2001 assumed different origins for the two herds.

All the literature also previously agreed that there were no further introductions of reindeer since 1925. However, the discovery (in Tønsberg library archives in 2006) of a further report in the Tønsberg Blåd newspaper (22/8/1928) has thrown this into doubt with the suggestion that 'four reindeer females will be sent to-day by BUSEN from Tønsberg to Husvik Harbour, South Georgia.' The article continues with the words: 'those reindeer who were sent a couple of years ago to Husvik Harbour are doing well'. This final comment indicates that this newspaper report had not confused the dates of the initial import but that there was indeed a further introduction.

The Shipping Register, held by the Falkland Island Government Archives, confirms that BUSEN did arrive on South Georgia on 20/09/1928 from Tønsberg but there is no mention of what her cargo was.

In summary, the evidence seems clear that there were seven females and three males introduced to the Barff peninsula from Filefjeld Reinlag, Hemsedal, Norway in 1911 and there have been no further additions to this herd.

There were reindeer around Leith Harbour from 1912 for at least five years but reports suggest there was no remnant of this herd following a snow slide. It seems very unlikely that any samples picked up in 2003 would have been descended from this original herd.

There were probably four females and three males introduced to Husvik Harbour in 1925 though there may have been up to ten reindeer in this founder group. These reindeer also came from Filefjeld Reinlag, Hemsedal, Norway. There may have been a further four female reindeer of unknown origin added to this herd in 1928.

Although there have been no previous genetic studies of the reindeer on South Georgia, considerable ecological, physiological and demographic data exists from extensive studies in the 1970s (Leader-Williams 1988).

Tissue samples have been collected from the present-day Barff and Husvik herds as well as from the source Filefjeld Reinlag herd. Extraction of DNA and subsequent amplification by the polymerase chain reaction (PCR) has yielded information based on thirteen microsatellite loci. Comparable

samples from each population has made it possible to make a direct comparison of the genetic variability of each of the post-bottleneck populations with the pre-bottleneck population.

Alleles are lost when a population experiences a bottleneck event; initially this is due to the sampling effect as the founder animals are selected out of the source population. At this stage, rare alleles are more likely to be lost than alleles that are found more frequently in the population (Nei *et al.* 1975). Further loss of diversity occurs due to inbreeding in small populations, and the smaller the population, the greater the rate of this loss (Wright 1951; Fuerst and Maruyama 1986) .

From the mathematical study of the decline of genetic variability it has been shown that the loss of alleles depends more on bottleneck size than on the subsequent rate of population growth (Nei *et al.* 1975). In contrast, the reduction of average heterozygosity depends not only on the size of the bottleneck but more significantly on the rate of post-bottleneck population growth (Nei *et al.* 1975). This is partly due to the fact that the rare alleles that were easily lost made a limited contribution to heterozygosity (Fuerst and Maruyama 1986). Indeed the proportions of polymorphic loci and the numbers of alleles per locus were shown to be more sensitive indicators than heterozygosity of the differences in genetic diversity between pre- and post-bottleneck populations in experimental populations of mosquitofish (Leberg 1992).

Computer simulation of the predicted consequences of a bottleneck allows comparison between a hypothetical population with the genetic variability of the actual post-bottleneck populations. GENELOSS (England and Osler 2001) uses Monte Carlo sampling of a hypothetical population bottleneck but the population size is kept constant during the bottleneck. BNSIM (Hoelzel *et al.* 1993) uses life-history parameters to simulate both the demographic growth and genetic consequences of the bottleneck. It does not take into account environmental stochasticity but the reiterative approach using random seeds does provide estimates of the variance associated with demographic stochasticity (Hoelzel 1999).

For loci which are neutral to the effects of selection in a natural population, the allele numbers and frequencies are affected by mutation and genetic drift. The mutation parameter (θ) of this mutation-drift equilibrium is dependant on the effective population size (N_e) and the mutation rate (u) so that $\theta=4N_eu$ (Tajima 1983; Watterson 1984).

Classically there were two models used to describe the extremes of mutation in variable number tandem repeats (VNTR). Under the Infinite Allele Model (IAM) a mutation involves any number of tandem repeats and always results in an allele that is not already encountered in the population. Under the Stepwise Mutation Model (SMM), an allele can only mutate by gaining or losing a single tandem repeat and so a mutation may result in a copy of an allele already present in the population; consequently, alleles of very different sizes will be more distantly related than alleles of similar sizes (Shriver *et al.* 1993; Valdes *et al.* 1993; Estoup *et al.* 1995).

A third model, the Two-Phase Mutation Model (TPM), which assumes that most mutations result in an increase or decrease of a single tandem repeat but that larger mutations can also occur, has been suggested as a better fit for microsatellite data (Di Rienzo 1994).

Populations which have experienced a recent bottleneck, or reduction in effective population size, show a reduction in allele numbers and heterozygosity. However, due to the rapid loss of rare alleles, allelic diversity is reduced faster than heterozygosity (Nei et al 1975). This means that observed heterozygosity is larger than the expected gene diversity that would be predicted by the mutational model from the number of alleles, were the population at mutation-drift equilibrium. Thus, populations which have experienced a recent population bottleneck would be expected to exhibit significant heterozygosity excess and this can be used as a molecular 'signature' of a recent bottleneck (Cornuet and Luikart 1996).

Typically in a population at equilibrium, alleles at low frequency are far more common than alleles at intermediate frequency (Chakraborty *et al.* 1980) so that a graph of allele frequency distributions is heavily skewed to the lower frequencies. Due to the greater probability that low frequency alleles are lost in the random sampling of the founder event, there is a redistribution of allele frequencies. This distortion can be modelled and detected as another 'signature' of a recent bottleneck (Luikart *et al.* 1998a).

A third 'bottleneck signature' is the magnitude of the ratio between the number of the alleles and the range in allele size. At the founder event there

is random sampling of alleles of all sizes so the number of alleles is reduced more than the range of allele sizes (Garza and Williamson 2001).

This study was able to use microsatellite data to directly assess the genetic impact of two parallel bottlenecks on a natural population of reindeer. The data was also used to validate the computer simulation models of the effect of bottlenecks and test the utility of commonly used 'bottleneck signatures'.

2.2 Materials and Methods

2.2.1 Collection of samples

The reindeer in the source Norwegian herd are gathered for the culling of the old stock in September and for the slaughter of calves in February. In September 2000 and February 2001, skin samples were collected by punching a circular hole (5mm by 5mm) out of the ear of freshly-slaughtered reindeer.

Some of the Husvik herd were gathered in January 2001 for the translocation of some calves from South Georgia to the Falkland Islands for commercial purposes. Ear punches were taken from 38 of the Husvik reindeer that were still alive on the Falkland Islands in February 2003.

The remainder of the South Georgia tissue samples were taken from the carcasses of reindeer found dead on the island. Where possible a punch of ear was taken (as described above); otherwise any piece of skin was collected. If present, the whole head or skull of the carcass was also collected for the morphometric study (see Chapter Three). Samples were

collected by expeditions to the Barff peninsula in January 2000 and January 2002 and expeditions to the Husvik area in February 2003 and December 2003.

All skin samples were stored in 20% DMSO/saturated NaCl solution (Amos and Hoelzel 1991).

When there had been a skull collected but there was no corresponding tissue sample, one of the teeth was removed for DNA extraction.

There were a total of 97 Norwegian samples, 63 Barff samples and 59 Husvik samples from which DNA was extracted.

2.2.2. DNA extraction and isolation from skin samples

A small sample (approximately $3mm^3$) was cut off the original specimen and finely chopped in an Eppendorf tube. Between each sample the scissors and forceps were kept in ethanol and then cleaned in distilled water to prevent cross contamination of samples. Samples were digested at 37°C overnight in 500µl of digestion buffer (50mMol Tris pH7.5, 1mM EDTA, 100mM NaCl, 1%w/v SDS) with 30µl Proteinase K (10mg μl^{-1}). The EDTA was used to chelate the divalent cations, the salt to stabilise the nucleic acid and the anionic detergent, SDS to increase the solubility of the cell membrane (Milligan 1998).

The DNA was extracted with two phenol stages to remove proteins and carbohydrates, and chloroform to remove the phenol. Sodium acetate was

added (volume 1:10) and 100% ethanol (volume 2:1) to precipitate the DNA. This was centrifuged for 15 minutes at 100,000 xg and the ethanol poured off the DNA pellet. 70% ethanol was used to clean this pellet and it was dried in a vacuum centrifuge. The dry pellet was resuspended in 250µl TE (1 x Tris EDTA buffer) at 65°C. The protein was precipitated by the addition of equal volumes of lithium chloride. The tube was inverted and kept at -20 °C for one hour. After 10 minutes in the centrifuge, the supernatant was drawn off. The DNA was precipitated out by addition of 100% ethanol. The pellet was again cleaned with 70% ethanol and dried in the vacuum centrifuge. The final DNA pellet was resuspended in 600µl TE at 65 °C and stored at -20 °C. Following extraction, 3µl of each DNA sample was run on a 0.8% agarose gel at 100W against a 1kB marker to allow the visualisation of the DNA.

It was necessary to undertake purification of some of the older South Georgia samples by further gel extraction. For this procedure 15µl DNA in TE buffer was loaded into a 1% agarose gel in a TAE buffer. Approximately 0.3g gel was excised and extraction carried out using the recommended protocol of the QIAquick Gel Extraction Kit™.

2.2.3 DNA extraction and isolation from teeth

DNA sampling from teeth followed standard protocols designed to avoid contamination. These procedures were performed in a laboratory which was not used for other PCR or post-PCR work. It was distant from the laboratory used for skin samples and, as a general rule, materials that had

been used in the skin sample laboratory were not subsequently taken to the teeth laboratory. Any materials that were taken into the teeth laboratory were wiped down with 10% bleach (sodium hypochlorite solution) and placed in front of the ultra violet lamp for at least 10 minutes.

Sampling from the teeth was performed in a laminar flow hood. All surfaces were wiped down with 10% bleach at the start and completion of each tooth sampling. Drill bits used for sampling were soaked in 10% bleach for 10 minutes and then placed in ultra violet light for a further 10 minutes. Water and other solutions prepared were filtered through a 0.2µm syringe filter and autoclaved. The ultra violet light was left on all the time that the laboratory was not being used. Pipette tips were certified sterile by the manufacturers. A laboratory coat, dedicated for use only in this laboratory, was worn with plastic gloves taped to the sleeves to avoid contamination from wrist skin. Controls carried out in parallel with all isolations and PCR reactions to monitor for any contamination (Milligan 1998).

Teeth were extracted from the skull and brushed with 10% bleach. They were placed in a solution of 10% bleach for six hours in the shaker in the incubator at 37°C. Each tooth was removed from the bleach, brushed and rinsed in distilled water, rinsed in 100% ethanol and left to dry in the incubator overnight.

The teeth were sampled by drilling through the proximal end into the area where the dental pulp had been located. A small hand-held drill and Dremmel™ drill bits were used. Designed for engraving, these drill bits have

rough spherical or pyramid-shaped tips which bore into the teeth creating a fine powder. Aluminium paper was folded to make a tray to collect the powder from drilling each tooth. The powder was collected into sterile 10 ml tubes and drilling continued until there was approximately 0.5ml powder. At the start of drilling, an empty tube was capped and placed under the hood, to be used as an extraction control.

A high ethylenediaminetetraacetic acid (EDTA) digestion buffer (0.425 M EDTA, 0.5% sodium dodecyl sulfate, 0.05 Tris) (Hagelberg and Clegg 1991) was mixed from certified sterile components. Once made up, the digestion buffer was exposed to UV irradiation for 10 minutes to destroy any contamination that occurred during preparation. Approximately 5 ml of buffer was added to 0.3ml powder and 8 ml buffer to 0.5 ml powder with a gradient for intermediate volumes. 3 ml digestion buffer was added to each of the empty control tubes and they were treated identically to the sample tubes.

20µl Proteinase K (50 µg μl^{-1}) was added to each tube and they were left in the shaker in the incubator at 37°C for between 48 and 96 hours. The tube was spun (60,000 xg) to remove the suspension and the supernatant drawn off to a new tube. Extraction proceeded following the protocol of the Qiagen™ QIAquick PCR Purification kit™.

2.2.4 Polymerase Chain Reaction

Table 2.1 Table to show details of microsatellite primers with their sequence, PCR conditions, relevant dye and reference.

Locus	Genebank accession number	Size range	Dye	Standard Taq (Bioline™)		Qiagen QTAQ ™	Primer sequence	Reference
				$MgCl_2$	Temp	Temp		
RT9	U90741	116-128	Blue D4	1.75	54°C	55°C	TGA AGT TTA ATT TCC ACT CT CAG TCA CTT TCA TCC CAC AT	(Wilson *et al.* 1997)
RT5	U90738	143-171	Blue D4	0.60	50°C	n/a	CAG CAT AAT TCT GAC AAG TG AAT TCC ATG AAC AGA GGA G	(Wilson *et al.* 1997)
RT1	U90737	222-240	Blue D4	0.75	50°C	n/a	TGC CTT CTT TCA TCC AAC AA CAT CTT CCC ATC CTC TTT AC	(Wilson *et al.* 1997)
RT13	U90743	293-314	Blue D4	0.75	50°C	53°C	GCC CAG TGT TAG GAA AGA AG CAT CCC AGA ACA GGA GTG AG	(Wilson *et al.* 1997)
NVHRT03	AF068204	112-126	Green D3	0.75	51°C	53°C	TGG AGA GCT GAG TAT GAA AG AGA AAT GCA GCT ACC TAA AAG	(Roed and Midthjell 1998)
NVHRT22	AF068208	142-168	Green D3	1.00	54°C	58°C	GTA TTC TTG CCA GGA AAA ACC GTT GCT TCA GTG CTC TCA GAT	(Roed and Midthjell 1998)
NVHRT73	AF068218	219-231	Green D3	1.00	54°C	53°C	CTT GCC CAT TTA GTG TTT TCT TGC GTG TCA TTG AAT AGG AG	(Roed and Midthjell 1998)
RT27	U90748	135-155	Black D2	0.75	50°C	55°C	CCA AAG ACC CAA CAG ATG TTG TAA CAC AGC AAA AGC ATT	(Wilson *et al.* 1997)
RT7	U90740	216-234	Black D2	1.00	50°C	55°C	CCT GTT CTA CTC TTC TTC TC ACT TTT CAC GGG CAC TGG TT	(Wilson *et al.* 1997)
BM848	G18477	355-401	Blue D4	2.25	60°C	55°C	TGG TTG GAA GGA AAA CTT GG CCT CTG CTC CTC AAG ACA C	(Cronin *et al.* 2003)

NVHRT71	AF068217	109-123	Green D3	1.00	57°C	55°C	GAG TTG GCA GGT GTA AAG G CAG TGG GGG AAA TGA TG	(Roed and Midthjell 1998)
Ca13	AY302222	213-225	Green D3	2.00	61°C	58°C	CAG AAA GTT GTG AGG CAC AG GTG GCC TCT GTT TCA GTG TA	Molecular Ecology Notes (in press)
CRH	M22853	229-251	Green D3	1.75	60°C	55°C	CTC GCT CAC CTG CAG AAG CAC C GCT GAG CAG CCG TCT AAG TTG C	(Cronin *et al.* 2003)
Ca71	AY302228	308-318	Green D3	1.50	*60°C	53°C	TGC ACA CCC CCA GTC TGG T GTC TCA CCT TTC CCA TCA GC	Molecular Ecology Notes (in press)
RT30	U90749	190-220	Black D2	1.25	51°C	55°C	CAC TTG GCT TTT GGA CTT A CTG GTG TAT GTA TGC ACA CT	(Wilson *et al.* 1997)

*Step down occurred so that there were two cycles at 62°C followed by 38 cycles at 60°C.

Fifteen published microsatellite loci were used. The loci are listed in table 2.1 together with their primer sequence, reference, optimum $MgCl_2$ concentration and annealing temperatures used for amplification.

PCR amplification was carried out in 20µl reactions. The reaction mix was 2µl Tris buffer, 2 µl dNTP mix (0.2mM concentration), 0.4-0.8 µl Bovine Serum Albumin (20mg ml^{-1}), 0.4 µl of each primer (0.5 µg $µl^{-1}$ in 20% TE), 0.08 µl TAQ (5 units $µl^{-1}$), 0.4µl DNA and magnesium chloride. The program started with 3 minutes at 95^0C for denaturing. The cycle profile continued for 45 seconds at 95^0C, 90 seconds at the annealing temperature and 90 seconds at the extension temperature, 72^0C. The cycle was repeated 35 times and then held at 72^0C for 8 minutes for the extension stage. Some of the more difficult samples were run for up to 42 cycles. 25% of one primer in each pair was labelled with a fluorescent dye at the 5' end of the oligonucleotide.

A number of the older skin samples from South Georgia showed good evidence of DNA on the gel run following extraction but proved extremely difficult to amplify despite the use of a number of different techniques. PCRs were attempted following dilution of the amount of DNA (by between ten and a hundred fold) and following further purification of the samples with Qiagen™ PCR Purification kit. 'Hot start' technique was tried as was stabilisation of the reaction by the addition of DMSO, glycerol or 1% Triton X. All these methods were unsuccessful. These difficult South Georgia samples and a number of the tooth extracted samples were amplified with Qiagen™ Multiplex PCR Master Mix containing HotstarTaq™ DNA Polymerase, Qiagen

buffer (containing 6mM $MgCl_2$) and dNTP mix of dATP, dCTP, dGTP and dTTP. The conditions are shown in the table 2.1.

Microsatellite PCR products were run through the capillary injection Beckman Coulter™ CEQ™ 8000 Fragment Analysis System with up to 10 PCR products multiplexed. The amount of each PCR product was varied depending on the dye label and the strength of signal following some trial runs.

Before loading the plates, it was necessary to clean excess salt from the PCR products. Between 0.5µl and 6µl of each dyed PCR product was mixed with 20µl water, 4µl Sodium Acetate and 100µl 100% ethanol. This was kept at -20 °C for 30 minutes and then spun at 20,000 xg for 15 minutes. The ethanol was removed by pipette leaving the pellet of DNA which was cleaned by two further steps with 70% ethanol. The pellet was resuspended in 40 µl CEQ™ Sample Loading Solution (SLS).

Loading buffer was made up of 320µl CEQ™ SLS and 2µl CEQ™ DNA size standard kit; this was divided into the 8 sample wells on the plate. 4µl of mixed and cleaned product was added to each sample well. This was overlaid with sticky aluminium foil sheet and vortexed to mix. A drop of mineral oil was added to each sample before running through the CEQ™ 8000 Fragment Analysis System by DBS Genomics (University of Durham). The amplification product signals were visualised using the CEQ™ Fragment Analysis software.

2.2.5 Analysis of results

The presence of null alleles would confound the results by underestimating the true level of heterozygosity (David 1998). The presence of null alleles, stuttering or large allele dropout was checked using the program Microchecker (http://www.microchecker.hull.ac.uk). This program uses a Monte Carlo simulation (bootstrap) method to calculate expected homozygote and heterozygote allele size difference frequencies and the Hardy-Weinberg theory of equilibrium to calculate expected allele frequencies. By comparison of the observed number of homozygotes with the expected, the program is able to detect the possible presence of null alleles.

The program GENEPOP (Raymond 1995) was used to conduct Hardy Weinberg Exact tests to consider the null hypothesis that gametes were in random union and Fisher Exact tests to consider the presence of linkage disequilibrium within populations and to test for both genotypic and allelic differentiation between populations. The program creates a contingency table for all pairs of loci and performs a probability test (Fishers Exact test) for each table using a Markov chain (Guo and Thompson 1992). In each case, the dememorization period for the Markov chain was 1000 steps long and there were 100 batches of 1000 iterations. Each set of results was tested for consistency by re-running the test with the number of iterations increased to 10000.

The number of observed alleles in a sample is highly dependant on the sample size (Nei 1987) and in this study unequal sample sizes may have been

a confounding factor in the comparison of allele numbers between the populations. For this reason allelic richness, which is a measure of the number of alleles independent of sample size, was calculated using FSTAT (Goudet 1995).

Differences between the populations in the numbers of alleles, allelic richness and the expected and observed heterozygosity were assessed using Wilcoxon's signed ranks test (as done by Whitehouse and Harley 2001, Le Page et al. 2000). The Wilcoxon signed-rank test is a non-parametric alternative to the paired student's t-test. Like the t-test, it compares differences between measurements but it does not require assumptions about the form of distribution of the measurements so it is more appropriate when the data is not distributed normally.

'Coancestry' or inbreeding due to differentiation among populations was considered by calculation of the fixation index, F_{st}. The program ARLEQUIN (Schneider 2000) calculated F_{st} by Slatkin's distance (Slatkin 1995b) and tested whether the results were significantly different from zero. F_{st} is a function of probabilities of identity (either by descent or by state) and thus has a lowered expectation when the mutation rate is high (as common for microsatellites) (Balloux and Goudet 2002)

Further quantification of population differentiation was undertaken by calculation of R_{ST} which, in contrast to of F_{ST}, assumes SMM and is little affected by mutations. R_{ST} is based on the variance in allele size (in terms of number of repeat units) that is between populations and thus accounts for

evolutionary distance between alleles (Slatkin 1995b). R_{ST}, computed using R_{ST}Calc (Goodman 1997), was estimated as Rho values to account for differences in sample sizes and was calculated following standardization of the data to correct for differences in the magnitude of variance in allele size across the loci (Goodman 1997). The significance of the R_{ST} values were tested by permutation tests (Hudson 1992) and bootstrapping (Efron 1979) with 1000 iterations in each case.

ARLEQUIN was also used to calculate average gene diversity over all loci. For this calculation the allowed level of missing data was set at 40% so that all thirteen loci were still considered for each population.

2.2.6 Bottleneck Simulation Programs

Monte Carlo sampling to simulate the effects of the bottleneck on gene diversity was used via two computer programs, GENELOSS and BNSIM. GENELOSS (England and Osler 2001) does not simulate population growth and thus does not consider demographic processes such as population growth rate and the effect that this may have on the survivorship of genetic diversity. The data used in the GENELOSS program assumed monogamy with equal numbers of male and female reindeer contributing their genes. BNSIM (Hoelzel *et al.* 1993) simulates population growth based on life-history parameters such as age-specific mortality and reproductive data and combines this information with genetic data to predict the effect of the bottleneck on genetic diversity.

Each of the simulation models requires input of the numbers of reindeer in the founder population. There is no controversy as to the numbers in the founder population for the Barff herd: seven females and three males. The Husvik herd may have been founded by as few as seven reindeer (four females and three males) though there may have been up to ten in the original group and four additional females added in 1928.

A dominance hierarchy exists amongst the male reindeer and they are highly polygynous with large breeding groups (Leader-Williams 1988). For this reason it would be reasonable to suspect that not all males in the founder groups were able to mate and thus contribute to the post-bottleneck gene pool.

GENELOSS requires data entry to state the number of breeding pairs with no flexibility for polygynous animals. The model was run twice with three and seven breeding pairs in the founder group and the expectation was that the actual results would lie somewhere in between these two extremes. BNSIM was run with ten reindeer in the founder group, as a exact representation of the Barff founder group and as an average of the possible numbers in the Husvik founder group.

2.2.6.1 Herd numbers used in demographic program

The total number of reindeer in each of the herds has been counted or estimated on a number of occasions. Table 2.2 and 2.3 display a summary of those results, the accuracy of which vary considerably. The numbers counted

by Leader Williams between 1972 and 1976 are probably the most reliable, as over this period considerable research of the reindeer was undertaken and detailed accounts of the counting methodology have been published.

Table 2.2 Number of reindeer in the Barff herd

Date	Number of deer in Barff herd	Reference
1911	10	(Leader-Williams 1978)
1916	45	SG file 650 (various)
1920	~120	SG file 650 (various)
1921-22	nearly 300	(Wilkins 1925 quoted in Leader-Williams 1988)
1928	400-500	(Olstad 1930)
1953	2000	SG file 650 (various)
1955-57	3000-5000	(Bonner 1958)
1964	approx similar nos to 1972	(Leader-Williams 1978)
1972	2100	(Leader-Williams 1988)

Table 2.3 Number of reindeer in the Husvik herd

Date	Number of deer in Husvik herd	Reference
1925	7	(Olstad 1930)
1953	40	SG file 650 (various)
1957	100-200	(Bonner 1958)
1973	785	(Leader-Williams 1988)
1976	~600	(Leader-Williams 1988)
1993	800	(Moen and MacAlister 1994)
2000	~1000	(Bell 2001)

The program BNSIM runs a model based on reproductive data and life history tables to simulate the population growth following a bottleneck. This program was used to simulate the actual population growth seen in both the Husvik and Barff herds. See figure 2.1 and 2.2.

Figure 2.1 Barff herd numbers

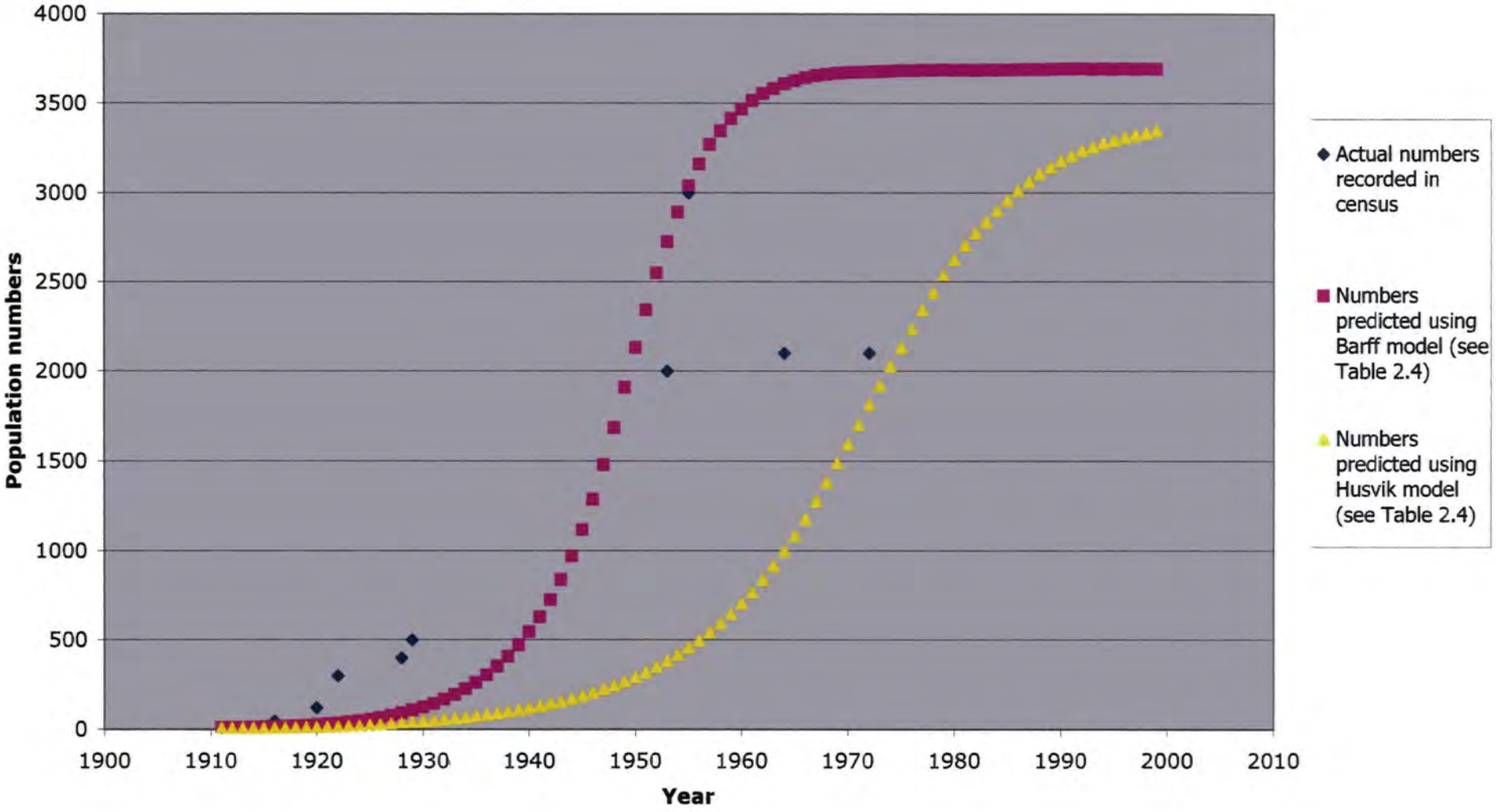
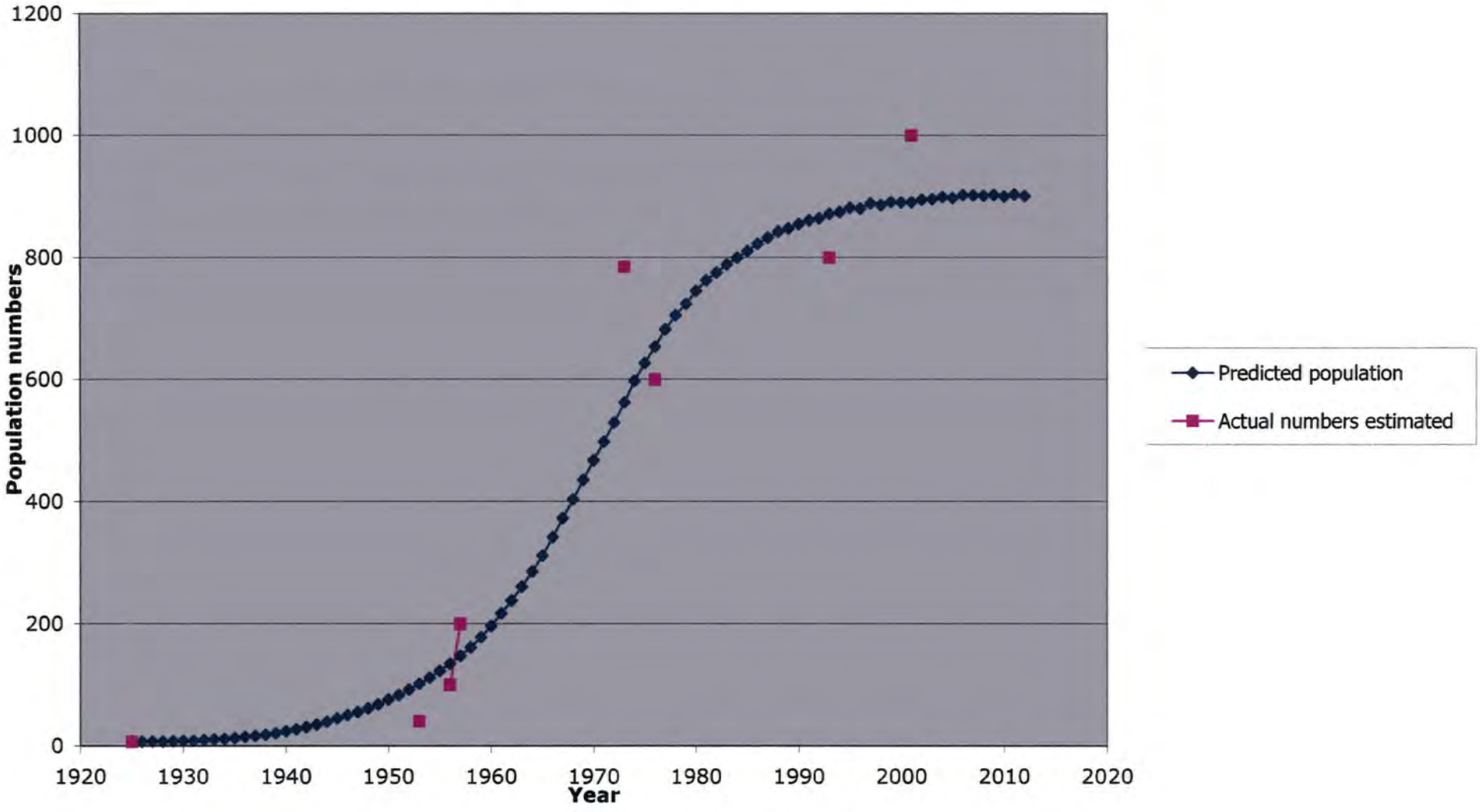


Figure 2.2 Husvik herd numbers



When run with the actual life history data calculated by Leader Williams for the Barff herd over the period between 1973 and 1976 (Leader-Williams 1988), the model populations did not grow at rates that resembled the actual population growth. For this reason, the survivability data calculated by Leader Williams were adjusted within a credible range so that the numbers predicted by the model correlated more closely with the actual numbers counted on the ground. See table 2.4 for the survivability data used.

The figures used for reproductive success for females were exactly the same as the pregnancy rates reported by Leader Williams (1988). The figures used for reproductive success of male reindeer should have been considerably lower than that of females to reflect the highly polygynous nature of reindeer. However, the figures that were actually used (table 2.5) were such that the model predicted population numbers that correlated as closely as possible with actual population growth.

Figure 2.1 Barff herd numbers – the pink line represents the predicted population growth (with Barff survivability data' see table 2.4) that was used in the model. The yellow line represents what the predicted population growth would be were the Husvik survivability data used. The blue dots represent actual counts or estimates.

Figure 2.2 Husvik herd numbers – the blue line represents the predicted population growth (with 'Husvik survivability data' see table 2.4)

that was used in the model. The pink dots represent actual counts or estimates.

The estimates of the Barff herd from 1916 to 1928 (various; Olstad 1930) appear to be greatly inflated as these figures would not be achievable even with 100% reproductive success and zero mortality up to the age of 12 years old!

Table 2.4 Survivability data reported by Leader Williams (1988) and adjusted figures used in BNSIM models.

Age group	Survivability (Leader-Williams 1988)		Survivability data used in Husvik model		Survivability data used in Barff model	
	Male	Female	Male	Female	Male	Female
0-1	0.53	0.50	0.59	0.57	0.68	0.65
1-2	0.58	0.97	0.63	0.97	0.73	0.97
2-3	0.65	0.96	0.70	0.96	0.80	0.97
3-4	0.53	0.90	0.58	0.90	0.68	0.97
4-5	0.43	0.82	0.48	0.87	0.58	0.97
5-6	0.31	0.72	0.36	0.77	0.46	0.87
6-7	0	0.61	0	0.66	0.15	0.76
7-8		0.52		0.57	0	0.67
8-9		0.42		0.47		0.57
9-10		0.33		0.38		0.48
10-11		0.27		0.32		0.42
11-12		0		0		0.15
12-13						0

The simulation model requires that the population exhibits density-independent growth. From field observations and estimations of numbers it appears that the Barff population reached a zenith in the late 1950s and then

reduced to approximately 2000 animals (Leader-Williams 1988). As can be seen from the graph, the parameters used in the model predicted a population that reached the maximum observed on the ground. It was not possible to further model the subsequent fall in population numbers. The holding population was set at 1500 for the Husvik and 5000 for the Barff population. The mutation rate was set at $1x10^{-4}$ (Approximate mean rate for mammal microsatellites summarised in Frankham 2002).

Table 2.5 Pregnancy rates reported by Leader Williams (1988) and reproductive success figures used in BNSIM models.

Age group	Pregnancy rates (Leader-Williams 1988)	Reproductive success figures used in both models	
		Male	Female
0-1	0	0	0
1-2	89	0.5	0.89
2-3	90	0.8	0.9
3-4	92	0.8	0.92
4-5	92	0.8	0.92
5-6	92	0.8	0.92
6-7	92	0.8	0.92
7-8	92		0.92
8-9	92		0.92
9-10	92		0.92
10-11	92		0.92
11-12	92		0.92

2.2.7 Bottleneck Signatures

The theory behind the infinite allele model (IAM), the stepwise mutation model (SMM) and the two-phased mutation model (TPM) has been explained in the introduction. The TPM model used for this study was composed of 70% IAM and 30% SMM model.

2.2.7.1 Heterozygosity excess

The program Bottleneck (Cornuet and Luikart 1996) was used to compute the distribution of gene diversity (H_{eq}) expected from the observed numbers of alleles (k), given the sample size (n). This works by simulating the coalescent process under the three mutation models described above, IAM, SMM and TPM assuming mutation-drift equilibrium and thus enabling computation of the average H_{eq} to be compared to the observed heterozygosity (H_{obs}). In a population at mutation-drift equilibrium, there would be approximately equal probability that a locus shows gene diversity excess or deficit.

