Relatedness Within and Among Sperm Whale (*Physeter macrocephalus*) Social

Units Assessed Using Single Nucleotide Polymorphisms

By
Wilfried Beslin

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
In Partial Fulfillment of the Requirements for
The Degree of Bachelor of Science,
Major and Honours Certificate in Biology.

April, 2013, Halifax, Nova Scotia

Approved:	Dr. Timothy Frasier	
	Supervisor	
Approved:	Dr. Jeremy Lundholm	
	Reader	
Approved:	Kate Crosby	
	Reader	

Date: April 2013

Table of Contents

List of Figures and Tables

Table 1: SNP loci characterized in *Physeter macrocephalus*. Corresponding forward and reverse primer sequences tested in this study are shown, along with the optimal annealing temperatures found for them. Bold loci represent linked SNPs.

Figure 1: Electropherograms of sperm whale DNA sequences at a SNP locus (DRD2Y679) for three individuals. The SNP is located at the 9th position from the left. The two alleles for this locus are C or T. The individual in a) is a C homozygote, the individual in b) is a T homozygote, and the individual in c) is a heterozygote.

Table 2: Sequences of 20 PCR amplification fragments containing sperm whale SNPs.

Figure 2: 2% agarose gel stained with SYBR Green I containing raw DNA from 18 sloughed skin samples. Note the faint presence of DNA and extensive smearing in most samples, indicative of a high level of degradation.
(Wells 1 and 11 contain a low mass ladder for comparing the amount and size of DNA. Wells 2, 3, and 20 contain positive controls, while well 19 contains a negative control. All other wells contain DNA from the samples of interest).

Figure 3: 2% agarose gel stained with EtBr containing PCR product from two samples (one of high quality and the other severely degraded) amplified with four SNP primer pairs using three annealing temperatures. Product in the top left was obtained with an annealing temperature of 48°C, the top right used 53°C, and the bottom left used 58°C. Note the high yield of amplified DNA obtained from all primers for both samples.

Table 3: Expected sequences of PCR amplification fragments used for genotyping using Sanger sequencing.

Table 4: Genotyping success of SNPs contained within PCR fragments sequenced on an ABI 3500xl.

Table 5: The final 8 loci used for relatedness network analysis.

Figure 4: Observed network of genetic relatedness among 72 sperm whales off the coast of Dominica based on pairwise relatedness from SNPs. Network analysis has detected three clusters of related individuals, which were found to be statistically insignificant (see text, **Figure 5**).

Figure 5: Histogram of the modularity distribution of 1000 simulated networks based on observed allele frequencies in 72 individuals. The modularity of the observed network is indicated by the red line, and is not significantly different from the simulated results.

Figure 6: Statistical power estimates of SNP and microsatellite markers for detecting three clusters of related individuals in a population with low overall genetic variation. The eight SNP loci include 12 SNPs in total, eight of which are linked. The 24 SNP loci include the original 8 multiplied three times (36 SNPs, total, 24 linked). The 11 microsatellite loci were those used by Arseneault (2011).

Abstract

Relatedness Within and Among Sperm Whale (*Physeter macrocephalus*) Social Units Assessed Using Single Nucleotide Polymorphisms

By Wilfried Beslin

Altruistic behavior, such as caring for the offspring of others, is an intriguing phenomenon observed in several mammalian species. One of the most successful theories for explaining how this behavior could have evolved is kin selection. According to this theory, altruism can be adaptive if the individuals being helped are related to the helper. Female sperm whales and their calves live in stable matrilineal families called "units" within which there is extensive altruism, notably in the form of alloparental care. Furthermore, two or more units will often temporarily associate into larger "groups" and hunt for prey together. Because units are largely composed of related individuals, it is likely that this unusual social system evolved by kin selection. To investigate this, the detailed relatedness patterns within and among units must be known. Sperm whales naturally shed pieces of skin that can be used to extract DNA for this purpose. However, the DNA within these samples is often heavily degraded. Microsatellites are informative DNA markers that are commonly used for such genetic analyses. However, microsatellite analysis is difficult with degraded DNA, because they tend to be relatively long sequences. Single nucleotide polymorphisms (SNPs) are another, increasingly popular DNA marker that are only one nucleotide in length, so they can be more easily analyzed in degraded DNA. The purpose of this study was to use SNPs to investigate kin selection in sperm whales from highly degraded DNA in sloughed skin. A total of 12 SNPs, eight of which were linked, were genotyped in 72 individuals, and a network analysis was performed. Results were unable to resolve any genetic structuring, owing to the poor informativeness of individual SNPs. It was concluded that microsatellites are still more useful for studying relatedness, and that biopsy sampling of whales for higher quality DNA may be required.

April 17th, 2013

Acknowledgements

INTRODUCTION

Fitness is an important concept in evolutionary biology. Though many interpretations exist, an individual's fitness can basically be defined as its ability to survive and reproduce in its environment (Orr 2009). This ability is not equal among individuals, and those that are more fit will produce a greater number of offspring than those who are less fit. The ability to survive and reproduce is granted by an individual's traits, which may be physical or behavioural. If these traits are heritable, then it is expected that future generations will contain a greater proportion of traits from those individuals who are most fit. This is the central principle of the theory of natural selection, which was originally described in Darwin's work *On The Origin of Species* (1859) and became formalized by Fisher in 1930. Thus natural selection explains how populations change, or evolve, with time such that their constituents become better adapted to their environments.

Caring for young is a behavioural trait that has evolved in many animal species, and it is easy to see why. By helping offspring to survive, the offspring have better opportunities for reproduction themselves. This means increased fitness for the offspring and ultimately the parent, making the parent's caring behaviour an adaptive trait by natural selection. However, most animals do not care for their young (Alcock 1998). The reason for this is that helping progeny to survive is a costly investment. It requires placing large amounts of time, energy, and other resources into offspring growth which could otherwise go towards the parent or into the production of other offspring. It may also expose the parent directly to danger if the young need to be defended (Alcock 1998). Thus, while parental care increases fitness by improving reproductive success, it also reduces it by lowering the parent's chance of survival. For most animals, the costs associated with parental care probably outweigh the benefits, so they do not show this behaviour (Alcock 1998).

Where an animal's environment makes reproductive success particularly difficult without parental care, it is in the interest of parents to help their young from a fitness perspective. However, there also exists species in which some individuals provide care for young they have not produced themselves. This is known as alloparental care (Woodroffe & Vincent 1994; Whitehead 1996; Gero *et al*. 2009). At first glance, this appears to be a form of altruism: behaviour in which an individual increases the fitness of others at the expense of its own. Since natural selection theory implies that traits of the fittest individuals propagate most into future generations, then altruism should not be adaptive. Because an individual displaying altruism reduces its own direct fitness, this trait should not be sustainable within populations.

The evolution of altruism has become a popular problem in sociobiology, and a number of theories have been proposed to explain why it is observed despite it being apparently counter-adaptive. The most common is kin selection, as described by Hamilton in 1964. In this scenario, altruism can be adaptive if the helpful acts are directed towards relatives. This is based on the concept of inclusive fitness introduced by Hamilton, which subdivides fitness into direct and indirect components. The direct fitness of an individual is its fitness achieved from directly producing surviving offspring, while its indirect fitness is that of offspring produced by relatives which survived because of the help it provided (Hamilton 1964; Alcock 1998). From a genetic perspective, an individual shares a certain proportion of genes with its relatives in addition to offspring. Therefore, by helping these relatives, it can increase its indirect fitness by helping those genes to survive in future generations. Altruistic behaviour can then evolve such that it maximizes inclusive fitness. Since direct fitness carries more weight than indirect fitness (direct offspring share more genes with the parent than do the offspring of its relatives), altruism based

on kin selection should only evolve if personal reproduction is difficult, and helping others is more beneficial than attempting to reproduce.

It should be noted that under a kin selection scenario, because organisms do obtain a benefit from helping in the form of inclusive fitness increase, the act is not "truly" altruistic. Many studies define altruism such that there is no or very little benefit whatsoever. However, for discussion in this study, "altruism" will be considered to be any helpful act that results in a direct fitness loss, regardless of whether or not indirect fitness is affected.

Kin selection has been very successful in explaining the evolution of helping behaviour in many species. However, altruism among non-kin is known to exist as well. This cannot be explained by kin selection, because there are no indirect fitness gains to raise inclusive fitness. Another theory, often seen as an alternative to kin selection for this scenario, is reciprocal altruism (Trivers 1971). In this model, an individual who conducts an altruistic act towards another individual is eventually repaid by that individual. In terms of fitness, this means one individual takes an immediate fitness loss to increase that of another individual, but this other individual reciprocates the altruism at a later time. For reciprocal altruism to work, the magnitude of the fitness gain achieved from receiving help must be greater than the cost of helping. In this way, both individuals can enjoy a net fitness benefit (Trivers 1971). Ultimately this may be seen as a type of mutualism, which differs from altruism in that two individuals help each other to increase their collective fitness. What makes reciprocal altruism different from true mutualism, however, is that there is a delay in the fitness gain of one individual, temporarily filled with a small fitness loss (Alcock 1998). Because of this delay, it would seem that the system is vulnerable to "cheating", where the individual that was helped first does not return the favour. Natural selection should intuitively favour such a cheater, because it receives a fitness

gain without incurring a loss. However, Trivers (1971) has argued that cheating is not an adaptive strategy, because if individuals reciprocate altruistic acts continually in the long term, they will gain a greater net fitness than cheaters, who will be left out of future opportunities to receive help. Thus reciprocal altruism can be an evolutionarily stable system.

A third model to explain the evolution of altruism is group selection, which was described in detail by Wynne-Edwards in 1962. According to this model, individuals belonging to a "group" (which can be anything from a small population to an entire species; Alcock 1998) may act altruistically towards other group members, kin or non-kin, because these acts are beneficial to the group as a whole. Groups with higher collective fitness outcompete other groups, and therefore they are the ones to propagate (Wynne-Edwards 1962). This used to be the dominant explanation for the evolution of altruism, until it was challenged by Williams in 1966 (Alcock 1998). Williams (1966) basically showed that natural selection at the level of the individual is a much more powerful force than at the level of the group. This is because individuals who reduce their fitness for the benefit of the group would reproduce less than those who act selfishly, and so the selfish traits would propagate more than the altruistic ones. Since then, altruism has largely been assessed through kin selection or reciprocal altruism rather than group selection.

Though group selection has fallen out of favour, it deserves to be mentioned here. It is important to note at this point that the reasoning behind explaining the evolution of altruism is dynamic, and this aspect of sociobiology is currently being debated. Kin selection and reciprocal altruism, as they are presented here, have been the standard ways of thinking since the rejection of group selection after the 1960s; however, recent research has begun to question this foundation (e.g. Wilson & Wilson 2007; Nowak *et al*. 2010). Specifically, this revision calls for

a reconsideration of the significance of group selection, and possibly unification of the various theories that make up sociobiology. This revision was proposed for numerous reasons, including claims that the alternatives to group selection are oversimplified and even dependant on the concepts of group selection, that more sophisticated simulations since the 1960s have been able to support group selection, and that there exists empirical evidence for it in organisms such as microbes and eusocial animals (Wilson & Wilson 2007; Nowak *et al.* 2010). These arguments have been met with considerable criticism in defence of kin selection (e.g. Clutton-Brock 2009; Abbot *et al.* 2011; Boomsma *et al.* 2011; Strassmann *et al.* 2011; Ferriere & Michod 2011; Herre & Wcislo 2011), and there is an ongoing debate on how to correctly interpret the evolution of helpful behaviour. This means conclusions derived from sociobiology should ideally be made and examined critically. With that said, the success of the traditional theories in explaining observed phenomena cannot be denied. Therefore, this thesis will continue to consider kin selection and reciprocal altruism as viable explanations for the evolution of altruism.