Three separate statistical tests were used: sign test, standardized differences test (Cornuet and Luikart 1996) and a Wilcoxon sign-rank test (Luikart and Cornuet 1998). The sign test determines whether the proportion of loci with heterozygosity excess is significantly larger than expected at equilibrium. This is a nonparametric test, which does not require further assumptions but it has low statistical power.

The standardized-differences test establishes whether the average of standardized difference between observed and expected heterozygosities is significantly different from zero. However this test relies on a Gaussian distribution of the statistic T_2 which requires a minimum of 20 polymorphic loci (Cornuet and Luikart 1996). There were only thirteen loci used in this study.

The Wilcoxon sign-rank test can be used with as few as four polymorphic loci and any number of individuals (Luikart and Cornuet 1998).

2.2.7.2 Mode-shift distortion in the distribution of allele frequencies

Luikart and co-workers (1998a) illustrate how populations that are at mutation-drift equilibrium have a large proportion of alleles which are at very low frequency. In contrast, recently bottlenecked populations characteristically show a mode-shift in the distribution of allele frequencies so that alleles of very low frequency (<0.1) are less abundant than alleles that occur more frequently. They claimed that this signature is apparent in samples of 5 – 20 polymorphic loci and approximately 30 individuals but that it only lasts for between 40 and 80 post bottleneck generations. For this reason the generation length of the reindeer on South Georgia was calculated.

Generation length

The generation length was calculated from published life tables and pregnancy rates of South Georgia reindeer (Leader-Williams 1988). Generation length (T) is the mean lapse of time between a female's date of birth and the mean date of birth of her offspring (Caughley 1966) so that

$$T = \frac{\Sigma \ell_x m_x x}{\Sigma \ell_x m_x}$$

where ℓ_x is the probability of surviving to age x and m_x is the number of female live births per female at age x. Generation length for South Georgia reindeer was calculated to be 4.2 years. There have been 22 generations since the foundation of the Barff herd and 19 generations since the foundation of the Husvik herd.

2.2.7.3 Garza's M Ratio

Allele frequency distributions contain information about both the frequency and total number of the alleles k, as well as the distance between the number of repeats and the overall range in allele size r. In a population that is reduced in size, there will be a reduction in k, but only the loss of the largest of smallest allele will cause a reduction in r. Thus it is expected that the ratio M=k/r will be smaller in those populations that have recently been reduced in size compared to populations at equilibrium (Garza and Williamson 2001).

2.3 Results

2.3.1 Basic parameters of genetic diversity

The presence of null alleles was highly likely ($p<0.0001$) for two loci (NVHRT71 and RT7) so no further analysis was performed with these two loci. There was no evidence of null alleles, stuttering or large allele drop-out with any of the other loci.

Hardy Weinberg Exact tests (sequential Bonferroni correction applied) showed that there was one locus (RT27) in the Husvik population in which there was significant deviation from the hypothesis that the gametes were in random union (at the level $p<0.05$). There was no significant deviation for any loci in either the Norwegian or Barff populations.

After application of the sequential Bonferroni correction, there was significant linkage (at the level $p<0.01$) between loci RT27/NVHRT22 and NVHRT22/NVHRT03 in the Husvik population. There was no significant linkage disequilibrium within the Norwegian or Barff populations.

Linkage disequilibrium may result from close physical linkage of loci (Hill 1977) but if this were the case here, one might expect evidence across each of the populations. It has been suggested that the presence of linkage disequilibrium may be characteristic of severely bottlenecked populations due to a natural homozygote 'stocking effect' (Clegg *et al.* 1980; Houlden *et al.* 1996), an artifact which occurs as a result of the elevated number of homozygotes in a bottlenecked population.

GENEPOP was used to perform a log-likelihood (G) based exact test for genotypic differentiation and a Fishers Exact test to give an unbiased estimate of the p-value for allelic differentiation between the populations. There was no significant genotypic or allelic differentiation between Norway and Barff for locus NVHRT73, between Barff and Husvik for locus CRH or between any of the populations for locus Ca71. However, there was significant differentiation between each population for all other loci. When all loci were considered together the value of X^2 was infinity which was highly significant (df=26, $p<0.001$)

Table 2.6 Basic genetic parameters for each population

	No of genotypes	No of Alleles	Allelic richness	Expected Hetero-zygosity	Observed Hetero-zygosity	Gene Diversity over all loci
Norway						
Average	92.8	11.2	9.4	0.75	0.75	0.661
SD	5.6	3.8	3.0	0.10	0.08	0.34
Range	78-99	4-17	3.4-13.6	0.44-0.85	0.54-0.84	
Barff						
Average	49.2	6.7	6.6	0.72	0.67	0.526
SD	6.3	1.8	1.8	0.09	0.07	0.28
Range	41-58	4-10	4.0-9.7	0.54-0.85	0.56-0.82	
Husvik						
Average	52.2	5.8	5.7	0.59	0.54	0.493
SD	4.0	2.1	1.9	0.20	0.19	0.26
Range	43-57	2-9	1.9-8.6	0.04-0.78	0.04-0.71	

Table 2.7 Proportion of pre-bottleneck genetic diversity present in post-bottleneck populations

Compared to Norway	% Numbers of alleles	% Heterozygosity	% Gene diversity
Barff	60.0	95.5	80.0
Husvik	53.1	78.4	74.6

Table 2.8 Comparison of genetic diversity between pre-bottleneck population and both post-bottleneck populations

Comparison between	Expected Heterozygosity		Observed Heterozygosity		Number of alleles		Allelic richness	
	z	p-value	z	p-value	z	p-value	z	p-value
Norway & Barff	-1.19	ns	-2.75	0.001	-3.07	0.002	-3.04	0.002
Norway & Husvik	-3.04	0.002	-3.11	0.002	-3.20	0.001	-3.18	0.001
Barff & Husvik	-2.62	0.009	-2.13	0.03	-1.62	ns	-1.85	ns

The numbers of alleles, allelic richness, heterozygosity and the overall gene diversity of the post-bottleneck populations were lower than that of the Norwegian population. The values were compared using the Wilcoxon Signed Ranks Test. Expected and observed heterozygosity and numbers of alleles were significantly different in Husvik compared to Norway (at the level $p<0.01$). Observed heterozygosity, numbers of alleles and allelic richness were significantly different in Barff compared to Norway (at the level $p<0.01$). There was a significant difference in the heterozygosities between the Barff and the Husvik populations but not in the number of alleles or allelic richness.

2.3.2 Private alleles

The tables of alleles present in each population in this study are presented in Appendix 1. There was one allele present at a low frequency in both the post bottleneck populations which was not found in the Norwegian population (Allele 160, locus NVHRT 22). There were three further private alleles in the Husvik population: allele 132, locus NVHRT03 and allele 228, locus NVHRT73, present at low frequencies and allele 176 which accounted for more than 20% of locus NVHRT22 in the Husvik population.

Alleles may be currently present in the Norwegian population but rare and consequently not found in this study due to sampling stochasticity. A binomial test was undertaken to assess the probability of missing a rare allele in the 97 Norwegian samples. It would be unlikely (at probability $p<0.05$) to miss a single allele present in the Norwegian population if it was present at more than 1.5% frequency. However, the cumulative probability of missing four different alleles makes sampling stochasticity an extremely unlikely explanation for the presence of four private alleles in the Husvik population and this situation is further explored in the discussion.

2.3.3. Measure of inbreeding due to differentiation among subpopulations

The fixation indices, F_{st} were significantly different from zero (at level $p<0.001$) which indicated that there was significant differentiation between each of the populations. Approximately 2% of the differences between Barff

and Norway and approximately 7% of the differences between Husvik and Norway or Husvik and Barff could be explained by differences between the populations rather than within each population.

Table 2.9 Fixation Indices for each population

F_{st} (Slatkin 1995)		
	Norway	Barff
Barff	0.0164***	
Husvik	0.0679***	0.0720***

*** indicates significance at the level $p<0.001$

Table 2.10 Rst (Rho value) for each population

R_{st} - Rho value averaged over variance components and loci (Goodman 1997)		
	Norway	Barff
Barff	0.0452***	
Husvik	0.0832***	0.0515***

*** indicates significance at the level $p<0.001$

There was significant differentiation between the three populations (at level $p<0.001$) when the R_{st} (Rho value) was tested by both a permutation test and bootstrapping. The assessment of the significance of the values by permutation to yield an unbiased p-value is a very powerful test (Balloux and Lugon-Moulin 2002).

2.3.4 GENELOSS program

GENELOSS was run with the allele frequencies of the Norway population and a bottleneck length of one generation, for both three breeding pairs and seven breeding pairs in the bottleneck. The predicted post bottleneck allele numbers and heterozygosity are displayed in table 2.11 and 2.12 for comparison with the actual post bottleneck figures.

The actual number of alleles and the heterozygosity of the post bottleneck populations show close agreement with the values predicted by GENELOSS.

Table 2.11 Comparison of numbers of alleles

Loci	ACTUAL			GENELOSS predicted post-bottleneck			
	pre-BN	post-BN	post-BN	3 BREEDING PAIRS		7 BREEDING PAIRS	
	Norway	Husvik	Barff	mean	SD	mean	SD
RT27	15	5	5	6.08	1.27	9.32	1.46
RT30	11	5	7	4.56	1.10	6.46	1.18
RT13	12	6	6	6.25	1.09	8.38	1.10
NVHRT22	13	7	7	5.34	1.18	7.66	1.35
CA13	6	5	5	4.19	0.69	4.76	0.60
CA71	4	3	4	2.67	0.59	3.06	0.36
NVHRT03	9	8	6	5.05	0.98	6.58	0.98
NVHRT73	12	7	7	4.18	0.96	5.63	1.16
CRH	6	4	5	4.00	0.77	4.86	0.76
RT9	13	2	7	5.24	1.20	7.73	1.24
BM848	17	9	9	5.68	1.23	8.34	1.51
RT5	14	8	9	5.48	1.06	7.41	1.18
RT1	13	7	10	6.05	1.09	8.28	1.23
average	**11.15**	**5.85**	**6.69**	**4.98**		**6.80**	
SD	3.85	2.08	1.80	1.03		1.80	
Range	4-17	2-9	4-10	2.7-6.3		3.1-9.3	

Table 2.12 Comparison of actual heterozygosity in each population and heterozygosity predicted by GENELOSS for three or seven breeding pairs in the founder population

Loci	ACTUAL			GENELOSS predicted post-bottleneck			
	pre-BN	post-BN	post-BN	3 BREEDING PAIRS		7 BREEDING PAIRS	
	Norway	Husvik	Barff	mean	stddev	mean	stdev
RT27	0.82	0.49	0.68	0.75	0.08	0.78	0.05
RT30	0.72	0.73	0.78	0.65	0.11	0.69	0.07
RT13	0.85	0.78	0.65	0.78	0.06	0.81	0.04
NVHRT22	0.78	0.65	0.71	0.71	0.08	0.75	0.05
CA13	0.75	0.49	0.67	0.68	0.07	0.72	0.04
CA71	0.44	0.40	0.54	0.40	0.15	0.42	0.09
NVHRT03	0.78	0.73	0.69	0.71	0.07	0.75	0.05
NVHRT73	0.72	0.71	0.77	0.65	0.08	0.69	0.05
CRH	0.73	0.62	0.71	0.67	0.07	0.70	0.04
RT9	0.73	0.03	0.62	0.67	0.12	0.70	0.08
BM848	0.81	0.61	0.84	0.73	0.07	0.77	0.05
RT5	0.80	0.72	0.81	0.73	0.08	0.76	0.05
RT1	0.84	0.70	0.85	0.76	0.06	0.80	0.04
average	**0.75**	**0.59**	**0.72**	**0.68**		**0.72**	
SD	0.10	0.20	0.09	0.10		0.10	
Range	0.44-0.85	0.03-0.78	0.54-0.85	0.40-0.78		0.42-0.81	

2.3.5 BNSIM

Table 2.13 Comparison of numbers of remaining allele predicted by BNSIM model with actual Barff population

	Model – Barff					Actual
			Number of alleles remaining post-bottleneck			numbers of alleles post-bottleneck
Name of loci	Populations which died	Populations which survived	Mean	Median	Range	
RT27	1	499	5.2	5	0-10	5
RT30	2	498	4.4	4	0-8	7
RT13	0	500	5.5	6	1-10	6
NVHRT22	0	500	4.8	5	1-9	7
Ca13	3	497	4.1	4	0-7	5
Ca71	1	499	2.8	3	0-5	4
NVHRT03	0	500	4.8	5	1-9	6
NVHRT73	0	500	4.1	4	1-8	7
CRH	1	499	4.0	4	0-7	5
RT9	0	500	4.7	5	1-11	7
BM848	3	497	4.9	5	0-9	9
RT5	3	497	4.9	5	0-9	9
RT1	5	495	5.3	5	0-9	10

For each of the thirteen loci, the BNSIM model was run for 500 simulations and it predicted that for the Barff population between 497 and 500 of the 500 simulation populations would survive. For each surviving population, the model predicted there would be between one and eleven alleles that survived with a mean between 2.8 and 5.5 and a median of between three and six. In reality there were between one and ten surviving alleles with a mean of 6.7 and a median of seven.

Table 2.14 Comparison of numbers of remaining allele predicted by BNSIM model with actual Husvik population

Model – Husvik						Actual numbers of alleles post-bottleneck
Name of loci	Populations which died	Populations which survived	Number of alleles remaining post-bottleneck Mean	Median	Range	
RT27	29	471	3.6	3	0-7	5
RT30	12	488	3.3	3	0-9	5
RT13	32	468	4.0	4	0-8	6
NVHRT22	15	485	3.7	4	0-8	7
Ca13	20	480	3.3	3	0-6	5
Ca71	12	488	2.3	2	0-4	3
NVHRT03	22	478	3.6	4	0-8	8
NVHRT73	21	479	3.2	3	0-6	7
CRH	10	490	3.2	3	0-6	4
RT9	20	480	3.4	3	0-8	2
BM848	23	477	3.7	4	0-8	9
RT5	15	485	3.7	4	0-8	8
RT1	24	476	3.8	4	0-9	7

The BNSIM model predicted that the Husvik population between 468 and 490 of the 500 simulation populations would survive. For each surviving population, the model predicted there would be between one and nine alleles that survived with a mean of between 2.3 and 4.0 and a median of between two and four. In reality there were between one and nine surviving alleles with a mean of 5.8 and a median of six.

Table 2.15 Comparison of heterozygosity predicted by BNSIM with actual Barff and Husvik populations

Loci	Barff			Husvik		
	Actual	BNSIM model predicted		Actual	BNSIM model predicted	
		Mean	Range		Mean	Range
RT27	0.68	0.44	0.003-0.56	0.49	0.26	0.001-0.49
RT30	0.78	0.46	0.003-0.59	0.73	0.29	0.003-0.50
RT13	0.65	0.44	0.05-0.57	0.78	0.26	0.001-0.49
NVHRT22	0.71	0.46	0.01-0.57	0.65	0.28	0.001-0.47
Ca13	0.67	0.47	0.08-0.57	0.49	0.30	0.01-0.50
Ca71	0.54	0.47	0.17-0.58	0.40	0.31	0.02-0.51
NVHRT03	0.69	0.48	0.01-0.57	0.73	0.28	0.01-0.47
NVHRT73	0.77	0.47	0.04-0.57	0.71	0.29	0.003-0.47
CRH	0.71	0.46	0.11-0.58	0.62	0.26	0.001-0.46
RT9	0.62	0.45	0.01-0.57	0.03	0.27	0.01-0.48
BM848	0.84	0.45	0.06-0.56	0.61	0.27	0.003-0.46
RT5	0.81	0.45	0.04-0.57	0.72	0.28	0.002-0.47
RT1	0.85	0.44	0.01-0.56	0.70	0.26	0.001-0.46

With the exception of one locus in the Husvik population, actual heterozygosity was found to be consistently higher than that predicted by the BNSIM model.

2.3.6 Bottleneck Signatures

2.3.6.1. Heterozygosity Excess

Table 2.16 a. Results from Bottleneck program showing the expected and actual numbers of loci with heterozygosity excess under the different mutation models. b.& c. Statistical tests to show how the predicted results compare with the actual Barff and Husvik populations.

a.

Sign Test	Number of loci with heterozygosity excess	I.A.M.	T.P.M.	S.M.M.
Norwegian population	Expected	7.83	7.76	7.63
	Actual	9	5	2
	Probability difference is significant	ns	0.10	0.010
Barff population	Expected	7.72	7.71	7.69
	Actual	13	12	6
	Probability difference is significant	0.0011	0.011	ns
Husvik population	Expected	7.31	7.62	7.60
	Actual	9	5	3
	Probability difference is significant	ns	ns	0.01

b.

Barff population			
Standardized Difference test for Barff T_2	3.17	1.71	-1.38
Probability of difference	0.00077	0.044	0.084
Wilcoxon Test on Barff population			
Probability (one tail for H excess)	0.00006	0.002	0.90
Probability (two tails for H excess or deficiency)	0.00012	0.004	0.216

c.

Husvik population			
Standardized Difference test for Husvik T_2	1.30	-.77	-4.8
Probability of difference	0.096	0.22	<0.001
Wilcoxon Test on Husvik population			
Probability (one tail for H excess)	0.055	0.77	0.99
Probability (two tails for H excess or deficiency)	0.11	0.50	0.003

In the case of the Norwegian population, not one of the statistical tests suggested that it had been though a recent bottleneck.

For the Barff population, each of the statistical tests suggested significantly more heterozygosity excess that that expected at equilibrium, if it were assumed that all the loci fitted the IAM or the TPM mutation-drift model. However, the occurrence of a previous bottleneck was not predicted for Barff if it were assumed that the loci fitted either the SMM model.

The Husvik population showed no evidence of heterozygosity excess regardless of which mutation-drift equilibrium model was used. Indeed if it were assumed that the loci fit the SMM models then this population showed a significant degree of heterozygosity deficit.

2.3.6.2 Mode-shift distortion in the distribution of allele frequencies

For each population in this study the allele frequency distribution was found to be normally L-shaped as expected under mutation-drift equilibrium and there was no evidence of a mode-shift.

2.3.6.3 Garza M ratio

Table 2.17 Calculation of Garza M ratio for thirteen loci for each of the populations and comparison with example populations from other studies.

	M ratio	Variance	Historical status	Reference
Norway reindeer	0.880	0.021	stable	This study
Barff reindeer	0.721	0.034	reduced	
Husvik reindeer	0.723	0.058	reduced	
Polar bear	0.919	0.016	stable	(Paetkau *et al.* 1997)
Brown bear	0.854	0.015	stable	
Brown bear	0.693	0.030	reduced	
Koala	0.835	0.015	stable	(Houlden *et al.* 1996)
Koala	0.599	0.050	reduced	
Northern wombat	0.618	0.081	reduced	(Taylor *et al.* 1994)

The M ratio calculated was tested against a model of 10000 simulations. The model was run twice with a different proportion of steps larger than one step and a different average size for these steps. When 20% of the mutations were set at larger than one step with an average size for these steps of 3.5, a lower value would be expected on 90.9% occasions for the Norwegian population at equilibrium, on 9.5% occasions for the Barff population or on 9.9% occasions for the Husvik population. When 12% of the mutations were set at larger than one step with an average size for these steps of 2.8, a lower value would be expected on 31.1% occasions for the Norwegian population, 0.03% occasions for the Barff population and 0.08% occasions for the Husvik population. Garza and Williamson reviewed a number

of data sets (some examples shown in table 2.16) and they considered that all populations that had not suffered a recent reduction in size had a value of M greater than 0.82 and that a value of M of less than 0.70 indicated that the population has recently gone through a reduction in size. The two reduced populations in this study gave an average M ratio of 0.72 and the stable Norwegian population gave an average M ratio of 0.88.

2.4 Discussion

2.4.1 Private alleles

Although rare alleles do not add measurably to genetic distance statistics, they are good indicators of gene flow (Slatkin 1985). Houlden and co-workers (1996) found a substantial number of private alleles in post-bottleneck populations but they discounted the importance of these due to the fact they were present at a frequency which made their differential detection a likely artefact of sample size.

The presence of private alleles in either or both post-bottleneck populations may be due to the presence of rare alleles in the Norwegian population at the time the South Georgia populations were founded that are now lost due to genetic drift. If one of these alleles was carried by the dominant male reindeer in a founder group, it is possible that it could be found in a high proportion of South Georgia samples despite no corresponding allele in the Norwegian population.

Mutations may have been responsible for the formation of private alleles though it is very unlikely that this could be the explanation for the existence of as many as four private alleles.

Alternative explanations for the presence of private alleles in the Husvik population would be that not all the reindeer from the 1912 introduction were killed in the snow slide, that the reindeer introduced to Husvik were from a different source or that there were further reindeer imported in addition to the initial founder group. These possibilities are discussed in detail in the introduction which concludes that it seems very unlikely that there would have been any remnants from the 1912 introduction and there is no reason to doubt the origin of the Husvik herd as Filefjeld Reinlag but there may have been up to four female reindeer of unknown origin added to the Husvik herd in 1928.

There are numerous different semi-domestic and wild reindeer herds in southern Norway, often geographically isolated in different mountain areas (Roed 1985; Roed 1998a). There is significant genetic heterogeneity between different wild herds even within the same mountain regions and higher genetic variation in wild herds compared to semi-domestic reindeer (Roed 1986). However, there is little genetic differentiation between semi-domestic herds and occasionally animals from different herds are deliberately mixed to prevent inbreeding (Roed 1985).

The most likely explanation for the presence of private alleles in the Husvik population is that there was indeed a further introduction of reindeer

and it seems very likely that this was the four females referred to in the newspaper in 1928. For practical reasons, it seems very likely that these were also from a semi-domestic herd from southern Norway and thus not significantly genetically differentiated from the original introduction.

2.4.2 Comparison of gene diversity

Most studies that measure loss of genetic variation due to bottlenecks consider heterozygosity (the proportion of heterozygous individuals at a locus) and actual numbers of alleles present at a locus as described originally by Nei and others (1975). However, these two measures reveal different information.

Assuming there is no selective difference between alleles, the expected proportion of original heterozygosity remaining after a bottleneck of one generation is 1-1/2N where N is the effective number of individuals in the bottleneck (Allendorf 1986). The polygynous nature of reindeer means that the effective population size is likely to be considerably lower than the actual numbers counted so that $N_e=4(N_f N_m)/(N_f +N_m)$ (Nei 1987).

A bottleneck has effects for more than one generation and genetic variation will continue to decrease at a rate that is inversely proportional to the population size. Heterozygosity is reduced to $(1-1/2N)^t$ where t is the number of generations of the bottleneck. After one generation of a bottleneck where the effective population size is ten individuals, the heterozygosity would have reduced to 95% of the original. If the effective population size remained at ten individuals for five generations the

heterozygosity would be 77% of the original. Allendorf (1986) showed that heterozygosity returns to equilibrium quickly following a bottleneck of short duration and within just a couple of generations of random mating.

A bottleneck causes loss of alleles firstly due to the initial sampling process and subsequently due to the small size of the resultant population (Fuerst and Maruyama 1986). This subsequent loss, due to genetic drift, can be shown to be approximately $n(n-1)/2N_e$ where n is the number of alleles remaining in the population and N_e is the effective population size (Kimura 1955).

An important consideration is the fate of rare alleles which are particularly susceptible to loss during the first generation of the bottleneck (Fuerst and Maruyama 1986). Rare alleles have very little effect on the levels of heterozygosity in comparison with alleles of intermediate frequencies which contribute greatly to heterozygosity. This explains why bottlenecks of short duration have relatively little effect on heterozygosity but can show severe reductions in the numbers of alleles present in the population. This phenomena led Allendorf (1986) to conclude that consideration solely of heterozygosity leads to an overly optimistic view of the genetic variation within the population. He argues that although heterozygosity provides a good measure of the capability of the population to respond to selection immediately following the bottleneck, it is the number of alleles remaining after the bottleneck that are important for the long-term response of the post-bottleneck population to selection and its subsequent survival.

In this study there was significant reduction in both the numbers of alleles and observed heterozygosity following bottlenecks in both populations. The expected heterozygosity for the Husvik population was significantly different from the Norway population but this was not the case for the Barff population. Observed heterozygosity in natural populations is usually much lower than the expected heterozygosity for neutral alleles when the current population size is considered (Nei 1987). There is a prolonged effect of the population size reduction (bottleneck effect) on observed average heterozygosity (Nei *et al.* 1975) but the expected heterozygosity reaches equilibrium quickly.

In the estimation of gene diversity there are two sampling processes: the sampling of loci from the genome and the sampling of individuals from the population (Nei 1987). Examination of sampling variances has suggested that expected heterozygosity is a better index of genetic variability than observed heterozygosity due to the distorting effects caused by small population size, natural selection and inbreeding (Nei and Roychoudhury 1974).

There was a considerable difference between the populations in terms of post-bottleneck demographics. The population size of the Barff population increased rapidly following the bottleneck so that there was a population of more than 500 reindeer after between 20 and 30 years (see figure 2.1). However the Husvik population remained small for much longer with a population of approximately 100 after 30 years (see figure 2.2). Although each population will have been affected by different stochastic influences, this

difference in post-bottleneck population growth is likely to have affected the reduction of average heterozygosity more than the loss of alleles (Nei *et al.* 1975).

2.4.3 Population differentiation

Interpretation of the actual values of F_{ST} (Wright 1951) and R_{ST} (Slatkin 1995a) can be difficult and misleading (Balloux and Lugon-Moulin 2002). When two subpopulations and a two-allele locus are considered, the value for F_{ST} reaches one when the two subpopulations are totally homozygous and fixed for different alleles. However, in a situation where there are two subpopulations, each with ten equifrequent alleles but where none are shared between the populations, the maximum value of F_{ST} is only 0.053. Even this seemingly low value indicates significant differentiation as it is clear that there is no gene exchange between subpopulations and genetic differentiation is as high as possible (Balloux and Lugon-Moulin 2002). In a study of the common shrew, Balloux and co-workers (2000) demonstrated how genetic differentiation based on F- and R-statistics were much lower for autosomal microsatellites than for all other genetic markers.

The levels of differentiation in this study (range of F_{ST} values from 0.016 to 0.072) would be considered low to moderate when compared to a number of recent studies (summarised in Lugon-Moulin *et al.* 1999). However, for the reasons discussed above as well as effects due to sampling and effective population size, these are significant levels of differentiation and

are an important measure of how both post-bottleneck populations are now distinctly differentiated from their source population.

Whereas F_{ST} assumes evolution of loci under IAM, R_{ST} assumes SMM. R_{ST} is little affected by mutations and generally may be more appropriate for microsatellite data (Balloux and Goudet 2002) though this may not be an important consideration for this study due to the fact that mutations are unlikely to be relevant in the short time-frame. The disadvantage of using only R_{ST} is its high associated variance (Slatkin 1995a; Balloux and Lugon-Moulin 2002).

Quantitative measures of genetic differentiation such as R_{ST} may be disrupted by the effect of a bottleneck and unequal population sizes so that they are no longer proportional to divergence times (Gaggiotti and Excoffier 2000). Consequently a high value may indicate the rapid genetic drift resulting from the bottleneck and small population size (Whitehouse and Harley 2001; Harley *et al.* 2005).

The fact that the values comparing Norway and Husvik are higher than those comparing Norway and Barff may be due to the fact that the Husvik population was smaller for a longer period of time than the Barff population which grew rapidly following the bottleneck (see figures 2.1 and 2.2). We believe that there were more animals in the Barff founder population (seven females and three males) than the Husvik (four females and three males) though the addition of four extra female animals to Husvik in 1928 may have complicated this. If these extra reindeer were from a different source

population (as might be suggested by the high number of private alleles in the Husvik population) then there would have been the further complicating effect of outbreeding and admixture.

2.4.4 Bottleneck Simulation programs

There was very close correlation between the bottleneck simulation program GENELOSS and the actual populations. GENELOSS does not consider life history traits or demographics so much of the variation in the model depends on the numbers in the founder group. Due to incomplete historical records the actual numbers in the founder group of the Husvik population was unclear but the results of GENELOSS suggest that there were effectively the equivalent of between three and seven breeding pairs. The polygynous nature of reindeer complicates the extrapolation of the model.

The fact that correlation with the GENELOSS model was much closer than with the BNSIM model suggests that any discrepancies may have resulted from inaccurate assumptions about rate of population recovery. The census data from between 1911 and 1928 did indeed indicate a very rapid increase in population though, as already explained, there is some doubt about the credibility of these figures.

In order to simulate demographics that mirrored the actual numbers observed on the ground, the survivability and reproductive figures used in the BNSIM model were more generous than life history data reported by Leader Williams (1988). Despite this, the model consistently predicted lower

numbers of surviving alleles than those observed in each of the post-bottleneck populations. Some of these discrepancies may be a consequence of the large standard errors expected due to demographic stochasticity and due to an underestimation of the mutation rate.

With the exception of one locus (RT9) in the Husvik population, predictions of heterozygosity by the model were consistently lower than those found in either population. A recent longitudinal study of heterozygosity in an isolated population of mouflon, which was founded by a single pair of individuals, found unexpectedly high levels of heterozygosity despite a time scale of only 46 years and no possibility of immigration (Kaeuffer *et al.* 2007). Several lines of evidence suggested that the increase in heterozygosity may be attributable to selection which was also suggested as a possible reason for high levels of genetic variation in brown bear that had experienced a severe population bottleneck (Hartl and Hell 1994). An alternative explanation would be that the females show active choice for males that are unrelated and more genetically diverse. Although this has been suggested in the case of the highly polygynous Antarctic fur seals (Hoffman *et al.* 2007), the mechanism by which the female might remotely assess male genetic diversity is not clearly defined.

2.4.5 Bottleneck signatures

Computer simulations with varying sample size, number of loci, bottleneck size and length of time since the bottleneck have shown that the most useful markers for bottleneck detection are those evolving under the

IAM (Cornuet and Luikart 1996). The use of 'heterozygosity excess' as a bottleneck signature relies on the fact that allelic diversity is reduced faster than heterozygosity during a bottleneck (Nei *et al.* 1975) so that there is a transient deficit in the number of alleles found in a sample of individuals (Maruyama and Fuerst 1985). The theory behind these statements relied on the assumption that all loci evolve according to the IAM (Cornuet and Luikart 1996). The reason for also considering evolution under the SMM was that a number of authors have suggested that microsatellites may follow the SMM more closely (Shriver *et al.* 1993; Valdes *et al.* 1993) though this is probably more relevant over a longer timescale than that represented by this study.

Cornuet and Luikart (1996) employed a computer simulation approach based on the coalescent process to track the change in heterozygosity for loci evolving under SMM. They showed that even following a dramatic reduction in population size, heterozygosity excess is consistently found to be lower under the SMM than the IAM, that maximum excess occurs for lower numbers of alleles and that even heterozygosity deficit can occur when the value of θ is high (where $\theta=4N_eu$ (Tajima 1983; Watterson 1984)).

The IAM model and the TPM model (70% IAM, 30% SMM) accurately predicted that there had been a bottleneck in the Barff population by showing significantly more heterozygosity excess that that expected at equilibrium, though this was not apparent under the SMM model. The Husvik population showed no evidence of heterozygosity excess regardless of which mutation-drift equilibrium models was used for the predictions.

In the study of a documented bottleneck in Bennett's wallabies in New Zealand, significant heterozygosity excess was found compared to mutation-drift equilibrium under the IAM but not under TPM or SMM. The authors suggested four reasons why these models did not predict the bottleneck: the disjunct allele size-distributions in the post-bottleneck population indicated a departure from SMM, too few microsatellite loci (five) were used to achieve sufficient power, the loci studied were not selectively neutral or that the presence of null alleles confounded the analysis (Le Page *et al.* 2000). In contrast to the Wallaby study, there were thirteen microsatellite loci used in the present study which is well in excess of the four polymorphic loci required by the Wilcoxon sign-rank test (Luikart and Cornuet 1998) and only loci which showed no evidence of null alleles (following bootstrap test via Microchecker) were used in the analysis. There were disjunct allele size-distributions in the populations in this study (see Appendix 1).

In a study of post-bottleneck elephant populations that had been fragmented by widespread hunting, there was no evidence of significant heterozygosity excess under any of the three mutation models (Whitehouse and Harley 2001). The authors suggested that this might have been due to immigration into one of the populations or due to the fact that the heterozygosity excess effect only lasts for a few generations. Immigration (of four females in 1928) into the Husvik population may have occurred but the inaccessibility of the island and the comprehensive record keeping suggests that there were no other instances of immigration. Bottleneck-induced heterozygosity excess is transient and is likely to be detectable only for 0.2-

4.0 N_e generations (where N_e is the bottleneck effective size) until a new equilibrium is reached between mutation and drift (Luikart and Cornuet 1998). In this study, N_e would have been considerably less than the apparent number of individuals due to the highly polygynous mating system of reindeer (Leader-Williams 1988; Luikart *et al.* 1998b). Between 19 and 22 generations have elapsed since the bottlenecks and it is possible that this is too long a timescale for there to still be evidence of heterozygous excess.

The mode-shift alteration in the distribution of allele frequencies is described as characteristic following a bottleneck in samples of 5-20 polymorphic loci, approximately 30 individuals and for between 40 and 80 generations (Luikart *et al.* 1998a). The two post-bottleneck populations in this study fit these criteria but there was no indication of a mode-shift in the distribution of allele frequencies.

Luikart et al suggest five reasons why a bottlenecked population might not show a mode-shift: 1) the bottleneck was not recent or small enough to be detected, 2) not enough individuals and/or loci were sampled, 3) the individuals sampled were not representative of the bottlenecked population, 4) a demographic bottleneck occurred but not a genetic bottleneck, 5) the bottlenecked population was not completely isolated and contains genes from immigrants. None of these reasons would seem relevant to the reason why there is no evidence of a mode-shift in this study.

Examination of the allele frequencies (listed in Appendix 1) suggests that there were considerable reductions in the numbers of rare alleles in the

post-bottleneck populations in all the loci tested apart from two (NVHRT03 and Ca13). In these two loci, as well as a couple of others (CRH and Ca71), there were surprising low numbers of rare alleles in the pre-bottleneck population which suggests that even at mutation drift equilibrium there would not have been a typical L-shaped distribution curve. To test whether the results from these four loci had obscured a mode-shift in the other loci, the tests were rerun without the data for NVHRT03, Ca13, CRH and Ca71. However, a mode-shift was still not evident.

There is clear evidence that there were indeed bottlenecks at the founding of both South Georgia populations but without this information, the lack of a mode-shift may have led to the conclusion that a recent bottleneck was unlikely. Indeed in the absence of a mode-shift, Harley et al (2005) were led to conclude that remnant populations of black rhinoceros had experienced low population size at mutation-drift equilibrium rather than a recent bottleneck. Whitehouse and Harley (2001) found no evidence of a significant mode-shift in allele frequencies in the elephant populations that had experienced known bottlenecks and there was a reduction in numbers of rare alleles in the bottlenecked Bennetts's wallabies but there was not a significant departure from the standard L-shape in the mode-shift test (Le Page *et al.* 2000). Bimodal distribution of allele frequencies were found in a post-bottleneck population of koalas (Houlden *et al.* 1996) and northern hairy-nosed wombats (Taylor *et al.* 1994).

A reduction in allele numbers compared to the reduction in range size is expected to last longer following a bottleneck than heterozygosity excess

(Garza and Williamson 2001). Indeed the value of Garza's M was reduced for both the post-bottleneck populations compared to the Norwegian population even if the actual values were not <0.68 which was the recommended limit of M expected for bottlenecked populations (Garza and Williamson 2001). The post-bottleneck values in this study were less than the values of M calculated for known bottlenecked elephant populations (Whitehouse and Harley 2001).

2.4.6 Conclusions

This study has allowed consideration of the genetic effects of bottlenecks in two parallel situations. As expected, both bottlenecks caused reductions in both heterozygosity and allele numbers though there was a greater effect on in the Husvik population. This might lead one to believe that there were indeed fewer animals in the Husvik founder group. However, the presence of as many as four private alleles in the Husvik population has leant weight to the possibility that there was indeed a further importation of reindeer following the initial founder event in 1925. If the newspaper report is to be believed and there were four females added to the initial group then in essence there were eight females and three males introduced to Husvik which is more in total than in the Barff introduction.