Altruism, such as alloparental care, is a social behaviour that obviously involves interactions between individuals. This often requires individuals to be in close proximity to one another. When these interactions occur over a long term, this means that individuals live in groups. Hinde (1976) has argued that groups are held together by a social structure built upon relationships between individuals. These relationships arise from repeated interactions between the individuals (Hinde 1976). This means that social interactions and group living are closely dependant on one another (this is also evident from the logic of group selection). Thus, understanding the evolution of social behaviours like altruism also means understanding the evolution of groups (Alexander 1974). This is significant because, like providing care to offspring, living in groups itself carries both costs and benefits to fitness. For example, an

individual living next to others may have lower chances of being preyed upon by predators (because of a dilution effect), but it may also be more likely to contract a contagious disease (Alexander 1974). The various interactions that occur in a group can also have advantageous or disadvantageous consequences to different individuals. This includes altruism, but there are others, including mutualism, where individuals cooperate and receive a mutual fitness gain; selfish behaviour, in which one individual increases its fitness at a cost to others; and spiteful behaviour, where an individual's actions reduce the fitness of both itself and others (Alcock 1998). There are also by-product interactions, such as when one individual benefits from the selfish actions of another (Sachs *et al*. 2004). From a Darwinian perspective, the evolution of each of these behaviours should be explainable through some ultimate selfish gain in fitness, as was shown here for altruism. This should then provide insight into the evolution of group formation itself. To summarize, exhibiting social behaviour means living in groups, and determining the evolutionary forces behind the development of social behaviour requires analysis of the complex interactions between costs and benefits that come with group living.

Groups exist in many different forms, varying in such things as population size, spatial distribution, longevity, relatedness among individuals, and types of interactions involved. The types of groups that form in a species are expected to be a result of the environmental and biological challenges it faces. Alexander (1974) has argued that the selective pressures leading to group evolution fall into one of three general categories: increased defense from predators, increased foraging efficiency, or resource limitations. These pressures would favour different types of behaviour; For instance, in groups that evolve mainly because of predator defence or foraging benefits, individuals would be expected to remain close together, and there may also be extensive cooperation. In contrast, groups that form mainly because of convergence to a limited

resource patch would have avoidance among individuals, and aggressive behaviour may be rampant (Alexander 1974).

Families can be seen as special types of groups that consist of related individuals. By definition, families arise when offspring delay dispersal and continue to live in proximity of their parents into adulthood (Emlen 1994). Two principal models to explain why this happens are the benefits of philopatry model, and the ecological constraints model (although it has been argued that the two are actually interdependent; Emlen 1994). The benefits of philopatry model focuses on the advantages of living in a family. This includes enhanced survival from group living, indirect fitness gains from helping family members, and better opportunities to secure high-quality breeding sites including the natal one itself (Emlen 1994). In contrast, the ecological constraints model stresses the detriments of dispersing. These include a lack of quality breeding resources (territories or mates), and low chance of reproductive success by inexperienced breeders (Emlen 1994). Like any group, families may benefit from containing multiple individuals, but also suffer automatic costs. However, these costs may be alleviated somewhat because it is possible to increase one's inclusive fitness when in a family. Therefore, families may be more stable than groups of unrelated individuals, and extensive altruism is likely to be observed within them (Emlen 1995).

Many mammals are known for being highly social, particularly the primates. Cetaceans are another order of mammals that consists of the whales, dolphins, and porpoises. Cetaceans also have complex social systems comparable to primates, but are not as well studied due to the difficulties in observing them (Mann *et al*. 2000). They are distinguished from other mammals in that they live entirely in water, and studies conducted on their behaviour are usually limited to the brief moments that they surface, such as when taking breaths (Mann *et al*. 2000). Their fully

aquatic lifestyle means they have evolved special morphological characters to exploit this environment, obvious to anyone who has seen a cetacean; however, it also has less obvious implications for their behaviour. Aquatic environments vary from terrestrial ones in several ways, such as in the lack of refuges from predators, in the efficiency of locomotion possible in the dense but buoyant medium, and in generally widespread and mobile prey (Mann *et al*. 2000; Connor 2000). This means that cetaceans are particularly vulnerable to predator attacks, the cost of locomotion is lower, and hunting may be less effective individually. These are clearly conditions that would seem to favour group formation: by associating into groups, cetaceans would be better able to defend themselves against predators and forage more efficiently, all while the costs of resource competition between them are alleviated by ease of movement. These factors are thought to be the primary forces behind the evolution of sociality in cetaceans (Connor 2000).

The sperm whale (*Physeter macrocephalus*) is a cetacean with highly complex social behaviour, which seems to include forms of alloparental care (Whitehead 1996). The evolution of behaviour in this species is particularly interesting, in part because of its extreme differences from other cetaceans. Morphologically, the sperm whale is the largest of the toothed whales, and it contains the largest brain of any animal. Its appearance is distinguished by a large, somewhat rectangular head which overhangs a long narrow jaw (Whitehead 2003). Within the head is a complex series of structures which largely includes the spermaceti organ: an oil-filled tissue mass unique to sperm whales. The primary advantage of the spermaceti organ is thought to be in the production of acoustic clicks for echolocation and communication (Whitehead 2003; Cranford 1999).

The sperm whale also has extreme morphological differences between the sexes, the greatest of any cetacean (Whitehead & Weilgart 2000). Adult males grow to be on average three times more massive and one-and-a-half times longer than females, and their spermaceti organs are proportionally larger as well (Whitehead 2003). Such sexual dimorphism is usually characteristic of uneven intrasexual competition between the sexes, presumably because of different selection pressures on male and female reproductive strategies (Alcock 1998). In addition to morphological differences, sperm whale males and females have great behavioural differences as well. Though they are distributed throughout the oceans worldwide as a species, sperm whales have a unique population structure in that the sexes are segregated. Females and their calves live in groups around equatorial waters (Whitehead & Weilgart 2000). Males however disperse from their natal groups after they are no longer dependent on milk, which seems to occurs at about six years of age (Richard *et al*. 1996). Soon after their dispersal, males travel to higher latitudes and usually join with other similar-aged males in loosely aggregated groups called "bachelor schools" (Best 1979). As these males age and grow larger, they generally continue to move to higher latitudes, the size of bachelor schools decreases, and ultimately the largest males may live solitarily near the poles (Best 1979; Rice 1989). This difference in sociality and preferred ecological niche between the sexes further suggests that males and females are under different selection pressures.

To better understand how the unique characteristics of the sperm whale could have evolved requires knowledge of the life history and ecology of the species. Sperm whales are long-lived (usually surviving to sixty or seventy years), and their fecundity and mortality rates are low (Rice 1989). Females reach sexual maturity between seven and thirteen years of age, after which they usually produce a single calf about every four to five years until they are about

forty years old (Rice 1989; Best *et al*. 1984; Whitehead & Weilgart 2000; Whitehead 2003). After forty years of age, the calving interval may drop to one calf every fifteen years (Whitehead 2003). Sex ratios are close to 1:1 at birth (Best *et al*. 1984). Gestation lasts for about fifteen months, and once a calf is born, it is dependent on lactating females for milk. Calves begin to eat solid food sometime before they are one year of age, but there is evidence that they continue to consume milk long after this; possibly up to eight years or longer (Best *et al*. 1984).

Sperm whales are deep-water feeders that usually dive to depths of 400-1200 m to catch their food (Watwood *et al*. 2006; Watkins *et al*. 2002; Papastavrou *et al.* 1989). Their diet is wide and variable, and seems to be dependent on the prey available in their geographic location and time of year. They are known to consume cephalopods (e.g. squid), fish, and crustaceans, but the vast majority of their prey are certain species of squid (Kawakami 1980; Rice 1989). Though sperm whales have large energetic requirements in general, their prey are often small and of poor nutritional value (Whitehead & Weilgart 2000). This means they need to consume a considerable amount of food. In the case of young calves, diving to catch food appears difficult, and they tend to eat smaller prey once they become proficient enough to hunt (Best *et al*. 1984). However, their metabolic demands are also significantly greater than that of adults (Best *et al*. 1984), possibly because of the costly development of the spermaceti organ (Lyrholm *et al*. 1999). Young whales may then need to rely on milk even after learning to catch prey to satisfy their energetic requirements. This would explain why lactation is long-lasting in sperm whales (Best *et al*. 1984).

There are several factors that could affect why male sperm whales are socially and morphologically different from females, based on sexual selection. From the time that a female becomes pregnant until her calf is born and then weaned, she will be unavailable to mate. In

addition, since sex ratios are equal and females reproduce once every 4-6 years, then it can be approximated that there are about 4-6 males for every female in a breeding season (Whitehead 2003). This means competition between males for access to mates should be relatively high, which could explain why they are less social than females. There is evidence that breeding males actively avoid each another, and occasionally fight (Whitehead 1993). According to sexual selection theory, females choose their mates based on traits they like, which are often representative of genetic quality (Alcock 1998). In sperm whales, a trait males may use to advertise to females is the emission of loud clicks. Acoustic signals are thought to be an important mode of communication in sperm whales (Whitehead & Weilgart 2000), and there is evidence that mature males emit a unique type of click especially when on breeding grounds (Weilgart & Whitehead 1988). Since the spermaceti organ plays an important role in sound production, its large size in males may have evolved from female choice (Cranford 1999). However, the development and maintenance of a large spermaceti organ must also require a tremendous amount of energy (Lyrholm *et al*. 1999; Cranford 1999). This could also explain why males prefer higher latitudes, as there is evidence that food can be larger and of better quality in these regions (Teloni *et al*. 2008).

The behaviour of females can also be explained by examining potential selective pressures. Unlike males, females and immature whales rarely travel to waters less than 15°C (Rice 1989; Lyrholm *et al*. 1999; Whitehead 2003). If productivity is greater at high latitudes, why should this be the case? One reason that has been proposed is that thermoregulatory limitations on females and/or calves prevents them from living at extreme latitudes (Lyrholm *et al*. 1999). Another hypothesis is that prey occur too deeply at high latitudes for young whales to dive and catch them (Best 1979; Lyrholm *et al*. 1999). It is also quite possible that predation has

some impact. For sperm whales and almost all other cetaceans, killer whales are a significant predator (Pitman *et al*. 2001; Connor 2000; Jefferson *et al*. 1991). These animals are known to attack other cetaceans of all sizes, but their focus and most successful attacks seem to be on juveniles (Connor 2000). Though killer whales are found virtually everywhere, they are more common in higher latitudes (Connor 2000), which could explain why female sperm whales and their calves prefer to live in tropical latitudes. Support for this hypothesis may come from the seasonal migration of baleen whales to the tropics for breeding (Corkeron & Connor 1999; but see Clapham 2001).

Females are more social than males, and normally live in various levels of association. In order of increasing size, these levels are termed "units", "groups", "aggregations" and "super-aggregations", and "concentrations" (Whitehead & Weilgart 2000). The largest associations, which contain about 40 animals or more in the case of aggregations and super aggregations, and 1000 animals in the case of concentrations, are spread over several kilometers and exist only temporarily. These associations are correlated with the presence of prey patches, which indicates they probably represent mere convergence to abundant food sources (Whitehead & Weilgart 2000). Groups, which contain roughly 20 animals, are more stable and may represent cooperative foraging (Whitehead *et al*. 1991). However, units, which typically contain around 10 animals, are extremely stable and show indications of communal care for calves (Whitehead & Weilgart 2000; Whitehead *et al*. 1991; Whitehead 1996). These units often (but not always) appear to consist of matrilineally related individuals (Whitehead *et al*. 1991; Richard *et al*. 1996; Christal *et al*. 1998), and are considered by Whitehead and Weilgart (2000) to be fundamental to sperm whale societies.

The alloparental care that occurs in units seems to serve two basic roles: protection of calves from predators, and assistance in meeting calves' high nutritional requirements. Whitehead (1996) has shown that foraging dive patterns are different between groups with calves and no calves, such that calves are rarely left alone at the surface. Since young calves are unable to dive deep, and mothers must still devote a considerable amount of time to feeding, then the diving coordination of other adults can reduce isolation of calves while still allowing mothers to meet their own nutritional requirements. This babysitting system may be important for protecting calves from predators like the killer whale (Whitehead 1996), and it is present in several other mammalian species (Reidman 1982). In events where a sperm whale unit does encounter predators, adult members have been observed to actively defend any calves present. When attacked, sperm whales will typically band close together in a circular formation, sometimes with their powerful tails pointing radially outwards in a structure resembling a "rosette" or "marguerite" (Arnbom *et al*. 1987; Pitman *et al*. 2001). Any calves in the unit remain in the center of these formations where it is most safe (Arnbom *et al*. 1987), and should they still be attacked, the adults have been observed to come to their aid (Caldwell & Caldwell 1966). In addition to providing protection from predators, the adult females in a sperm whale unit may also communally assist calves in meeting their nutritional requirements through allonursing. In some units, more lactating females than calves have been found (Best *et al*. 1984), and there is evidence that calves attempt to suckle from multiple females (Gero *et al*. 2009).

Given their vulnerability and high energetic requirements, the alloparental care of calves was likely important in the evolution of social sperm whale units. However, as discussed earlier, alloparental care is a type of altruistic behaviour that is immediately costly for the individuals displaying it, and is therefore expected to have some benefit to these individuals in order to be

adaptive. Among sperm whales, it is possible for these benefits to arise through reciprocity. By raising another's calf, an adult female could engage in a reciprocal relationship with the parent and have the favour returned once she becomes a mother herself. However, since sperm whale units often appear to consist of matrilineal families, then it is more likely that their alloparental care system evolved primarily through kin selection (Whitehead 1996). If this is the case, then one would expect that the individuals who provide the most care for a calf are also the ones who are most closely related to the mother. A study by Gero *et al*. (2008) following a unit of seven whales showed that individuals within the unit had preferred associations with certain unit members, and that these associations appeared to reflect kinship. This suggests that kin selection may indeed be acting on this scale.

Kin selection could also be acting at the group level. The larger groups of about 20 animals are thought to consist of typically two matrilineal units coming together (Whitehead *et al*. 1991). The associations of different units into groups are fairly stable and must therefore be beneficial, otherwise they would likely not be seen. The main benefit of these groups, as stated earlier, is thought to be increased foraging efficiency (Whitehead *et al*. 1991). Therefore it is possible that kin selection is acting here too, such that matrilineal units with close relatedness associate into groups to increase their collective inclusive fitness. There is some evidence to support this hypothesis, as Richard *et al*. (1996) have found some groups that contained related individuals from different matrilines. In contrast, a recent study by Ortega-Ortiz *et al*. (2012) found no significant correlation between preferred associations and kinship among units. Clearly the details behind what selective force is acting at this level have yet to be fully understood.

To determine the relatedness between individuals requires the use of genetic analysis. This technique works because of the fact that there is variation in the genetic code among

different organisms, arising from mutations that occur over long stretches of time (Sunstad & Simmons 2012). Since organisms inherit portions of their genetic code from their parents, it is expected that individuals who share a recent common ancestor (and are therefore closely related) will have a similar code. Genetic information can be obtained from macromolecules that exist within cells, specifically proteins and nucleic acids. Protein variants, or allozymes, were discovered in the 1960's and were the first molecular markers used (DeWoody 2005; Avise 1994). Though still useful for certain studies, allozymes have largely been replaced by deoxyribonucleic acid (DNA) markers, because they can be difficult to obtain (usually requiring to be extracted from fresh tissues), and they have a low level of variation (DeWoody 2005). Most studies assessing genetic variation in animals now rely on two types of DNA: mitochondrial DNA (mtDNA), and nuclear (or genomic) DNA (DeWoody 2005), with nuclear DNA being the most useful for relatedness analyses.

Nuclear DNA is divided into several linear chromosomes in most eukaryotic species, and it is inherited from both parents in those that reproduce sexually (Sunstad & Simmons 2012). Each individual possesses two versions of each chromosome: one from its father and another from its mother. During sexual reproduction, the two homologous parental chromosomes also recombine to produce unique pairs in the offspring (Sunstad & Simmons 2012). This bi-parental mode of inheritance makes nuclear DNA particularly useful for studies of relatedness, parentage, and individual identification (Sunnucks 2000).

Microsatellites are perhaps the most common type of nuclear DNA marker used in population genetics (Freeland 2005). Also known as short tandem repeats or STRs, microsatellites are repeating sequences of 1-6 base pairs that mostly occur in non-coding regions of the nuclear genome (Ellegren 2004), which means they are generally not influenced by

selective pressures. Microsatellites mutate rapidly and tend to be extremely variable, with some loci exhibiting over 50 different alleles in animals (Ellegren 2004; DeWoody 2005). Polymorphism in microsatellites does not occur as differences in the nucleotide sequences, but rather as variations in the number of tandem repeats (Ellegren 2004). It is not entirely clear how microsatellites mutate, but the mechanism generally invoked is that of replication slippage: a case in which one DNA strand dissociates and rehybridizes out of alignment with the other during replication, causing a lengthening or shortening of the overall sequence (Ellegren 2004). The fact that microsatellites are highly variable, mutate quickly, occur at multiple loci within the genome, and reside in non-coding regions makes them very informative markers for relatedness and parentage analysis (Freeland 2005; DeWoody 2005).

Single nucleotide polymorphisms (SNPs) are another type of nuclear DNA marker that are starting to increase in popularity. A SNP is simply a single base pair within the genome that is variable between different individuals (Freeland 2005). SNPs may occur in both coding and non-coding regions, and they are much more common throughout the genome than microsatellites in many species (Morin *et al*. 2004). In humans, it was found that about 90% of genomic variation occurs in the form of SNPs (Collins *et al*. 1998). The rate of mutation for SNPs is relatively low: on the order of 10^{8}-10^{9} mutations per generation, compared to 10^{4} for microsatellites (Brumfield *et al*. 2003). As a result, variation per SNP locus is also low. Theoretically, SNPs can have a maximum of four alleles per locus (corresponding to the four different nucleotide bases in DNA: A, C, G or T). However, their low mutation rate makes this unlikely, and so they are typically bi-allelic (Brumfield *et al*. 2003). Because their allelic diversity is lower than that of microsatellites, individual SNPs are not as informative. However, since SNPs are also much more abundant in the genome than microsatellites, they can potentially

provide even more information when large numbers of them are analyzed (Brumfield *et al*. 2003; Giordano *et al*. 1999). SNPs have become very popular in human genetics studies, and are increasingly being used for ecological and evolutionary work on wild populations (Vignal *et al*. 2002; Brumfield *et al*. 2003; Aitken *et al*. 2004). Some researchers even advocate that SNPs may become a "standard" molecular marker for population genetics in the future, potentially replacing microsatellites as the most used marker (e.g. Morin *et al*. 2004; Brumfield *et al*. 2003).

There are three ways in which tissue samples can be obtained for extraction of DNA: destructive sampling, which involves the killing of the specimen to obtain fresh tissue; nondestructive sampling, which involves the removal of fresh tissue from a living specimen; and noninvasive sampling, which involves the collection of samples left behind by the specimen such as feces, hair, and sloughed skin (Taberlet *et al*. 1999). For obvious reasons, destructive sampling is no longer common practice. Nondestructive and noninvasive sampling are the two most popular methods of sampling, with noninvasive sampling often seen as superior to nondestructive sampling because animals do not need to be captured or disturbed (Taberlet *et al*. 1999). However, noninvasive sampling often returns little or poor quality tissue with many dead cells (Taberlet *et al*. 1999). After cells die, the DNA within them is subject to a number of chemical processes that break it down over time (Deagle *et al*. 2006; Lindahl 1993). This results in fragmentation of DNA strands into smaller pieces.

The analysis of nuclear DNA markers such as microsatellites usually involves the amplification of the DNA sequences at the marker loci through the polymerase chain reaction (PCR). For PCR to be successful, the template sequences of interest must be intact. Therefore, applied to PCR, DNA fragmentation can reduce the number of successful reactions because the sequences being amplified can be discontinuous (Goldenberg *et al*. 1996). Sequences that are

longer have a greater chance of being fragmented, and therefore have a lower successful amplification rate than smaller sequences. Microsatellites can have target sequences as long 300-400 bp long, which is enough to be problematic if DNA is heavily degraded (Butler 2005). This means that microsatellites may not be a very reliable marker in this case. SNPs on the other hand are just one base pair in length, so their target fragments do not have to be long; typically between 50-150 bp. Therefore, SNPs have a higher chance of amplifying successfully in PCR even for highly degraded DNA samples (Morin & McCarthy 2007).

Sperm whales and other cetaceans can lose thin pieces of skin naturally as they swim and rub against one another. This sloughed skin provides a source of tissue for DNA extraction that can be noninvasively sampled, and is popular in genetic studies of cetaceans (Amos *et al*. 1992). However, while easy to collect in the field, sloughed skin typically contains heavily degraded DNA that does not amplify well for large sequences, including microsatellites (Taberlet *et al*. 1999). Previous work with assessing relatedness in sperm whale units from microsatellites and sloughed skin suffered from low resolution because of this (Arseneault 2011).

The goal of this study is to reveal the relatedness patterns within and among female sperm whale units found off the coast of Dominica at high resolution, using single nucleotide polymorphisms to obtain reliable information from degraded DNA contained in sloughed skin. These relatedness data can also be complemented with photo identification and unit assignment of individuals, and observations of unit grouping and alloparental behaviour. By combining these data, it should be possible to unravel the details behind kin selection in sperm whale units.

It should be noted that this study does not attempt to quantify fitness and determine whether or not selection is acting. Previous studies have already suggested that sperm whale social behaviour is under the influence of selective pressures, most notably kin selection

(Whitehead 1996; Gero *et al.* 2008; Gero *et al.* 2009). The purpose of this study is to investigate patterns of relatedness within and among units, and see how this compares with the heterogeneous helping and grouping exhibited by these whales. Where fitness becomes relevant is in the potential impact of these behaviours on survival and reproduction: helping and group living are known to incur automatic costs, so the fact that they are observed means that they must have offsetting benefits which are under positive selection pressures. Therefore, if preferential helping and group association in sperm whales are found to coincide with kinship, then it would likely imply that kin selection is the main force driving the evolution of these behaviours. Alternatively, if there is no significant correlation between kinship and helping, or kinship and grouping, then the behaviours are likely driven by another selective force, probably reciprocity or other types of mutualism.

I hypothesize that sperm whales can assess how related they are to other individuals, and offer more help based on the strength of this genetic relationship on both group and unit scales. From this hypothesis, I predict groups will consist of matrilineal units which are closely related to one another. Within units, I predict that closely related individuals will provide more help to each other or their offspring than they would for others. I also expect that the use of SNPs should produce more informative results than microsatellites given the degraded state of DNA in the sloughed skin samples available.

MATERIALS AND METHODS

Sample Collection and Field Data

Sloughed skin samples were collected from sperm whales off the coast of Dominica between 2005 and 2011. Because of the long lifespan, low reproductive rate, and site fidelity of female sperm whales, the population is not likely to have changed significantly over this timescale, and samples from all years can be analyzed together. In addition, the whales living in this area have been observed extensively: individuals can be identified, it is known who belongs to which unit, and there is considerable information on who interacts with whom. Upon collection, skin samples were assigned a field code and stored either in ethanol or a 20% DMSO solution saturated with NaCl (Seutin *et al.* 1991). Whenever possible, individual whales from which skin was collected were photographed and identified. This permitted the use of the observational data to determine which units the whales in this study belonged to, how much each individual helped other unit members, and which units associated into groups.

DNA Extraction

DNA was extracted from a total of 128 skin samples using a standard phenol:chloroform protocol (Sambrook & Russell 2001). A maximum of 40mg of tissue was taken from each sample for extraction. The tissue was mixed with 100 µL of lysis buffer (10 mM Tris-base, 10 mM ethylenediaminetetraacetic acid (EDTA), 2% sodium dodecyl sulfate (SDS), 0.1 M sodium chloride (NaCl), and 40 mM dithiothreitol (DTT); Budowle *et al.* 2000), frozen using liquid nitrogen, and ground to a fine powder with a mortar and pestle. The powder was transferred to a 1.5 mL Eppendorf tube and mixed with an additional 300 µL of lysis buffer. The tubes were incubated at room temperature for five to six days with periodic mixing, after which an aliquot of

proteinase-K at a concentration of 0.5 U/mg of tissue was added. Approximately 24 hours later, proteinase-K was added again, the tube was incubated in a 65°C water bath for 60 minutes, and then at room temperature for another 60 minutes. This was immediately followed by another addition of proteinase-K that incubated overnight at room temperature, for a total of three 0.5 U/mg proteinase-K aliquots.