If this indeed was the case then the difference in present day genetic diversity between the two South Georgia populations could not be accounted for by the actual numbers in the founder groups. In this study the comparison of genetic diversity of the two post-bottleneck populations with

concurrent consideration of their post-bottleneck demographics, has emphasised the effects of stochasticity on the outcome of a bottleneck.

Considering the small numbers of individuals at each founder event, the subsequent genetic drift expected due to low population numbers and the complete isolation between the two populations, it is perhaps surprising that 70% of post-bottleneck alleles are found in common in both post-bottleneck populations (see Appendix 1). Presumably this reflects the reality that, despite stochasticity, there is an increased chance that common alleles will survive the founder event than rare ones.

In common with the Western Carpathian brown bear (Hartl and Hell 1994), Bennett's wallabies (Le Page *et al.* 2000) and island mouflon (Kaeuffer *et al.* 2007), the South Georgia reindeer have shown considerable resilience and maintenance of genetic diversity despite each population experiencing a severe population bottleneck.

The varying reliability with which each of the different tests of bottleneck signatures were able to accurately predict the known presence of a previous bottleneck in each of these populations gives further reason to doubt the reliability of these widely used tests for bottleneck signatures.

Chapter Three

Direct comparison of the morphometrics between pre-bottleneck and post-bottleneck populations

3.1 Introduction

Small numbers of reindeer were introduced onto the island of South Georgia at the beginning of the twentieth century and these populations have subsequently been isolated. The founder event, subsequent isolation and different environment will have affected the populations. This chapter reports a study which investigated whether these factors have had significant quantitative effects on the morphometrics of the populations by making direct comparisons with the source herd in Norway.

Inbreeding may result in the interruption of pleiotropic interactions. As a consequence, inbred populations may show high levels of morphometric variation (Waddington 1942; Lerner 1954) and a number of studies have shown an inverse relationship between heterozygosity and morphological variance (Robertson and Reeve 1952; Mitton 1978; Eanes 1981; Fleischer 1983; Leary 1983). Levels of morphometric variation were measured and directly compared among populations in this study.

In the ideal, or the most developmentally stable situation, one would expect there to be perfect symmetry between opposite sides of the same individual for bilateral traits. The disruption of non-additive gene interactions, especially dominance and epistasis, affects the developmental stability of a

trait and thereby affect its bilateral symmetry (Leamy and Klingenberg 2005). Fluctuating asymmetry (FA) is a random pattern of between-side variation that is not directionally biased. FA is commonly used to estimate the effects of minor developmental accidents or 'noise' (Van Valen 1962; Bryant 1986; Palmer and Strobeck 1986).

In a sample of individuals, FA is a pattern of between-side variation which reflects a compromise between two opposing processes – that of developmental noise and that of developmental stability. Developmental noise occurs due to a combination of processes which tend to disrupt precise development. These processes may include small random differences in the rates of either cell division and growth or the physiological processes within cells or the effects of thermal noise on enzymatic processes.

Developmental stability depends on the combination of processes that tend to resist or buffer the disruption of precise development. These processes may include negative feedback systems to regulate enzymatic reactions within or between cells or the central nervous or hormonal control required for homeostasis (Palmer 1994).

As a measure of developmental stability, fluctuating asymmetry may show no change or increase with increasing 'stress'. This 'stress' may be intrinsic (predominately genetic in origin) or extrinsic (due to environmental factors) (Palmer and Strobeck 1986; Palmer 1994). Indeed, extreme temperatures, parasites, nutritional or chemical stress such as pollution have all been found to increase FA (Rasmuson 2002). As an example, wing FA was

shown to increase over time in flies captured from a population that was being eradicated by poisoning over a five year period (Tsubaki 1998).

Genetic stress can arise from intense selection, inbreeding, mutations or outbreeding. There are a number of studies that have shown significant associations between increased heterozygosity and low FA (Soule 1979; Vrijenhoek and Lerman 1982; Leary 1983). Indeed it is considered that heterozygosity generally stabilizes the phenotype, possibly due to one of two mechanisms – that of overdominance or the concealing of deleterious alleles (Lerner 1954). A third mechanism by which genotype has been thought to influence developmental stability is that of genetic coadaption (Markow 1995).

An important assumption underlying the usefulness of using FA as a measure of developmental stability is the fact that these small random deviations from bilateral symmetry do not have a heritable basis (Palmer and Strobeck 1986). It is identical genes that influence development of both sides of bilaterally symmetrical traits. It is important to check that the asymmetry is genuinely fluctuating as only FA has been suggested to result from poorly co-adapted gene complexes. Both antisymmetry and directional asymmetry have heritable elements and the inability to partition out the genetic basis of these asymmetries makes them less useful as indicators of developmental stability (Palmer 1994).

A number of studies have shown a correlation between an increase in FA and a reduction in fitness (eg Beardmore 1960) especially where the trait measured is one that directly affects performance. Meta-analyses have

found a weak but significant negative relationship between FA and various fitness components though there is ongoing controversy concerning the direct link between FA and fitness components (see reviews in Clarke 1995; Leung and Forbes 1996; Moller 1997a; Clarke 1998a; Leamy and Klingenberg 2005).

It seems that the extent that FA is able to predict fitness depends on the character chosen for analysis. When the normal functioning of an organism directly depends on the symmetry of a certain character, that character can be expected to be so well buffered or canalized that asymmetry is extremely rare. However if the functioning of the organism is completely independent of the symmetry of a different highly labile character, high levels of asymmetry may occur in that character with no direct link with levels of fitness (Clarke 1995). Indeed developmental stability has been shown to be highly dependant on character and often specific to taxon and population (Clarke 1998b; Clarke 1998c).

The comprehensive review by Moller (1997) of both published and unpublished studies claimed that there was clear evidence of a negative relationship between developmental instability and fitness components but the accuracy of this work has been heavily criticised (Clarke 1998a). There is a general consensus amongst other review authors that any relationships between symmetry and fitness components are weak, heterogenous and equivocal (Leung and Forbes 1996; Clarke 1998a). Relationships have been more obvious in the comparison of populations and the results have not been supported by analysis on an individual level.

The ambiguity of results may be due to a mismatch between the genetic architecture behind FA and that behind different fitness components. It appears that FA has a predominantly nonadditive genetic basis with substantial dominance and especially epistasis. The genes involved in the epistatic interactions are most likely to be character-specific and involved in some way in the formation of the character (Leamy and Klingenberg 2005).

The use of FA as a reliable measure of developmental stability is contentious and it is considered to be an imperfect tool. However, much of the controversy has been due either to the inaccurate reporting of studies or the drawing of unreliable or overzealous conclusions. These should not be allowed to detract from further studies, such as this one, which aim to use FA with caution, as a measure of developmental stability describing variation within individuals to be compared both among (this chapter) and within (Chapter Four) populations.

Whereas fluctuating asymmetry, as a measure of developmental stability, describes variation within individuals, canalization is better measured by phenotypic variation among individuals (Clarke 1998c). Clarke found that both canalization and developmental stability were character-dependent and there was significant correlation between character variation and FA in 9 out of 11 species of invertebrate. He suggested that the mechanisms responsible for both inter and intra-individual variation affect individual characters in the same manner so that characters which display low levels of inter-individual variation also display relatively low levels of intra-individual variation.

Unlike FA, morphometric variation has an additive genetic component as well as a component due to random genetic accidents (Leary *et al.* 1985) so a trait may show high morphometric variation due to developmental instability or due to the accurate expression of genetic variability in the phenotype (Kieser and Groeneveld 1991). This suggests that FA is a better measure of developmental stability than morphometric variation.

The collection of reliable morphological data in this study has allowed the comparison of phenotypic variation among individuals and the direct comparison of overall size of individual skulls both among and within populations.

The tendency for island populations to differ in body size from their mainland relatives has been well documented (Foster 1964; Lomolino 1985; Palkovacs 2003). The direction of body size trends are generally different for different vertebrate families but as a general rule, large mammals tend towards smaller insular forms and small mammals tend towards larger insular forms (Lomolino 1985). Reviews of previous studies have found that artiodactyls tend towards smaller forms on islands (Foster 1964; Case 1978).

Island body size of individuals depends on the interaction of primary causal factors: competition, predation, resource availability and physiological efficiency (Heaney 1978; Polkovacs 2003). For example, reduced interspecific competition would lead to an expected increase in body size though reduced resource availability would lead to an expected decrease in body size. Reduced predation may lead to an increase in body size for small mammals

that escape predation by hiding or a decrease in body size for larger mammals that escape predation by running away or fighting. Physiological factors of relevance may include the efficiency of thermoregulation and locomotion as well as gestation length and litter size (Heaney 1978; Lawlor 1982).

Island populations often experience a decrease in the number of both competitor and predator species (Heaney 1978; Lomolino 1985) which leaves the island population free to evolve towards the body size most advantageous for exploiting resources of energy in the diet (Damuth 1993). Damuth claims that medium sized animals have the most advantage in controlling energy resources and this is the explanation for small animals to be larger on islands and large animals to be smaller.

Polkavacs (2003) stresses the importance of the balance between individual growth rate and the reaction norm which determines both age and size at maturity. The age-size reaction norm relates to the set of phenotypes expressed by a single genotype over a range of environmental conditions. It is primarily affected by the genetics of the organism but there will also be a plastic response to the two common environmental effects of island living, reduced extrinsic mortality and reduced resource availability. These two factors can affect body size directly via age at maturity and individual growth rate and indirectly via the effects of altered population density (Polkovacs 2003).

Many of the well documented cases of large mammals (elephants, hippopotami, deer and goats) that exhibited dwarfism were isolated on islands during the Pleistocene (summarised in Heaney 1978; Lister 1989; Raia and Meiri 2006) but there are reports that significant changes in mammalian body size may occur in much shorter time periods such as 70 years in the case of an island population of house mice (Berry 1964; Berry and Jakobson 1975) or 175 years for various wild mammals in Denmark purportedly as a result of habitat fragmentation (Schmidt and Jensen 2003). In a comparison of otter populations in Europe the populations in decline showed some increase in FA and a decrease in skull size compared to the more viable populations (Pertoldi 2000).

Studies of reindeer have shown that density dependant resource limitation has a major effect on size of wild reindeer (Reimers 1972; Skogland 1983; Skogland 1988) and as reindeer do not grow during the winter when feeding on lichens, it is the quality of summer grazing that has a major effect on body size (references in Skogland 1983). Skull measurements (jawbone or diastema length) have been found to reflect long term range quality, specifically during the growth phase, compared to dressed body weight which was found to reflect short term resource availability (Reimers 1972; Skogland 1983). For three separate wild Norwegian herds, it was possible to compare reindeer that had migrated to richer habitats as a result of increasing grazing pressure with their herds of origin. Although the migrations had only occurred 20-30 years before the study, jawbone size was significantly greater

in each of the herds on the richer habitats compared to the herd of origin (Skogland 1983).

In this study skull morphology was measured to compare developmental stability (estimated as FA), morphometric variation and body size. In each case, the objective was to quantify and assess the impact of the population bottlenecks on morphology, and interpret these data in the context of theories about the underlying mechanisms.

Hypotheses

1. Some of the traits will show greater FA in the post-bottleneck populations than the pre-bottleneck population.
2. The post-bottleneck populations will show a greater magnitude of morphometric variance than the pre-bottleneck population.
3. The overall size of the skulls from the post-bottleneck populations will be smaller than the skulls from the pre-bottleneck population.

3.2 Materials and Methods

3.2.1 Collection of samples

Complete heads from culled animals were taken from the source herd in Norway on two occasions. The initial load of 40 calf skulls (of approximately 7 months old) were packed in sealed barrels at room temperature for transport. The load of 41 adult skulls were frozen on collection and moved directly into storage at -20°C.

Samples were collected from 64 carcasses on the Barff Peninsula and 37 carcasses in the Busen area of South Georgia. There were 41 complete skulls from the Barff herd and 23 from the Husvik herd. These skulls were hung in the sea to remove most of the flesh and packed with salt in sealed barrels for transport back to the UK.

3.2.2 Preparation of skulls

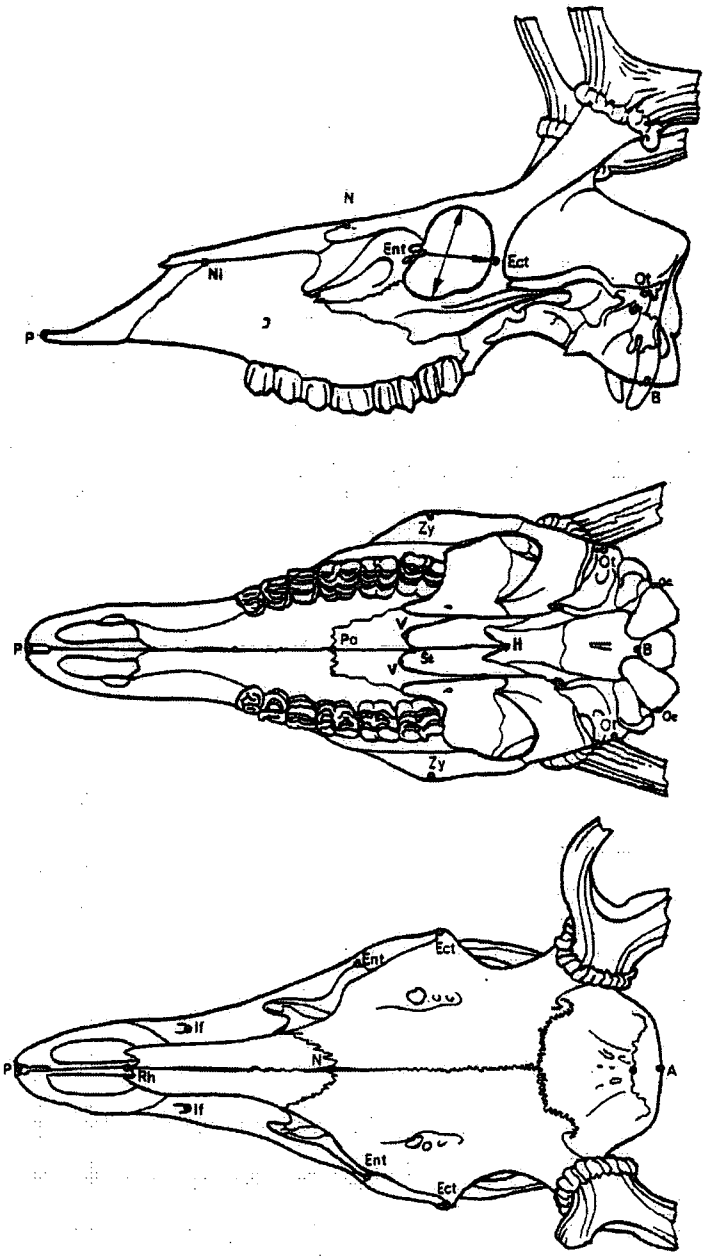
The 40 Norwegian calf samples deteriorated due to natural decomposition. The soft tissue was removed and the skulls were dried and placed with demestids. The 41 Norwegian adult heads were removed from -20°C, skinned and placed in boiling water for 2 hours to loosen the soft tissue. Most of the soft tissue on the South Georgia skulls was in an advanced state of decomposition following transportation. Any substantial bits of soft tissue that remained were removed by hand with a sharp domestic knife.

All biological waste was double bagged and incinerated in accordance with the requirements of the import licences.

All skulls were placed in a strong solution of biological washing powder (phosphates, surfactants, soap, carboxy-methylcellulase) at room temperature for up to two weeks to remove remaining flesh. The clean skulls were placed in 100% ethanol for ten minutes and left to dry overnight at room temperature.

3.2.3 Measurements

Figure 3.1 Measurements taken



A	Akrokranion
P	Prosthion
Rh	Rhinion
N	Nasion
Ent	Entorbitale
Ect	Ectorbitale
Ni	Nasointermaxillare
If	Infraorbitale
Ot	Otion
B	Basion
Po	Palatinoorale
St	Staphylion
Zyg	Zygion
H	Hormion
Vom	Vomar notch

(after von den Driesch 1976)

Measurements of overall size:

Lengths	Widths
A-P	Zyg-Zyg
A-Rh	Ot-Ot
A-N	Oc-Oc
B-P	
B-Po	Height
B-St	B-Nuch

Figure 3.1 Measurements taken (cont)

Bilateral Measurements for assessment of asymmetry:

Ent-P	Ni-Ot	If-Ent	Ot-Ect	Ot-N	B-If
B-Ent	B-Zyg	H-Zyg	St-Zyg	Ni-P	Vom-Po
N-Ent	Po-Zyg		OrbLgth = Ent-Ect		
OrbHght = height of Orbitale			ChkTth = length of cheek teeth row		

Description of points on the skull

A	Akrokranion	Most aboral point on the vertex of the cranium in the medial plane
P	Prosthion	Median point of the line joining the most oral point of the premaxillae
Rh	Rhinion	Most oral point of the nasales on the dorsal aspect
N	Nasion	Median point of the nasofrontal suture
Ent	Entorbitale	Most inner angle of the orbit (the nasomedial indentation)
Ect	Ectorbitale	Most lateral point of the frontal bone on the occipital side of the orbit
Ni	Nasointermaxillare	Most aboral point of the premaxilla on the facial surface
If	Infraorbitale	Dorsal point of the foramen infraorbitale
Ot	Otion	Most lateral point of the mastoid region (dorsal to external auditory meatus)
B	Basion	Orobasal border of foramen magnum in the median plane
Po	Palatinoorale	Median point of the palatine maxillary suture
St	Staphylion	Most aboral point of the horizontal part of the palate
Zyg	Zygion	Most lateral point of the zygomatic arch
H	Hormion	Aboral border of the vomar in the median plane
Vom	Vomar notch	Most oral point of vomar

Three sets of 44 measurements were taken on each skull. Seven measurements were taken sagitally along the midline of each skull and three further measurements of the width of various parts of the cranium. The symmetry of bilateral characters was tested using 17 measurements on each side of the skull. See figure 3.1.

3.2.3.1 Measurement Error

Measurement error (ME) is of great importance in studies of fluctuating asymmetry because ME, like FA, is expected to be normally distributed with a mean of zero.

Differences in FA are likely to be small relative to the size of the trait and are measured by their variance so the results are likely to be confounded by further variance, such as that of ME (Palmer and Strobeck 1986; Swaddle 1994). For these reasons, care was taken to control for measurement error.

All the skulls were measured by the same person (FML) using precision Vernier callipers. On each skull, the complete set of traits on the dorsal aspect were measured on the left side and then the right side. The skull was turned over and the complete set of measurements were taken on the left of the ventral aspect and then the right of the ventral aspect. When all the skulls had been measured once the procedure was repeated twice more with the subsequent measurements taken 'blindly', that is with the measurer having no knowledge of the previous measurements. Several days elapsed between each complete set of measurements on each skull as this has been shown to give the most reliable estimate of measurement error (Palmer 1994)

If it was not possible to measure a trait accurately due to damage to the skull, a 'missing value' was entered in the data.

The variance due to measurement error was partitioned out of the total variance by use of a two-way mixed ANOVA (Palmer and Strobeck 1986) with the sides as fixed and the individuals as random factors. This tests for whether the variation between sides is significantly greater than the variation due to measurement error.

Measurement error was described by three different indices:

ME2 is the standard deviation of repeated measurements so that $ME2=\sqrt{MS_m}$ where MS_m is the error mean square from the sides x individuals ANOVA (Palmer 1994; Palmer and Stobeck 2003).

ME3 expresses the average difference between repeat measurements as a percentage of the average difference between sides so that

$ME3 = 100^* MS_m / MS_{interaction}$. (Palmer 2003)

ME5 is a measure of repeatability which expresses variation due to asymmetry as a proportion of the total between-sides variation (which includes ME). It is a dimensionless number which ranges from −1 to +1.

Repeatability, $r = (MS_{interaction} - MS_m)/ (MS_{interaction}+(n-1) MS_m)$ (Palmer 2003)

3.2.3.2 Determination of age

Only adult skulls were used in the analysis to avoid inaccuracies resulting from uneven growth and development. Two different criteria were used to define the skulls as adult or calf:

1. Measurements: 47 Norwegian skulls were measured of which six were known to be calves. Examination of these measurements allowed the definition of an adult skull if trait A-P was larger than 28.0cm and A-Rh larger than 20.5cm.
2. Examination of the teeth: There is published information on the approximate age that the mandibular premolars and molars erupt and become established (Bergerud 1970; Leader-Williams 1979) (see table 3.1).

Table 3.1 Published information on the approximate age that mandibular premolars and molars erupt and become established

Permanent mandibular teeth	Barff reindeer (Leader-Williams 1979)		Scandinavian reindeer (Bergerud 1970)
	Eruption begins (Age in months)	Teeth established (Age in months)	Teeth established (Age in months)
P2	24	31	30
P3	26	31	30
P4	26	34	28
M1	3	6	4
M2	11	15	15
M3	19	32	29

Mandibles were available for all 47 Norwegian skulls but not for a large proportion of the South Georgia skulls. Comparison of the mandibular and maxillary teeth of the Norwegian population confirmed that there is very little difference between the timings of eruption and establishment of the corresponding teeth. For this reason, the skulls were aged approximately on the eruption and establishment of the maxillary premolars and molars.

There was no discrepancy between the two criteria used for allocating skulls by age. Further confirmation of age was possible as a large proportion of the Norwegian reindeer had been ear-tagged with a code that related to their year of birth so that those in the 1-2 year age category were known to be 20 months old at the time of the cull.

Table 3.2 Age profile of the skulls measured for this study

Age in years	Norway	Barff	Husvik
0-1	6	6	0
1-2	25	2	3
2-3	0	1	0
3+	16	32	20
Total	47	41	23

The Mann-Whitney *U* statistic was used to test the difference in magnitude of FA between the 1-2 year olds (25 skulls) and the aged skulls (16 skulls) within the Norwegian population. The consistency of the results of the comparisons among populations was checked by comparing a subsample of just the aged skulls from each population (Norwegian n=16, Barff n=32 and Husvik n=20).

There were 25 skulls in the Norwegian population that were aged 1-2 years where the molars were not fully established. The length of the cheek teeth row was one of the traits measured for calculation of the degree of asymmetry. However due to continued tooth eruption the asymmetry result was considered unreliable as an indicator of developmental instability. For this reason, the results for this trait are not included in the analysis but the values measured are included in Appendix 2 for completeness.

3.2.3.3 Determination of sex by polymerase chain reaction

Some of the skulls were identified by sex in the field and further sexing was possible by amplification of a region of the Y chromosome by PCR of previously identified primers, SRY1 (5'-CTT CAT TGT GTG GTC TCG TG-3') and SRY2 (5'-CGG GTA TTT GTC TCG GTG TA-3') (Wilson and White 1998). These primers correspond to nucleotide positions 2-20 and 58-78 of a published bovine SRY sequence (Payen and Cotinot 1993) and amplify a product of approximately 180 base pairs (Wilson and White 1998). The SRY primers were used in multiplex with the primer, BM848 (5'-TGG TTG GAA GGA AAA CTT GG-3' and 5'-CCT CTG CTC CTC AAG ACA C-3') (Cronin *et al.* 2003) which amplifies a product of approximately 360 base pairs. Microsatellite locus BM848 was used as a positive control so that female samples were expected to only show a 360 base pair product and male samples were expected to yield products at 180 and 360 base pairs.

PCR amplification was carried out in 20µl reactions. The reaction mix was 2µl Tris buffer, 2µl dNTP mix (0.2mM concentration), 0.8µl Bovine Serum Albumin (20mg ml^{-1}), 0.4µl of each primer (0.5µg μl^{-1} in 20% TE), 0.08µl TAQ (5 unit μl^{-1}), 0.4µl DNA and 3.0M magnesium chloride. The program started with 5 minutes at 94^0C for denaturing. The cycle profile continued for 60 seconds at 94^0C, 60 seconds at the annealing temperature of 60^0C and 60 seconds at the extension temperature, 72^0C. The cycle was repeated 38 times and then held at 72^0C for 5 minutes for the extension stage. A proportion of the Barff samples were amplified with Qiagen™ Multiplex PCR

Master Mix containing HotstarTaq™ DNA Polymerase, Qiagen buffer (containing 6mM $MgCl_2$) and dNTP mix at the annealing temperature of 55^0C. 5µl of the PCR products were run on a 1.2% agarose gel at 100W against a 100 bp marker to allow visualisation. (See figure 3.2.)

Figure 3.2 Photograph of the agar gel of individuals typed at BM848 (360 bp) and SRY (180 bp). Photograph shows (left to right) 100 base pair ladder, individual SG2 (female), SG7 (male), SG8 (sex unclear), B20 (male), Na4 (male), Na14 (sex unclear), Na32 (female), negative control.

Table 3.3 Sex and age profile of the skulls measured for this study

Age category	Norway			Barff			Husvik		
	Male	Female	Sex not known	Male	Female	Sex not known	Male	Female	Sex not known
0-1 years			6	2	1	3			
1-2 years	25			2			1	1	1
2-3 years				1					
3+years	5	10	1	11	12	9	15	4	1

The Mann-Whitney *U* statistic was used to test the difference in magnitude of FA between male and female skulls within each population and across all populations. The consistency of the results of the comparison among populations was checked against a single-sex subsample of just male skulls (30 Norwegian, 14 Barff and 16 Husvik) to allow for sexual dimorphism.

3.2.4 Fluctuating Asymmetry

3.2.4.1 Tests for presence of directional asymmetry, antisymmetry and size dependence

Fluctuating asymmetry is assumed if the signed differences between paired structures are normally distributed with a mean of zero (Soule 1967)

Directional asymmetry (DA) occurs when one side of a character is consistently greater than the other side (e.g. mammalian heart (Van Valen 1962)). In traits that are directionally asymmetrical the mean difference between right and left side will be normally distributed but significantly greater or less than zero. Certain FA indices (including FA1) are artificially inflated by the presence of DA (Palmer 1994).

Antisymmetry occurs where the distribution of the difference between right and left is not normal but it is characterised by a platykurtic (broad peaked) or bimodal distribution of the difference between right and left side about a mean of zero. In this situation, there is asymmetry but there is no bias as to which side is the greater (Van Valen 1962; Palmer 1994). All FA indices are artificially inflated by the presence of antisymmetry (Palmer 1994). The difference between sides (R-L) was used as the measure of asymmetry to test for both directional asymmetry and antisymmetry.

The presence of directional asymmetry was tested by two different methods: a two-tailed one sample t-test against a mean of zero and the two-way mixed model ANOVA (Palmer 1994; Palmer and Stobeck 2003).

Scatter plots were examined for evidence of platykurtosis or bimodality and the Kolomogorovo-Smirnov test was used to test the significance of any departure from normality.

When the magnitude of asymmetry is dependent on the size of the trait, spurious differences in FA arise due to variation in overall size. The Spearman coefficient of rank correlation between the absolute value of FA(|R-L|) and the average of the sum of both sides was used to test for size-dependence of FA. Spearman's coefficient was used as it is a non parametric test of association which does not assume homogeneity of variance and is not influenced by a few extreme observations (Palmer 1994).

Independence across traits was tested by the Spearman coefficient measuring correlation of FA across traits.

3.2.4.2 Measures of Fluctuating Asymmetry on individual traits

Three different measures of FA were used for each trait.

FA1 is the mean of the absolute difference between the right and left sides. This index is inflated by the presence of DA or antisymmetry, it is sensitive to size-dependence of FA and it lacks statistical power with small sample sizes (Palmer 1994). Due to the highly skewed nature of the absolute values, a non parametric test, the Mann-Whitney *U* statistic was used to compare the magnitude of FA1 for the different populations

FA4 is the variance of the difference between the right and left sides. It is more efficient than FA1 at estimating between-sides variation and it is

not biased by DA. It is very biased by the presence of antisymmetry and sensitive to size-dependence of FA (Palmer 1994).

Palmer (2003) suggests a rule of thumb for deciding whether to include traits when there is DA present but too much data would be lost if these traits were excluded. If DA, as mean (R-L), is less than FA4a (where FA4a = $0.798\sqrt{FA4}$) then the predisposition towards one side is less than the average deviation about the mean (R-L) (Palmer 2003). For this reason FA4a was calculated (values not shown).

FA10 is the difference between the mean square of interaction and the error mean square divided by the number of repeat measurements undertaken. It was calculated from the mean-squares obtained from the two-way ANOVA with factors of sides (fixed) and individuals (random). FA10 was calculated as it is the only index that allows measurement error to be partitioned out of the total between-sides variance.

The F-test (Lehmann 1959; Palmer 1994) was used to compare the variances (FA4 and FA10) between populations.

3.2.4.3 Composite measures of Fluctuating Asymmetry

The combination of information from multiple traits has been argued to be a more reliable estimate of the underlying developmental instability than the use of fluctuating asymmetry of single traits (Leung *et al.* 2000; Palmer and Stobeck 2003). In this study two different composite measures of FA were used. The term 'CFA 1' was used for the measure calculated as the mean of the absolute FA values for individual traits; this measure

corresponded to CFA 1 in Leung et al. (2000) and Index 11 in Palmer and Strobeck (2003). The term 'CFA 3' (Leung *et al.* 2000) was used for the measure that was calculated by ranking individuals in order of the magnitude of absolute FA and summing the ranks for each individual. This method standardizes the magnitude of FA values across traits.

3.2.5 Variance of cranial traits between populations

The level of variability of cranial traits was compared between populations. The comparisons were made using regression analysis to correct for variation in skull size. Each trait was regressed against a measure of skull length, the akrokranion to the rhinion (A-Rh). This measure was used instead of the overall length (A-P) as there were a number of skulls which had lost nasal bones and had missing values recorded for all measures that involved the prosthion (P). The measure A-Rh was highly correlated to A-P (R=0.946 for 85 skulls).

Variance was measured by the residuals which are the distance a point is from the regression line. These residuals were compared as the residual mean square (RMS) using an *F*-statistic (Hoelzel 1999).

3.2.6 Size comparison between populations

Skull length has been used in previous studies as a reliable indicator of overall body size (eg. Heaney 1978; Smith 1992). Although most of the overall measurements were normally distributed there was significant kurtosis

in the distribution of lengths in the Husvik skulls and there was not homogeneity of variance. For these reasons the non-parametric Kruskall Wallis test was used to test for differences in overall size of skulls between the three populations.

3.2.7 Table-wide probability of type 1 error

Due to the fact that measurements and statistical analyses were undertaken for a large number of traits from each skull, the Bonferroni correction was applied to each set of related tests (Rice 1989).

To prevent committing a type I error, only those results significant at $p<0.05$ after application of the Bonferroni correction were considered as significant. However, those results which were significant before application of the Bonferroni correction were indicated on the results tables as the application of the Bonferroni correction increases the likelihood of committing type II error.

3.3 Results

3.3.1 Measurement error

The average value for the standard deviation of repeated measurements, ME2, was 0.61 (range 0.40-1.10). The average value for the difference between repeat measurements as a percentage of difference between sides, ME3, was 12.6% (range 1.9%-38.7%). The average value for

the measure of repeatability (ME5) which expresses variation due to asymmetry as a proportion of total between-side variation was 0.71 (range 0.35-0.94).

The two-way ANOVA was used to test for whether the variation due to between-sides differences was significantly greater than the variation due to measurement error. Every trait was tested for each population and in all cases the result was highly significant ($p<0.001$, sequential Bonferroni correction applied) which suggests that the asymmetry estimates in this study were highly repeatable.

3.3.2 Directional Asymmetry

Some evidence of directional asymmetry was found in the following traits: Ni-P, Ot-N, OrbHgt and Po-Zyg in the Norway population; If-Ent, Ot-N and Po-Zyg in the Barff population and B-Ent, B-If, Ent-P, If-Ent and Vom-Po in the Husvik population. However, after the application of the sequential Bonferroni correction, this DA was only significant for trait Orb-Hgt ($p<0.001$) and Ni-P ($p<0.05$) in the Norwegian population, trait Po-Zyg ($p<0.001$) and trait Ot-N ($p<0.05$) in the Barff population and trait If-Ent ($p<0.05$) in the Husvik population. (See table 3.4.)

In the Norwegian population, Ni-P (the length of the nares) was significantly greater on the left side than the right and OrbHgt (the height of the orbit) was significantly greater on the right side than the left. However, neither of these traits showed any degree of DA in either of the post-bottleneck populations. Both Ot-N (right>left) and Po-Zyg (left<right)

showed a low level of DA in the Norwegian population (though this was no longer significant after application of Bonferroni) and the level of DA in both these traits was greater in the Barff population (significant at the levels $p<0.05$ and $p<0.001$ respectively). The mean value for Po-Zyg was even higher in the Husvik but due to high standard error in this population, it did not show significant directional asymmetry. The mean value for Ot-N was very low for the Husvik population with no significant DA.

There was no directional asymmetry for trait If-Ent in the Norwegian population but on average the left was greater than the right measurement for this trait in the Barff population (not still significant after application of Bonferroni) and in the Husvik population (at the level $p<0.05$ after Bonferroni). There were three further traits (Ent-P, B-If and Vom-Po) in which the left side was greater than the right and one trait (B-Ent) in which the right was greater than the left in the Husvik population (though not significant after Bonferroni) though these traits showed no such evidence of DA in either the Norwegian or the Barff population.

There was no difference in the significance of these results between the two methods used to calculate DA.

Out of the seventeen traits measured, there were five traits that showed significant DA in one or more of the populations although there was no case in which DA, as mean (R-L), was greater than FA4a. These traits did not consistently show DA in all the populations and it was decided that FA comparisons would only be made between two populations where there was no significant DA in either population. Excluding traits which showed

significant DA is an approach recommended by Palmer (1994, 2003) but may be considered conservative compared to some recent studies (Hutchison and Cheverud 1995; Sonne *et al.* 2005) which have compared levels of FA despite finding significant levels of DA.

Composite FA was calculated to include only the traits which showed no significant DA in any of the populations.

3.3.3 Antisymmetry

The Kolmogorovo-Smirnov test was used to test the significance of any departure from normality. There was some departure from normality in traits Ni-P and ChkTth in the Norway population, traits N-Ent and B-Ent in the Barff population and trait Ent-P in the Husvik population. However, not one of these traits showed significant departure from normality after application of the sequential Bonferroni correction. (See table 3.5.)

3.3.4 Size dependence

The Spearman coefficient of rank correlation between the absolute value of FA(|r-L|) and the average of the sum of both sides was used to test for size-dependence of FA. Spearman's coefficient was used as it is a non parametric tests of association which does not assume homogeneity of variance and is not influenced by a few extreme observations. There was some correlation between FA and size of trait for Ot-N in the Barff population and B-Zyg in the Norway population but after application of the sequential Bonferroni correction there was no significant correlation between magnitude of FA and size of any trait. (See table 3.5.)

Table 3.4 Tests for Directional Asymmetry. ns - not significant, ^significant before application of Bonferroni correction, *p<0.05 (0.0029 is the threshold after application of Bonferroni), **p<0.01, ***p<0.001, Bonferroni correction applied.

Trait	Norway					Barff					Husvik				
	n	mean (mm)	t	ANOVA F	p-value	n	mean (mm)	t	ANOVA F	p-value	n	mean (mm)	t	ANOVA F	p-value
Ent-P	40	-0.04	-0.15	0.03	ns	24	-0.74	-1.62	3.31	ns	12	-1.43	-2.52	6.09	0.03^
Ni-Ot	41	0.59	1.77	3.69	ns	34	-0.27	-0.84	1.42	ns	22	0.35	0.64	0.55	ns
If-Ent	41	0.04	0.20	0.13	ns	33	-0.86	-2.65	8.54	0.006^	23	-1.56	-3.58	14.33	0.001*
Ot-Ect	41	-0.10	-0.38	0.19	ns	35	0.15	0.39	0.05	ns	21	1.01	2.08	3.57	ns
Ot-N	41	0.53	2.04	6.99	0.01^	35	0.91	3.52	11.91	0.002*	22	-0.09	-0.14	0.01	ns
N-Ent	41	0.16	0.71	0.71	ns	34	0.44	1.56	1.45	ns	23	-0.23	-0.87	0.46	ns
Ni-P	40	-0.57	-3.52	12.59	0.001*	24	0.13	0.69	0.17	ns	11	0.02	0.08	0.00	ns
OrbLgh	41	0.16	1.49	2.92	ns	34	-0.09	-0.67	1.68	ns	22	-0.10	-0.83	0.61	ns
OrbHgt	41	0.50	4.04	15.84	<0.001**	33	0.35	1.69	2.06	ns	21	0.18	1.05	2.32	ns
B-If	41	-0.37	-1.84	2.09	ns	34	-0.36	-1.49	2.31	ns	22	-0.61	-2.42	5.98	0.02^
B-Ent	41	-0.12	-0.96	0.45	ns	34	-0.01	-0.04	0.00	ns	22	0.58	2.36	4.94	0.04^
B-Zyg	41	0.61	2.80	3.95	ns	32	0.39	1.53	1.60	ns	20	-0.44	-1.16	0.91	ns
H-Zyg	41	0.13	0.62	0.76	ns	32	0.06	0.33	0.02	ns	20	-0.26	-0.97	0.59	ns
St-Zyg	41	-0.07	-0.29	0.22	ns	30	-0.35	-1.43	1.58	ns	20	0.49	1.29	1.81	ns
Po-Zyg	41	-0.61	-2.36	4.67	0.04^	32	-1.02	-4.27	17.72	<0.001***	21	-1.78	-2.10	4.12	ns
ChkTth	41	0.07	0.29	0.49	ns	34	0.21	1.12	0.41	ns	22	-0.10	-0.25	0.07	ns
Vom-Po	38	0.26	0.89	0.77	ns	33	0.07	0.12	0.05	ns	22	-2.30	-3.19	10.60	0.004^

Table 3.5 Tests for antisymmetry and size dependence ($p<0.0029$ is the threshold after Bonferroni at the level $p<0.05$)

Trait	Norway			Barff			Husvik			
	n	Kolmogorov-Smirnov statistic	Spearman rho	n	Kolmogorov-Smirnov statistic	Spearman rho	n	Kolmogorov-Smirnov statistic		Spearman rho
Ent-P	40	0.07	ns	ns	24	0.08	ns	ns	12	0.11	0.01^	ns
Ni-Ot	41	0.08	ns	ns	34	0.08	ns	ns	22	0.12	ns	ns
If-Ent	41	0.09	ns	ns	33	0.07	ns	ns	23	0.15	ns	ns
Ot-Ect	41	0.11	ns	ns	35	0.09	ns	ns	21	0.12	ns	ns
Ot-N	41	0.09	ns	ns	35	0.08	ns	0.002^	22	0.17	ns	ns
N-Ent	41	0.10	ns	ns	34	0.17	0.02^	ns	23	0.20	ns	ns
Ni-P	40	0.13	ns	ns	24	0.17	ns	ns	11	0.16	ns	ns
Orb Lgth	41	0.10	ns	ns	34	0.15	0.04^	ns	22	0.11	ns	ns
Orb Hgt	41	0.11	ns	ns	33	0.13	ns	ns	21	0.12	ns	ns
B-If	41	0.14	0.04^	ns	34	0.12	ns	ns	22	0.09	ns	ns
B-Ent	41	0.12	ns	ns	34	0.14	ns	ns	22	0.18	ns	ns
B-Zyg	41	0.06	ns	0.007^	32	0.10	ns	ns	20	0.14	ns	ns
H-Zyg	41	0.09	ns	ns	32	0.11	ns	ns	20	0.11	ns	ns
St-Zyg	41	0.13	ns	ns	30	0.09	ns	ns	20	0.13	ns	ns
Po-Zyg	41	0.09	ns	ns	32	0.10	ns	ns	21	0.11	ns	ns
ChkTth	41	0.16	0.01^	ns	34	0.11	ns	ns	22	0.13	ns	ns
Vom-Po	38	0.12	ns	ns	33	0.12	ns	ns	22	0.28	ns	ns

3.3.5 Test of independence across traits

Correlation of FA across traits using the Spearman coefficient showed that there was significant correlation between trait Ent-P and If-Ent (r^2=0.35, n=76) at the $p<0.001$ level and between trait Vom-Po and Po-Zyg (r^2=0.11, n=89) at the $p<0.05$ level after the Bonferroni correction was applied. Due to evidence of DA in traits If-Ent and Po-Zyg, the asymmetry results for these traits were already considered with caution and these traits were excluded in the calculation of composite FA. There was no further significant interdependence of FA between any other traits.

3.3.6 Influence of age and sex

There was no significant difference in FA when skulls of aged animals (n=16) were compared with skulls of 1-2 year old reindeer (n=25) within the Norway population. When a subsample of only aged skulls (Norway n=16, Barff n=32, Husvik n=20) were considered in the comparison of FA between the populations, the pattern of significant results was the same as when all skulls were considered.

There was no significant difference when FA of female skulls were compared with male skulls within any of the populations. When a subsample of just male skulls (Norway n=30, Barff n=14, Husvik n=16) was considered in comparison of FA between populations, the pattern of significant results was the same as when all skulls were considered. For these reasons, all the reported results include all adult skulls with no division on the basis of age or sex.

Table 3.6 Basic fluctuating asymmetry statistics for traits that show true FA. See section 3.2.4.3 for method of calculation of CFA1. *The traits marked with an asterisk show significant DA in one of the populations so were not included in the calculation of CFA1. Statistics for all traits are shown in Appendix 2.

Trait	Norway				Barff				Husvik			
		FA1		FA4		FA1		FA4		FA1		FA4
	n	mean (cm)	range	variance	n	mean (cm)	range	variance	n	mean (cm)	range	variance
Ent-P	40	0.12±0.09	0-0.34	0.022	24	0.19±0.14	0-0.56	0.049	12	0.22±0.08	0.1-0.34	0.039
Ni-Ot	41	0.18±0.12	0-0.5	0.045	34	0.15±0.11	0-0.45	0.035	22	0.19±0.17	0-0.68	0.067
Ot-Ect	41	0.13±0.12	0-0.51	0.031	35	0.18±0.15	0-0.75	0.056	21	0.17±0.17	0-0.58	0.050
N-Ent	41	0.12±0.09	0-0.36	0.021	34	0.15±0.08	0-0.39	0.027	23	0.10±0.08	0-0.3	0.016
Orb Lgh	41	0.05±0.05	0-0.17	0.005	34	0.06±0.05	0-0.23	0.006	22	0.04±0.04	0-0.13	0.003
B-If	41	0.10±0.09	0-0.38	0.016	34	0.12±0.08	0-0.28	0.019	22	0.11±0.07	0-0.22	0.014
B-Ent	41	0.07±0.05	0-0.17	0.006	34	0.13±0.13	0-0.68	0.034	22	0.10±0.08	0-0.29	0.013
B-Zyg	41	0.12±0.09	0-0.31	0.020	32	0.11±0.10	0-0.41	0.021	20	0.14±0.09	0-0.35	0.028
H-Zyg	41	0.10±0.09	0-0.53	0.018	32	0.08±0.05	0-0.25	0.010	20	0.10±0.06	0-0.25	0.014
St-Zyg	41	0.12±0.10	0-0.49	0.025	30	0.11±0.08	0-0.31	0.018	20	0.13±0.12	0-0.44	0.029
Vom-Po	38	0.13±0.12	0-0.44	0.031	33	0.25±0.22	0-0.83	0.114	22	0.34±0.21	0-0.8	0.114
CFA1	41	0.11±0.03	0.06-0.22		35	0.14±0.05	0.07-0.3		23	0.15±0.05	0.07-0.3	
*If-Ent	41	0.11±0.09	0-0.36	0.020	33	0.16±0.12	0-0.5	0.035				
*Ot-N	41	0.13±0.11	0-0.43	0.028					22	0.25±0.15	0-0.57	0.088
*Ni-P					24	0.07±0.06	0-0.25	0.008	11	0.06±0.04	0-0.14	0.006
*Orb Hgt					33	0.09±0.09	0-0.47	0.014	21	0.06±0.05	0-0.17	0.006
*Po-Zyg	41	0.15±0.10	0-0.38	0.028					21	0.37±0.21	0.1-0.78	0.151

Table 3.7

Trait	Norway compared to Barff						Norway compared to Husvik					
	|FA1|		FA4		FA10		|FA1|		FA4		FA10	
	z	p-value	F	p-value	F	p-value	z	p-value	F	p-value	F	p-value
Ent-P	-2.10	0.04^	2.34	0.005^	2.04	0.02^	-3.27	0.001*	1.77	ns	1.77	ns
Ni-Ot	-1.16	ns	1.25	ns	1.17	ns	-0.41	ns	1.48	ns	1.68	ns
Ot-Ect	-1.74	ns	1.78	0.04^	1.85	0.03^	-0.67	ns	1.60	ns	1.92	ns
N-Ent	-1.91	ns	1.33	ns	1.32	ns	-0.71	ns	1.31	ns	1.59	ns
Orb Lgth	-0.01	ns	1.13	ns	1.18	ns	-0.84	ns	1.48	ns	1.73	ns
B-If	-1.55	ns	1.22	ns	1.10	ns	-1.36	ns	1.16	ns	1.07	ns
B-Ent	-2.45	0.01^	5.31	<0.0001***	6.35	0.0001***	-1.45	ns	2.07	0.02^	2.36	0.009^
B-Zyg	-0.28	ns	1.10	ns	1.08	ns	-1.02	ns	1.45	ns	2.20	0.01^
H-Zyg	-0.34	ns	1.93	0.03^+	2.01	0.02^+	-0.79	ns	1.24	ns	1.37	ns
St-Zyg	-0.28	ns	1.36	ns	1.38	ns	-0.08	ns	1.17	ns	1.14	ns
Vom-Po	-2.37	0.02^	3.76	<0.0001***	3.60	0.0001***	-4.11	<0.0001***	3.67	0.0002**	3.95	<0.0001***
If-Ent	-2.01	0.04^	1.77	0.04^	1.58	ns						
Ot-N							-2.99	0.003*	3.18	0.0007*	3.62	0.0002*
Po-Zyg							-4.28	<0.0001***	5.43	<0.0001***	6.64	<0.0001***

Trait	Barff compared to Husvik					
	|FA1|		FA4		FA10	
	z	p-value	F	p-value	F	p-value
Ent-P	-0.99	ns	1.32	ns	1.15	ns
Ni-Ot	-0.48	ns	1.85	ns	1.96	0.04^
Ot-Ect	-0.50	ns	1.12	ns	1.07	ns
N-Ent	-2.26	0.02^	1.75	ns	2.10	0.04^
Orb Lgth	-0.84	ns	1.31	ns	2.59	0.01^
B-If	-0.11	ns	1.42	ns	1.18	ns
B-Ent	-0.72	ns	2.57	0.01^+	2.68	0.009^
B-Zyg	-1.26	ns	1.32	ns	2.38	0.01^
H-Zyg	-1.25	ns	1.55	ns	1.47	ns
St-Zyg	-0.18	ns	1.59	ns	1.58	ns
Vom-Po	-1.57	ns	1.02	ns	1.10	ns
Ni-P	-0.20	ns	1.07	ns	1.23	ns
Orb Hgt	-1.05	ns	2.32	0.02^+	3.56	0.001*+

Table 3.7 Comparison of true fluctuating asymmetry among populations. Mann-Whitney test comparing the absolute value of |R-L| to give Z value and an F-statistic comparing the variance. ns - not significant, ^ significant before application of Bonferroni correction. After application of Bonferroni correction *p<0.05, **p<0.01, ***p<0.001. In all cases Norway showed more FA than Barff or Husvik and Barff showed more FA than Husvik unless marked with + to indicate the cases where Norway showed less FA than Barff or Husvik and Barff showed less FA than Husvik.

The only traits shown are those which did not show significant DA in either of the two population. Comparison of all traits are shown in Appendix 2.

3.3.7 Differences in single trait FA among the three populations

Basic statistics showing the mean and range of FA1 and the value for FA4 for traits which showed true FA in each of the populations, are shown in the table 3.6. The comparisons of FA1, FA4 and FA10 for all traits showing true FA among each of the populations is shown in table 3.7. Versions of tables 3.6 and 3.7 which include all traits measured are shown in Appendix 2. The significant results are highlighted in the following text.

Out of the twelve bilateral traits which showed true FA, there were two traits, Vom-Po and B-Ent ($p<0.001$) in which there was significantly more FA in the Barff population than the Norwegian population and a further two traits (Ent-P and Ot-Ect) where there appeared to be more asymmetry in the Barff population than the Norwegian population though this was not significant after the Bonferroni correction. Only in one trait (H-Zyg) was there more indication of FA in the Norwegian population than the Barff population but this weak difference was not apparent in all measures of FA and may be a false positive (Type I error) because after the Bonferroni correction, the results were no longer significant.

Thirteen traits showing true FA were compared between the Norwegian and Husvik populations. When the absolute difference FA1 was considered, there was significantly more FA in traits Vom-Po ($p<0.001$), Po-Zyg ($p<0.0001$), Ot-N ($p<0.05$) and Ent-P ($p<0.05$, Bonferroni correction applied) in the Husvik than the Norwegian population. When the variances (FA4) were compared, there was a significant difference in the comparison of Vom-Po ($p<0.01$), Po-Zyg ($p<0.001$) and Ot-N ($p<0.05$, Bonferroni correction

applied). There were no traits which showed greater FA in the Norwegian population compared to the Husvik population.

Scatter plots comparing right and left measurements for Vom-Po, Po-Zyg and Ot-N (Husvik compared to Norway) and Vom-Po and B-Ent (Barff compared to Norway) are shown in figure 3.3.

3.3.8 Differences in composite value for multiple trait FA among the three populations

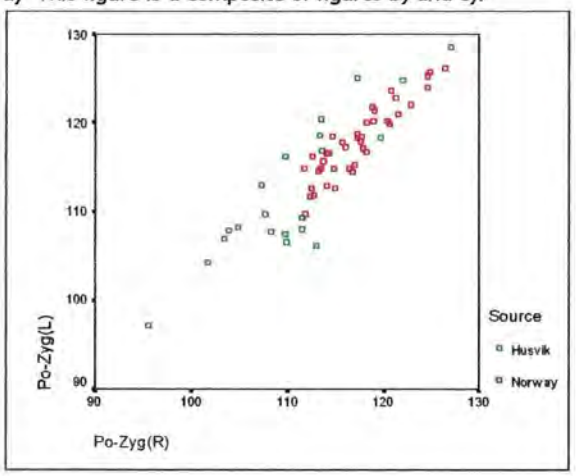
The composite FA value for multiple traits, CFA considered the values of asymmetry averaged across the traits which showed true FA (excluding traits If-Ent, Ot-N, Ni-P, Orb Hgt & Po-Zyg which showed significant DA in one of the populations and excluding Chk Tth which was considered unreliable due to continued eruption of teeth). Due to the exclusion of If-Ent and Po-Zyg, there was no significant interdependence between the traits that were included in CFA.

When CFA was compared between populations there was a significant difference between Norway and Barff at the level $p<0.005$ for both CFA1 and CFA3 and a significant difference between Norway and Husvik at the level $p<0.005$ for CFA1 and $p<0.05$ for CFA3. There was no significant difference between the two bottlenecked populations (see table 3.8).

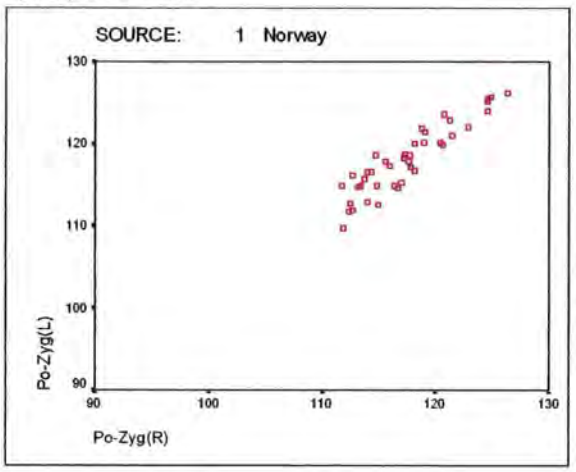
Table 3.8	CFA 1		CFA 3	
	z	p-value	z	p-value
Norway compared to Barff	-3.13	0.0018**	-3.05	0.002**
Norway compared to Husvik	-3.17	0.0015**	-2.46	0.014*
Barff compared to Husvik	-0.83	ns	-0.68	ns

Fig 3.3 Scatter plots of left against right side for traits that showed significantly more FA in South Georgia population than Norway.

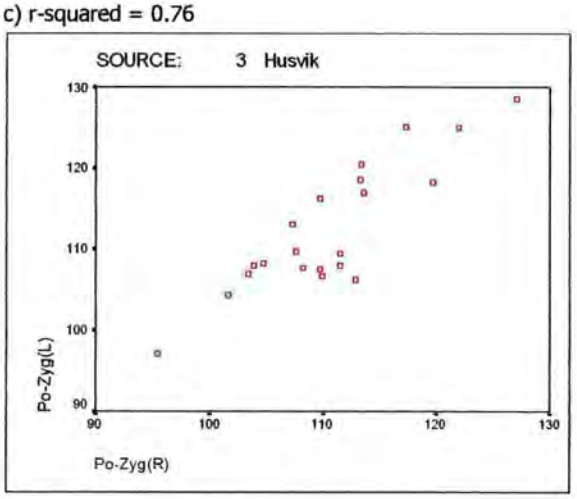
a) This figure is a composite of figures b) and c).



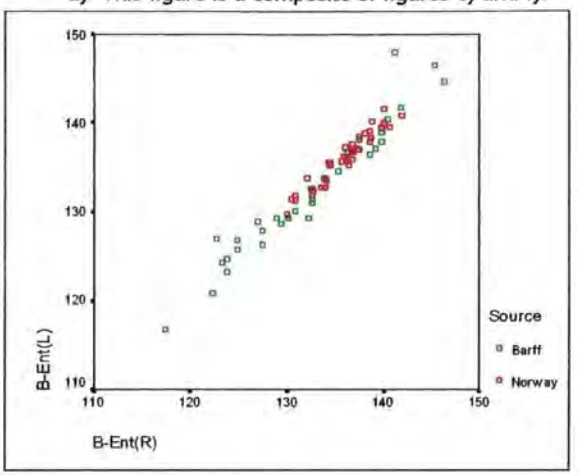
b) r-squared = 0.85



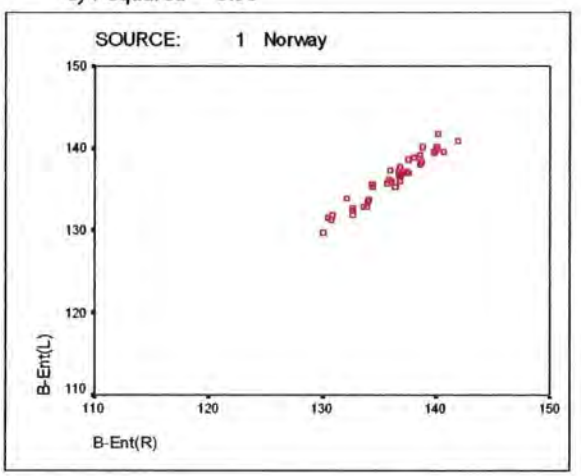
c) r-squared = 0.76



d) This figure is a composite of figures e) and f).



e) r-squared = 0.93



f) r-squared = 0.94

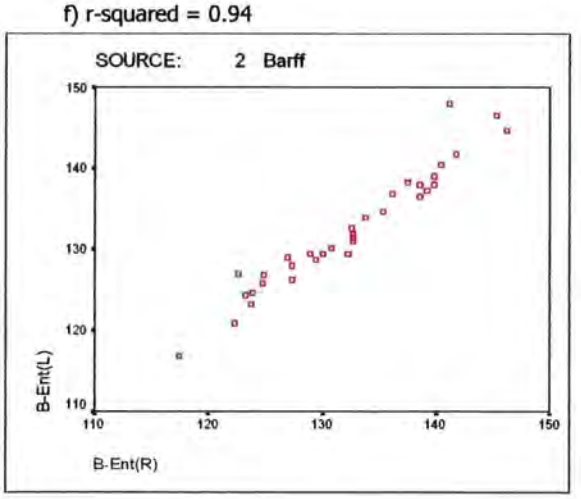
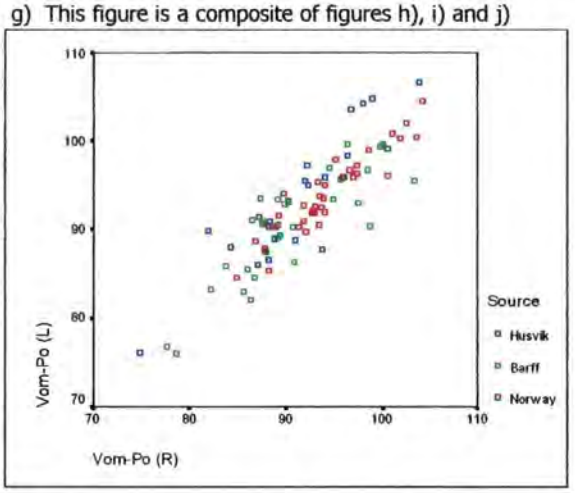
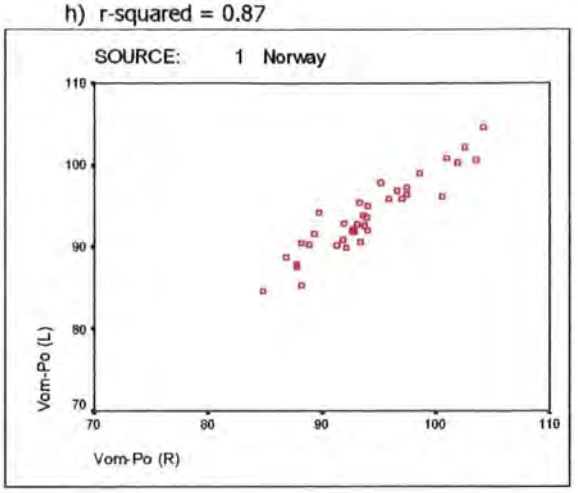


Fig 3.3 Scatter plots of left against right side for traits that showed significantly more FA in South Georgia population than Norway.

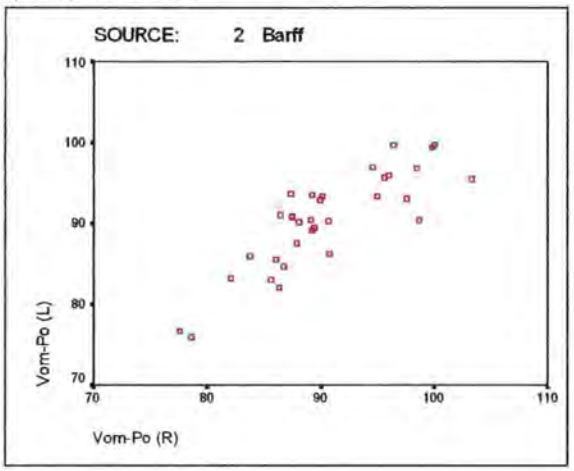
g) This figure is a composite of figures h), i) and j)



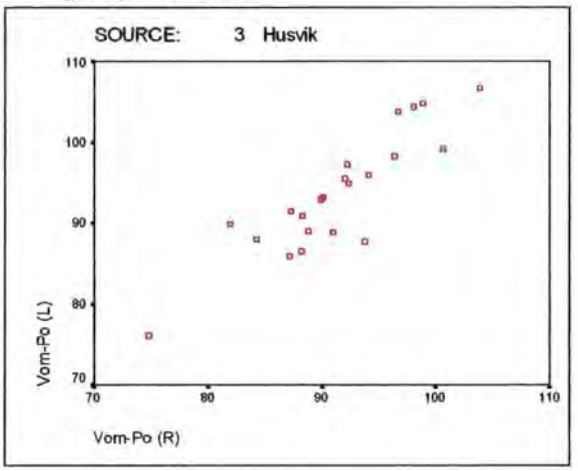
h) r-squared = 0.87



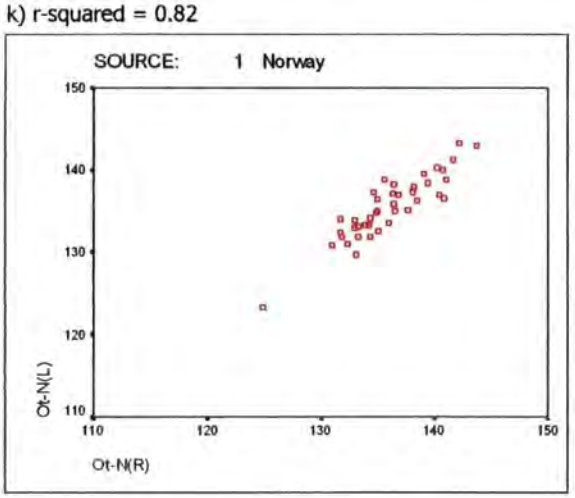
i) r-squared = 0.72



j) r-squared = 0.79



k) r-squared = 0.82



l) r-squared = 0.87

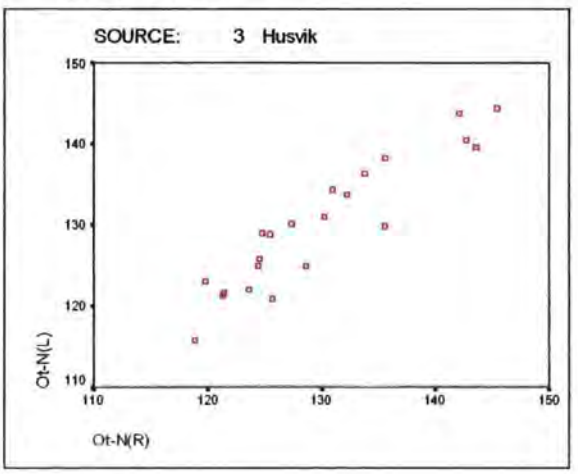
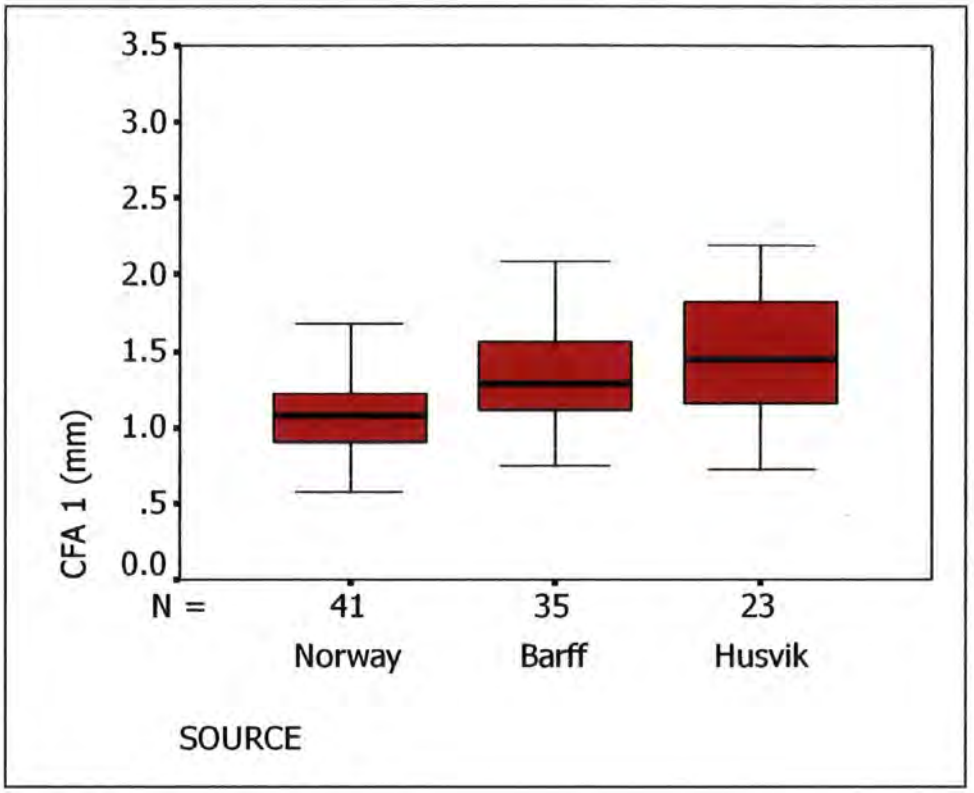


Figure 3.4
Box and whisker plots to compare CFA1 and CFA3 for each of the populations



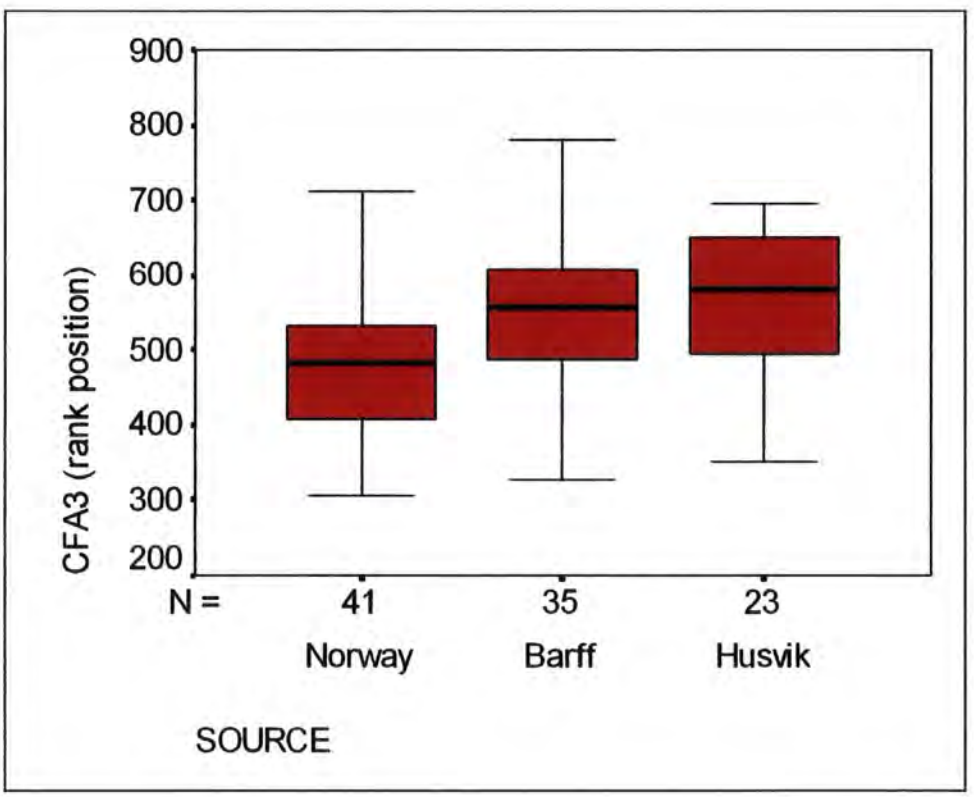


Figure 3.5 Histograms to show residual mean squared for 42 cranial traits regressed against a measure of skull size for each population.
***p<0.001, **p<0.01, *p<0.05

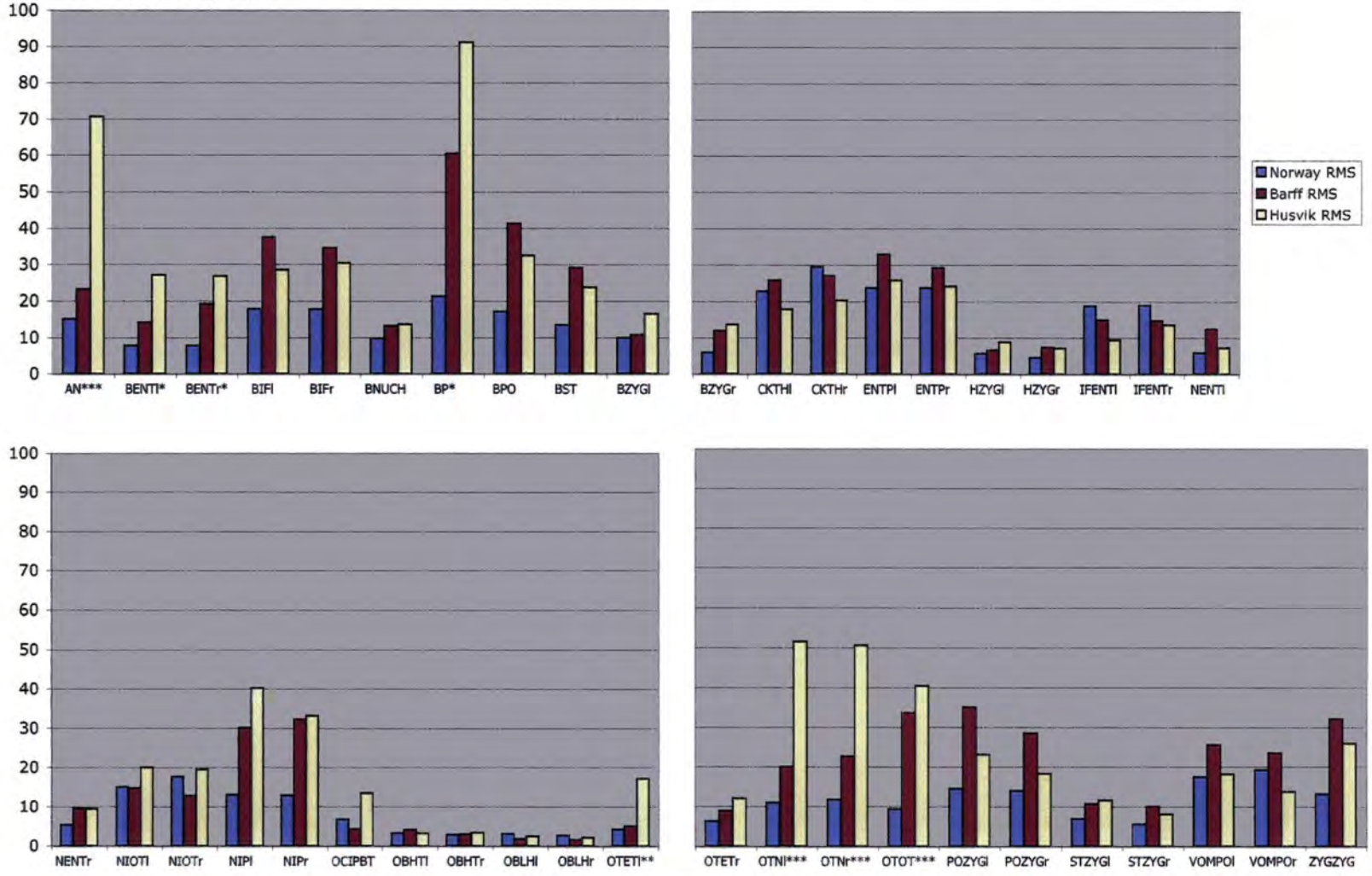
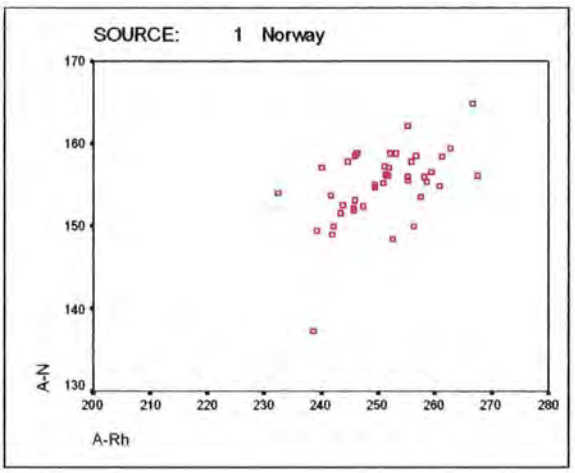
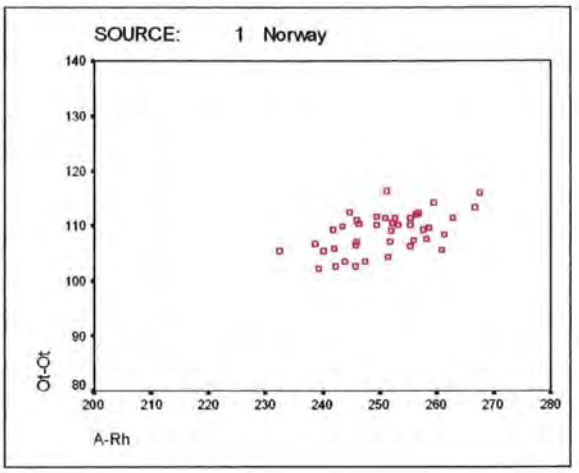
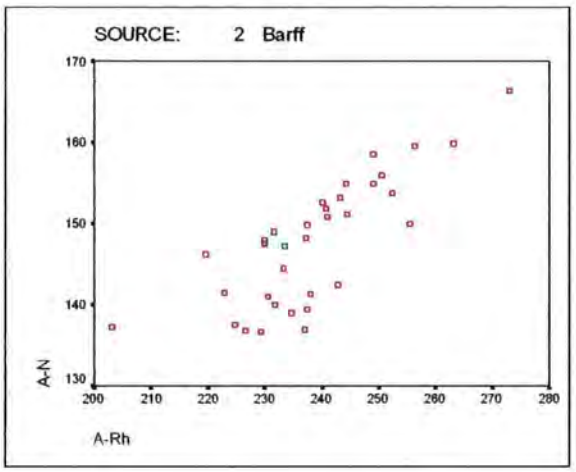
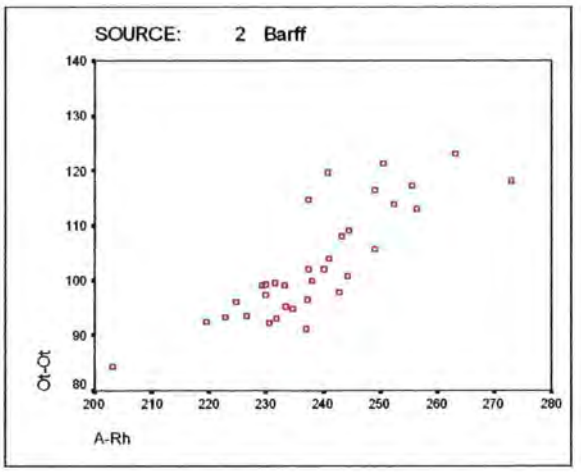


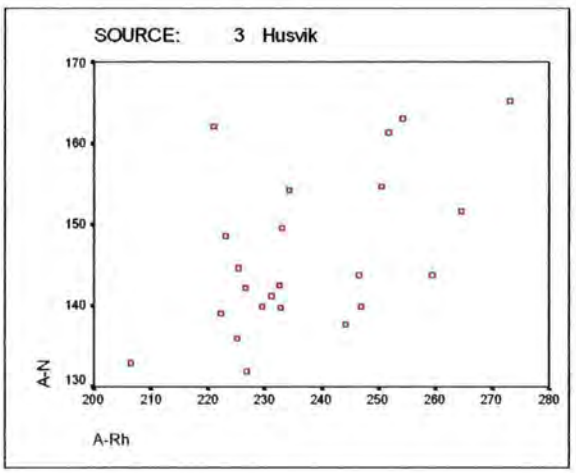
Figure 3.6 Scatter plots of traits, A-N and Ot-Ot against A-Rh

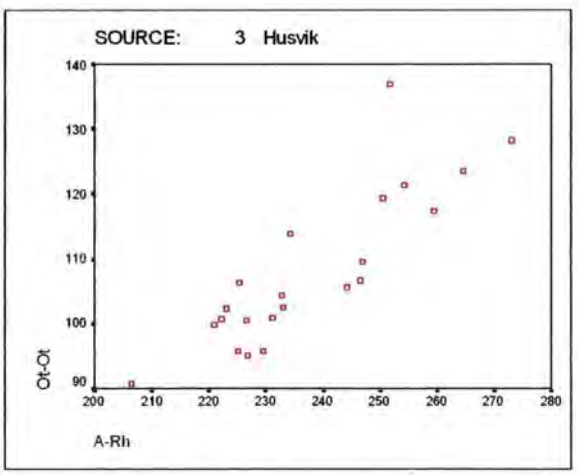












The comparison of CFA 1 and CFA 3 are shown graphically as box and whisker plots in figure 3.4.