The samples were then transferred to a fume hood for phenol:chloroform extraction. An equal volume (400 µL) aliquot of phenol:chloroform solution was added to each sample. The tubes were then upended by hand for five minutes, and centrifuged for four minutes at 12 000 x g. Centrifugation resulted in the separation of organic and aqueous layers. The aqueous layer, expected to contain DNA, was removed and transferred to a new 1.5 mL Eppendorf tube. A second round of 400 µL of phenol:chloroform was added to the samples, and the tubes were again upended for five minutes followed by centrifugation for four minutes at 12 000 x g. The resulting aqueous layer was transferred to a new 1.5 mL Eppendorf tube, to which 400 µL of chloroform was added. The samples were further upended for five minutes and centrifuged for four minutes at 12 000 x g. The resulting aqueous layer was again transferred to a new 1.5 mL Eppendorf tube.

The DNA obtained was then precipitated using ethanol. Briefly, ammonium acetate was added to each tube to a concentration of 2 M, followed by 800 µL of ice-cold 95% ethanol. The samples were then stored at -20°C for approximately 24 hours to allow the DNA to precipitate.

Following the -20°C storage, the samples were centrifuged for ten minutes at 12 000 x g. This resulted in the formation of a pellet at the bottom of each tube expected to contain DNA. The ethanol was decanted, with excess carefully removed using a Kimwipe such that the pellet was not disturbed. A 100 µL aliquot of 70% ethanol was gently added to each tube, and the

samples were centrifuged again for ten minutes at 12 000 x g. The ethanol was decanted again, with excess removed using a Kimwipe while avoiding the pellet. The tubes were then left at room temperature with lids open for 15 to 30 minutes to allow any remaining ethanol to evaporate. The pellets were dissolved in 150 µL or 200 µL of $TE_{0.1}$ (10 mM Tris, 0.1 mM EDTA) depending on their size. The tubes were mixed vigorously to ensure full dissolution of the pellets in solution. All DNA samples were then kept in a -20°C freezer for storage.

DNA Quantification and Quality Assessment

The amount of DNA in each sample was estimated based on spectrophotometry, with the use of a NanoDrop 2000 (Thermo Scientific Inc.). Two µL of $TE_{0.1}$ were loaded on the NanoDrop as a blank, and four standards of calf thymus DNA (2 µL each of 1ng/µL, 5ng/µL, 10ng/µL, and 50ng/µL standards) were used to calibrate the instrument. To determine the quantity of DNA for each sample, 2 µL of DNA were loaded on the NanoDrop twice, and the estimated concentration was taken to be the average of the two readings. From this concentration, 5ng/µL working aliquots of DNA were prepared using $TE_{0.1}$ and stored at -20°C.

Agarose gel electrophoresis was used to assess the quality of DNA. About 20 ng of DNA (based on the NanoDrop concentration) were combined with Orange-G dye. This mixture was then loaded onto 2% agarose gels, and the DNA was visualized via staining with SYBR Green I (Life Technologies). Voltage was set to 5-8 V/cm separating the cathode from the anode, and run for roughly one hour. DNA separated by molecular weight in the agarose gel was then visualized on a UV transilluminator.

SNP PCR Optimization

A total of 38 SNPs have been characterized in sperm whales to date by Morin *et al.* (2007) and Mesnick *et al.* (2011), 20 of which are closely associated and therefore linked. Forward and reverse primers designed for 20 of these 38 SNP loci were obtained (**Table 1**). The primers were tested at three annealing temperatures: 48°C, 53°C, and 58°C. The optimal annealing temperature for each primer pair was determined to be the lowest for which there was a high yield of amplified DNA and no evidence of nonspecific binding, and was used in all subsequent PCRs (**Table 1**). Two DNA samples were used for primer testing: one which showed high molecular weight on an initial agarose gel, and another which showed a high degree of degradation. This allowed for an evaluation of how successfully each SNP-targeting fragment could amplify in poor quality DNA.

Initial PCRs were conducted to determine optimal conditions for amplification of SNP-targeting fragments. PCR cocktails were made for 10 µL reaction volumes with the following Ingredients: 1X PCR buffer, 0.2mM each dNTPs, 1.5 mM $MgCl_2$, 0.3 mg/mL bovine serum albumin, 0.3 µM forward primer, 0.3 µM reverse primer, and 0.05 U/µL Taq polymerase. Reactions were tested using 10 ng and 20 ng of template DNA. The number of denaturing-annealing-extension cycles was also tested at 30 and 35. The combination of these factors which resulted in the highest yields and specific primer binding was used in all subsequent PCRs.

PCR products were size-separated and visualized on 2% agarose gels and ethidium bromide (EtBr) staining. Four µL of PCR product for each reaction was mixed with 2 µL of Orange-G loading dye and loaded on an agarose gel. Voltage was set to 5-8 V/cm separating the cathode from the anode, and run for roughly one hour. Amplified DNA separated by molecular weight in the agarose gel was then visualized on a UV transilluminator.

SNP Genotyping

SNPs were initially to be genotyped using the Amplifluor SNPs Genotyping System (EMD Millipore Corporation, Billerica, MA, USA). This method was chosen because it was the method previously used for genotyping these loci in sperm whales (Morin *et al.* 2007; Mesnick *et al.* 2011). Amplifluor is a PCR-based system that uses technology described by Myakishev *et al.* (2001). Tailed primers specific to a particular allele anneal to pre-amplified fragments containing a SNP of interest, and particular universal primers containing fluorescent dyes anneal to the tails of these primers. The dyes are inhibited from fluorescing when in the presence of "quencher" molecules. The Amplifluor universal primers contain both dye and quencher molecules, and form a hairpin loop which brings the two together. During a successful PCR, the hairpin structure is disrupted, separating the dye and quencher, and allowing the dye to fluoresce. This marks a successful PCR, and also the genotype of the SNP based on the colour of fluorescence. Two types of dyes must be used, for example one that fluoresces green and another that fluoresces red, depending on which allele is amplified. Since SNPs are largely bi-allelic with three possible genotypes, the primers are designed such that either of the two homozygous genotypes will result in either green or red fluorescence, and heterozygotes will result in a mix of red and green (Myakishev *et al.* 2001). This fluorescence can be detected in real-time using quantitative PCR. Though it is theoretically possibly to assay many SNPs simultaneously in multiplex reactions (Garvin *et al.* 2010), the Amplifluor system does not allow for this, requiring each SNP to be amplified individually.

The Amplifluor system was tested using six samples and 10 SNPs. A cocktail for 10 µL reaction volumes was made with the following Amplifluor ingredients: 1X PCR buffer, 0.25 mM each dNTP, 1X JOE primer, 1X FAM primer, 0.3 µM each allele-specific forward primer, 0.3

μM reverse primer, and 0.05 U/μL Taq polymerase. The cocktail was added to two μL of PCR

product for each sample. Fluorescence from the Amplifluor reactions was detected using

quantitative PCR on a Roche LightCycler® 480 (located at the Marine Gene Probe lab at

Dalhousie University).

Due to difficulties in applying the Amplifluor system (see results), SNPs were actually

genotyped by sequencing the PCR fragments using the Sanger method (Sanger *et al.* 1977).

Nucleotide sequences can be visualized as different colour peaks on an electropherogram.

Because SNPs are single base-pair variations, individuals can be typed based on the colour and

number of peaks at these sites. For homozygotes, there should be one peak per site which may

change colour between different individuals depending on which allele they have, while in the

case of heterozygotes, two peaks of different colour should overlap (**Figure 1**). This is the

method by which SNPs are typically discovered in a genome (Vignal *et al.* 2002).

Prior to sequencing, the expected sequence of each PCR fragment was determined from

the primer sequences and sequences of the sperm whale SNP regions published on GenBank

(**Table 2**). From these expected fragment sequences, primers were selected for sequencing based

on their likelihood of producing a sequence from which SNPs could be typed. Sequencing is

known to produce messy and unreliable sequences for the first ~30 base pairs. Therefore,

forward or reverse primers were selected to be used in sequencing reactions based on how distant

the target SNPs were from them. For fragments so small that SNPs were within 24 bp of both

primers, sequencing was not performed.

After the initial PCR, excess primers and dNTPs were removed from the solutions using

an ExoSAP protocol. Twenty ng of PCR product (as determined from EtBr-stained agarose gel

electrophoresis) were subject to ExoSAP for all samples to be sequenced. A cocktail was made

containing 0.00614 µL Exonuclease I, 0.02 µL Antarctic phosphatase, and 0.129 µL Antarctic phosphatase buffer per 1 µL reaction volume. Cocktail was added to each sample in a 96-well plate, and the samples were heated in a PCR thermal cycler for 15 minutes at 37°C followed by another 15 minutes at 80°C. Following ExoSAP treatment, samples were stored at -20°C until they were retrieved for sequencing reactions.

Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies). A cocktail was made for 15 µL reaction volumes with the following concentrations: 1X sequencing buffer, 0.25X ready reaction mix, and 0.33 µM forward or reverse primer. Cocktail was added to each ExoSAP-treated sample, and the samples were placed in a PCR thermal cycler with the following protocol: one cycle at 96°C for 2 minutes, followed by 30 cycles of 96°C denaturing for 20 seconds, 50°C annealing for 20 seconds, and 60°C extension for 4 minutes. Following the sequencing reactions, samples were stored at -20°C until they were retrieved for ethanol precipitation.

Ammonium acetate was added to each sample to a final concentration of 0.68 M. A 46.2 µL aliquot of 95% ethanol was then added. The sample plates were sealed and centrifuged for 35 minutes at 2550 x g. Following centrifugation, the ethanol was gently decanted from each well. Excess ethanol was removed by taping Kimwipes on the top of each plate and spinning them inverted to 300 rpm. A 100 µL aliquot of 70% ethanol was then added to each sample, and the plates were centrifuged for two minutes at 4550 x g. The ethanol was decanted again, with excess removed by inverted centrifugation to 300 rpm. Following this, the sequencing fragments in each well were redissolved in 10 µL of HiDi Formamide for use with capillary gel electrophoresis on an ABI 3500 XL Genetic Analyzer (Applied Biosystems).

Data Analysis

Samples for which more than four loci could not be typed were excluded from subsequent analyses. The program CERVUS 3.0 (Kalinowski *et al.*2007; Marshall *et al.* 1998) was used to identify samples with the same genotype across all SNPs for which they could be typed. In the case of a match, the samples were considered to be from the same individual, and only one of them was used in subsequent analyses.

Some pairs of SNPs genotyped in this study were linked, which means that they are likely to be inherited together and therefore cannot be treated independently. These linked SNPs were analyzed by collapsing them into single multiallelic loci. The alleles for these collapsed loci were considered to be the combination of the first allele for each locus as one allele, and a combination of the second allele of each locus as the other allele. However, this method is problematic for double heterozygotes, since there is no way of knowing which combination of alleles are actually linked on the same chromosome. Therefore, for double heterozygotes, one of the two linked SNPs was removed from analysis.

Patterns of genetic relatedness were assessed using network analysis. Networks are models of individual components (nodes) that are interconnected with one another (Wey *et al.* 2008). The theory behind networks originates from mathematical graph theory, and they have proved to be a powerful means of visualizing and analyzing interactions between components in many fields, including sociobiology (Wey *et al.* 2008). In this analysis, a network is constructed from a matrix of pairwise relatedness based on SNP allele frequencies. The nodes in this network represent individual sperm whales, and their connections represent relatedness. The length of the connections between nodes reflects the degree of relatedness, with shorter connections indicating close relatives.

The program GeNetwork was used to estimate allele frequencies for the SNP loci used, and to construct a matrix of pairwise relatedness between each individual. This program estimates relatedness using the methods described by Li *et al.* (1993), with each locus weighted as described by Lynch and Ritland (1999) and Van de Casteele *et al.* (2001). The relatedness matrix was then analyzed as a network using the *igraph* package (Csardi & Nepusz 2006) for the statistical program R (R Development Core Team 2008). GeNetwork was also used to produce 1000 simulated populations based on the observed allele frequencies. These simulations represent randomized populations with no structure, and were used to test the significance of clustering within the observed data by comparing modularity (a measure of clustering). If the modularity of the observed data was greater than the modularity of at least 95% of the simulations, then the observed data were considered to be significantly structured.

Pairwise relatedness could also be compared with helping and group association data to assess which types of selection operate on sperm whale units. However, this analysis was excluded from this study (see results).