3.3.9 Variance

There were a total of 42 traits compared and in each of the post-bottleneck populations 34 traits showed increased variability compared to the Norwegian population. Of these 34 traits, one from the Barff population (Ot-Ot) was significantly different ($p<0.001$) and eight from the Husvik population (AN, Ot-Ot, Ot-NR and Ot-NL at the level $p<0.001$, Ot-EctL at $p<0.01$ and B-P, B-EntL and B-EntR at $p<0.05$) after the sequential Bonferrroni correction was applied.

When only aged-skulls (Norway n=16, Barff n=32, Husvik n=20) were compared the result were largely similar with equivalent F-values. However, most likely due to a smaller sample size none of these results were significant after Bonferroni correction.

Histograms of the residual mean squares for 42 cranial traits regressed against A-Rh as a measure of skull size are shown for each population in figure 3.5. Scatter plots of traits A-N and Ot-Ot against A-Rh are shown in figure 3.6.

Both the left and right measures of Ot-N showed significantly more variability in the Husvik population than the Norwegian population; this trait also showed significantly more FA in Husvik compared to Norway. Both left and right measures of B-Ent showed significantly more variability in the Husvik population than the Norwegian population; this trait also showed

significantly more FA in the Barff population compared to Norway and more FA in Husvik compared to Norway though not significantly so after the application of Bonferroni.

3.3.10 Comparison of overall size of skulls

The measurements of the different lengths, heights and width of the skulls were used to compare the overall size of the skulls in the three populations (see table 3.9 and 3.10)

For every trait except Zyg-Zyg the mean value was greater for the Norwegian population than either of the South Georgia populations. The Norwegian skulls were significantly longer and higher than the South Georgia skulls.

Table 3.9 Basic measurements of different lengths taken along the midline, widths and heights of male skulls to compare the three populations.

Trait	Norway		Barff		Husvik	
	n	mean (mm)	n	mean (mm)	n	mean (mm)
A-P (length)	29	325.8±6.5	8	319.8±20.5	8	314.0±23.6
A-Rh (length)	30	252.9±6.5	14	241.5±17.4	16	237.8±17.4
A-N (length)	30	156.1±3.6	14	150.0±8.8	16	146.3±10.4
B-P (length)	29	292.9±5.3	8	290.1±18.37	7	287.9±21.4
B-Po (length)	30	150.0±4.5	14	143.1±9.2	15	144.9±11.4
B-St (length)	30	95.3±3.6	13	91.8±7.0	14	89.6±8.3
B-Nuch (height)	30	78.8±3.0	14	75.6±5.6	15	76.0±6.4
Zyg-Zyg (width)	30	127.2±3.5	14	132.0±10.3	15	127.4±9.0
Ot-Ot (width)	30	109.8±3.0	14	105.9±11.2	15	108.1±13.2
Occipbrth (width)	30	63.0±2.8	14	59.7±2.7	15	61.6±4.7

Table 3.10 Comparison of the overall size of skulls between the three populations using the Kruskal Wallis test. Significance: ns - not significant, ^significant before application of Bonferroni correction, *p<0.05, **p<0.01, ***p<0.001, Bonferroni correction applied.

	Male skulls only		Female skulls only	
	X^2	p-value	X^2	p-value
A-P (length)	6.69	0.04^	8.43	0.01^
A-Rh (length)	13.86	0.001*	5.83	ns
A-N (length)	13.06	0.001*	4.69	ns
B-P (length)	3.46	ns	5.34	ns
B-Po (length)	8.71	0.013^	3.60	ns
B-St (length)	9.56	0.008^	3.45	ns
B-Nuch (height)	5.15	ns	11.42	0.003*
Zyg-Zyg (width)	2.50	ns	4.55	ns
Ot-Ot (width)	2.92	ns	6.25	0.04^
Occipbrth (width)	10.38	0.006^	15.46	0.0004**

A direct comparison was made between aged female skulls from Norway and aged female skulls from Barff (see table 3.11). There were insufficient female skulls from the Husvik population (n=4) for these results to be added to this comparison.

Table 3.11 Comparison of means by t-test between aged female skulls from Norway and aged female skulls from Barff. Significance: ns not significant, ^significant at the level $p<0.05$ before application of the Bonferroni correction ** $p<0.01$, Bonferroni correction applied.

Trait	Aged female skulls Norway		Aged female skulls Barff		Comparison of populations
	mean (mm) ± SD	n	mean (mm) ± SD	n	t
A-P (length)	322.6 ± 9.7	10	309.5 ± 11.4	10	2.8^
A-Rh (length)	245.6 ± 9.7	10	236.0 ± 6.6	12	2.8^
A-N (length)	151.7 ± 5.6	10	147.0 ± 5.6	12	1.9 ns
B-P (length)	290.5 ± 8.6	10	281.6 ± 10.7	10	2.1 ns
B-Po (length)	146.3 ± 5.8	10	141.5 ± 7.5	12	1.6 ns
B-St (length)	92.4 ± 3.9	10	88.0 ± 5.5	12	2.1^
B-Nuch (height)	74.3 ± 2.3	10	72.1 ± 3.6	10	1.7 ns
Zyg-Zyg (width)	124.4 ± 4.1	10	129.4 ± 6.9	12	-2.0 ns
Ot-Ot (width)	106.8 ± 4.1	10	102.1 ± 8.8	12	1.5 ns
Occipbrth (width)	61.0 ± 2.1	10	57.9 ± 1.7	12	3.8**

Table 3.12 Comparison of 20 month old male Norwegian skulls with aged Barff and Husvik skulls. Significance of the difference between the post bottleneck populations and the Norwegian population: ns not significant, ^p<0.05 before application of Bonferroni correction, *p<0.05, **p<0.01, ***p<0.001, Bonferroni correction applied,

	Young male Norway	Aged male Barff	Aged male Husvik
	mean (mm) ± SD	mean (mm) ± SD	mean (mm) ± SD
A-P (length)	325.7 ± 6.3	324.5 ± 16.8 ns	314.0 ± 23.6 ^
A-Rh (length)	253.3 ± 5.9	248.5 ± 10.8 ns	238.9 ± 17.4 ***
A-N (length)	156.2 ± 3.1	152.7 ± 7.8 ns	146.7 ± 10.5 ***
B-P (length)	292.7 ± 5.4	294.4 ± 14.8 ns	287.9 ± 21.4 ns
B-Po (length)	150.2 ± 4.6	146.5 ± 6.8 ns	144.9 ± 11.8 ns
B-St (length)	95.7 ± 3.6	93.5 ± 6.2 ns	89.9 ± 8.6 *
B-Nuch (height)	79.6 ± 2.7	77.0 ± 5.5 ns	76.1 ± 6.6 ^
Zyg-Zyg (width)	127.5 ± 3.5	135.3 ± 8.7 *	127.8 ± 9.2 ns
Ot-Ot (width)	109.8 ± 2.3	110.0 ± 8.3 ns	108.6 ± 13.5 ns
Occipbrth (width)	63.6 ± 2.5	60.5 ± 2.5 *	61.5 ± 4.9 ns

3.4 Discussion

3.4.1 Influence of Age and Sex

Sexual dimorphism is marked in reindeer (Leader-Williams 1988) and there was no surprise to see that in this study there was a significant difference between the overall size of the male and female skulls. Previous studies suggest there is also a significant difference in the rate of growth

between males and females. In a comparison of growth between herds in Norway, relating diastema length to range quality, there was a steep increase in size for both sexes up to one year. After this stage, female growth leveled out while in males growth continued to increase (Reimers 1972). By 12 months old reindeer have achieved approximately 80-90% of their adult skeletal size (Leader-Williams 1988) though Bergerud (1964) was unable to distinguish male and female mandibles that were less than three years old due to the fact that they were still growing.

For these reasons it was important to consider subsamples of single sex and age classes particularly in the comparison of overall skull size but also in the comparison of FA. Were the sample sizes large enough it would have been prudent to only use aged skulls in the comparison of FA. However, to ensure as large a sample size as possible, skulls from animals above one year old were used. No difference in FA was found between different age and sex classes. This is as expected, since the phenomenon of developmental instability occurs *in utero,* and is a stochastic process.

3.4.2 Directional Asymmetry

Directional asymmetry is the propensity of one side to develop more than the other (Van Valen 1962). As an adaptive asymmetry, it has a significant but unknown genetic basis (Palmer and Strobeck 1986; Moller 1997b) and has thus been argued as not suitable as a measure of developmental stability. However, a study of the effect of the stress of toxic

chemicals on *Drosophila* found that at the highest concentration of the toxin there was a transition from FA to DA (Graham *et al.* 1993). Though not ideal as a measure of developmental instability, the presence of DA in any of the reindeer populations would be of interest if it indicated an even greater stress than that causing FA. The other reason the presence of DA is of interest is that there have been few previous reports of asymmetry in reindeer.

This study showed that there were two traits in the Norwegian population, two in the Barff and one in the Husvik population that showed significant evidence of directional asymmetry. The fact that there was no consistency as to whether the right or the left was the larger side suggests that the skulls are somehow skewed rather than uniformly larger on one side.

Reindeer antlers are frequently asymmetrical in the numbers of tines and more likely to have an extra or larger tine on the left antler than the right (Davis 1973). Behavioural observations suggested that the function of this extra tine was to scrape away snow to access food in the winter and to act as a shield to protect the male's eyes whilst fighting. On the basis of the observation that the dominant tine may be on different sides on alternate years, conclusions were drawn that there was no heritable basis as to the side of dominance (Davis 1973; Davis 1974). However, the reason that directional asymmetry is not useful as a measure of developmental asymmetry is due to the very fact that it does generally have an heritable basis (Palmer and Strobeck 1986). More recent studies of asymmetries in reindeer antler have shown that the numbers of tines and total antler length showed deviations from normality and a mean of zero but that there were no

indications of directional asymmetry in the length of the main beam of the antler or the jaw length (Lagesen and Folstad 1998). A different study showed apparently no indication of DA in either antler length or jaw length (Markusson and Folstad 1997). Although the traits measured in this study did not include either antler length or jaw length specifically, there were a number of skull traits which did show some evidence of DA.

There are many examples of DA or handed-bias naturally occurring in a number of taxa (summary in Moller 1997b) and these are often associated with particular unidirectional behaviours. For example, the beak of the wry-billed plover is consistently bent up to 12° to the right associated with a 'right-handed' feeding technique (Neville 1976).

There is a 'handed-bias' associated with behaviour of reindeer that may have relevance to the finding of DA. When gathered or disturbed, northern hemisphere reindeer will generally circle in an anticlockwise direction (Dieterich – personal communication). Due to the uniformity of this behaviour, the corrals used for gathering are designed to allow for anticlockwise circling and this design may encourage consistency in the direction of circling. However, there are no previous reports of this sort of directional behaviour having skeletal morphometric effects in reindeer and due to the inconsistency of both trait and direction of DA, this association is suggested very tentatively.

3.4.3 Fluctuating Asymmetry

Small random deviations from bilateral symmetry have little or no measurable genetic heritability and FA provides an appealing measure of developmental noise (Palmer and Strobeck 1986).

Previous work on asymmetry in reindeer suggested that reindeer antlers showed FA and that high FA was correlated with larger antler size (Markusson and Folstad 1997) and measures of immunity to parasites (Lagesen and Folstad 1998). Neither of these studies found a relationship between jaw length FA and either antler FA or other measures of fitness. They concluded that ornamental character (antler) FA did not reflect overall body FA due to the different time span over which the antlers developed compared to the rest of the body. All the traits measured in this study were 'body' measurements rather than measurements of sexual ornament such as antlers.

The post-bottleneck populations in this study showed significantly greater FA than the pre-bottleneck population – in two out of 12 traits compared between Barff and Norway and in four out of 13 traits compared between Husvik and Norway. In a comparable study of a population of northern elephant seals which had experienced an extreme bottleneck due to over-hunting, Hoelzel (2002) measured four traits and found there was increased FA in one trait of the post bottleneck population compared to the pre-bottleneck population and in three traits of the post-bottleneck population

compared to a population of southern elephant seals which had not experienced a bottleneck.

The range of absolute values of FA in this study were comparable to results of previous studies of FA in cheetah (Kieser and Groeneveld 1991), tamarins (Hutchison and Cheverud 1995), red squirrels (Wauters *et al.* 1996), red jungle fowl (Kimball *et al.* 1997) and elephant seals (Hoelzel 2002).

Based on mathematical algorithms, the relation between FA and quality or FA and stress is expected to be weak and the relationships are expected to be variable or even absent depending on the characters assessed (Leung and Forbes 1997). For this reason composite values of FA are considered to be more reliable indicators of developmental instability or stress (Leung *et al.* 2000). In this study two composite values of FA were used, CFA1 which was the average FA of all traits that showed true FA in all populations and CFA3 which was a standardised value based on the rank position of FA for each individual. For both parameters there was significantly more FA in the post-bottleneck populations compared to the Norwegian population. There was no significant difference between the two post-bottleneck populations.

As a measure of developmental stability, fluctuating asymmetry may show no change or increase with increasing 'stress', either intrinsic stress of predominately genetic origin or extrinsic stress due to environmental factors (Palmer and Strobeck 1986; Palmer 1994).

Each post-bottleneck population experienced the genetic stress of inbreeding resulting from the low founder numbers at the bottleneck and

subsequently the low herd numbers and isolation of each population. Direct comparison of the genetic effect of the bottleneck has been made in Chapter Two.

Previous studies into the effects of genetic stress on FA have found inconsistent results. Wayne et al (1986) reported that the cheetah which exhibits very low levels of heterozygosity, showed significantly higher levels of asymmetry than three other feline species, known to be polymorphic for genetic variation. However, this work was heavily criticised for not adequately considering measurement error or type of asymmetry and subsequent studies, albeit measuring different traits, showed the cheetah to have comparable levels of asymmetry to other felids (Kieser and Groeneveld 1991). Hutchison and Cheverud (1995) showed a negative correlation between heterozygosity and levels of cranial FA in three taxa of tamarins. They suggested that previous studies failed to demonstrate the use of FA to identify populations endangered by reduced genetic variability and/or under stress due to small sample sizes, lack of repeat measurements and lack of appropriate reference populations.

A study that compared the FA of meristic characters (number of foramina on the skull and mandible) of both dama and dorcas gazelles that were introduced in small herd sizes in Spain found only one trait to show a significant increase in FA associated with inbreeding. However, when a composite measure of foramina from five skull regions was considered there was a significant increase in FA with increased inbreeding (Alados *et al.* 1995).

This study showed clearly that there was increased levels of FA in both post-bottleneck populations of reindeer compared to the pre-bottleneck population in certain individual traits and in the comparison of composite FA. However, it is not simple to disentangle the effects of the genetic stress of the bottleneck from the different environmental effects on each population.

Studies of the relationship between FA and inbreeding have been most easily undertaken in the laboratory where the environmental influences are more easily controlled. Indeed, following studies of inbred and outbred laboratory mice, Leamy and co-workers (2001) concluded that FA was not generally a sensitive proxy measure for fitness but that it could be associated with fitness reductions for certain genetic stressors and no significant difference in FA was found when inbred and crossed lines of yellow dung flies were compared supporting the notion that FA levels are stress, trait and taxon-specific (Hosken *et al.* 2000).

A number of studies have looked at FA as an indicator of environmental stress. Red squirrels in small fragmented woodlands were shown to have slightly higher levels of FA and smaller body size than those squirrels in larger forests (Wauters *et al.* 1996) and in a comparison of otter populations in Europe the populations in decline showed some increase in FA and a decrease in skull size compared to the more viable populations (Pertoldi 2000). FA levels were found to be significantly greater in at least one trait in five out of the six most abundant species of birds living in forest fragments compared to unfragmented Brazilian rain forests (Anciaes and Marini 2000). Indeed it has been suggested that FA can be used as an early warning of a

decrease in survivability. In a study of three populations of Taita thrushes in three differentially disturbed forest fragments, tarsus FA and survival were considered. Increased asymmetry but comparable survivability was found in the moderate compared to the least disturbed habitat but there was both increased asymmetry and reduced survivability in the population in the highly deteriorated habitat (Lens *et al.* 2002).

Direct comparison of the environmental factors affecting each of the three reindeer populations is not simple. The Norwegian herd is handled very little with no supplementary feeding given to the herd and no routine application of medication. The herd is gathered twice a year to slaughter calves and cull surplus yearling males and old females. Though limited, this interference may exert some selection pressure on the herd.

There has been no controlled management of the South Georgia herds and availability of food and the type of terrain on the South Georgia is comparable between the two herds. The Barff herd increased in size much more rapidly than the Husvik herd which may suggest that they were fitter or that the environmental pressures limiting reproduction were less for the Barff herd. There is some indication that the release of the original Husvik reindeer close to three active whaling stations meant that there was considerable poaching of the reindeer up until the 1950s when they became protected (Leader-Williams 1988).

3.4.4 Variance

In this study there were increased levels of character variation in the Husvik compared to the Barff population and in the Barff population compared to the Norwegian population. This appeared to reflect the same pattern as that shown by the levels of FA with significantly higher levels in both the post-bottleneck populations compared to the Norwegian population and greater levels in the Husvik population. As well as the apparent link between FA and character variation on a population level, there was apparently also a link by character. These results would tend to support the suggestion that there may be a relationship between the levels of canalization and developmental stability across a range of characters (Clarke 1998c).

In a comparison of skulls from pre and post-bottleneck populations of northern elephant seals, Hoelzel (1999) showed increase in variability in the post-bottleneck population in 22 out of 25 characters measured, with ten traits significant at the level $p<0.01$. This increase in variability between the pre- and post-bottleneck populations of elephant seals was comparable to that shown in this study.

3.4.5 Size comparison

Comparison of the overall size of Norwegian skulls with South Georgia skulls showed that for the majority of traits the Norwegian skulls were consistently larger than the South Georgia skulls. This trend was even apparent when 20-month old male Norwegian skulls were compared to aged

South Georgia skulls despite the fact that growth in males continues up to six years old (Bergerud 1964; Reimers 1972; Skogland 1983).

This difference in size between the mainland Norwegian skulls and the island South Georgia skulls may indicate the tendency in these reindeer of the island populations to be smaller than their mainland counterparts in common with other artiodactyls. The rationale behind this evolutionary trend is that resource limitation (intensified by high population densities and intraspecific competition on species-poor islands) and release from predation confers high fitness on smaller individuals as they require less energy to survive and reproduce (Lomolino 2005).

There are no predators on South Georgia. Wolves, lynx and wolverine may predate on European reindeer (references given in Leader-Williams 1988) and in the 2005/6 hunting year, seven lynx and three wolverines were killed under licence in the two Norwegian regions, Oppland and Buskerud, which are closest to the area where the Filefjell Reinlag herd ranges (www.ssb.no/rovdyravg_en). However, the owner of the Filefjell Reinlag herd does not consider that predation is a significant cause of losses within his herd (Asgrim Opdal – personal communication).

In this study, the Barff skulls were found to be consistently larger than the Husvik skulls. This was in direct contrast to the findings of studies in the seventies when measurements of crown to tail length, jaw length and hind foot lengths were taken on reindeer from the Husvik (n=99), Barff (n=300), and Royal Bay herds (n=100) (Leaderwilliams and Ricketts 1982). The Royal

Bay herd was formed from a group of Barff reindeer that spread across the front of the Cook Glacier in the 1960s. Leader Williams and Ricketts showed that Husvik reindeer were significantly larger than those from the Barff herd but on some traits, such as jaw length, the Husvik herd were not larger than the Royal Bay herd. The fact that Leader Williams and Ricketts found differences between the Barff and Royal Bay reindeer indicates that significant changes in skeletal size of reindeer of the same genetic stock is possible within a timescale of less than 15 years. They attributed the differences in body size as primarily due to the differences in resource availability and diet.

Resource limitation is a well documented reason for smaller body size in reindeer (Reimers 1972; Skogland 1983). South Georgia reindeer are unique amongst reindeer in their dependence on one species of grass, *Poa flabellata* for their winter diet. In Norway, lichen form the predominant part of the winter diet and further north, arctic reindeer depend on mosses and graminoids (references in Leader-Williams 1988). There is evidence of overgrazing in both the Barff and Husvik areas (Leaderwilliams *et al.* 1981; Leaderwilliams *et al.* 1987). Although census data suggests that the current size of the Husvik herd is considerably greater than it was in the 1970s (see table 2.3) there is also evidence that the herd has expanded over a greater range (Moen and MacAlister 1994). There have been no recent counts of the Barff herd.

As suggested in the introduction, there are a number of different factors influencing skull size and it is not simple to partition out the genetic impact of the bottleneck with the environmental differences between the

populations. Considering primarily the direct comparison of the results from this study with those found by Leader Williams and Ricketts (1982) suggests that the difference in skull size between the herds is most likely to be a plastic response due to differences in resources. Human selection of the Norwegian herd would also have had an effect as each year reindeer are culled and some form of selective breeding takes place to improve carcass size and quality.

3.4.6 Conclusions

Quantifiable morphological differences were found between the three reindeer populations in this study. In common with the study of Northern elephant seals (Hoelzel 2002), the post bottleneck populations in this study showed significantly greater levels of fluctuating asymmetry and greater levels of morphological variance than the pre-bottleneck population.

These results are relevant and useful in the midst of numerous studies of FA which apparently show conflicting results. This study considered two separate populations which had undergone different bottlenecks of known size at known dates. It was possible to make direct comparisons between both post-bottleneck populations and with the pre-bottleneck population whilst rigorously eliminating the influence of measurement error or the presence of asymmetries other than true FA.

Although there was some overlap in the traits which showed significantly greater FA in each of the post-bottleneck populations, there was not a consistent pattern between the two. This detail is important as it

emphasises the stochastic nature of the bottleneck event and the specificity of results between different traits although within the same original population.

In common with studies of red squirrels (Wauters *et al.* 1996) and otters (Pertoldi 2000) in fragmented environments, the post-bottleneck populations in this study showed decreased skull size compared to the pre-bottleneck population. However, as already discussed, the reason for this effect is difficult to disentangle from the complex interaction of genetic and environmental pressures.

In order to partition out the genetic effects of the bottleneck from the environmental effects on each of the population it is necessary to make a comparison of the size of the skulls and the level of FA on an individual basis within each of the populations and to investigate how these measures vary with individual measures of genomic diversity. This will be addressed in Chapter Four.

Chapter Four

Relationship between indirect measures of fitness and genomic diversity at the level of the individual.

4.1 Introduction

The standard measure of the degree of inbreeding is that of Wright's coefficient of inbreeding, f, and is generally interpreted as the probability that two alleles at a locus in an individual are identical by descent (Wright 1921). Wright calculated inbreeding coefficients by studying detailed pedigree information but this is not easy with wild populations due to the lack of data. In studies which consider allozyme marker loci, individual average heterozygosity (the proportion of heterozygous loci for each individual) has often been used to estimate parental relatedness but this parameter has been described as crude (Amos *et al.* 2001). This is due to the fact that it only considers identity or non-identity of allelic phenotype at each locus (Coltman *et al.* 1998) and in allozyme studies which have with small numbers of polymorphic loci or few alleles, there will be a number of loci that are homozygous in the absence of inbreeding (Coltman *et al.* 1998; Coulson *et al.* 1998).

In studies of microsatellites, it has been suggested that there are more sensitive genetic measures of inbreeding based on the mutational difference between microsatellite alleles and the time since coalescence. Microsatellites consist of tandem repeats with different alleles defined by the number of

repeats. They show high polymorphism due to a relatively high mutation rate (ranging from 10^{-2} to 10^{-5}) with most observed mutations consisting of an increase or decrease of one repeat (Weber and Wong 1993; Di Rienzo 1994) though this depends on the locus. This forms the basis of the stepwise mutation model (Valdes *et al.* 1993) which suggests that new alleles are likely to have been derived from alleles one repeat unit different from themselves. Thus it is argued that allele lengths contain historical information and that microsatellite mutation dynamics can be used at the individual level to estimate both recent inbreeding and the degree of historical outbreeding (Coltman *et al.* 1998; Coulson *et al.* 1998; Coulson *et al.* 1999). Mean d^2 is a measure of the genetic distance between the gametes that formed the individual and it is calculated as the squared difference in length between the alleles at a locus averaged across all the loci considered (Coulson *et al.* 1998).

A further measure of inbreeding, internal relatedness (IR) considers the frequency of every allele and thus allows the sharing of rare alleles to be weighted more than the sharing of common alleles (Queller and Goodnight 1989; Amos *et al.* 2001). When calculated over several loci, the value is approximately normally distributed and more or less centered on zero for individuals born to 'unrelated' parents.

The use of heterozygosity or mean d^2 as valid indicators of the level of inbreeding and its consequences has been the subject of much debate. In studies of a red deer population from the Isle of Rum mean d^2 was found to be positively correlated with birth weight but no such relation was found between heterozygosity and birth weight (Coulson *et al.* 1998).

Also in harbor seals, mean d^2 explained more trait variance than heterozygosity (Coltman *et al.* 1998). It was argued that mean d^2 was a more appropriate measure than heterozygosity for considering the fitness consequences of inbreeding due to the opportunity it gives to distinguish 'highly outbred' with 'moderately outbred' individuals (Coulson *et al.* 1999). However, a number of further studies looking at the adult breeding success of the red deer (Slate *et al.* 2000), survival and parasite resistance in Soay sheep (Coltman *et al.* 1999) and the known inbreeding coefficient of a captive population of wolves (Hedrick *et al.* 2001) all suggested that mean d^2 was less powerful than heterozygosity at detecting the fitness consequences of inbreeding.

Indeed a much larger study of the red deer on Rum, using a panel of 71 microsatellites on 364 individuals, found that there was no correlation of individual d^2 scores across loci and there was no correlation of mean d^2 with either birth weight or juvenile survival (Slate and Pemberton 2002). The authors of this study concluded that heterozygosity was a much more robust measure than any of the mean d^2 based measures with which to detect inbreeding depression and heterosis. One of the reasons for this conclusion was that there is no obvious way to distinguish between different reasons for high variance in locus-specific d^2 measures. High variance may represent information about coalescent times between parental alleles or may be an artifact of either a high mutation rate or a non-stepwise mutation process (Slate & Pemberton 2002). Furthermore, they argue that as evidence is presented to suggest that microsatellites do not evolve in a simple stepwise

process (Ellegren 2000), the rationale behind using d^2 measures may be flawed.

Tsitrone et al (2001) undertook a theoretical investigation of how heterozygosity and d^2 are correlated with fitness by constructing models with different scenarios. They compared recent or close inbreeding (mating between close relatives) with deep inbreeding (mating between individuals from two isolated populations which were founded from the same ancestral population) (Tsitrone *et al.* 2001; Goudet and Keller 2002). For low mutation rates, they found that heterozygosity and d^2 were equivalent in their correlation with fitness. Under the close inbreeding model they found that fitness was more highly correlated with heterozygosity than with d^2 and that this was especially the case when markers with high mutation rates, such as microsatellites, were used. Under the deep inbreeding model they found that the relationship between the measures of inbreeding and fitness depended on two other parameters, that of mutation rate, μ and size of the subpopulations, N. The product of these two parameters determines the number of mutations per generation in the subpopulations. If $N\mu < 1$ then heterozygosity is a better indicator of fitness than d^2. This is due to the fact that most individuals in the subpopulations are homozygous due to fixing of alleles but after admixture of populations most individuals are heterozygous. Thus heterozygosity provides more useful information than d^2 because including information on the difference in allele sizes adds unnecessary noise to the estimation of inbreeding (Goudet and Keller 2002).

However, if the mutation rate is high and both the divergence time and subpopulation size are also high then most individuals are heterozygous whether they are parents or hybrids. In this situation heterozygosity does not reveal enough information but d^2 is more useful as individuals with parents from different subpopulations are more likely to have larger differences between the size of alleles than individuals with parents from the same subpopulation (Goudet and Keller 2002). Tsitrone *et al* (2001) concluded that there is little theoretical reason to use d^2 instead of heterozygosity when considering correlations between microsatellite genotype and fitness even in long-term inbreeding scenarios.

Meta-analyses of a number of studies, both published and unpublished, also suggested that heterozygosity was more strongly correlated with life-history traits than was mean d^2. It was also found that morphometric traits were less reliable than life-history traits in detecting inbreeding depression (Coltman and Slate 2003).

In this study the indirect measures of fitness that were investigated were fluctuating asymmetry (FA), body size, growth rate and longevity. The relationship between these indirect measures of fitness and the measures of genomic diversity were investigated at the level of the individual reindeer within each of the populations.

Fluctuating asymmetry has been defined as a population parameter (Van Valen 1962; Palmer and Strobeck 1986) as the levels of asymmetry fluctuate within the population. Indeed it is only possible to describe the

nature of subtle asymmetries in an individual with reference to the asymmetries of other individuals within the sample. Without reference to the rest of the sample, it is not possible to determine that the asymmetry seen is random FA rather than directional asymmetry (DA) or antisymmetry, both of which may have an heritable basis (Palmer 1994). However, individuals within the population will display a certain level of asymmetry as a reflection of how well their genotype can express the ideal phenotype under specific conditions (Moller 1997b).

Even in a sample which shows high FA, the mean of the frequency distribution of the difference between right and left side is still zero so, by chance, there are still many individuals which will not differ significantly from symmetry (Palmer 1994). Mathematical modeling of the relationship between FA with stress and fitness has indicated that unsigned FA has little predictive power at low levels of FA due to this category containing both high quality individuals and low quality ones that happen to have low FA by chance. In contrast, high FA values do reliably indicate low quality individuals (Leung and Forbes 1997).

Body weight in reindeer has been shown to be correlated with other measures of fitness. Female body weight directly affects calf body weight and hence survival and it is also correlated with age of sexual maturity and conception rate (Roed 1987). Male body size directly affects their fighting success for access to females and hence their reproductive success (Roed 1987).

Protein heterozygosity has been linked with growth rate and hence body size since studies in oysters (eg Singh and Zouros 1978) and the tiger salamander (Pierce and Mitton 1982) where associations were found between heterozygosity with body mass or snout to vent length respectively. Further studies in white tailed deer, sheep, pigs and humans (summarised in Mitton and Grant 1984) have confirmed these findings with higher growth rates and heavier body weight in more heterozygous individuals.

However, as discussed in Chapter Three, there is not a clear positive relationship between body size and fitness and on islands large herbivores tend towards a smaller size, thought to be due to the selection pressure of reduced resources and reduced predation. A plastic response of the phenotype to the environment together with a combination of selective forces in different directions may mask any direct relationship between body size and measures of genomic diversity.

Although heavily influenced by environmental factors and stochastic variation, longevity is usually associated with measures of total fitness (Kruuk *et al.* 2000). Studies of life-history traits of big-horn sheep showed that there was significant positive correlation between longevity and a other life-history traits, such as lifetime fecundity or reproductive success measured by survival of offspring (Berube *et al.* 1999; Reale and Festa-Bianchet 2000). One of these studies also showed significant positive correlation between longevity and body mass in midlife (Berube *et al.* 1999). Other studies, such as those in damselflies (Fincke 1988) and dragonflies (McVey 1988), have shown that lifespan was a good predictor of reproductive success. However, studies in

red deer have shown that although reproductive life span was highly correlated with summer calf survival, it was not correlated with other measures of fitness such as fecundity, yearling survival or winter calf survival rates (Cluttonbrock 1988). These null findings should not be surprising considering the inevitable influence of stochastic environmental factors on longevity.

The hypotheses tested in this study were:

1. Levels of FA within individuals would be negatively correlated with levels of heterozygosity and d^2 measures and positively correlated with levels of internal relatedness.
2. Body size in mature individuals would be negatively correlated with levels of FA.
3. Body size of individuals of mature age would be positively correlated with levels of heterozygosity and d^2 measures and negatively correlated with levels of internal relatedness.
4. Out of the Norwegian male reindeer culled at 20 months old, the smaller ones (of lower growth rate) would show lower levels of heterozygosity, lower mean d^2 and higher levels of internal relatedness than those that had grown faster.
5. Individuals that died of natural causes at a young age should show lower levels of heterozygosity, lower mean d^2 and higher levels of internal relatedness than those that survived to old age.

4.2 Materials and Methods

4.2.1 Sample collection

Tissue and skull samples were collected from the pre-bottleneck herd in Norway at the time the herd was gathered for culling and slaughter. Forty six Norwegian skulls were measured.

South Georgia samples were collected from carcasses of reindeer that had died from natural causes. Forty one Barff skulls and 23 Husvik skulls were measured. Further details can be found in sections 2.2.1 and 3.2.1.

4.2.2 DNA extraction and isolation from skin samples and teeth

Details of the methods used can be found in sections 2.2.2 and 2.2.3.

4.2.3 Polymerase Chain Reaction

Details of the methods used can be found in section 2.2.4

4.2.4 Skull preparation and measurements

Details of the methods used can be found in section 3.2.2 and 3.2.3.

4.2.5 Indices used to measure genomic diversity

In this study the following measures of genetic diversity were used: multilocus individual heterozygosity, standardised heterozygosity, mean d^2, scaled mean d^2, standardised mean d^2 and internal relatedness.

Due to the fact that not all individuals were typed for all the marker loci and to ensure that all individuals were measured on an identical scale, standardized heterozygosity (SH) was used (Coltman *et al.* 1999):

SH = proportion of heterozygous typed loci/mean heterozygosity of typed loci

Mean $d^2 = {}^1/_n\Sigma (i_a - i_b)^2$ where i_a and i_b refer to the lengths of each allele at locus i, averaged over n typed loci. Because d^2 values vary greatly between loci, a simple arithmetric mean gives undue weighting to loci that have a large range of allele size. 'Standardised mean d^2' is calculated by dividing the d^2 values by the maximum observed value at that locus and averaging the results across loci. The resulting values are all less than one (Amos *et al.* 2001).

'Scaled mean d^2' is calculated as the average of the d^2 scores once they have been scaled by the variance at that locus. This controls for an effect in which highly polymorphic loci would contribute more to the overall score of mean d^2 so it allows all loci to contribute equally (Coulson *et al.* 1999). Scaled mean $d^2 = {}^1/_n\Sigma ((i_a - i_b)^2 / \sigma^2_i)$

Internal relatedness is calculated as $(2H-\Sigma f_i)/(2N-\Sigma f_i)$ where H is the number of loci that are homozygous, N is the number of loci and f_i is the frequency of the ith allele contained in the genotype. When calculated over several loci, the value is approximately normally distributed and more or less centred on zero for individuals born to 'unrelated' parents. Negative values suggest relatively 'outbred' individuals and high positive values suggest inbreeding (Amos *et al.* 2001).

All of the measures of genomic indices were normally distributed except for scaled mean d^2 which was both leptokurtic and positively skewed. Correlation across indices was tested using the Pearson coefficient to

determine independence among each other. Due to the fact that 'scaled mean d^{2}' was not normally distributed, Spearman's Rho was used to test how this parameter correlated with the other indices.

Differences among the three populations were tested for by an ANOVA for all the indices of genomic diversity except for scaled mean d^2 which was compared with a Kruskall Wallis test.

A greater number of samples were measured for genetic diversity than were measured for morphometric differences as there was not an intact skull for every genetic sample. For this reason the ANOVA and tests for correlation were done twice: once for all samples and once for the samples that had corresponding morphometric measurements.

4.2.6 Indirect measures of fitness

4.2.6.1 Determination of fluctuating asymmetry

Details of the methods used can be found in section 3.2.4. The measures of FA that were used for this study were those that were determined in chapter III to be true FA, showing no evidence of either DA or antisymmetry.

4.2.6.2 Body Size

The measure of body size that was used was A-Rh, a measure of skull length which was highly correlated with total skull length, A-P so that R=0.946 for 85 skulls. Within each population, the values for A-Rh were normally distributed.

Each of the traits that showed significantly more FA in one of the South Georgia populations was regressed against A-Rh to test for a relationship between body size and FA.

The relationship between mature body size and the measures of genomic diversity was tested using linear correlations; the Pearson coefficient was used for all measures except for scaled mean d^2 where Spearman's Rho was used due the fact that scaled mean d^2 was not distributed normally. This analysis was performed separately on each sex of aged reindeer from each of the populations.

4.2.6.3 Growth Rate

This investigation used the skulls of the male Norwegian reindeer that were known to be killed at the same age (20 months old). Seven measurements of skull traits that were known to be correlated with skull length were used as a measure of growth rate up to that age (see table 4.1).

Table 4.1 Correlation of measures of skull size with total skull length within the Norwegian population (n=40)

Correlation with total length A-P	A-Rh (length)	A-N (length)	B-Po (length)	B-P (length)	B-Nuch (height)	Ot-Ot (width)
Pearson coefficient	0.80	0.48	0.68	0.90	0.42	0.53
p-value	<0.0001	0.0008	<0.0001	<0.0001	0.003	0.0002

The relationship between skull size and the measures of genomic diversity was tested using linear correlations; the Pearson coefficient was

used for all measures except for scaled mean d^2 where Spearman's Rho was used.

4.2.6.4 Longevity

The South Georgia skulls (which were all from reindeer that had died from natural causes) were placed into age categories (see section 3.2.3.2). To further categorise the skulls in which all the molars had erupted , the teeth were examined for evidence of wear. This allowed the 'aged' skulls (older than 3 years) to be divided into 'young aged' in which molars were all erupted but there was minimal evidence of wear and 'old aged' in which the molars showed considerable evidence of wear. See table 4.1.

Table 4.1 Age and sex profile of skulls measured for this study.

Age category	Norway			Barff			Husvik		
	Male	Female	Sex not known	Male	Female	Sex not known	Male	Female	Sex not known
0-1 years			6	2	1	3			
1-2 years	25			2			1	1	1
2-3 years				1					
Young aged				5		3	5	1	
Old aged	5	10	1	6	12	6	10	3	1

A comparison of the measures of genomic diversity was made between reindeer that died young (either some molars not yet erupted or molar teeth erupted but showing minimal sign of wear) and those that survived to an old age (molar teeth showing considerable wear).

For this comparison, no distinction was made between reindeer from Barff or Husvik or between different sexes. A simple 2x2 contingency test

was used to compare reindeer that died young (n=26) with those that survived to old age (n=38).

4.2.7 Table-wide probability of type I error

As six different genetic variables were used, the sequential Bonferroni correction (Rice 1989) was performed to reduce the risk of Type I error caused by multiple comparisons. This correction is likely to be conservative as the genetic variables are nonindependent (Slate and Pemberton 2002).

4.3 Results

4.3.1 Measures of genomic diversity

The mean results of the different indices of genomic diversity are shown in table 4.2 for all the genetic samples (a) and for the samples that correspond to a skull (b). The three populations are compared in table 4.3.

Table 4.2a Summary of the results of the different indices of genomic diversity for all genetic samples

All samples	Norway		Barff		Husvik	
	n	mean	n	mean	n	mean
Heterozygosity	99	0.753 ± 0.12	56	0.680 ± 0.14	57	0.547 ± 0.14
Standardised heterozygosity	99	1.059 ± 0.25	56	1.068 ± 0.21	57	1.105 ± 0.16
Mean d^2	99	58.50 ± 27.29	56	45.41 ± 23.71	57	47.81 ± 25.38
Scaled mean d^2	99	0.017 ± 0.01	56	0.021 ± 0.02	57	0.044 ± 0.13
Standardised mean d^2	99	0.182 ± 0.07	56	0.172 ± 0.08	57	0.178 ± 0.08
Internal relatedness	99	-0.011 ± 0.15	56	0.042 ± 0.18	57	0.190 ± 0.37

Table 4.2b Summary of the results of the different indices of genomic diversity for all samples that correspond to a skull.

Skull samples	Norway		Barff		Husvik	
	n	mean	n	mean	n	mean
Heterozygosity	41	0.746 ± 0.13	35	0.682 ± 0.15	23	0.581 ± 0.14
Standardised heterozygosity	41	1.085 ± 0.19	35	1.119 ± 0.25	23	1.102 ± 0.26
Mean d^2	41	58.93 ± 23.93	35	44.25 ± 24.06	23	39.06 ± 21.52
Scaled mean d^2	41	0.023 ± 0.01	35	0.026 ± 0.02	23	0.101 ± 0.10
Standardised mean d^2	41	0.214 ± 0.08	35	0.194 ± 0.10	23	0.233 ± 0.10
Internal relatedness	41	0.013 ± 0.17	35	0.045 ± 0.20	23	0.070 ± 0.23

Table 4.3 Comparison of the measures of genomic diversity among the three populations

	All samples		Skull samples only	
	$F_{(2.209)}$	Significance	$F_{(2.98)}$	Significance
Heterozygosity	46.17	***	10.23	***
Standardised heterozygosity	0.82	ns	0.81	ns
Mean d^2	5.68	**	6.46	**
Standardised mean d^2	0.31	ns	0.21	ns
Internal relatedness	13.14	***	0.50	ns
Scaled mean d^2	X^2=1.20	ns	X^2=11.17	p=0.004*

The Norwegian population showed the highest level of heterozygosity and the Husvik population showed the lowest level. There was a significant difference in the heterozygosity (at the level $p<0.001$) but no such difference

in standardised heterozygosity which may suggest that the fact that not all individuals were typed for all loci may have affected the results.

The Norwegian population showed the highest level of mean d^2 and the Husvik population showed the lowest level. There was a significant difference in the mean d^2 (at the level $p<0.01$) between populations but not in standardised mean d^2 which may indicate that loci with large ranges of allele size had a effect on the mean d^2 result. The effect of highly polymorphic loci was taken into consideration with scaled mean d^2 and when just the samples with corresponding skulls were considered, there was a significant difference in scaled mean d^2 between populations. This was not apparent when all the samples were considered which suggests that there may be an effect of sampling stochasticity.

The importance of each single locus to the overall mean d^2 effect was tested by dropping each locus from the mean d^2 calculation in turn. The comparison of mean d^2 between populations remained highly significant ($p<0.01$) after removal of all loci except RT9. When RT9 was removed there was no longer a significant difference between the populations ($p=0.07$) which suggests that the fact that the Husvik population was virtually monomorphic for this locus had a large influence on the comparison of mean d^2.

The Norwegian population showed the lowest level of internal relatedness and the Husvik population showed the highest level. There was a significant difference ($p<0.001$) in internal relatedness when all the samples

were considered though this difference was not significant when just the samples with corresponding skulls were considered. This was probably due to lower statistical power with the smaller sample size. The negative value of internal relatedness for the Norwegian population indicated that these individuals were relatively outbred.

An ANOVA was also used to compare the genetic samples which corresponded to a skull with the samples that had no corresponding skull within each population. There was no significant difference between the genomic indices for the different groups of samples except for standardised heterozygosity where there was a significant difference between the skull samples and the non-skull samples within the Norway population ($F_{(1,97)}$= 7.04). This was no longer significant after the application of the Bonferroni correction.

The high correlations between some indices, especially heterozygosity and internal relatedness, indicate that they are not independent from each other. (See table 4.4.)

Table 4.4 This table shows linear correlations of the six indices of genomic diversity. The Pearson coefficient was used for all the comparisons except those involving scaled mean d^2 where the coefficient used was Spearman's rho. Significance: ns - not significant, ^significant at the level $p<0.05$ before application of Bonferroni. *$p<0.05$, **$p<0.01$, ***$p<0.001$, Bonferroni correction applied. n=99

	Heterozygosity	Standardised heterozygosity	Internal relatedness	Mean d^2	Standardise d mean d^2
Standardised heterozygosity	0.87 ***				
Internal relatedness	-0.92 ***	-0.94 ***			
Mean d^2	0.32 **	0.20 ^	-0.17 ns		
Standardised mean d^2	0.35 **	0.41 ***	-0.34 ***	0.72 ***	
Scaled mean d^2	0.15 ns	0.32 **	-0.25 ^	0.21 ^	0.66 ***

4.3.2 Indirect measures of fitness

4.3.2.1 Fluctuating asymmetry

If inbreeding had caused an individual to show higher levels of FA then the relationship between FA and measures of genomic diversity would be negative for all the measures except internal relatedness where the relationship would be positive. Each of the correlations present (see table 4.5) are in the direction that would be expected although not one of these is still significant after the application of the Bonferroni correction.

Table 4.5 Linear correlation of fluctuating asymmetry with the indices of genomic diversity using Spearman's rho. Significance: ns - not significant, ^ significant at the level $p<0.05$ before the application of the Bonferroni correction.

	Heterozygosity	Standardised heterozygosity	Mean d^2	Scaled mean d^2	Standard mean d^2	Internal relatedness
Norway (n=41)						
Ni-P	ns	ns	-0.33^	-0.35^	-0.40^	ns
CFA & individual FA of all other traits	ns	ns	ns	ns	ns	ns
Barff (n=35)						
B-Ent	-0.41^	-0.41^	ns	ns	ns	0.39^
St-Zyg	ns	ns	ns	-0.38^	-0.41^	ns
CFA 3	ns	-0.34^	ns	-0.35^	ns	0.40^
CFA1 & individual FA of all other traits	ns	ns	ns	ns	ns	ns
Husvik (n=23)						
po-Zyg	-0.52^	-0.54^	ns	ns	ns	ns
CFA & individual FA of all other traits	ns	ns	ns	ns	ns	ns

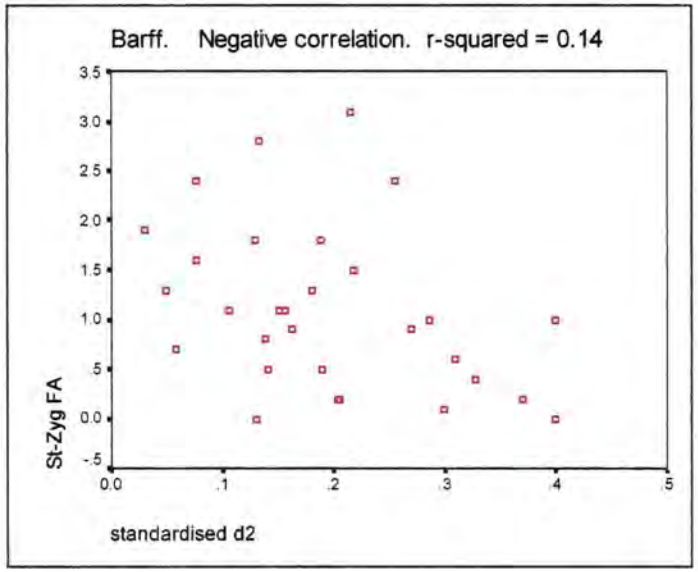
In the Norwegian population there was an association between the three measures of mean d^2 and the asymmetry of the trait Ni-P. However, this cannot be used as a measure of FA as this trait showed significant DA in this population. There were no associations between composite FA or FA of

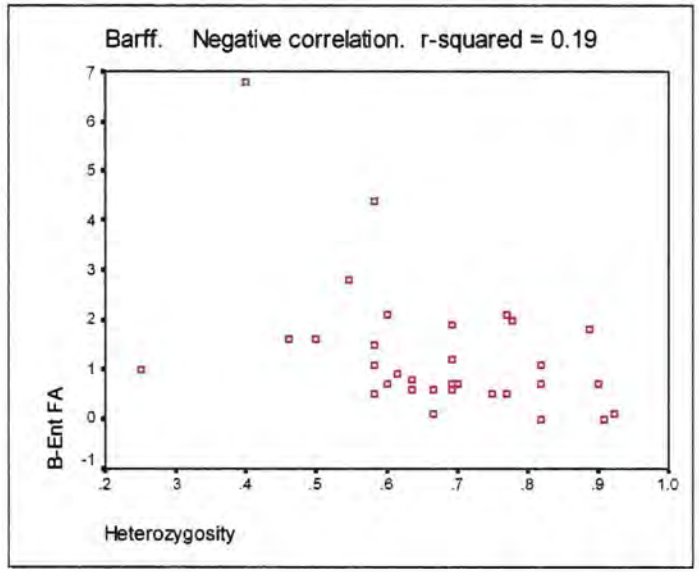
any of the individual traits with any of the measures of genomic diversity in the Norwegian population.

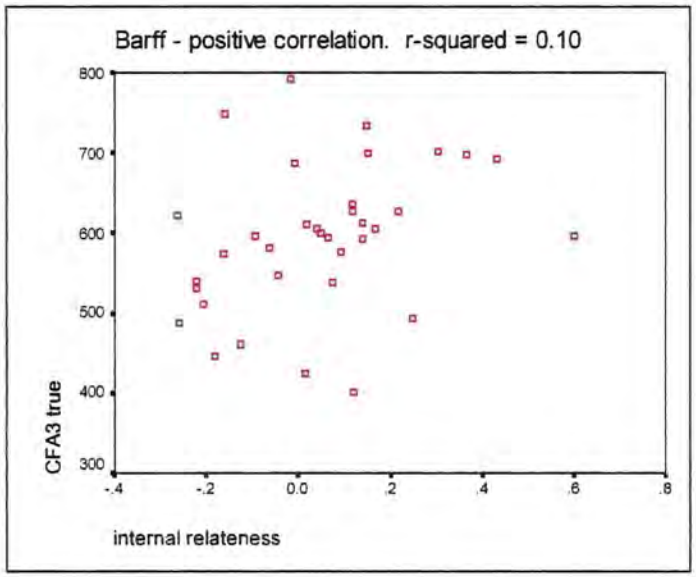
Within the Barff population there was a correlation between the FA of B-Ent with heterozygosity, standardised heterozygosity and internal relatedness. There was also correlation between the FA of St-Zyg with scaled mean d^2 and standardised mean d^2. The composite value of FA based on rank values, CFA3 was correlated with standardised heterozygosity, scaled mean d^2 and internal relatedness in the Barff population. After application of the Bonferroni correction, not one of these correlations were still statistically significant, but this may have been as a result of low statistical power due to the small sample size (35 individuals).

The sample size of the Husvik population was even smaller (23 individuals) and although there appeared to be a strong correlation between FA of trait Po-Zyg and both heterozygosity and standardised heterozygosity, this was not significant after application of the Bonferroni correction.

The scatter plots of some of these relationships are shown in figure 4.1.







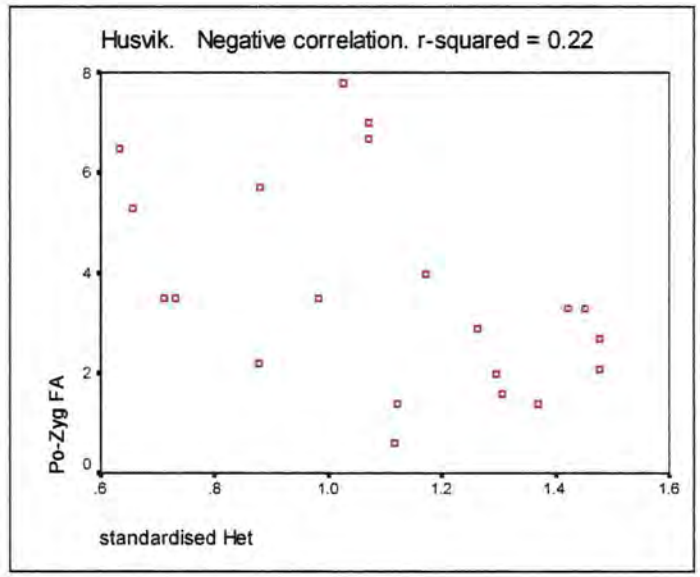
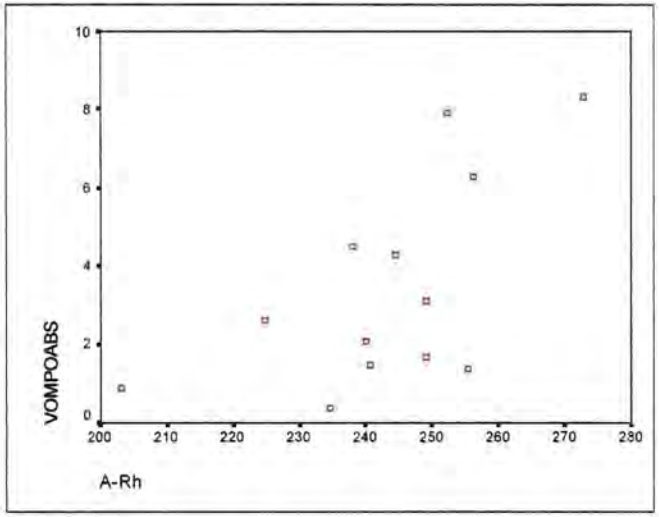


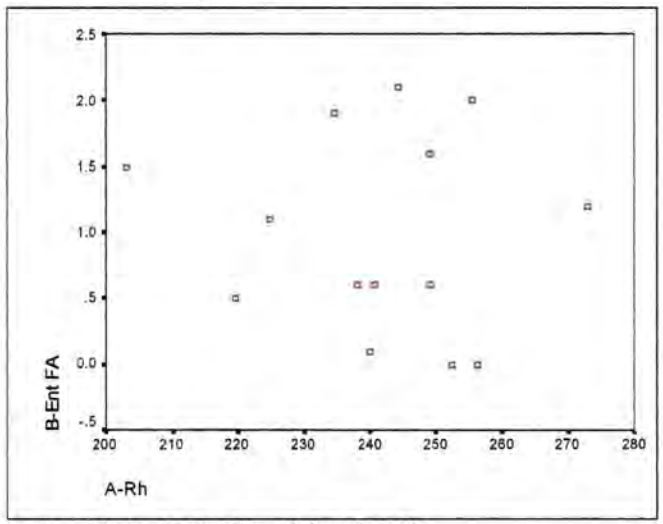
Figure 4.1
Graphical representation of correlations found between FA and indices of genomic diversity

Each of illustrated correlations were significant at the level $p<0.05$ before Bonferroni.

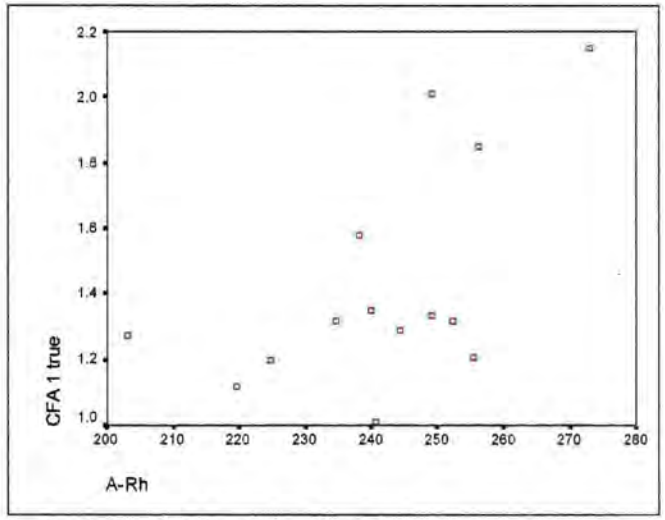
Figure 4.2 BARFF MALE SKULLS



a. Correlation of Vom-Po FA with length A-Rh
r-squared = 0.39 Spearman coefficient = 0.56^



c. Correlation of B-Ent FA with length A-Rh
r-squared = 0.01. Spearman coefficient = -0.07 ns



b. Correlation of CFA1 with length A-Rh
r-squared = 0.34. Spearman coefficient = 0.54^

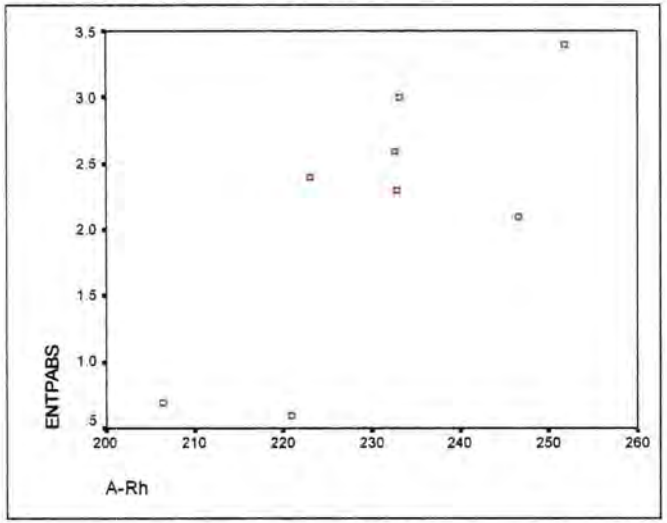
Figure 4.2

4.2.a, b & c Barff male skulls
CFA1, Vom-Po & B-Ent showed significantly more FA in Barff than Norway
d, e, f & g Husvik male skulls
CFA1, Ot-N, Ent-P & Po-Zyg showed sig more FA in Husvik than Norway

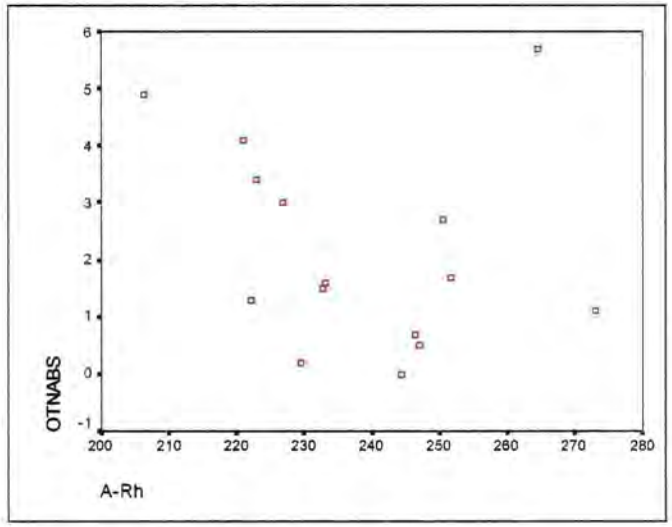
A-Rh is a measure of length. It is highly correlated to total length, A-P (R= 0.946 for 85 skulls)

Significance: ^ is significant at the level $p<0.05$ before Bonferroni, ns is not significant

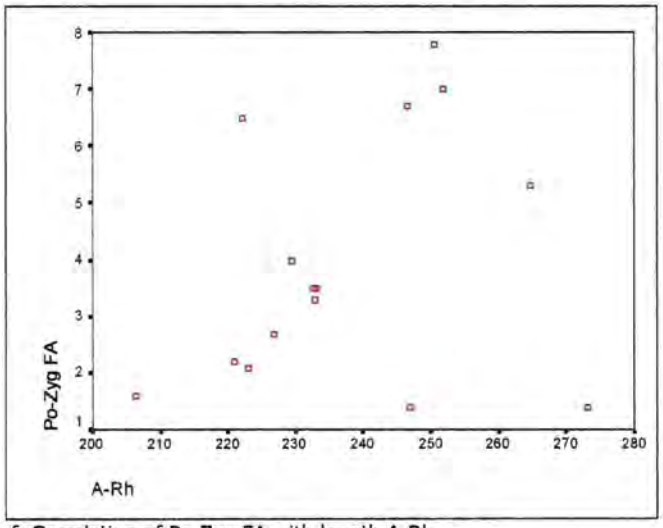
Figure 4.2 (cont) HUSVIK MALE SKULLS



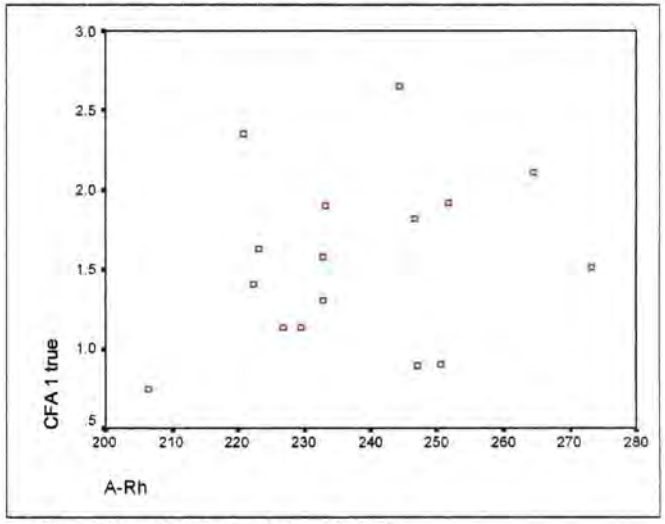
d. Correlation of Ent-P FA with length A-Rh
r-squared = 0.58. Spearman coefficient = 0.66 ns



e. Correlation of Ot-N FA with length A-Rh
r-squared = 0.05. Spearman coefficient = -0.25 ns

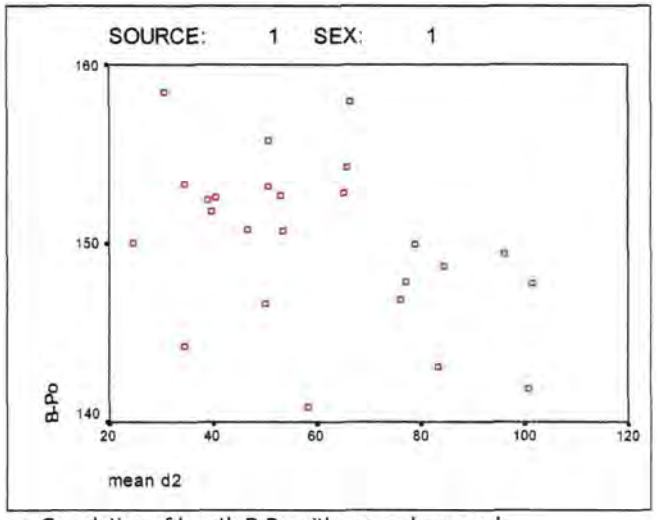


f. Correlation of Po-Zyg FA with length A-Rh
r-squared = 0.07. Spearman coefficient = 0.25 ns

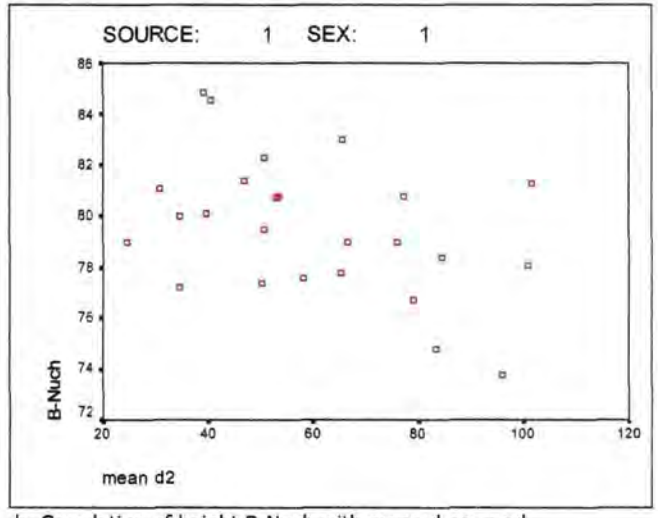


g. Correlation of CFA 1 with length A-Rh
r-squared = 0.06. Spearman coefficient = 0.21 ns

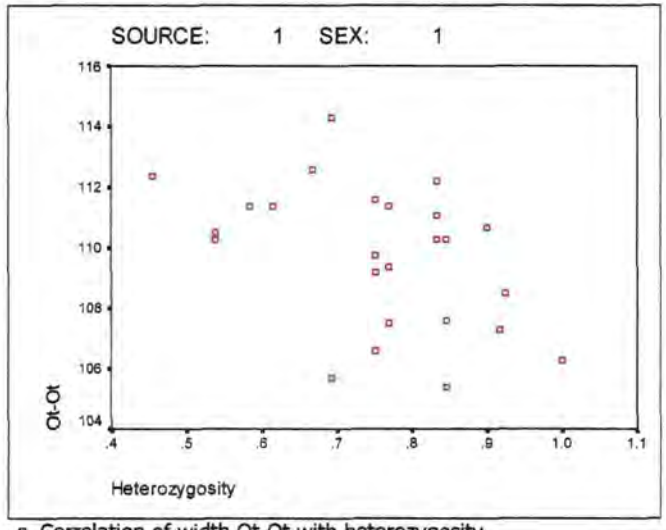
Figure 4.3 Scatterplots of skull size against measures of genomic diversity in male Norwegian reindeer shot at 20 months old



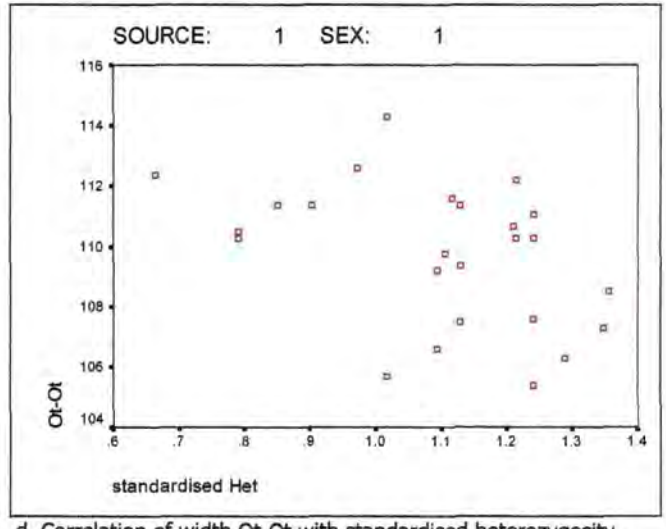
a. Correlation of length B-Po with mean d-squared
r-squared = 0.17 Spearman coefficient = -0.42^



b. Correlation of height B-Nuch with mean d-squared
r-squared = 0.18 Spearman coefficient = -0.42^



c. Correlation of width Ot-Ot with heterozygosity
r-squared = 0.20 Spearman coefficient = -0.45^



d. Correlation of width Ot-Ot with standardised heterozygosity
r-squared = 0.19 Spearman coefficient = -0.43^

4.3.2.2 Body size

There was no significant negative relationship between FA and body size within any of the populations. Indeed there was some indication that FA increased within increasing body size with positive trends when FA traits were regressed against A-Rh which was the measure of skull size used. Positive trends were seen for trait Ent-P in the Husvik population though these were not significant at the level $p<0.05$ and for Vom-Po FA and CFA1 in the Barff population (significant at the level $p<0.05$ before application of Bonferroni). See figure 4.2. To test how much these trends were affected by continued growth of the individuals, a subsample of 'old aged skulls' (ie those with very worn teeth) was tested and the positive trend was still apparent (though not significant due to reduced sample size).

There was no significant relationship between the trait A-Rh (used as a measure of overall skull size) and any of the measures of genomic diversity for aged skulls of either sex in any of the populations. There were no trends apparent to suggest that the larger skulls were the ones that were more heterozygous, had greater mean allelic distances or showed lower levels of internal relatedness.

4.3.2.3 Growth Rate

Table 4.6 Linear correlations of skull size of 25 male Norwegian reindeer (killed at 20 months old) with measures of genomic diversity. The Pearson coefficient was used for all the comparisons except those involving scaled mean d^2 where the coefficient used was Spearman's rho. Even when not significant, the value of the coefficient has been shown where it exceeds |0.1| to show the direction of the relationship. Significance: ^significant at $p<0.05$ before Bonferroni, ns not significant.

	Heterozygosity	Standardised heterozygosity	Mean d^2	Scaled mean d^2	Standard mean d^2	Internal relatedness
A-P (length)	-0.17 ns	-0.22 ns	-0.32 ns	ns	-0.22 ns	0.18 ns
A-Rh (length)	ns	ns	-0.27 ns	-0.13 ns	-0.18 ns	ns
A-N (length)	ns	ns	ns	ns	-0.19 ns	ns
B-P (length)	-0.16 ns	-0.15 ns	-0.32 ns	ns	-0.23 ns	ns
B-Po (length)	-0.16 ns	-0.21 ns	-0.42^	-0.15 ns	-0.23 ns	ns
B-Nuch (height)	-0.29 ns	-0.32 ns	-0.42^	-0.23 ns	-0.41^	0.26 ns
Ot-Ot (width)	-0.45^	-0.43^	-0.10 ns	-0.29 ns	-0.31 ns	0.37 ns

The group of 25 male Norwegian reindeer that were shot at 20 months old was tested for correlations between skull size and each of the measures of genomic diversity. Unexpectantly, there was a distinct pattern of trends so that the larger skulls showed lower levels of heterozygosity, lower measures of mean d^2 and higher levels of internal relatedness. As can be seen in table

4.6, although some of these correlations were weak and not significant, they were all in same direction. At the level $p<0.05$, there was a significant negative relationship between the width Ot-Ot with heterozygosity and standardised heterozygosity, between the length B-Po with mean d^2 and between the height B-Nuch with mean d^2 and standardised mean d^2. These correlations were no longer significant after application of the Bonferroni correction. Scatter plots are shown in figure 4.3.