Statistical power analysis was conducted using the program POWSIM (Ryman & Palm 2006) to assess the ability of the SNPs used to detect significant levels of clustering in the Dominica sperm whale population. This program uses the F_{ST} metric as a measure of genetic differentiation. Sperm whales in the Atlantic are known to have overall low levels of genetic variation (Engelhaupt *et al.* 2009); therefore power estimates were calculated for F_{ST} values ranging from 0 to 0.05. For the Dominica population sampled, it was assumed that F_{ST} is likely around 0.01. Power estimates were made for the eight loci used, as well as the three times those loci to approximate the power that could have been obtained if all SNPs characterized for the

sperm whale were used. Microsatellite data from Arseneault (2011) were also used to compare

the relative power of SNPs versus microsatellites.

RESULTS

DNA quality

Of the 128 samples extracted, 34 showed no signs of DNA on SYBR Green I stained agarose gels, signifying that the DNA from these samples is too heavily degraded to be visualized. Of the 94 samples for which the presence of DNA could be seen, 66 showed significant smearing on the gels, indicating a high degree of degradation (**Figure 2**). This means that overall, only 28 samples (21.9%) contained intact, high molecular weight strands of DNA.

Optimal PCR conditions for SNP amplification

Primers for the 20 SNPs to be analyzed resulted in amplification fragments ranging from 55-174 bp. EtBr-stained agarose gels showed that all primers successfully amplified the desired fragments for both the high molecular weight and highly degraded samples used for testing (**Figure 3**). Desired fragments were found to amplify best with 20 ng of template DNA. The optimal PCR cycles for high yield of the target fragments were found to be the following: one cycle at 94°C for 5 minutes, followed by 35 cycles of 94°C denaturing for 30 seconds, annealing for one minute, and 72°C extension for 1 minute, and one cycle at 60°C for 45 minutes. Optimal annealing temperatures varied between primer pairs and are shown in **Table 1**. PCR product yielded on average 10 ng/μL of DNA.

Genotyping using Amplifluor

None of the six samples tested at 10 loci with the Amplifluor SNP genotyping kit resulted in significant fluorescence signals during qPCR, meaning that no genotype could be obtained. Following this failure, the Amplifluor kit was tested again using the same samples subject to

different treatments with two loci. The treatments consisted of variable dilutions of PCR product including 1:10, 1:20, 1:50, and 1:100 dilutions with $TE_{0.1}$. Additionally, two experiments were run with these dilutions: one treated with ExoSAP, and one that was not. None of these test resulted in significant fluorescence for any sample. Therefore, the approach of genotyping samples via the Amplifluor kit was abandoned in favour of sequencing.

Genotyping using Sequencing

Some of the initial primers designed to target the 20 SNPs to be analyzed produced fragments that contained up to three linked SNPs. Unlike Amplifluor, sequencing does not require the use of allele-specific primers. Therefore, for cases in which there were multiple SNPs in a fragment, only one primer pair was used for genotyping all of them. There was also one instance where forward primers designed for two different SNPs with close linkage were combined in a single reaction to produce a slightly larger (105 bp) fragment that contained both SNPs (CSF2K552 and CSF2Y589) (**Table 3**). In total, this resulted in the sequencing of 13 fragments and 19 SNPs. Of these 19 SNPs, only 12 could be confidently typed for most samples (**Table 4**). A total of 117 samples out of 128 (91.4%) were successfully typed at more than seven of these 12 SNPs.

Data analysis

Using CERVUS to identify samples with identical genotypes resulted in the removal of 45 samples from the 117 typed. This means that 72 samples (56.3% of all samples) could be analyzed in a network. Of the 12 SNPs that could be typed, eight of them were linked. This

resulted in a final effective panel of eight loci, half of which were bi-allelic, while the other half could contain up to four alleles (**Table 5**).

The observed network of relatedness among sperm whales based on pairwise relatedness at the eight loci analyzed is shown in **Figure 4**. Analysis of this network suggests the presence of three distinct clusters of individuals. However, the modularity of this network was only 0.1818, which was not significantly higher than the expected modularity values from a randomized set of unrelated individuals. Comparison with 1000 simulated networks based on the observed allele frequencies showed that the level of clustering detected in the observed data is not statistically significant ($p = 0.139$; **Figure 5**). This means that a single large population of equally related whales was found. With this being the case, it would be meaningless to compare the results with observational data, because correlation between preferential helping or association and degree of relatedness could not be determined. Therefore, analysis of social behaviour was excluded.

Power analysis showed that the SNPs used are lacking in their ability to resolve structuring at the low levels of genetic differentiation expected (**Figure 6**). Power at $F_{ST} = 0.01$ was only around 0.4, which is not reliable. Tripling the number of SNP loci resulted in a somewhat more reliable power, but not at the level achieved with a panel of 11 microsatellites. A larger number of SNPs would be needed just to match the information content of these microsatellites.

DISCUSSION

DNA from 78.1% of all sloughed skin samples used showed signs of high degradation. This illustrates the fact that DNA retrieved from noninvasively acquired tissue samples is often of poor quality. However, the 91.4% success rate of genotyping using SNPs shows that information can still be obtained from them using these types of markers that only require short fragments of DNA. Compared to microsatellites, SNPs are clearly superior in terms of amplification success of degraded DNA. Previous work in our lab using sloughed sperm whale skin was able to type 11 microsatellites for only about 33% of samples (Arseneault 2011). The length of the amplification fragments for these microsatellites ranged from roughly 120-220 bp, which was sometimes large enough to be problematic for heavily degraded samples. In contrast, the generally shorter (55-174 bp) SNP fragments proved small enough to be within intact segments of degraded DNA.

Of the 117 samples scored at most loci, only 72 individuals were identified. This could be due to a number of factors. One may be redundancy during field sampling of sloughed skin. Though it is often possible to identify which whales leave behind which pieces of skin, this can be difficult and error-prone if the whales are associating close together and rubbing against one another (Valsecchi *et al.*1998). Indeed most samples were collected when the whales were in their associated units, which therefore made it impossible to know specifically from which whale the sample originated. In this case, the same individuals could have been sampled multiple times. Another issue may simply have been lack of genetic resolution. Samples were considered to be from the same individual if they had the same genotype for all SNPs for which they could be typed. However, it is possible that different individuals could have the same genotype at the few 8-12 SNP loci that were used, with their genotypes then being falsely collapsed into one

individual. An increase in the number of SNPs used would increase resolution, and therefore increase the number of individuals identified. Also, sexing of the samples could have differentiated "identical" individuals if they were found to differ in their gender. Molecular sexing for some these samples was done previously using the method described by Gilson *et al.* (1998). However, only 13.6% of samples resulted in a scorable sex, likely a consequence of severe DNA degradation once again. Future work could try using other sexing methods more likely to yield results in poor quality DNA, such as that described by Morin *et al.* (2005).

Results from network analysis did not reveal any significant structuring, implying that all whales sampled near Dominica were equally related. Sperm whales are known to have low genetic diversity overall, likely because of the roaming habits of males among multiple female groups for mating, and perhaps also because of a recent bottleneck (Engelhaupt *et al.* 2009). It is possible, then, that sperm whale groups really are all closely related, because the whales could share the same paternal genes. Because of the long lifespan of sperm whales, it is also possible that there is generational overlap in relatedness, since males returning to the female groups could potentially mate with their own progeny. However, this is unlikely, because the absence of any genetic clustering at all within sperm whale populations is inconsistent with the results of most other studies (e.g. Richard *et al*. 1996; Pinela *et al*. 2009), who have found that relatedness is greater within groups then among them. Additionally, it is inconsistent with what was found previously using microsatellites. This is almost certainly due to the lack of resolution for this study as was shown from the power analysis, rather than a true reflection of the genetic structure of sperm whales. Because SNPs are typically only bi-allelic and therefore less informative than individual microsatellites, the panel of 11 microsatellites used previously was substantially more powerful than the eight loci (12 SNPs in total) used here, even though the sample size was

greater in this study. Linkage of eight SNPs analyzed must also have contributed to a loss of resolution, as linked loci cannot be treated as independent data points.

Since no significant structuring could be detected, conclusions on what type of selection is acting in sperm whale social systems cannot be made. Such analyses would require that the behaviour of individuals who are more closely related be compared with that of individuals who are not as related, and this cannot be done if all individuals are just as related. Detail is needed for such analyses, and it could not be obtained here.

What can be concluded from these results is that while SNPs can be more successfully analyzed in highly degraded DNA than microsatellites, they can suffer from a lack of power unless a large number of them are available. Unfortunately, the discovery of new SNPs in non-model species (that is, species for which there are no large-scale genome projects; Morin *et al.* 2004) is not a trivial process. Since SNPs are located throughout the genome, discovering them often requires large-scale genome sequencing efforts for many individuals in the species of interest (Morin *et al.* 2004). With the current conventional technology, such endeavors can be extremely time-consuming and expensive (Garvin *et. al.* 2010), making high-resolution SNP analysis an unfeasible goal for most wildlife biologists. However, it is possible that this situation will change in the future. New advances in sequencing technologies, collectively called Next Generation Sequencing (NGS), are being developed which are potentially capable of identifying SNPs on large scales in a relatively efficient and cost-effective manner (Ekblom & Galindo 2011; Garvin *et al.* 2010). The potential of NGS in wildlife biology is looking promising, as the technology is improving rapidly, and an increasing number of ecological studies have begun using it since 2007 (Ekblom & Galindo 2011). It is quite possible that SNPs will become a more commonly used marker through NGS, and future work on sperm whale relatedness could look

into these methods. However, time and financial resources available could still be a limiting factor.

Another problem encountered with SNPs is cross-lab reproducibility and ease of amplification. One proclaimed benefit of SNPs is fewer cross-lab protocol discrepancies than microsatellites (Seeb *et al.* 2011; Mesnick *et al.* 2011). However, that was not the case with this study. The Amplifluor SNP genotyping system to be used was selected because it was the method used by Morin *et al.* (2007) and Mesnick *et al.* (2011) for genotyping the same SNP loci in sperm whales. While those studies claimed to be successful with the Amplifluor system, these protocols were not successful in our laboratory, even though the same protocol was followed. This represents an inconsistency with this method of genotyping that needs to be addressed. Sequencing proved to be an effective means of genotyping, but it is far from efficient. In addition to being costly and time-consuming, sequencing resulted in the loss of some potential SNPs because only a certain portion of the PCR fragments could be reliably sequenced. Genotyping from nucleotide sequences can also be more error-prone, as it is possible to confuse the double peaks of a heterozygote with a noisy signal. SNP analysis would benefit from the development of more reliable and efficient means of genotyping. Future methods could also look at the potential of multiplexing to improve efficiency when working with a large number of SNPs. Another proposed advantage of SNPs over microsatellites is that they can theoretically be multiplexed to greater numbers than microsatellites, potentially up to tens of thousands or even millions of SNPs (DeWoody 2005; Garvin *et al.* 2010). However the technology for multiplexing this many SNPs is cumbersome and, for non-model organisms (such as the sperm whale), currently needs to be custom-designed (Garvin *et al.* 2010). More conventional SNP genotyping assays also exist which are currently capable of multiplexing about 2-50 SNPs (Garvin *et al.* 2010).

All things considered, microsatellites are probably more useful markers than SNPs for relatedness analyses at present time. The main problem is dealing with degraded DNA. Noninvasively obtained samples are useful, because they do not require the study animals to be captured or disturbed. For sociobiology studies, noninvasive sampling also reduces the likelihood that animals will react in a manner that alters their normal behaviour (e.g. Whitehead *et al.* 1990). However, noninvasive sampling clearly suffers from poor quality DNA. If high-resolution analysis of relatedness is desired, the best option may be to accept more invasive nondestructive sampling. In the case of sperm whales and other cetaceans, nondestructive sampling typically involves the collection of small skin and blubber samples obtained by firing a biopsy dart into the whale's skin using bows or airguns (Noren & Mocklin 2012). Many studies have been conducted on the effects of this sampling method on whales' behaviour and condition. The consensus is that biopsy darting is a relatively benign sampling method, with only short-term effects (Noren & Mocklin 2012). Also, when conducted under the right conditions, whales often respond only slightly or not at all to biopsy attempts (Noren & Mocklin 2012). Therefore, it may be possible to obtain reliable behavioural data shortly after biopsy sampling.

In situations where biopsy sampling is restricted, sloughed skin could still provide an acceptable level of genetic information (Valsecchi *et al.*1998). In such cases, it could be possible to increase resolution by gathering a larger sample size. However in conclusion, the best current solution for resolving relatedness on a fine scale within and among sperm whale social units is likely to use microsatellite analysis from biopsy samples.

Table 1: SNP loci characterized in *Physeter macrocephalus*. Corresponding forward and reverse primer sequences tested in this study are shown, along with the optimal annealing temperatures found for them. Bold loci represent linked SNPs.

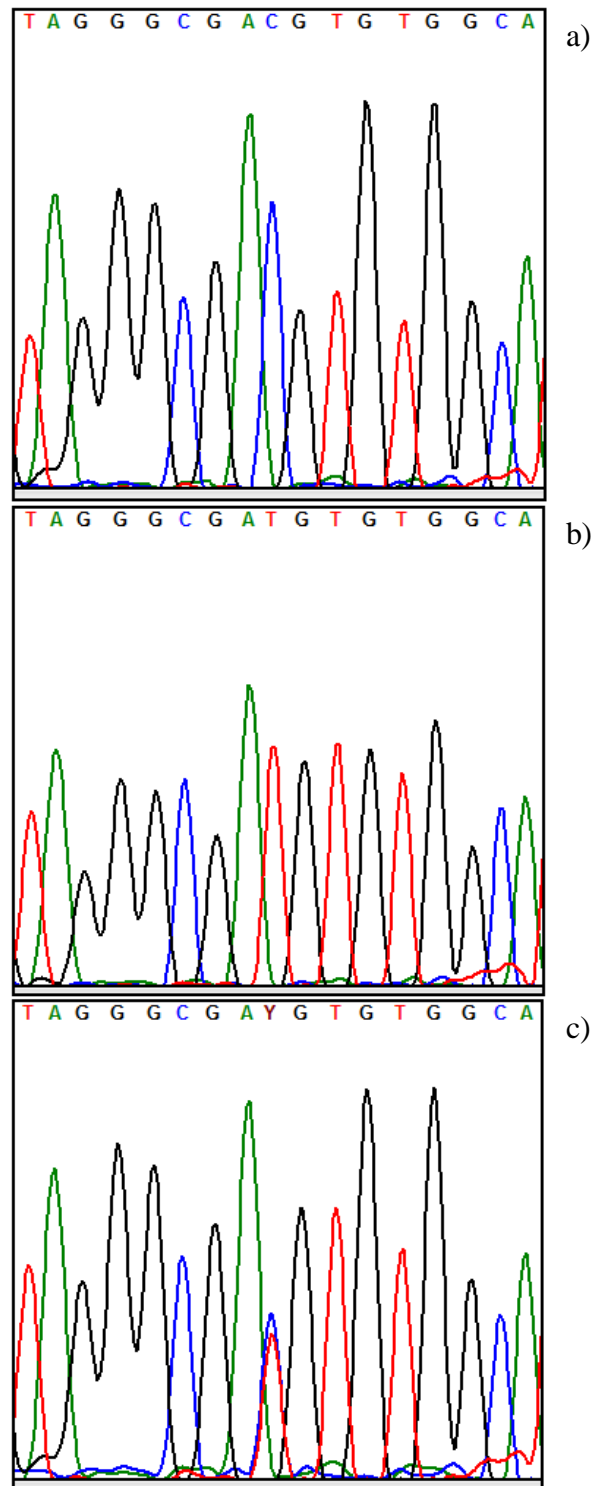| Locus	Study	Forward Primer	Reverse Primer	Optimized Annealing Temperature
ACTCM392	Morin *et al.* (2007)	—	—	—
ADRBK1R131	Morin *et al.* (2007)	—	—	—
CATR262	Mesnick *et al.* (2011)	TGTTTTCTTTTTTAAGTGTCTTTGC	CCCTCATTCACAGGCTAGGAAA	48°C
CATR456	Morin *et al.* (2007)	TCCCTTCCAGTTCCTGTTTG	GTGAGTGCATTGGTTGAAACTTT	48°C
CHRNA1Y111	Morin *et al.* (2007)	GTCAGACCAGGAGTCCAATAACG	GGCACATCTTACGTCAGTAGCT	48°C
CHRNA1R76	Mesnick *et al.* (2011)	GGCACATCTTACGTCAGTAGCT	GTCAGACCAGGAGTCCAATAACG	48°C
CKK273	Morin *et al.* (2007)	CCACCCCTATGTTCAAGCAC	ACAGGTGGTGGCAGGATT	58°C
CSF2K552	Mesnick *et al.* (2011)	CCCCTTGAATGCTAGGTCTG	CTTTGCCCCTGTGGTAGT	48°C
CSF2R278	Morin *et al.* (2007)	CTTTGCTCACACAGCAGGTC	AGCACAGCCACATTCCCTT	58°C
CSF2Y589	Mesnick *et al.* (2011)	GCCAGAAGGAGGGTTGCT	GACTACCACAGGGGCAAA	53°C
DRD2Y679	Morin *et al.* (2007)	GGGTTAGGTCTCGTTCAGCA	TACACCCTTTGCCGGATTCTT	53°C
ELN40K209	Morin *et al.* (2007)	TCCCTTAAACTGTGGCTCGT	AGCGGCACCTGAGAAGAG	53°C
EPOR237	Mesnick *et al.* (2011)	GAAATGCACAAGCCTGGAGT	AAGATCCAGGAACCTGGAGT	58°C
EPOY292	Morin *et al.* (2007)	—	—	—
F9Y80	Morin *et al.* (2007)	CGAAAAAGAAGACAGGTTAATGG	GGGAACCATACTTGCCTTTGGAA	53°C
GRPY190	Morin *et al.* (2007)	TGCAGCAGGACATCAAGTTC	CCCCCCCCCCACTTTCTTTTTT	53°C
IFNGY234	Morin *et al.* (2007)	—	—	—
INTS368	Morin *et al.* (2007)	GAACCTCAGCAAAGGCCC	CAAATCTCTGCGGGACAGTT	58°C
PKMY237	Morin *et al.* (2007)	CTTCCTTAGCAGAGCGTCTCA	AGACTTGGCCAGCCCTCTAT	53°C
PmABHD5M274	Mesnick *et al.* (2011)	TCTGTCCTAGAGCCACAGCA	ACCTCTAACAATAGGGCCAACA	53°C
PmABHD5R671	Mesnick *et al.* (2011)	—	—	—
PmABHD5Y447	Mesnick *et al.* (2011)	—	—	—
PmBH92S122	Mesnick *et al.* (2011)	—	—	—
PmBH92Y172	Mesnick *et al.* (2011)	GGCCTATGGTAGGTATCCTGTC	GCTCAACCCAATAAAAAAAAGCTA	48°C
PmCHYR304	Mesnick *et al.* (2011)	—	—	—
PmDDX5R109	Mesnick *et al.* (2011)	—	—	—
PmHSPA9Y220	Mesnick *et al.* (2011)	—	—	—
PmLAPTM4R553	Mesnick *et al.* (2011)	—	—	—
PmMYL4R358	Mesnick *et al.* (2011)	—	—	—
PmPHGDHM200	Mesnick *et al.* (2011)	—	—	—
PmPHGDHR223	Mesnick *et al.* (2011)	—	—	—
PmPHGDHS172	Mesnick *et al.* (2011)	AGGGAGCAGTGAGACCAGAA	CAACCTCACCAGCTCCAAA	58°C
PmPHGDHY321	Mesnick *et al.* (2011)	—	—	—
PNDR111	Morin *et al.* (2007)	—	—	—
RDSK456	Morin *et al.* (2007)	—	—	—
RYR2R327	Morin *et al.* (2007)	CACGTATCTCTAGGGAGCAGC	GGTGGAAAGGATGAGCAGAAT	58°C
SPTBN1S279	Morin *et al.* (2007)	CCAAGCAGCACTCAAGTACG	GAGATGCAGTAGGGTAGCCTT	53°C
SPTBN1Y753	Mesnick *et al.* (2011)	—	—	—

Figure 1: Electropherograms of sperm whale DNA sequences at a SNP locus (DRD2Y679) for three individuals. The SNP is located at the 9th position from the left. The two alleles for this locus are C or T. The individual in a) is a C homozygote, the individual in b) is a T homozygote, and the individual in c) is a heterozygote.

Table 2: Sequences of 20 PCR amplification fragments containing sperm whale SNPs.

Locus	Amplification Fragment	Fragment Length (bp)
CATR262	TGTTTTCTTTTTTAAGTGTCTTTGCAACTAATTAAAAAAAAATTGTGGTCAAGCRTTCGTAAGTTGTATACAAAACACAGTGGTACCACTTAAAGAGTTTCCTAGCCTGTGAATGAGGG	118
CATR456	TCCCTTCCAGTTCCTGTTTGATGAGTTTACTAAGTTCATCTGGGTGGCCTGATATATTGTTATTAGCAGGGAACAAATTTTGATGAGCTGATATACTTTTGCCCAGGGAAAGACTCARTGTTTACTGTTTACTGTTGTGAAAGTTTCAACCAATGCACTCAC	162
CHRNA1Y111	GTCAGACCAGGAGTCCAATAACGTAAGCTCTGTGGCTTGAGATCTGCACCTCCTGTTTAAATGGTCAAATGCTCRAGCACAGAGGGATGGGTTTGGCTTGATGGGAAGGYTGGCGTTCAAGGGGCAGCTACTGACGTAAGATGTGCC	147
CHRNA1R76	GTCAGACCAGGAGTCCAATAACGTAAGCTCTGTGGCTTGAGATCTGCACCTCCTGTTTAAATGGTCAAATGCTCRAGCACAGAGGGATGGGTTTGGCTTGATGGGAAGGYTGGCGTTCAAGGGGCAGCTACTGACGTAAGATGTGCC	147
CKK273	CCACCCCTATGTTCAAGCACCGCCCCTAAGAACGCGCCTTCCCKGCGTCCTCAAGCCCAGCCCCCTCAGAATCCTGCCACCACCTGT	87
CSF2K552	CCCCTTGAATGCTAGGTCTGTCCATGGCTGGTCAGCTAATAAAGKAGCTCACGTGACTACCACAGGGGCAAAG	73
CSF2R278	AGCACAGCCACATTCCCTTGGCCTGGCATCTCAAGAGGGTCCCAGTCCCCRGAAAACAGAGGTCACTGTCCTCCCAGCATTCAGGGGGAAGTAGGAGGGGGCTCTCGCCCAGACCTGCTGTGTGAGCAAAG	131
CSF2Y589	GACTACCACAGGGGCAAAGCCAGTGGCYTCCTGAGCAACCCTCCTTCTGGC	51
DRD2Y679	TACACCCTTTGCCGGATTCTTGGAAGATGAATGGAACTAATGTGTCAGAAGTGCTGGGTAGGGCGAYGTGTGGCATCATAAACAATTACCATTGTTATTATTTTCTGCTGAACGAGACCTAACCC	125
ELN40K209	TCCCTTAAACTGTGGCTCGTCCTGACCCCCATCCKCCCTCTTCTCAGGTGCCGCT	55
EPOR237	AAGATCCAGGAACCTGGAGTCCCCCCCKGCACTTGGTTCAGGGGTGGAAGCGGGGAGCCAGAGCCCCCRGTAAAAGTAATAAGAGTGGTGGTCCCCAGCGCACACCTGGAAGCTAGGTAAGGGGYGAAGCCAGCAGGAGCCCCTGGGGCCCCTGACTCCAGGCTTGTGCATTTC	174
F9Y80	CGAAAAAGAAGACAGGTTAATGGTACTATTTAAAATGTTTTCGATTTGATTCCTTCTCTATTGTAACATTYATAGAGCACGTGTATGTATAGATATTTTAAAATTCCAAAGGCAAGTATGGTTCCC	126
GRPY190	TGCAGCAGGACATCAAGTTCTTGCCTTTCAAGGTTCGACCTGGTYTTTCTCTTCTAAATAGAAAGTGAGAAAATAAAAAAGAAAGTGGGGGGGGG	95
INTS368	CAAATCTCTGCGGGACAGTTCGCSCCTCCCTTCCTCTGTGCGCGGCTAGGCCTGGGCCTTTGCTGAGGTTC	71
PKMY237	AGACTTGGCCAGCCCTCTATGCTCTCCCTGAAACCCAGCCAGATGTCCAACTACAGAATTCTGGYTGTACAGAGCATGTCAAAAAGACCCTTCCCTAGGTCTCTGAGACGCTCTGCTAAGGAAG	124
PmABHD5M274	TCTGTCCTAGAGCCACAGCATCTCTGCAGTGGGAGAGACTGAAAGAGGGACAGACAGGTCCAATAGCACTGMGACCACCTCCCAAGGCAACACATTCTTAACTTTGTTCAGTGTTTCTGTTGGCCCTATTGTTAGAGGT	139
PmBH92Y172	GGCCTATGGTAGGTATCCTGTCTGTGAGTGTCATTTAGCASAGTGCAAGTAAAGTGAAATACAGTCAACTGATAGAAGTGGTTTATAGTAYGTAGCTTTTTTTTATTGGGTTGAGC	116
PmPHGDHS172	GCAACCTCACCAGCTCCAAAGGCSCTCTCTTCCTCCCACCAGCTCTTCCCCMCACACCTCCTTTCTGGGGGCCCRCGGTCTTTCTGGTCTCACTGCTCCCT	101
RYR2R327	GGTGGAAAGGATGAGCAGAATATGTACCAAGCCTCTCAGAGAGCTTTGGCRTGTTCTCCATTCCGAACAGGAATTCTCTAGCTGCTCCCTAGAGATACGTG	101
SPTBN1S279	GAGATGCAGTAGGGTAGCCTTCCATTATATTTTAACTSTGTCTTGCTATTCTGAAATGGACCTGTCTTCTCCTTCCGTACTTGAGTGCTGCTTGG	95