4.3.2.4 Longevity

Table 4.7 Comparison of the mean of the measures of genomic diversity for the categories of reindeer that died young compared to the ones that survived to an old age. Significance: ^ means significant at the level $p<0.05$ before application of the Bonferroni correction

Measure of genomic diversity	Age at death	n	Mean	Significance
Heterozygosity	died young	26	0.62 ± 0.16	ns
	died old	38	0.65 ± 0.14	
Standardised Heterozygosity	died young	26	1.06 ± 0.25	ns
	died old	38	1.13 ± 0.24	
Mean d^2	died young	26	42.50 ± 22.33	ns
	died old	38	40.54 ± 22.80	
Scaled mean d^2	died young	26	0.03 ± 0.04	p=0.04^
	died old	38	0.06 ± 0.09	
Standardised mean d^2	died young	26	0.18 ± 0.09	ns
	died old	38	0.21 ± 0.10	
Internal Relatedness	died young	26	0.10 ± 0.20	ns
	died old	38	0.04 ± 0.20	

The South Georgia skulls (which were all from natural deaths) were divided into those that died young and those that survived to old age

The reindeer that survived to an old age showed higher levels of heterozygosity, standardised heterozygosity, scaled mean d^2 and standardised mean d^2 and lower levels of internal relatedness than those that died at a young age. Although these differences were largely not significant, there was significance at the level $p<0.05$ for scaled mean d^2 though this was not still significant after application of the Bonferroni correction. Mean d^2 was the only measure of genomic diversity that did not follow the expected trend. The results for scaled mean d^2 and standardised mean d^2 were in the expected direction which may indicate that there may have been distortion for mean d^2 due to the undue weighting of highly polymorphic loci or of loci with a large range of allele size.

4.4 Discussion

Although the correlations between FA and the measures of genomic diversity were not significant, there were distinct associations in the directions expected. These results support the weakly significant negative associations found between heterozygosity and FA in a meta-analysis undertaken by Britten (1996).

The sample sizes in this study were low compared to other studies that have shown a significant relationships between measures of fitness and measures of genomic diversity. Studies that have shown significant relationships include 650 red deer calves, 275 harbour seal pups (Pemberton 1999) and 356 harbour porpoises (de-Luna-Lopez 2005). In the latter study,

the author suggested that it was due to the statistical power provided by the sample size of 356 which allowed a significant relationship to be apparent in this population but not in other populations in the same study in which the sample sizes were ten-fold smaller and more comparable to the sample sizes in this study.

When 12 different traits are to be tested for correlations, the sample size should be at least 83 to give an 80% chance of detecting a rho-squared value of 0.2 at a significance of $p<0.05$ (Lenth 2006).

It was not possible to say from this study which of the genomic diversity indices were more useful in predicting high FA though mean d^2 performed the least reliably with not one of the measures of FA in any of the populations showing a relationship with it. There were however associations between measures of FA and both standardised and scaled versions of mean d^2 in the Barff population.

The post-bottleneck populations of reindeer were shown to have significantly smaller skulls than the pre-bottleneck population in Chapter Three. Investigation of the relationship between body size and measures of genomic diversity at the level of the individual within each population removes the confounding environmental factors.

Body size has been shown to be positively correlated with other measures of fitness in reindeer (Roed 1987) and particularly in male red deer (Cluttonbrock 1988). However, investigation at the level of the individual in this study failed to show that those with smaller skull size were also those

with high levels of FA. This study also failed to show that the more inbred individuals had lower growth rates or smaller mature skull size.

In a meta-analysis of studies of this sort there was found to be a weak positive association between heterozygosity and growth rate (Britten 1996). The results from this study of the reindeer on South Georgia has not supported this finding. There were no obvious trends or associations between mature skull size and levels of inbreeding and there were even indications that in the Norwegian herd the more inbred individuals grew more quickly than those that were outbred. Anecdotal evidence (Asgrim Opdal - personal communication) suggests that predators are not a significant risk to the Filefjeld herd. Perhaps, without the threat of predators, the less fit individuals keep themselves outside the main social structure of the herd and thus have access to better food resources. Meanwhile the fitter individuals face greater competition within the herd structure and thus have less opportunity for grazing. In a managed herd, however extensively, there will not be a direct relationship between survival and fitness due to the influence of human selection in the culling of reindeer. This may have confounded these results.

Due to the many environmental and stochastic factors which would affect longevity (Kruuk *et al.* 2000) it is expected that any genetic influence on longevity would be difficult to detect due to the noise of these other influences. For this reason the results in this study have been reported despite the fact that the differences seen between reindeer that died young

and those that survived to old age were weak and not significant. It would be interesting to pursue this investigation with a larger sample size.

The order of magnitude of the difference in the means between the group of reindeer that died young and those that lived to an old age in this study was used to calculate required sample size for a future study. 200 samples in each group would give more than 80% chance of detecting an effect of the same order of magnitude as that found in this study for standardised heterozygosity, scaled mean d^2, standardised mean d^2 and internal relatedness at a significance of $p<0.05$ (Lenth 2006).

This study of the correlation between indirect measures of fitness and measures of genomic diversity at the level of the individual has added more detail to the general patterns seen at population level (reported in Chapters Two and Three). Although only tentative conclusions can be drawn due to the limits of the small sample sizes, it was particularly interesting to see distinct associations in some traits of increasing FA with decreasing genomic diversity.

Chapter Five

5.1 Synthesis of results and conclusions

In this study a direct comparison of two post-bottleneck populations of reindeer and their common source population was made using both genetic and morphological data. Although reindeer are not endangered, the populations investigated in this study can be considered as models for other wild populations that may be endangered. The inaccessibility and geography of the island have ensured that the South Georgia reindeer herds have been isolated from immigrants and separated from each other. These factors, coupled with good historical records, have allowed this study to clearly investigate the genetic and morphometric consequences of two populations that experienced extreme bottlenecks in parallel situations. Each of the post-bottleneck populations showed decreased genetic diversity and increased evidence of developmental instability compared to the pre-bottleneck population.

Low genetic diversity has been shown in numerous other studies of wild populations that experienced genetic bottlenecks, such as island moose in Canada (Broders *et al.* 1999) or black robins in New Zealand (Ardern and Lambert 1997). Considerable morphometric data (measurement of FA in 27 non-metric and six metric traits) was collected in a study of brown hares (Hartl 1995; Suchentrunk 1998) though there was no significant relationship found between FA and allozyme heterozygosity at the level of population or individual.

However there have been a number of previous studies of wild populations that have shown an association between low levels of genetic variation and increased fluctuating asymmetry such as those of gazelles in Spain (Alados *et al.* 1995), three different taxa of tamarins (Hutchison and Cheverud 1995), the Northern Elephant Seal (Hoelzel *et al.* 1993; Hoelzel 1999; Weber *et al.* 2000; Hoelzel 2002) and an island population of black-footed rock wallabies (Eldridge *et al.* 1999).

As demonstrated by the conflicting results from studies on wild populations of cheetah (Wayne 1986; Modi *et al.* 1987; Kieser and Groeneveld 1991; Merola 1994), it is of great importance for studies of this kind to be undertaken with very careful attention to possible measurement error or the use of asymmetries other than fluctuating asymmetry, which may produce spurious results. The rigorous protocols used in this study in the handling of the data addressed the risk of such inadvertent errors. Another strength of this study was the direct access to samples from the two separate post-bottleneck populations and the source population. There are not many previous studies that have had direct access to the pre-bottleneck population and consequently they have compared post-bottleneck populations with closely related populations that in some cases have even been of different species (eg. Wayne 1986; Kieser and Groeneveld 1991; Ardern and Lambert 1997).

The Barff herd was known to have been founded by seven females and three males from Filefjell Reinlag, Norway in 1911. There was no further immigration into this herd which grew rapidly to a peak of around 3000 deer

in the late 1950s. This population showed significant differentiation and reduced genetic diversity compared to the pre-bottleneck population.

The Barff herd also showed significantly greater levels of fluctuating asymmetry (FA) than the Norwegian herd in two out of thirteen individual traits, significantly more composite FA (based on twelve traits) and greater phenotypic variation. The length of the Barff skulls was significantly smaller than the Norwegian skulls though they were significantly wider at the cheekbones.

The Husvik population was thought to have been founded by four females and three males from Filefjell Reinlag, Norway in 1925. It seems likely that there was a further introduction of up to four female reindeer in 1928, but it is not known whether or not these animals survived. This herd grew relatively slowly and currently numbers approximately 1000.

Compared to the pre-bottleneck herd, the Husvik herd showed significant genetic differentiation and reduced genetic diversity. There were also significantly greater levels of FA in four out of fourteen individual traits, significantly more composite FA (based on twelve traits) and significantly more phenotypic variation in the post-bottleneck population. The overall size of all traits measured on the Husvik skulls were significantly less than the Norwegian skulls.

Comparisons at the level of the population are confounded by a number of factors, not least the different environmental effects on each of the populations. The investigation of the relationship between indirect measures

of fitness and measures of genomic diversity at the level of the individual within each population was designed to partition out the different environmental influences on each population.

Within the post-bottleneck populations there were associations between FA at some traits with heterozygosity, scaled mean d^2, standardised mean d^2 and internal relatedness. Although these relationships were weak, probably as a result of low sample sizes, they were interesting as they indicate the direct influence of genetics on the levels of FA, independent of the different environmental influences on each population. Empirically these results have demonstrated the non-additive genetic basis to fluctuating asymmetry.

There were no relationships between body size and measures of genomic diversity at the level of the individual. This suggested that the highly significant differences in body size between the different populations were more the result of a plastic response to environmental pressures than due to genetic differences between the populations.

There were seven females and three males that founded the Barff herd and between four and eight females and three males that founded the Husvik herd. However as demonstrated by the range of results predicted by the simulation models, the impact of the bottleneck has been shown to depend on more than simply the numbers in the founder group. The simulation program models demographic stochasticity, which in turn is affected by survival and reproductive success of individuals. As demonstrated by the

model, the variance among outcomes is very high. The severity of the bottleneck is also affected by polygony (due to reduction in the effective number of males) and life history characteristics which have a direct effect on population growth rate and the period of time over which the population remains small. Further factors which complicate investigations of the level of impact of a bottleneck include iteroparity and overlapping generations (Halley and Hoelzel 1996). In addition to the high degree of variance associated with these factors, environmental stochasticity contributes further variance to the genetic effects of the bottleneck.

An example of the potentially devastating consequences of environmental stochasticity was demonstrated by the circumstances surrounding the second introduction of reindeer on to South Georgia in 1912 (see Chapter Two for details). This herd at Leith Harbour increased from five to 17 individuals by 1917 but the whole herd was then killed in one snow slide (Olstad 1930). This demonstrates how a population is particularly vulnerable to environmental stochasticity in the early years following a bottleneck when the population size is still very low.

A further consequence of stochasticity results from the chance disruptions of relevant gene interactions and the impact that this has on developmental stability. Three genetic mechanisms have been proposed as explanations for associations between heterozygosity and fitness (for reviews see Mitton and Grant 1984; Britten 1996; David 1998; Slate and Pemberton 2002). The first hypothesis, true overdominance, assumes that the marker loci are themselves functional so this cannot generally be applied to

microsatellite loci due to the fact they are predominantly present in non-coding genetic regions (unless closely linked to a functional gene). Thus we must assume distinction between the marker (microsatellite) loci and the agent loci, which directly contribute to the observed phenotypic variation (David 1998). One possible mechanism in this case is that of a local effect, described as 'associative overdominance' where there is a genetic correlation between marker loci and agent loci, either through linkage disequilibrium (non-random association of alleles in gametes) or through identity disequilibrium (non-random association of genotypes resulting from inbreeding in the population) (Weir and Cockerham 1973). The other possible mechanism is more general, assuming that heterozygosity at marker loci reflects genome-wide heterozygosity.

Although this study was limited by small sample sizes, the results may usefully help to distinguish which mechanism may be responsible for the association seen between FA and measures of genomic diversity. There were no associations found in the Norwegian population which is large and assumed to be randomly mating, though due to the lack of power larger sample numbers would be required to prove no association. Weak associations were found in both post-bottleneck populations which were also shown to have experienced significant levels of inbreeding. These findings, as well as the inconsistency between the populations in the traits that showed increased FA, provides some support for a genome-wide rather than local cause for the association. This is consistent with the conclusions of Slate and

Pemberton (2002) following a study of a large panel of microsatellite loci in a wild population of red deer.

It is interesting to note that the Husvik herd experienced a more severe genetic effect from the bottleneck than the Barff herd with greater reductions in all the measures of genetic variation. Morphometric indicators of reduced developmental stability were also greater in the Husvik herd than the Barff herd with greater within-individual variation (FA) and among-individual variation. The Husvik herd were also found to be smaller in size than the Barff herd. In the simulation model of herd demographics, the life history parameters needed to most closely reflect the growth of the Husvik herd were lower than those needed to closely reflect the growth of the Barff herd. The slower growth of the Husvik population may have been due to reduced fitness or increased environmental or genetic stress on this population. Whatever the cause, the slower demographic growth will have further contributed to the loss of genetic variation by increased inbreeding, due to a longer duration at low population numbers.

The inconsistency between the traits which showed increased FA between the two post-bottleneck populations in this study emphasized the random nature of FA and the fact that individual traits are generally poor predictors of organism-wide FA (Leung and Forbes 1997) due to low FA occurring either by chance or as a result of high quality. However the fact that there was some overlap between the two post-bottleneck populations in the high FA traits supports the theory that high values of FA are reliable indicators of low quality (Leung and Forbes 1997; Leung *et al.* 2000). The

significant differences between levels of composite FA (both at the level of the population and of individuals in the case of the Barff herd) emphasized the better reliability of using CFA instead of individual traits (as suggested by Leary and Allendorf 1989; Leung *et al.* 2000).

Data from this study has allowed investigation into the quantifiable distortion of allele frequencies (see Appendix 1) following the two bottlenecks. Although as little as 75 and 90 years respectively elapsed from the time of the Husvik and Barff bottlenecks to the time that samples were collected, the measures of genetic differentiation, F_{ST} and R_{ST}, indicated that there was significant differentiation between both post-bottleneck populations and the pre-bottleneck population. This result was consistent with the significant differentiation found between pre- and post-bottleneck elephant (Whitehouse and Harley 2001), moose (Broders *et al.* 1999) and koala (Houlden *et al.* 1996) populations following bottlenecks and subsequent isolation over a comparable timescale to this study.

In common with a number of recent studies, this study has demonstrated the limitations of various methods that are commonly used to detect bottleneck signatures. As also demonstrated by the study of Bennett's wallabies (Le Page *et al.* 2000), significant heterozygosity excess in the Barff population showed how the IAM and TPM models better represented the evolution of these microsatellite loci over the timescale following this bottleneck than the SMM model. However the negative results from the Husvik herd despite the same source population and broadly similar time-scale, conditions and environment emphasizes the unreliability of using this

bottleneck signature even with these mutation models. Results from both populations in this study as well as studies in elephants (Whitehouse and Harley 2001), rhinos (Harley *et al.* 2005) and wallabies (Le Page *et al.* 2000) suggest that detecting the presence of a mode-shift in allele frequency distributions appears to be notoriously unreliable in predicting a previous bottleneck.

In common with most of the aforementioned studies, there was a reduction in the Garza "M" ratio in both post-bottleneck populations in this study. Although it might be suggested that this is one of the more reliable detection methods, the absolute values should be interpreted with caution with the application of a higher threshold than that suggested by Garza and Williamson (2001).

The findings of this study have contributed further evidence of the need to use a number of different methods for detecting recent bottlenecks and to interpret negative results with extreme caution. The implications of these findings are relevant to management decisions and conservation strategies that may otherwise be based on inaccurate interpretation of genetic patterns when inferring population structure and history.

Previous studies of the South Georgia reindeer assumed different origins of the two herds due to very different behaviour (Bonner 1958; Leader-Williams 1978; Leader-Williams 1988). This study showed quantifiable differences in measures of body size, fluctuating asymmetry and morphometric variation as well as significant genetic differences between the

two herds. However the historical evidence has shown that these differences cannot be attributed to different origins. Instead these differences are a demonstration of the effects of stochasticity both at the founder events and subsequently during the establishment of the new populations.

5.2 Recommendations in the context of the Environmental Management Plan for South Georgia

Reindeer from Husvik (26 females and 31 males) were translocated to the Falkland Islands in 2001 (Bell 2001). The purpose of this translocation was two-fold: to assist in the diversification of the economy of the Falkland Islands and to conserve 'the genetics of the South Georgia population' in the face of eradication (McIntosh and Walton 2000; Bell 2001). As a long-term policy the Government of South Georgia seeks to remove non-indigenous flora and fauna, as far as is practicable from South Georgia (McIntosh and Walton 2000).

It should be noted from this study that both the herds are genetically unique and that they are significantly differentiated from each other as well as from their source population so it is not possible to talk about South Georgia reindeer as if it were made up of one homogenous population.

Although it is the stated intention of the Government of South Georgia to remove the herds of reindeer as a matter of priority (McIntosh and Walton 2000), the practicalities mean that this is likely to be undertaken in phases. If a decision had to be taken as to which herd of the two original herds to

remove, it should be noted from this study that the Barff population has been shown to be both more genetically diverse and based on measures of FA, suggested to be more developmentally stable than the Husvik population.

5.3 Recommendations for further work

The collection of further skulls would allow better statistical power in the comparison of indirect measures of fitness and measures of genomic diversity at an individual level. Eighty three samples from each of the post-bottleneck populations would give an 80% chance of detecting a rho-squared value of 0.2 at a significance of $p<0.05$ if twelve traits were measured. However, if only seven traits were measured then there would only have to be 68 samples from each population to give an 80% chance of detecting a rho-squared value of 0.2 at a significance of $p<0.05$ (Lenth 2006). The seven traits that would be most usefully measured in a future study would be Ent-P, If-Ent, Ot-Ect, Ot-N, B-Ent, Po-Zyg and Vom-Po as there were significant differences between the pre-bottleneck and post-bottleneck populations in each of these traits.

Although this study has emphasized the importance of the distortion of both allele frequencies and phenotypes as a result of bottleneck events, it was not possible to directly link these distortions to measures of fitness due to the lack of power resulting from small sample sizes. However, the findings and associations that are apparent from this study and the identification of specific traits that showed significant increases of FA in one or both of these

post-bottleneck populations can usefully inform future similar studies. Robust data that links increases in FA to measures of fitness would have very useful implications in the management and conservation of endangered species in the sense of using FA as an early warning system (as suggested by Lens *et al.* 2002).

If large numbers of the South Georgia reindeer were to be killed in the eradication program, there may be the opportunity to collect freshly killed samples which would yield DNA that should readily amplify. Examination of whole carcasses instead of just skulls would simplify the determination of sex and allow for further measurements to be taken. Incisors could be collected for accurate aging (Reimers and Nordby 1968). There would be the opportunity to undertake *post mortem* examinations to add information to the studies undertaken in the 1970s (Leader-Williams 1988) and to update life tables with current rates of reproductive success and survivorship.

It would be particularly interesting to investigate the importance of different life history characteristics and factors such as iteroparity, generation length and the extent of generations overlapping on the outcome following a bottleneck.

Appendix 1 – Frequency of alleles at each loci for each population

Loci	Alleles	Frequency (%) Norway	Barff	Husvik
CRH	227	3.4		
	229	5.7	2.2	
	231	33.3	14.1	8.8
	233	29.3	32.6	47.4
	235	26.4	39.1	38.6
	237	1.7	12.0	5.3
RT9	103	0.5		
	105	0.5		
	107	1.0	0.9	1.8
	109	48.0	55.2	98.2
	111	7.6	5.2	
	115	9.1	26.7	
	117	5.1	5.2	
	119	2.0		
	121	5.6	3.4	
	123	4.5		
	125	12.6	3.4	
	127	1.5		
	129	2.0		
BM848	354	1.5		2.8
	356	5.7	6.7	0.9
	358	1.0		
	360	1.0		
	362	1.5	4.4	1.9
	364	34.0	16.7	59.3
	366	5.2	20.0	
	368	0.5		2.8
	370	2.6	5.6	
	372	2.6	3.3	
	374	1.5	5.6	2.8
	376	17.5	28.9	5.6
	378	19.6	8.9	20.4
	380	1.0		
	382	0.5		
	386	3.1		3.7
	388	1.0		

Loci	Alleles	Frequency (%) Norway	Barff	Husvik
RT5	141	1.1		
	143	2.7		
	145	1.1		
	151	1.1	7.0	4.5
	153	0.5		
	155	4.3	2.6	1.8
	158	0.5	6.1	
	160	1.1	2.6	0.9
	162	11.3	11.4	4.5
	164	36.0	35.1	38.4
	166	11.8	6.1	34.8
	168	21.0	13.2	13.4
	170	7.0	15.8	1.8
	172	0.5		
RT1	220	10.8	18.2	35.3
	222	2.7	3.6	
	224	4.8		
	226	1.6	2.7	
	228	0.5		
	230	17.2	15.5	3.9
	232	4.3	1.8	2.0
	234	2.7	2.7	2.9
	236	30.6	26.4	41.2
	238	8.1	13.6	3.9
	240	14.5	4.5	10.8
	242	1.6	10.9	
	244	0.5		
CA13	199	18.6	3.9	9.1
	201	35.6	24.5	10.9
	203	27.3	47.1	70.0
	205	13.9	23.5	5.5
	207	4.1		4.5
	209	0.5	1.0	

Loci	Alleles	Frequency (%) Norway	Barff	Husvik
RT27	122	3.2		
	124	3.2		
	126	5.1		
	128	4.5	3.7	
	132	23.7	32.9	23.3
	134	2.6		
	136	33.3	43.9	67.4
	138	0.6		
	142	1.3		
	144	4.5	15.9	2.3
	148	3.8	3.7	2.3
	150	7.7		
	152	1.9		
	154	3.2		4.7
	156	1.3		
RT30	184	1.1	3.6	
	186	24.7	32.1	32.7
	188	45.7	27.4	27.9
	190	1.6		
	192	1.1		
	194	2.2		
	196	5.9	1.2	
	198	9.7	9.5	3.8
	206	1.6	17.9	6.7
	208	5.9	8.3	28.8
	218	0.5		
RT13	289	1.1		
	291	3.3	2.2	13.7
	293	9.9		
	295	28.0	3.3	25.5
	298	7.1	4.3	3.9
	300	13.2	14.1	16.7
	302	14.3	52.2	33.3
	304	5.5		
	306	13.7	23.9	6.9
	308	1.6		
	310	1.6		
	312	0.5		

Loci	Alleles	Frequency (%) Norway	Barff	Husvik
NVHRT22	140	0.5		
	144	0.5		
	146	2.2		
	148	2.2	2.7	
	150	4.9		
	152	3.8		3.2
	154	3.3	14.3	4.3
	156	37.5	42.9	54.3
	158	17.4	4.5	10.6
	160		0.9	3.2
	162	20.7	29.5	3.2
	164	2.2		
	168	1.6		
	170	3.3	5.4	
	176			21.3
CA71	302	17.3	15.6	9.0
	308	9.7	16.7	15.0
	310	0.5	3.1	
	314	72.4	64.6	76.0
NVHRT03	114	0.5		
	116	11.9	8.3	12.0
	118	2.1	1.9	6.5
	120	5.7	13.9	0.9
	122	18.6	8.3	20.4
	124	18.0	17.6	6.5
	126	36.6	50.0	45.4
	128	3.1		7.4
	130	3.6		
	132			0.9
NVHRT73	222	2.2	6.1	3.8
	224	22.0	26.8	48.1
	226	1.1	2.4	1.0
	228			5.8
	230	42.9	29.3	18.3
	232	23.1	26.8	12.5
	234	2.2	6.1	10.6
	236	3.3		
	238	1.1		
	240	0.5		
	242	0.5		
	244	0.5		
	248	0.5	2.4	

Appendix 1 (cont)

Loci	Number of Alleles						
	Norway	Barff	Husvik	Shared between Norway & Barff	Shared between Norway & Husvik	Shared between Barff & Husvik	Found in pre-BN population but not in either post-BN population.
RT27	15	5	5	5	5	4	9
*RT30	11	7	5	7	5	5	4
*RT13	12	6	6	6	6	6	6
NVHRT22	13	7	7	6	5	5	6
CA13	6	5	5	5	5	4	0
*CA71	4	4	3	4	3	3	0
NVHRT03	9	6	8	6	7	6	2
NVHRT73	12	7	7	7	6	6	5
*CRH	6	5	4	5	4	4	1
*RT9	13	7	2	7	2	2	6
BM848	17	9	9	9	9	6	5
*RT5	14	9	8	9	8	8	5
*RT1	13	10	7	10	7	7	3
Total number of alleles	145	87	76	86	72	66	52

Across the 13 loci, there were a total of 145 alleles in the Norwegian population. 93 of these alleles were found in at least one of the post-bottleneck populations and 65 of these alleles (70%) were found in both the post-bottleneck populations. There were 87 alleles found in the Barff population and 76 in the Husvik population. One Barff and four Husvik alleles were not also found in the Norwegian population. In seven of the 13 loci (marked with an asterisk), each of the alleles in the Husvik population were also present in both the Barff population and Norwegian population.

Appendix 2 Basic fluctuating asymmetry statistics for all traits. The figures in grey correspond to traits where the FA results were not included in the analysis due to significant directional asymmetry in that population or due to the unreliability of measuring accuracy in the case of chk tth.

Trait	Norway				Barff				Husvik			
		FA1		FA4		FA1		FA4		FA1		FA4
	n	mean (cm)	range	variance	n	mean (cm)	range	variance	n	mean (cm)	range	variance
Ent-P	40	0.12±0.09	0-0.34	0.022	24	0.19±0.14	0-0.56	0.049	12	0.22±0.08	0.1-0.34	0.039
Ni-Ot	41	0.18±0.12	0-0.5	0.045	34	0.15±0.11	0-0.45	0.035	22	0.19±0.17	0-0.68	0.067
If-Ent	41	0.11±0.09	0-0.36	0.020	33	0.16±0.12	0-0.5	0.035	23	0.21±0.14	0-0.57	0.043
Ot-Ect	41	0.13±0.12	0-0.51	0.031	35	0.18±0.15	0-0.75	0.056	21	0.17±0.17	0-0.58	0.050
Ot-N	41	0.13±0.11	0-0.43	0.028	35	0.14±0.10	0-0.43	0.023	22	0.25±0.15	0-0.57	0.088
N-Ent	41	0.12±0.09	0-0.36	0.021	34	0.15±0.08	0-0.39	0.027	23	0.10±0.08	0-0.3	0.016
Ni-P	40	0.08±0.08	0-0.35	0.010	24	0.07±0.06	0-0.25	0.008	11	0.06±0.04	0-0.14	0.006
Orb Lgth	41	0.05±0.05	0-0.17	0.005	34	0.06±0.05	0-0.23	0.006	22	0.04±0.04	0-0.13	0.003
Orb Hgt	41	0.07±0.06	0-0.23	0.006	33	0.09±0.09	0-0.47	0.014	21	0.06±0.05	0-0.17	0.006
B-If	41	0.10±0.09	0-0.38	0.016	34	0.12±0.08	0-0.28	0.019	22	0.11±0.07	0-0.22	0.014
B-Ent	41	0.07±0.05	0-0.17	0.006	34	0.13±0.13	0-0.68	0.034	22	0.10±0.08	0-0.29	0.013
B-Zyg	41	0.12±0.09	0-0.31	0.020	32	0.11±0.10	0-0.41	0.021	20	0.14±0.09	0-0.35	0.028
H-Zyg	41	0.10±0.09	0-0.53	0.018	32	0.08±0.05	0-0.25	0.010	20	0.10±0.06	0-0.25	0.014
St-Zyg	41	0.12±0.10	0-0.49	0.025	30	0.11±0.08	0-0.31	0.018	20	0.13±0.12	0-0.44	0.029
Po-Zyg	41	0.15±0.10	0-0.38	0.028	32	0.14±0.10	0-0.36	0.018	21	0.37±0.21	0.1-0.78	0.151
ChkTth	41	0.11±0.12	0-0.67	0.025	34	0.08±0.07	0-0.28	0.012	22	0.13±0.11	0-0.42	0.031
Vom-Po	38	0.13±0.12	0-0.44	0.031	33	0.25±0.22	0-0.83	0.114	22	0.34±0.21	0-0.8	0.114

Appendix 2 (cont)

Trait	Norway compared to Barff						Norway compared to Husvik					
	|FA1|		FA4		FA10		|FA1|		FA4		FA10	
	z	p-value	F	p-value	F	p-value	z	p-value	F	p-value	F	p-value
Ent-P	-2.10	0.04^	2.34	0.005^	2.04	0.02^	-3.27	0.001*	1.77	ns	1.77	ns
Ni-Ot	-1.16	ns	1.25	ns	1.17	ns	-0.41	ns	1.48	ns	1.68	ns
If-Ent	-2.01	0.04^	1.77	0.04^	1.58	ns	-2.92	0.003*	2.16	0.02^	2.36	0.04^
Ot-Ect	-1.74	ns	1.78	0.04^	1.85	0.03^	-0.67	ns	1.60	ns	1.92	ns
Ot-N	-0.77	ns	1.14	ns	1.12	ns	-2.99	0.003*	3.18	0.0007*	3.62	0.0002*
N-Ent	-1.91	ns	1.33	ns	1.32	ns	-0.71	ns	1.31	ns	1.59	ns
Ni-P	-0.96	ns	1.83	0.04^+	1.72	ns	-0.38	ns	1.71	ns	2.11	0.03^+
Orb Lgth	-0.01	ns	1.13	ns	1.18	ns	-0.84	ns	1.48	ns	1.73	ns
Orb Hgt	-0.34	ns	2.30	0.006^	2.62	0.002*	-0.78	ns	1.01	ns	1.36	ns
B-If	-1.55	ns	1.22	ns	1.10	ns	-1.36	ns	1.16	ns	1.07	ns
B-Ent	-2.45	0.01^	5.31	<0.0001***	6.35	0.0001***	-1.45	ns	2.07	0.02^	2.36	0.009^
B-Zyg	-0.28	ns	1.10	ns	1.08	ns	-1.02	ns	1.45	ns	2.20	0.01^
H-Zyg	-0.34	ns	1.93	0.03^+	2.01	0.02^+	-0.79	ns	1.24	ns	1.37	ns
St-Zyg	-0.28	ns	1.36	ns	1.38	ns	-0.08	ns	1.17	ns	1.14	ns
Po-Zyg	-0.65	ns	1.47	ns	1.71	ns	-4.28	<0.0001***	5.43	<0.0001***	6.64	<0.0001***
ChkTth	-1.33	ns	2.07	0.02^+	2.92	0.001*+	-1.19	ns	1.23	ns	1.45	ns
Vom-Po	-2.37	0.02^	3.76	<0.0001***	3.60	0.0001***	-4.11	<0.0001***	3.67	0.0002**	3.95	<0.0001***

Trait	Barff compared to Husvik					
	|FA1|		FA4		FA10	
	z	p-value	F	p-value	F	p-value
Ent-P	-0.99	ns	1.32	ns	1.15	ns
Ni-Ot	-0.48	ns	1.85	ns	1.96	0.04^
If-Ent	-1.36	ns	1.22	ns	1.21	ns
Ot-Ect	-0.50	ns	1.12	ns	1.07	ns
Ot-N	-2.54	0.01^	3.63	0.0004**	4.04	0.0001***
N-Ent	-2.26	0.02^	1.75	ns	2.10	0.04^
Ni-P	-0.20	ns	1.07	ns	1.23	ns
Orb Lgth	-0.84	ns	1.31	ns	2.59	0.01^
Orb Hgt	-1.05	ns	2.32	0.02^+	3.56	0.001*+
B-If	-0.11	ns	1.42	ns	1.18	ns
B-Ent	-0.72	ns	2.57	0.01^+	2.68	0.009^
B-Zyg	-1.26	ns	1.32	ns	2.38	0.01^
H-Zyg	-1.25	ns	1.55	ns	1.47	ns
St-Zyg	-0.18	ns	1.59	ns	1.58	ns
Po-Zyg	-4.29	<0.0001***	7.99	<0.0001***	11.37	<0.0001***
ChkTth	-1.69	ns	2.54	0.007^	4.21	0.0001**
Vom-Po	-1.57	ns	1.02	ns	1.10	ns

Appendix 2 (cont) Comparison of fluctuating asymmetry for all traits among populations. Mann-Whitney test comparing the absolute value of |R-L| to give Z value and an F-statistic comparing the variance. ns - not significant, ^ significant before application of Bonferroni correction. After application of Bonferroni correction *p<0.05, **p<0.01, ***p<0.001. In all cases Norway showed more FA than Barff or Husvik and Barff showed more FA than Husvik unless marked with $^+$ to indicate the cases where Norway showed less FA than Barff or Husvik and Barff showed less FA than Husvik.

The figures in grey correspond to traits in which there was significant directional asymmetry in one or both of the populations in the comparison.

References

Alados, C. L., J. Escos and J. M. Emlen (1995). "Fluctuating asymmetry and fractal dimension of the sagittal suture as indicators of inbreeding depression in dama and dorcas gazelles." Canadian Journal Of Zoology-Revue Canadienne De Zoologie **73**(10): 1967-1974.

Allendorf, F. W. (1986). "Genetic Drift And The Loss Of Alleles Versus Heterozygosity." Zoo Biology **5**(2): 181-190.

Amos, W. and A.R. Hoelzel (1991). I.W.C. Special Issue **13**(99).

Amos, W., J. W. Wilmer, K. Fullard, T. M. Burg, J. P. Croxall, D. Bloch and T. Coulson (2001). "The influence of parental relatedness on reproductive success." Proceedings Of The Royal Society Of London Series B-Biological Sciences **268**(1480): 2021-2027.

Anciaes, M. and M. A. Marini (2000). "The effects of fragmentation on fluctuating asymmetry in passerine birds of Brazilian tropical forests." Journal of Applied Ecology **37**(6): 1013-1028.

Ardern, S. L. and D. M. Lambert (1997). "Is the black robin in genetic peril?" Molecular Ecology **6**(1): 21-28.

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

Ballou, J. D. (1997). "Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations." Journal Of Heredity **88**(3): 169-178.

Balloux, F., H. Brunner, N. Lugon-Moulin, J. Hausser and J. Goudet (2000). "Microsatellites can be misleading: An empirical and simulation study." Evolution **54**(4): 1414-1422.

Balloux, F. and J. Goudet (2002). "Statistical properties of population differentiation estimators under stepwise mutation in a finite island model." Molecular Ecology **11**(4): 771-783.

Balloux, F. and N. Lugon-Moulin (2002). "The estimation of population differentiation with microsatellite markers." Molecular Ecology **11**(2): 155-165.

Beardmore, J.A. (1960). "Developmental Stability in constant and fluctuating temperatures." Heredity **14**: 411-422.

Bell, C (2001). "South Georgia Reindeer Translocation, Department of Agriculture, Falkland Island Government." not published.

Bergerud, A. T. (1964). "Relationship of mandible length to sex in Newfoundland caribou." Journal of Wildlife Management **28**(1): 54-56.

Bergerud, A. T. (1970). "Eruption of Permanent Premolars and Molars for Newfoundland Caribou." Journal of Wildlife Management **34**(4): 962-&.

Berry, R. J. (1964). "Evolution Of Island Population Of House Mouse." Evolution **18**(3): 468-&.

Berry, R. J. and M. E. Jakobson (1975). "Ecological Genetics Of An Island Population Of House Mouse (Mus-Musculus)." Journal Of Zoology **175**(APR): 523-540.

Berube, C. H., M. Festa-Bianchet and J. T. Jorgenson (1999). "Individual differences, longevity, and reproductive senescence in bighorn ewes." Ecology **80**(8): 2555-2565.

Bonnell, M. L. and R. K. Selander (1974). "Elephant Seals - Genetic-Variation And Near Extinction." Science **184**(4139): 908-909.

Bonner, W.N. (1958). "The introduced reindeer of South Georgia." F.I.D.S Scientific Reports **No 22**.