Blue: Forward primer binding site

Green: Reverse primer binding site

Red: Target SNP

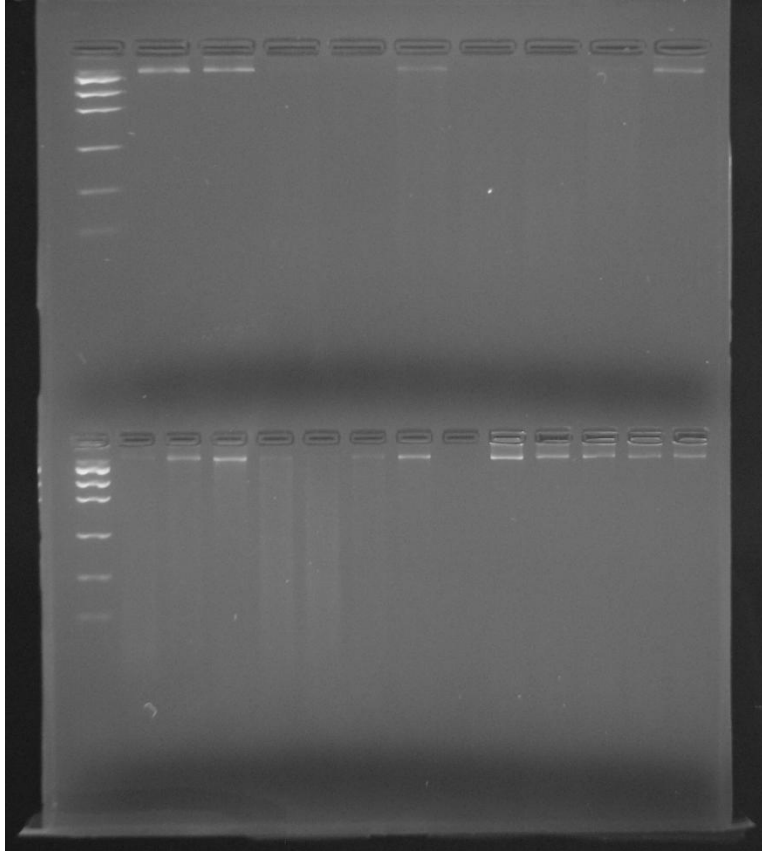Yellow: Other nontarget linked SNPs within the fragment

Figure 2: 2% agarose gel stained with SYBR Green I containing raw DNA from 18 sloughed skin samples. Note the faint presence of DNA and extensive smearing in most samples, indicative of a high level of degradation.

(Wells 1 and 11 contain a low mass ladder for comparing the amount and size of DNA. Wells 2, 3, and 20 contain positive controls, while well 19 contains a negative control. All other wells contain DNA from the samples of interest).
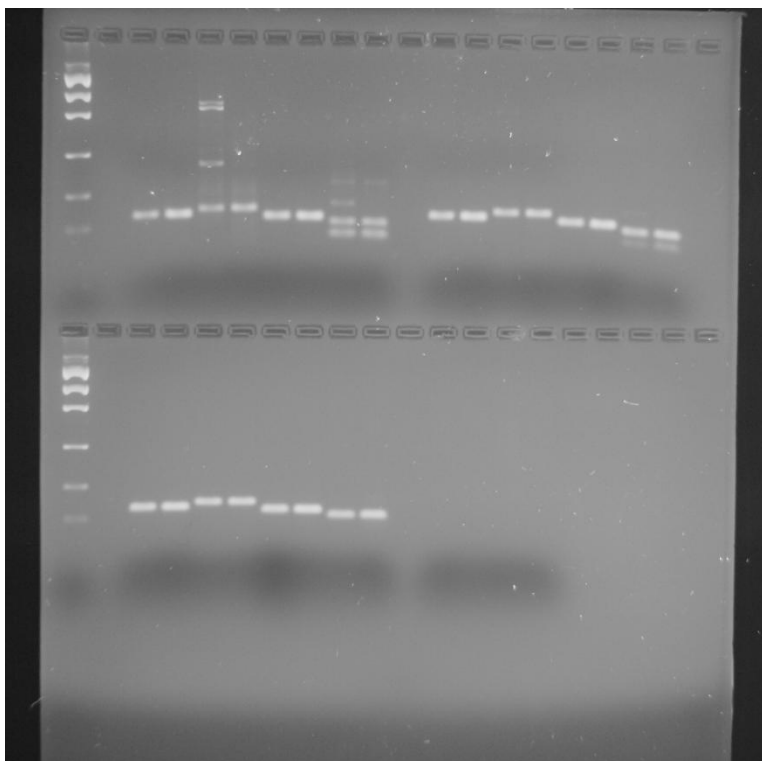
Figure 3: 2% agarose gel stained with EtBr containing PCR product from two samples (one of high quality and the other severely degraded) amplified with four SNP primer pairs using three annealing temperatures. Product in the top left was obtained with an annealing temperature of 48°C, the top right used 53°C, and the bottom left used 58°C. Note the high yield of amplified DNA obtained from all primers for both samples.

Table 3: Expected sequences of PCR amplification fragments used for genotyping using Sanger sequencing.

Locus/Region	Amplification Fragment	Fragment Length (bp)
CATR262	TGTTTTCTTTTTTAAGTGTCTTTGCAACTAATTAAAAAAAAATTGTGGTCAAGCRTTCGTAAGTTGTATACAAAACACAGTGGTACCACTTAAAGAGTTTCCTAGCCTGTGAATGAGGG	118
CATR456	TCCCTTCCAGTTCCTGTTTGATGAGTTTACTAAGTTCATCTGGGTGGCCTGATATATTGTTATTAGCAGGGAACAAATTTTGATGAGCTGATATACTTTTGCCCAGGGAAAGACTCARTGTTTACTGTTTACTGTTGTGAAAGTTTCAACCAATGCACTCAC	162
CHRNA1	GTCAGACCAGGAGTCCAATAACGTAAGCTCTGTGGCTTGAGATCTGCACCTCCTGTTTAAATGGTCAAATGCTCRAGCACAGAGGGATGGGTTTGGCTTGATGGGAAGGYTGGCGTTCAAGGGGCAGCTACTGACGTAAGATGTGCC	147
CKK273	CCACCCCTATGTTCAAGCACCGCCCCTAAGAACGCGCCTTCCCKGCGTCCTCAAGCCCAGCCCCCTCAGAATCCTGCCACCACCTGT	87
CSF2*	CCCCTTGAATGCTAGGTCTGTCCATGGCTGGTCAGCTAATAAAGKAGCTCACGTGACTACCACAGGGGCAAAGCCAGTGGCYTCCTGAGCAACCCTCCTTCTGGC	105
CSF2R278	AGCACAGCCACATTCCCTTGGCCTGGCATCTCAAGAGGGTCCCAGTCCCCRGAAAACAGAGGTCACTGTCCTCCCAGCATTCAGGGGGAAGTAGGAGGGGGCTCTCGCCCAGACCTGCTGTGTGAGCAAAG	131
DRD2Y679	TACACCCTTTGCCGGATTCTTGGAAGATGAATGGAACTAATGTGTCAGAAGTGCTGGGTAGGGCGAYGTGTGGCATCATAAACAATTACCATTGTTATTATTTTCTGCTGAACGAGACCTAACCC	125
EPO	AAGATCCAGGAACCTGGAGTCCCCCCCKGCACTTGGTTCAGGGGTGGAAGCGGGGAGCCAGAGCCCCCRGTAAAAGTAATAAGAGTGGTGGTCCCCAGCGCACACCTGGAAGCTAGGTAAGGGGYGAAGCCAGCAGGAGCCCCTGGGGCCCCTGACTCCAGGCTTGTGCATTTC	174
F9Y80	CGAAAAAGAAGACAGGTTAATGGTACTATTTAAAATGTTTTCGATTTGATTCCTTCTCTATTGTAACATTYATAGAGCACGTGTATGTATAGATATTTTAAAATTCCAAAGGCAAGTATGGTTCCC	126
PKMY237	AGACTTGGCCAGCCCTCTATGCTCTCCCTGAAACCCAGCCAGATGTCCAACTACAGAATTCTGGYTGTACAGAGCATGTCAAAAAGACCCTTCCCTAGGTCTCTGAGACGCTCTGCTAAGGAAG	124
PmBH92	GGCCTATGGTAGGTATCCTGTCTGTGAGTGTCATTTAGCASAGTGCAAGTAAAGTGAAATACAGTCAACTGATAGAAGTGGTTTATAGTAYGTAGCTTTTTTTTTATTGGGTTGAGC	116
PmPHGDH	GCAACCTCACCAGCTCCAAAGGCSCTCTCTTCCTCCCACCAGCTCTTCCCCMCACACCTCCTTTCTGGGGGCCCRCGGTCTTTCTGGTCTCACTGCTCCCT	101
SPTBN1S279	GAGATGCAGTAGGGTAGCCTTCCATTATATTTTAACTSTGTCTTGCTATTCTGAAATGGACCTGTCTTCTCCTTCCGTACTTGAGTGCTGCTTGG	95

Blue: Forward primer binding site

Green: Reverse primer binding site

Red: Target SNPs to be genotyped

* The CSF2 region fragment was obtained by mixing the forward primers for CSF2K552 and CSF2Y589

Table 4: Genotyping success of SNPs contained within PCR fragments sequenced on an ABI 3500xl.

Locus	Successfully Typed
CATR262	Y
CATR456	Y
CHRNA1Y111	Y
CHRNA1R76	Y
CKK273	N
CSF2K552	N
CSF2R278	Y
CSF2Y589	Y
DRD2Y679	Y
EPOR237	Y
EPOY292	Y
F9Y80	Y
PKMY237	N
PmBH92S122	N
PmBH92Y172	Y
PmPHGDHS172	Y
PmPHGDHM200	N
PmPHGDHR223	N
SPTBN1S279	N

Table 5: The final 8 loci used for relatedness network analysis.