Borisov, V. I., A. V. Valetsky, I. L. Dmitrieva, N. L. Krushinskaya and V. M. Zakharov (1997). "Inbreeding and outbreeding impact on developmental stability of laboratory rat Rattus norvegicus." Acta Theriologica: 67-72.

Britten, H. B. (1996). "Meta-analyses of the association between multilocus heterozygosity and fitness." Evolution **50**(6): 2158-2164.

Broders, H. G., S. P. Mahoney, W. A. Montevecchi and W. S. Davidson (1999). "Population genetic structure and the effect of founder events on the genetic variability of moose, Alces alces, in Canada." Molecular Ecology **8**(8): 1309-1315.

Bryant, E. H., L. M. Combs and S. A. McCommas (1986). "Morphometric Differentiation among Experimental Lines of the Housefly in Relation to a Bottleneck." Genetics **114**(4): 1213-1223.

Bryant, E. H., McCommas, S. A., Combs, L. M. (1986). "The Effect of an Experimental Bottleneck Upon Quantitative Genetic-Variation in the Housefly." Genetics **114**(4): 1191-1211.

Carson, H. L. and A. R. Templeton (1984). "Genetic Revolutions in Relation to Speciation Phenomena - the Founding of New Populations." Annual Review of Ecology and Systematics **15**: 97-131.
Case, T. J. (1978). "General Explanation For Insular Body Size Trends In Terrestrial Vertebrates." Ecology **59**(1): 1-18.
Caughley, G. (1966). "Mortality Patterns in Mammals." Ecology **47**(6): 906-&.
Caughley, G. (1994). "Directions in conservation biology." Journal of Animal Ecology **63**: 215-244.
Chakraborty, R., P. A. Fuerst and M. Nei (1980). "Statistical Studies On Protein Polymorphism In Natural-Populations .3. Distribution Of Allele Frequencies And The Number Of Alleles Per Locus." Genetics **94**(4): 1039-1063.
Chakraborty, R. and M. Nei (1977). "Bottleneck Effects on Average Heterozygosity and Genetic Distance with Stepwise Mutation Model." Evolution **31**(2): 347-356.
Charlesworth, D. and B. Charlesworth (1987). "Inbreeding Depression And Its Evolutionary Consequences." Annual Review Of Ecology And Systematics **18**: 237-268.
Clarke, G. M. (1995). "Relationships Between Developmental Stability And Fitness - Application For Conservation Biology." Conservation Biology **9**(1): 18-24.
Clarke, G. M. (1998a). "Developmental stability and fitness: The evidence is not quite so clear." American Naturalist **152**(5): 762-766.
Clarke, G. M. (1998b). "The genetic basis of developmental stability. IV. Individual and population asymmetry parameters." Heredity **80**: 553-561.
Clarke, G. M. (1998c). "The genetic basis of developmental stability. V. Inter- and intra-individual character variation." Heredity **80**: 562-567.
Clarke, G. M., Oldroyd, B. P., Hunt, P. (1992). "The Genetic-Basis Of Developmental Stability In Apis-Mellifera - Heterozygosity Versus Genic Balance." Evolution **46**(3): 753-762.
Clegg, M. T., J. F. Kidwell and C. R. Horch (1980). "DYNAMICS OF CORRELATED GENETIC SYSTEMS .5. RATES OF DECAY OF LINKAGE DISEQUILIBRIA IN EXPERIMENTAL POPULATIONS OF DROSOPHILA-MELANOGASTER." Genetics **94**(1): 217-234.
Cluttonbrock, T. H., Albon, S.D. and Guinness, F.E. (1988). Reproductive Success in Male and Female Red Deer. Reproductive Success. T. H. Cluttonbrock. Chicago, The University of Chicago Press.
Coltman, D. W., W. D. Bowen and J. M. Wright (1998). "Birth weight and neonatal survival of harbour seal pups ape positively correlated with genetic variation measured by microsatellites." Proceedings Of The Royal Society Of London Series B-Biological Sciences **265**(1398): 803-809.
Coltman, D. W., J. G. Pilkington, J. A. Smith and J. M. Pemberton (1999). "Parasite-mediated selection against inbred Soay sheep in a free-living, island population." Evolution **53**(4): 1259-1267.
Coltman, D. W. and J. Slate (2003). "Microsatellite measures of inbreeding: A meta-analysis." Evolution **57**(5): 971-983.
Cornuet, J. M. and G. Luikart (1996). "Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data." Genetics **144**(4): 2001-2014.
Coulson, T., S. Albon, J. Slate and J. Pemberton (1999). "Microsatellite loci reveal sex-dependent responses to inbreeding and outbreeding in red deer calves." Evolution **53**(6): 1951-1960.
Coulson, T. N., J. M. Pemberton, S. D. Albon, M. Beaumont, T. C. Marshall, J. Slate, F. E. Guinness and T. H. Clutton-Brock (1998). "Microsatellites reveal heterosis in red deer." Proceedings Of The Royal Society Of London Series B-Biological Sciences **265**(1395): 489-495.
Cronin, M. A., J. C. Patton, N. Balmysheva and M. D. MacNeil (2003). "Genetic variation in caribou and reindeer (Rangifer tarandus)." Animal Genetics **34**(1): 33-41.
Damuth, J. (1993). "Copes Rule, The Island Rule And The Scaling Of Mammalian Population-Density." Nature **365**(6448): 748-750.
David, P. (1998). "Heterozygosity-fitness correlations: new perspectives on old problems." Heredity **80**: 531-537.
Davis, T.A. (1974). "Further notes on asymmetry of reindeer antlers." FORMA FUNCTIO 7: 55-58.
Davis, T.A.

(1973). "Asymmetry of reindeer antlers." FORMA FUNCTIO 6 **6**: 373-382.
de-Luna-Lopez, C.J. (2005). The relationship between developmental stability, genomic diversity adn environmental stress in the eastern North Atlantic population of harbour porpoises. Biological Sciences. Durham, Durham. **PhD**.
Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M, Freimer, M.B. (1994). "Mutational processes of simple-sequence repeat loci in human populations." Proceedings of the National Academy of Sciences of the United States of America **91**: 3166-3170.
Dietrich, W., H. Katz, S. E. Lincoln, H. S. Shin, J. Friedman, N. L. Dracopoli and E. S. Lander (1992). "A Genetic Map of the Mouse Suitable for Typing Intraspecific Crosses." Genetics **131**(2): 423-447.
Eanes, W. F. (1981). "Enzyme Heterozygosity And Morphological Variance." Nature **290**(5807): 609-610.
Efron, B. (1979). "Bootstrap methods: Another look at the jack-knife." Annals of Statistics **7**: 1-26.
Eldridge, M. D. B., J. M. King, A. K. Loupis, P. B. S. Spencer, A. C. Taylor, L. C. Pope and G. P. Hall (1999). "Unprecedented low levels of genetic variation and inbreeding depression in an island population of the black-footed rock-wallaby." Conservation Biology **13**(3): 531-541.
Ellegren, H. (2000). "Microsatellite mutations in the germline: implications for evolutionary inference." Trends In Genetics **16**(12): 551-558.
England, P. R. and G. H. R. Osler (2001). "GENELOSS: a computer program for simulating the effects of population bottlenecks on genetic diversity." Molecular Ecology Notes **1**(1-2): 111-113.
Estoup, A., L. Garnery, M. Solignac and J. M. Cornuet (1995). "Microsatellite Variation In Honey-Bee (Apis-Mellifera L) Populations - Hierarchical Genetic-Structure And Test Of The Infinite Allele And Stepwise Mutation Models." Genetics **140**(2): 679-695.
Fincke, O.M. (1988). Sources of Variation in Lifetime Reproductive Success in a Nonterritorial Damselfly (*Odonata: Coenagrionidae*). Reproductive Success. T. H. Cluttonbrock. Chicago, University of Chicago Press.
Fleischer, R. C., Johnston, R. F., Klitz, W. J. (1983). "Allozymic Heterozygosity and Morphological Variation in House Sparrows." Nature **304**(5927): 628-630.
Foster, J. B. (1964). "Evolution Of Mammals On Islands." Nature **202**(492): 234-&.
Fowler, K. and M. C. Whitlock (1994). "Fluctuating Asymmetry Does Not Increase with Moderate Inbreeding in Drosophila-Melanogaster." Heredity **73**: 373-376.
Frankham, R. (1995). "Conservation genetics." Annual Review Of Genetics **29**: 305-327.
Frankham, R. (1997). "Do island populations have less genetic variation than mainland populations?" Heredity **78**: 311-327.
Frankham, R. (1998). "Inbreeding and extinction: Island populations." Conservation Biology **12**(3): 665-675.
Frankham, R., Ballou, J.D. and Briscoe, D.A. (2002). Introduction to Conservation Genetics. Cambridge, Cambridge University Press.
Fuerst, P. A. and T. Maruyama (1986). "Considerations On The Conservation Of Alleles And Of Genic Heterozygosity In Small Managed Populations." Zoo Biology **5**(2): 171-179.
Gaggiotti, O. E. and L. Excoffier (2000). "A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic distances." Proceedings Of The Royal Society Of London Series B-Biological Sciences **267**(1438): 81-87.
Garza, J. C. and E. G. Williamson (2001). "Detection of reduction in population size using data from microsatellite loci." Molecular Ecology **10**(2): 305-318.
Goodman, S. J. (1997). "R-ST Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance." Molecular Ecology **6**(9): 881-885.
Goodnight, C. J. (1987). "On the Effect of Founder Events on Epistatic Genetic Variance." Evolution **41**(1): 80-91.
Goudet, J. (1995). "FSTAT (Version 1.2): A computer program to calculate F-statistics." Journal Of Heredity **86**(6): 485-486.
Goudet, J. and L. Keller (2002). "The correlation between inbreeding and fitness: does allele size matter?" Trends In Ecology & Evolution **17**(5): 201-202.
Graham, J.H., K.E. Roe and T.B. West (1993). "Effects of lead and benzene on the developmental stability of *Drosophila melangastor*." Ecotoxicology **2**: 185-195.

Guo, S. W. and E. A. Thompson (1992). "Performing The Exact Test Of Hardy-Weinberg Proportion For Multiple Alleles." Biometrics **48**(2): 361-372.
Hagelberg, E. and J. B. Clegg (1991). "Isolation And Characterization Of Dna From Archaeological Bone." Proceedings Of The Royal Society Of London Series B-Biological Sciences **244**(1309): 45-50.
Halley, J. and A.R. Hoelzel (1996). Simulation models of bottleneck events in natural populations. Molecular Genetic Approaches in Conservation. T.B. Smith and R.K. Wayne. Oxford, Oxford University Press: 347-364.
Harley, E. H., I. Baumgarten, J. Cunningham and C. O'Ryan (2005). "Genetic variation and population structure in remnant populations of black rhinoceros, Diceros bicornis, in Africa." Molecular Ecology **14**(10): 2981-2990.
Hartl, G. B. and P. Hell (1994). "Maintenance Of High-Levels Of Allelic Variation In Spite Of A Severe Bottleneck In Population-Size - The Brown Bear (Ursus-Arctos) In The Western Carpathians." Biodiversity And Conservation **3**(6): 546-554.
Hartl, G. B., Suchentrunk, F., Willing, R. and Petznek, R. (1995). "Allozyme heterozygosity and fluctuating asymmetry in the brown hare (Lepus europaeus): A test of the developmental homeostasis hypothesis." Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **350**(1334): 313-323.
Headland, R (1984). The Island of South Georgia, Cambridge University Press.
Heaney, L. R. (1978). "Island Area And Body Size Of Insular Mammals - Evidence From Tri-Colored Squirrel (Callosciurus-Prevosti) Of Southeast-Asia." Evolution **32**(1): 29-44.
Hedrick, P., R. Fredrickson and H. Ellegren (2001). "Evaluation of (d)over-bar(2), a microsatellite measure of inbreeding and outbreeding, in wolves with a known pedigree." Evolution **55**(6): 1256-1260.
Hill, W. G. (1977). "Correlation Of Gene-Frequencies Between Neutral Linked Genes In Finite Populations." Theoretical Population Biology **11**(2): 239-248.
Hoelzel, A. (1999). "Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal." Biological Journal of the Linnean Society 68, no **1-2**: 23 (17 pages).
Hoelzel, A. R., Fleischer, R. C., Campagna, C., Le Boeuf, B. J. and Alvord, G. (2002). "Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal." Journal of Evolutionary Biology **15**(4): 567-575.
Hoelzel, A. R., J. Halley, S. J. O'Brien, C. Campagna, T. Arnbom, B. Le Boeuf, K. Ralls and G. A. Dover (1993). "Elephant Seal Genetic Variation and the Use of Simulation Models to Investigate Historical Population Bottlenecks." The journal of heredity **84**(6): 443-449.
Hoffman, J. I., J. Forcada, P. N. Trathan and W. Amos (2007). "Female fur seals show active choice for males that are heterozygous and unrelated." Nature **445**(7130): 912-914.
Hosken, D. J., W. U. Blanckenhorn and P. I. Ward (2000). "Developmental stability in yellow dong flies (Scathophaga stercoraria): fluctuating asymmetry, heterozygosity and environmental stress." Journal Of Evolutionary Biology **13**(6): 919-926.
Houlden, B. A., P. R. England, A. C. Taylor, W. D. Greville and W. B. Sherwin (1996). "Low genetic variability of the koala Phascolarctos cinereus in south-eastern Australia following a severe population bottleneck." Molecular Ecology **5**(2): 269-281.
Hudson, R.R., Boos, D.D. and Kaplan, N.L. (1992). "A statistical test for detecting geographical subdivision." Molecular Biology and Evolution **9**: 138-151.
Hutchison, D. W. and J. M. Cheverud (1995). "Fluctuating Asymmetry In Tamarin (Saguinus) Cranial Morphology - Intraspecific and Interspecific Comparisons between Taxa with Varying Levels of Genetic Heterozygosity." Journal of Heredity **86**(4): 280-288.
Kaeuffer, R., D. W. Coltman, J. L. Chapuis, D. Pontier and D. Reale (2007). "Unexpected heterozygosity in an island mouflon population founded by a single pair of individuals." Proceedings of the Royal Society B-Biological Sciences **274**(1609): 527-533.
Keller, L. F., P. Arcese, J. N. M. Smith, W. M. Hochachka and S. C. Stearns (1994). "Selection Against Inbred Song Sparrows During A Natural-Population Bottleneck." Nature **372**(6504): 356-357.
Keller, L. F. and D. M. Waller (2002). "Inbreeding effects in wild populations." Trends In Ecology & Evolution **17**(5): 230-241.

Kieser, J. A. and H. T. Groeneveld (1991). "Fluctuating Odontometric Asymmetry, Morphological Variability, And Genetic Monomorphism In The Cheetah Acinonyx-Jubatus." Evolution **45**(5): 1175-1183.
Kimball, R. T., J. D. Ligon and M. MerolaZwartjes (1997). "Fluctuating asymmetry in red junglefowl." Journal Of Evolutionary Biology **10**(4): 441-457.
Kimura, M. (1955). "Stochastic Processes And Distribution Of Gene Frequencies Under Natural Selection." Cold Spring Harbor Symposia On Quantitative Biology **20**: 33-53.
Kruuk, L. E. B., T. H. Clutton-Brock, J. Slate, J. M. Pemberton, S. Brotherstone and F. E. Guinness (2000). "Heritability of fitness in a wild mammal population." Proceedings of the National Academy of Sciences of the United States of America **97**(2): 698-703.
Lacy, R. C. (1997). "Importance of genetic variation to the viability of mammalian populations." Journal Of Mammalogy **78**(2): 320-335.
Lagesen, K. and I. Folstad (1998). "Antler asymmetry and immunity in reindeer." Behavioral Ecology and Sociobiology **44**(2): 135-142.
Lawlor, T. E. (1982). "The Evolution Of Body Size In Mammals - Evidence From Insular Populations In Mexico." American Naturalist **119**(1): 54-72.
Le Page, S. L., R. A. Livermore, D. W. Cooper and A. C. Taylor (2000). "Genetic analysis of a documented population bottleneck: introduced Bennett's wallabies (Macropus rufogriseus rufogriseus) in New Zealand." Molecular Ecology **9**(6): 753-763.
Leader-Williams (1978). "The history of the introduced reindeer of South Georgia." Deer **4**(5): 256-261.
Leader-Williams (1988). Reindeer on South Georgia, Cambridge University Press.
Leader-Williams, N (1979). "Age determination of Reindeer introduced into South Georgia." Journal of the Zoological Society of London **188**: 501-515.
Leaderwilliams, N. and C. Ricketts (1982). "Growth And Condition Of 3 Introduced Reindeer Herds On South Georgia - The Effects Of Diet And Density." Holarctic Ecology **5**(4): 381-388.
Leaderwilliams, N., T. A. Scott and R. M. Pratt (1981). "Forage Selection By Introduced Reindeer On South Georgia, And Its Consequences For The Flora." Journal Of Applied Ecology **18**(1): 83-106.
Leaderwilliams, N., R. I. L. Smith and P. Rothery (1987). "Influence Of Introduced Reindeer On The Vegetation Of South-Georgia - Results From A Long-Term Exclusion Experiment." Journal Of Applied Ecology **24**(3): 801-822.
Leamy, L. J. and C. P. Klingenberg (2005). "The genetics and evolution of fluctuating asymmetry." Annual Review Of Ecology Evolution And Systematics **36**: 1-21.
Leamy, L. J., S. Meagher, S. Taylor, L. Carroll and W. K. Potts (2001). "Size and fluctuating asymmetry of morphometric characters in mice: Their associations with inbreeding and t-haplotype." Evolution **55**(11): 2333-2341.
Leary, R. F. and F. W. Allendorf (1989). "Fluctuating Asymmetry as an Indicator of Stress - Implications for Conservation Biology." Trends in Ecology & Evolution **4**(7): 214-217.
Leary, R. F., F. W. Allendorf and K. L. Knudsen (1985). "Developmental Instability And High Meristic Counts In Interspecific Hybrids Of Salmonid Fishes." Evolution **39**(6): 1318-1326.
Leary, R. F., Allendorf F. W., Knudsen, K. L. (1983). "Developmental Stability and Enzyme Heterozygosity in Rainbow- Trout." Nature **301**(5895): 71-72.
Leberg, P. L. (1992). "Effects Of Population Bottlenecks On Genetic Diversity As Measured By Allozyme Electrophoresis." Evolution **46**(2): 477-494.
Lehmann, E.L. (1959). Testing Statistical Hypotheses. New York.
Lens, Luc, Stefan Van Dongen and Erik Matthysen (2002). "Fluctuating Asymmetry as an Early Warning System in the Critically Endangered Taita Thrush." Conservation biology : the journal of the Society for Conservation Biology **16**(2): 479 (9 pages).
Lenth, R.V. (2006). Java Applets for Power and Sample Size [Computer software]. http://www.stat.uiowa.edu/~rlenth/Power. Retrieved 14/7/07.
Lerner, I.M (1954). Genetic Homeostasis. New York, Oliver & Boyd.
Leung, B. and M. R. Forbes (1996). "Fluctuating asymmetry in relation to stress and fitness: Effects of trait type as revealed by meta-analysis." Ecoscience **3**(4): 400-413.
Leung, B. and M. R. Forbes (1997). "Modelling fluctuating asymmetry in relation to stress and fitness." Oikos **78**(2): 397-405.

Leung, B., M. R. Forbes and D. Houle (2000). "Fluctuating asymmetry as a bioindicator of stress: Comparing efficacy of analyses involving multiple traits." American Naturalist **155**(1): 101-115.
Lister, A. M. (1989). "Rapid Dwarfing Of Red Deer On Jersey In The Last Interglacial." Nature **342**(6249): 539-542.
Lomolino, M. V. (1985). "Body Size Of Mammals On Islands - The Island Rule Reexamined." American Naturalist **125**(2): 310-316.
Lomolino, M. V. (2005). "Body size evolution in insular vertebrates: generality of the island rule." Journal Of Biogeography **32**(10): 1683-1699.
Lugon-Moulin, N., H. Brunner, A. Wyttenbach, J. Hausser and J. Goudet (1999). "Hierarchical analyses of genetic differentiation in a hybrid zone of Sorex araneus (Insectivora : Soricidae)." Molecular Ecology **8**(3): 419-431.
Luikart, G., F. W. Allendorf, J. M. Cornuet and W. B. Sherwin (1998a). "Distortion of allele frequency distributions provides a test for recent population bottlenecks." Journal of Heredity **89**(3): 238-247.
Luikart, G. and J. M. Cornuet (1998). "Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data." Conservation Biology **12**(1): 228-237.
Luikart, G., W. B. Sherwin, B. M. Steele and F. W. Allendorf (1998b). "Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change." Molecular Ecology **7**(8): 963-974.
Madsen, T., B. Stille and R. Shine (1996). "Inbreeding depression in an isolated population of adders Vipera berus." Biological Conservation **75**(2): 113-118.
Markow, T. A. (1995). "Evolutionary Ecology And Developmental Instability." Annual Review Of Entomology **40**: 105-120.
Markusson, E. and I. Folstad (1997). "Reindeer antlers: Visual indicators of individual quality?" Oecologia **110**(4): 501-507.
Maruyama, T. and P. A. Fuerst (1985). "Population Bottlenecks and Nonequilibrium Models in Population- Genetics .3. Genic Homozygosity in Populations Which Experience Periodic Bottlenecks." Genetics **111**(3): 691-703.
Mather, K. (1953). "Genetical Control Of Stability In Development." Heredity **7**(3): 297-336.
McIntosh, E and D.W.H. Walton (2000). Environmental Management Plan for South Georgia, British Antarctic Survey on behalf of the Government of South Georgia and the South Sandwich Islands.
McVey, M.E. (1988). The opportunity for sexual selection in a territorial dragonfly, *Erythemis simplicicollis*. Reproductive Success. T. H. Cluttonbrock. Chicago, University of Chicago Press.
Merola, M. (1994). "Reassessment Of Homozygosity Acid The Case For Inbreeding Depression In The Cheetah, Acinonyx-Jubatus - Implications For Conservation." Conservation Biology **8**(4): 961-971.
Milligan, B.G. (1998). Total DNA Isolation. Molecular Genetic Analysis of Populations. A.R. Hoelzel. Oxford, Oxford University Press: 36-38.
Mitton, J. B. (1978). "Relationship between heterozygosity for enzyme loci adn variation of morphological characters in natural populations." Nature **273**: 661-662.
Mitton, J. B. and M. C. Grant (1984). "Associations Among Protein Heterozygosity, Growth-Rate, And Developmental Homeostasis." Annual Review Of Ecology And Systematics **15**: 479-499.
Modi, W. S., R. K. Wayne and S. J. Obrien (1987). "Analysis Of Fluctuating Asymmetry In Cheetahs." Evolution **41**(1): 227-228.
Moen, J. and H MacAlister (1994). "Continued range expansion of introduced reindeer on South Georgia." Polar Biology **14**: 459-462.
Moller, A. P. (1997a). "Developmental stability and fitness: A review." American Naturalist **149**(5): 916-932.
Moller, A.P., Swaddle, J.P. (1997b). Asymmetry, Developmental Stability, and Evolution, Oxford University Press.
Nei, M. (1987). Molecular Evolutionary Genetics. New York, Columbia University Press.
Nei, M., T. Maruyama and R. Chakraborty (1975). "Bottleneck Effect and Genetic-Variability in Populations." Evolution **29**(1): 1-10.
Nei, M. and A. K. Roychoudhury (1974). "Sampling Variances Of Heterozygosity And Genetic Distance." Genetics **76**(2): 379-390.

Neville, A.C. (1976). Animal Asymmetry. London, Edward Arnold.
Olstad, O (1930). "Rats and reindeer in the Antarctic " Scientific results of the Norwegian Antarctic Expeditions **4**: 1-19.
Paetkau, D., L. P. Waits, P. L. Clarkson, L. Craighead and C. Strobeck (1997). "An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations." Genetics **147**(4): 1943-1957.
Palkovacs, E.P. (2003). "Explaining adaptive shifts in body size on islands: a life history approach." Oikos **103**(1): 37-44.
Palmer, A. R. (1994). Fluctuating Asymmetry Analyses - A Primer. Dordrecht, Netherlands, Kluwer.
Palmer, A. R. and C. Strobeck (1986). "Fluctuating Asymmetry - Measurement, Analysis, Patterns." Annual Review of Ecology and Systematics **17**: 391-421.
Palmer, A.R. and C. Stobeck (2003). Fluctuating Asymmetry Analyses Revisited. Developmental Instability (DI): Causes and Consequences. M. Polak, Oxford University Press: 279-319.
Payen, E. J. and C. Y. Cotinot (1993). "Comparative Hmg-Box Sequences Of The Sry Gene Between Sheep, Cattle And Goats." Nucleic Acids Research **21**(11): 2772-2772.
Pemberton, J. M., Coltman, D.W., Coulson, T.N. and Slate, J. (1999). Using microsatellites to measure the fitness consequences of inbreeding and outbreeding. Microsatellites, Evolution and Application. Goldstein and Schlotterer. Oxford, Oxford University Press.
Pertoldi, C., T. N. Kristensen, D. H. Andersen and V. Loeschcke (2006). "Developmental instability as an estimator of genetic stress." Heredity **96**(2): 122-127.
Pertoldi, C., Loeschcke, V., Braun, A., Madson, A.B. and Randi, E. (2000). "Craniometrical variability and developmental stability. Two useful tools for assessing the population viability of Eurasian otter (Lutra lutra) populations in Europe." Biological Journal of the Linnean Society **70**: 309-323.
Pierce, B. A. and J. B. Mitton (1982). "Allozyme Heterozygosity And Growth In The Tiger Salamander, Ambystoma-Tigrinum." Journal Of Heredity **73**(4): 250-253.
Polkovacs, E.P. (2003). "Explaining adaptive shifts in body size on islands: a life history approach." Oikos **103**(1): 37-44.
Queller, D. C. and K. F. Goodnight (1989). "Estimating Relatedness Using Genetic-Markers." Evolution **43**(2): 258-275.
Queney, G., N. Ferrand, S. Marchandeau, M. Azevedo, F. Mougel, M. Branco and M. Monnerot (2000). "Absence of a genetic bottleneck in a wild rabbit (Oryctolagus cuniculus) population exposed to a severe viral epizootic." Molecular Ecology **9**(9): 1253-1264.
Raia, P. and S. Meiri (2006). "The island rule in large mammals: Paleontology meets ecology." Evolution **60**(8): 1731-1742.
Rasmuson, M. (2002). "Fluctuating asymmetry - indicator of what?" Hereditas **136**(3): 177-183.
Raymond, M. and Rousset, F. (1995). "GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism." Journal of Heredity **86**: 248-249.
Reale, D. and M. Festa-Bianchet (2000). "Quantitative genetics of life-history traits in a long-lived wild mammal." Heredity **85**(6): 593-603.
Reimers, E. (1972). "Growth in Domestic and Wild Reindeer in Norway." Journal of Wildlife Management **36**(2): 612-&.
Reimers, E. and O. Nordby (1968). "Relationship between Age and Tooth Cementum Layers in Norwegian Reindeer." Journal of Wildlife Management **32**(4): 957-&.
Rice, W. R. (1989). "Analyzing Tables Of Statistical Tests." Evolution **43**(1): 223-225.
Robertson, F. W. and E. C. R. Reeve (1952). "Heterozygosity, Environmental Variation And Heterosis." Nature **170**(4320): 286-286.
Roed, K. H. (1985). "Genetic-Variability in Norwegian Semi-Domestic Reindeer (Rangifer-Tarandus L)." Hereditas **102**(2): 177-184.
Roed, K. H. (1986). "Genetic-Variability in Norwegian Wild Reindeer (Rangifer- Tarandus L)." Hereditas **104**(2): 293-298.
Roed, K. H. (1987). "Transferrin Variation And Body Size In Reindeer, Rangifer-Tarandus L." Hereditas **106**(1): 67-71.
Roed, K. H. (1998a). "Influence of selection and management on the genetic structure of reindeer populations." Acta Theriologica: 179-186.

Roed, K. H. (1998b). "Microsatellite variation in Scandinavian Cervidae using primers derived from Bovidae." Hereditas **129**(1): 19-25.
Roed, K.H. and L. Midthjell (1998). "Microsatellites in reindeer, Rangifer tarandus, and their use in other cervids." Molecular Ecology **7**(12): 1773-1776.
Rooney, A. P., R. L. Honeycutt, S. K. Davis and J. N. Derr (1999). "Evaluating a putative bottleneck in a population of bowhead whales from patterns of microsatellite diversity and genetic disequilibria." Journal Of Molecular Evolution **49**(5): 682-690.
Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W., Hanski, I. (1998). "Inbreeding and extinction in a butterfly metapopulation." Nature **392**(6675): 491-494.
Schmidt, N. M. and P. M. Jensen (2003). "Changes in mammalian body length over 175 years - Adaptations to a fragmented landscape?" Conservation Ecology **7**(2).
Schneider, S., Roessli, D. and Excoffier, L. (2000). "Arlequin ver 2.000: A software for population genetics data analysis."
Shriver, M. D., L. Jin, R. Chakraborty and E. Boerwinkle (1993). "Vntr Allele Frequency-Distributions under the Stepwise Mutation Model - a Computer-Simulation Approach." Genetics **134**(3): 983-993.
Singh, S. M. and E. Zouros (1978). "Genetic-Variation Associated With Growth-Rate In American Oyster (Crassostrea-Virginica)." Evolution **32**(2): 342-353.
Skogland, T. (1983). "The Effects Of Density Dependent Resource Limitation On Size Of Wild Reindeer." Oecologia **60**(2): 156-168.
Skogland, T. (1988). "Tooth Wear by Food Limitation and Its Life-History Consequences in Wild Reindeer." Oikos **51**(2): 238-242.
Slate, J., L. E. B. Kruuk, T. C. Marshall, J. M. Pemberton and T. H. Clutton-Brock (2000). "Inbreeding depression influences lifetime breeding success in a wild population of red deer (Cervus elaphus)." Proceedings of the Royal Society of London Series B-Biological Sciences **267**(1453): 1657-1662.
Slate, J. and J. M. Pemberton (2002). "Comparing molecular measures for detecting inbreeding depression." Journal of Evolutionary Biology **15**(1): 20-31.
Slatkin, M. (1985). "Rare Alleles As Indicators Of Gene Flow." Evolution **39**(1): 53-65.
Slatkin, M. (1995a). "A Measure Of Population Subdivision Based On Microsatellite Allele Frequencies." Genetics **139**(1): 457-462.
Slatkin, M. (1995b). "A Measure Of Population Subdivision Based On Microsatellite Allele Frequencies (Vol 139, Pg 457, 1995)." Genetics **139**(3): 1463-1463.
Smith, F. A. (1992). "Evolution Of Body Size Among Woodrats From Baja-California, Mexico." Functional Ecology **6**(3): 265-273.
Sonne, C., F. F. Riget, R. Dietz, M. Kirkegaard, E. W. Born, R. Letcher and D. C. G. Muir (2005). "Trends in fluctuating asymmetry in East Greenland polar bears (Ursus maritimus) from 1892 to 2002 in relation to organohalogen pollution." Science Of The Total Environment **341**(1-3): 81-96.
Soule, M (1967). "Phenetics of Natural Populations. II. Asymmetry and evolution in a lizard." The American Naturalist **101**(918): 141-160.
Soule, M. E. (1979). "Heterozygosity And Developmental Stability - Another Look." Evolution **33**(1): 396-401.
Suchentrunk, F., Hartl, G. B., Flux, J. E. C., Parkes, J., Halden, A. and Tapper, S. (1998). "Allozyme heterozygosity and fluctuating asymmetry in brown hares Lepus europaeus introduced to New Zealand: Developmental homeostasis in populations with a bottleneck history." Acta Theriologica: 35-52.
Swaddle, J.P., Witter, M.S., Cuthill, I.C. (1994). "The analysis of fluctuating asymmetry." Animal Behaviour **48**: 986-989.
Tajima, F. (1983). "EVOLUTIONARY RELATIONSHIP OF DNA-SEQUENCES IN FINITE POPULATIONS." Genetics **105**(2): 437-460.
Taylor, A. C., W. B. Sherwin and R. K. Wayne (1994). "Genetic-Variation Of Microsatellite Loci In A Bottlenecked Species - The Northern Hairy-Nosed Wombat Lasiorhinus-Krefftii." Molecular Ecology **3**(4): 277-290.
Tebb, G. and J. M. Thoday (1954). "Stability In Development And Relational Balance Of X-Chromosomes In Drosophila-Melanogaster." Nature **174**(4441): 1109-1110.
Templeton, A. R. (1980). "The Theory of Speciation Via the Founder Principle." Genetics **94**(4): 1011-1038.
Tsitrone, A., F. Rousset and P. David (2001). "Heterosis, marker mutational processes and population inbreeding history." Genetics **159**(4): 1845-1859.

Tsubaki, Y. (1998). "Fluctuating asymmetry of the Oriental fruit fly (Dacus dorsalis) during the process of its extinction from the Okinawa Islands." Conservation Biology **12**(4): 926-929.
Valdes, A. M., M. Slatkin and N. B. Freimer (1993). "Allele Frequencies at Microsatellite Loci - the Stepwise Mutation Model Revisited." Genetics **133**(3): 737-749.
Van Valen, Leigh (1962). "A study of fluctuating asymmetry." Evolution **XVI**(2): 125-142.
various "SG file 650 - Box of correspondence and reports held in archive room at Scott Polar Institute, Cambridge."
Vollestad, L. A., K. Hindar and A. P. Moller (1999). "A meta-analysis of fluctuating asymmetry in relation to heterozygosity." Heredity **83**: 206-218.
Vrijenhoek, R. C. and S. Lerman (1982). "Heterozygosity And Developmental Stability Under Sexual And Asexual Breeding Systems." Evolution **36**(4): 768-776.
Waddington, C.H. (1942). "Canalization of development and the inheritance of acquired characters." Nature **150**(3811): 563-565.
Watterson, G. A. (1984). "Allele Frequencies After A Bottleneck." Theoretical Population Biology **26**(3): 387-407.
Wauters, L. A., A. A. Dhondt, H. Knothe and D. T. Parkin (1996). "Fluctuating asymmetry and body size as indicators of stress in red squirrel populations in woodland fragments." Journal Of Applied Ecology **33**(4): 735-740.
Wayne, R. K., Modi, W. S. and Obrien, S. J. (1986). "Morphological Variability and Asymmetry in the Cheetah (Acinonyx-Jubatus), a Genetically Uniform Species." Evolution **40**(1): 78-85.
Weber, D. S., B. S. Stewart, J. C. Garza and N. Lehman (2000). "An empirical genetic assessment of the severity of the northern elephant seal population bottleneck." Current Biology **10**(20): 1287-1290.
Weber, J. L. and C. Wong (1993). "Mutation Of Human Short Tandem Repeats." Human Molecular Genetics **2**(8): 1123-1128.
Weir, B.S. and C.C. Cockerham (1973). "Mixed self and random mating at two loci." Genetic Research **21**: 247-262.
Whitehouse, A. M. and E. H. Harley (2001). "Post-bottleneck genetic diversity of elephant populations in South Africa, revealed using microsatellite analysis." Molecular Ecology **10**(9): 2139-2149.
Whitlock, M. (1993). "Lack Of Correlation Between Heterozygosity And Fitness In Forked Fungus Beetles." Heredity **70**: 574-581.
Wilson, G.A., C. Strobeck, L. Wu and J.W. Coffin (1997). "Characterization of microsatellite loci in caribou Rangifer tarandus, and their use in other artiodactyls." Molecular Ecology **6**(7): 697-699.
Wilson, P. J. and B. N. White (1998). "Sex identification of elk (Cervus elaphus canadensis), moose (Alces alces), and white-tailed deer (Odocoileus virginianus) using the polymerase chain reaction." Journal Of Forensic Sciences **43**(3): 477-482.
Wooten, M. C., Smith, M. H. (1986). "Fluctuating Asymmetry and Genetic-Variability in a Natural- Population of Mus-Musculus." Journal of Mammalogy **67**(4): 725-732.
Wright, S. (1951). "The genetical structure of populations." Annals of Eugenics **15**: 323-354.
Wright, Sewall (1921). "SYSTEMS OF MATING." Genetics **6**(2): 111-178.