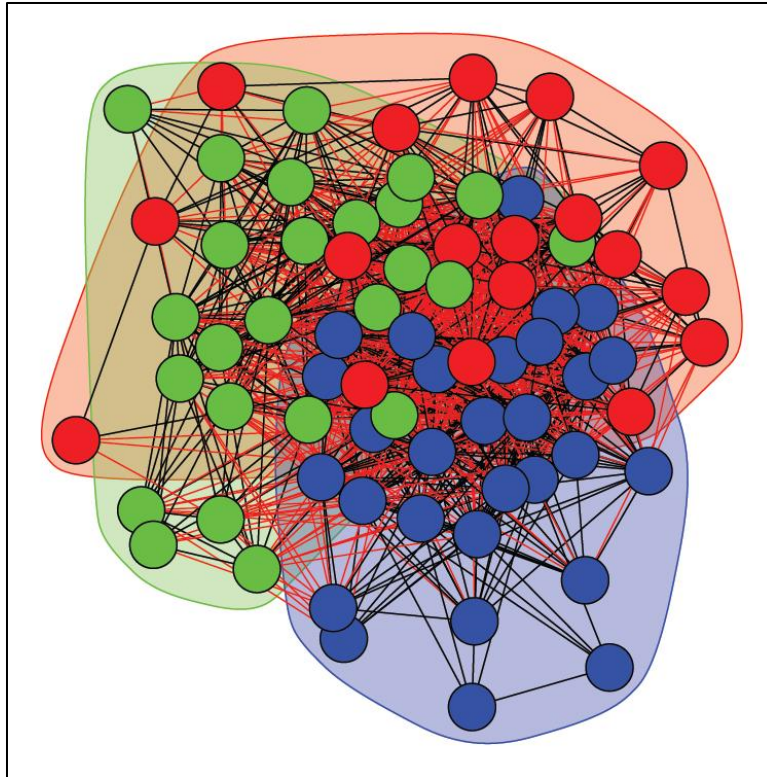| Locus	Number of Alleles	Component Linked SNPs
CAT	4	CATR262, CATR456
CHRNA1	4	CHRNA1Y111, CHRNA1R76
CSF2	4	CSF2R278, CSF2Y589
DRD2Y679	2	-
EPO	4	EPOR237, EPOY292
F9Y80	2	-
PmBH92Y172	2	-
PmPHGDHS172	2	-

Figure 4: Observed network of genetic relatedness among 72 sperm whales off the coast of Dominica based on pairwise relatedness from SNPs. Network analysis has detected three clusters of related individuals, which were found to be statistically insignificant (see text, **Figure 5**).
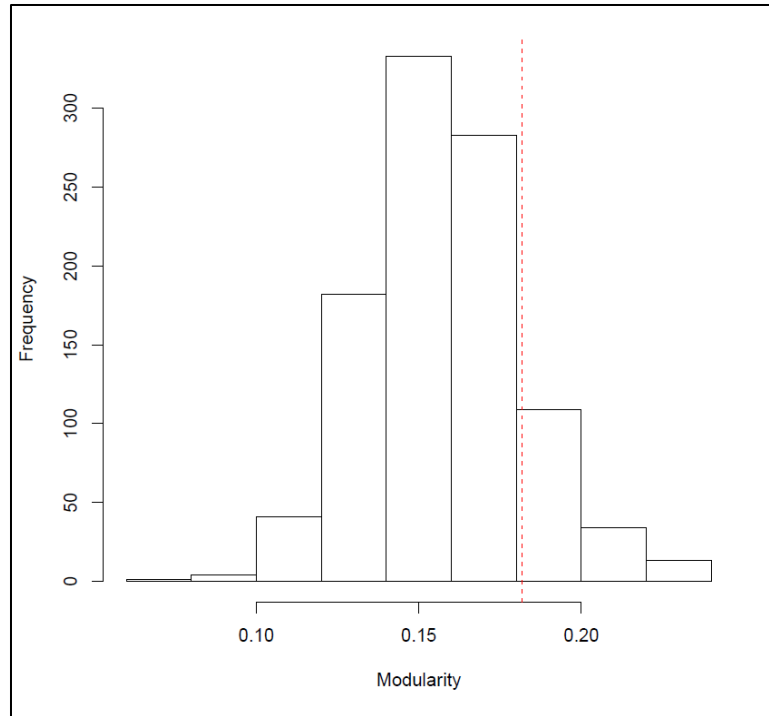
Figure 5: Histogram of the modularity distribution of 1000 simulated networks based on observed allele frequencies in 72 individuals. The modularity of the observed network is indicated by the red line, and is not significantly different from the simulated results.
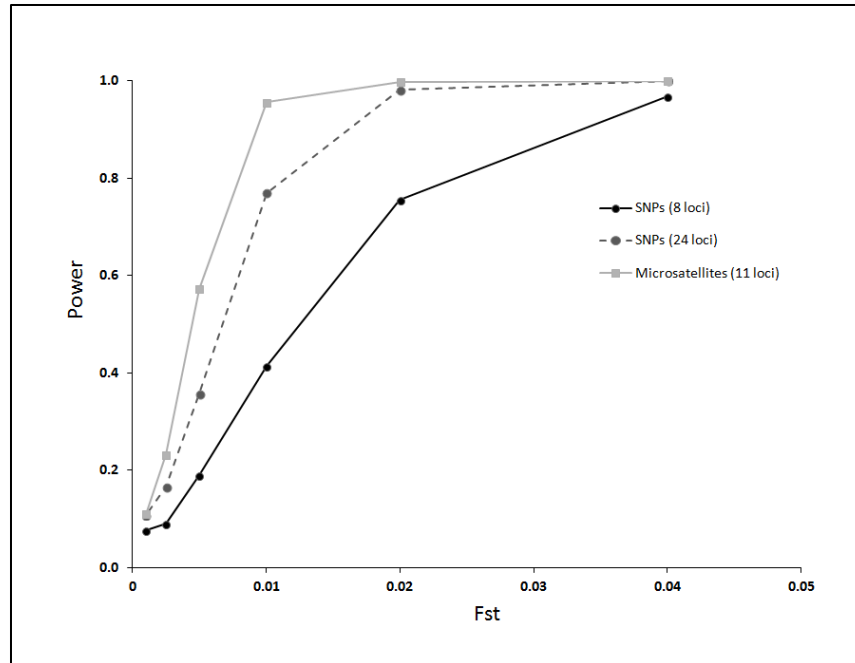
Figure 6: Statistical power estimates of SNP and microsatellite markers for detecting three clusters of related individuals in a population with low overall genetic variation. The eight SNP loci include 12 SNPs in total, eight of which are linked. The 24 SNP loci include the original 8 multiplied three times (36 SNPs, total, 24 linked). The 11 microsatellite loci were those used by Arseneault (2011).

REFERENCES

Abbot P, Abe J, Alcock J, Alizon S, Alpedrinha JAC, Andersson M, Andre JB, van Baalen M, Balloux F, Balshine S, Barton N, Beukeboom LW, Biernaskie JM, Bilde T, Borgia G, Breed M, Brown S, Bshary R, Buckling A, Burley NT, Burton-Chellew MN, Cant MA, Chapuisat M, Charnov EL, Clutton-Brock T, Cockburn A, Cole BJ, Colegrave N, Cosmides L, Couzin ID, Coyne JA, Creel S, Crespi B, Curry RL, Dall SRX, Day T, Dickinson JL, Dugatkin LA, Mouden CE, Emlen ST, Evans J, Ferriere R, Field J, Foitzik S, Foster K, Foster WA Fox CW, Gadau J, Gandon S, Gardner A, Gardner MG, Getty T, Goodisman MAD, Grafen A, Grosberg R, Grozinger CM, Gouyon PH, Gwynne D, Harvey PH, Hatchwell BJ, Heinze J, Helantera H, Helms KR, Hill K, Jiricny N, Johnstone RA, Kacelnik A, Keirs ET, Kokko H, Komdeur J, Korb J, Kronauer D, Kümmerli R, Lehmann L, Linksvayer TA, Lion S, Lyon B, Marshall JAR, McElreath R, Michalakis Y, Michod RE, Mock D, Monnin T, Montgomerie R, Moore AJ, Mueller UG, Noë R, Okasha S, Pamilo P, Parker GA, Pedersen JS, Pen I, Pfennig D, Queller DC, Rankin DJ, Reece SE, Reeve HK, Reuter M, Roberts G, Robson SKA, Roze D, Rousset F, Rueppell O, Sachs JL, Santorelli L, Schmid-Hempel P, Schwarz MP, Scott-Phillips T, Shellmann-Sherman J, Sherman PW, Shuker DM, Smith J, Spagna JC, Strassmann B, Suarez AV, Sundström L, Taborsky M, Taylor P, Thompson G, Tooby J, Tsutsui ND, Tsuji K, Turillazzi S, Úbeda F, Vargo EL, Voelkl B, Wenseleers T, West SA, West-Eberhard MJ, Westneat DF, Wiernasz DC, Wild G, Wrangham R, Young AJ, Zeh DW, Zeh JA, Zink A (2011). Inclusive fitness theory and eusociality. *Nature* 471: E1-E4

Aitken N, Smith S, Schwarz C Morin PA (2004). Single nucleotide polymorphism discovery in mammals: a targeted-gene approach. *Molecular Ecology* 13: 1423-1431

Alcock J (1998). Animal Behaviour (6th ed). Sunderland, Massachusetts: Sinauer Associates.

Alexander RD (1974). The evolution of social behaviour. *Annual Review of Ecology and Systematics* 5:325-383

Amos W, Whitehead H, Ferrari MJ, Glockner-Ferrari DA, Payne R, Gordon J (1992). Restrictable DNA from sluoghed cetacean skin; it's potential for use in population analysis. *Marine Mammal Science* 8: 275-283

Arnbom T, Papastavrou V, Weilgart LS, Whitehead H (1987). Sperm whales react to an attack by killer whales. *Journal of Mammology* 68: 450-453

Arseneault K (2011). Relatedness within and between sperm whale (*Physeter macrocephalus*) social groups [Honours thesis]. Halifax NS: Saint Mary's University

Avise JC (1994). Molecular Markers: Natural History and Evolution. New York, NY: Chapman and Hall.

Best PB (1979). Social organization in sperm whales, *Physeter macrocephalus*. In: *Behaviour of Marine Animals, vol. 3* (eds. Winn HE, Olla BL), 227-289. New York: Plenum Press.

Best PB, Canham PAS, Macleod N (1984). Patterns of reproduction in sperm whales, *Physeter macrocephalus*. *Reports of the International Whaling Commision (special issue)* 6: 51-79

Boomsma JJ, Beekman M, Cornwallis CK, Griffin AS, Holman L, Hughes WOH, Keller L, Oldroyd BP, Ratnieks FLW (2011) Only full-sibling families evolved eusociality. *Nature* 471: E4-E5

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* 18: 249-256

Budowle B, Smith J, Moretti T, DiZinno D (2000). DNA Typing Protocols: Molecular Biology and Forensic Analysis. Natick, MA: Eaton Publishing.

Butler JM (2005). Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers (2nd ed). Burlington, MA: Elsevier Academic Press.

Caldwell MC, Caldwell DK (1966). Epimeletic (care-giving) behaviour in Cetacea. In: *Whales, Dolphins, and Porpoises* (ed. Norris KS), Berkley: University of California Press.

Christal J, Whitehead H, Lettevall E (1998). Sperm whale social units: variation and change. *Canadian Journal of Zoology* 76: 1431-1440

Clapham P (2001). Why do baleen whales migrate? A response to Corkeron and Connor. *Marine Mammal Science* 17: 432-436

Clutton-Brock T (2009). Structure and function in mammalian societies. *Philosophical Transactions of the Royal Society B* 364: 3229-3242

Collins FS, Brooks LD, Chakravarti A (1998). A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research* 8: 1229-1231

Connor RC (2000). Group living in whales and dolphins. In: *Cetacean Societies: Field Studies of Dolphins and Whales* (eds. Mann J, Connor RC, Tyack PL, Whitehead H), Chicago: University of Chicago Press.

Corkeron PJ, Connor RC (1999). Why do baleen whales migrate? *Marine Mammal Science* 15: 1228-1245

Cranford TW (1999). The sperm whale's nose: sexual selection on a grand scale? *Marine Mammal Science* 15: 1133-1157

Csardi G, Nepusz T (2006). The igraph software package for complex network research. InterJournal, Complex Systems 1695. URL http://igraph.sf.net

Darwin C (1859). The Origin of Species. London: Murray

Deagle BE, Eveson JP, Jarman SN (2006). Quantification of damage in DNA recovered from highly degraded samples - a case study on DNA in faeces. *Frontiers in Zoology* 3: 11

DeWoody JA (2005). Molecular approaches to the study of parentage, relatedness, and fitness: practical applications for wild animals. *The Journal of Wildlife Management* 69: 1400-1418

Ekblom R, Galindo J (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107: 1-15

Ellegren H (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* 5: 435-445.

Emlen ST (1994). Benefits, constraints and the evolution of the family. *Trends in Ecology and Evolution* 9: 282-285

Emlen ST (1995). An evolutionary theory of the family. *Proceedings of the National Academy of Sciences of the United States of America* 92: 8092-8099

Engelhaupt D, Hoelzel AR, Nicholson C, Frantzis A, Mesnick S, Gero S, Whitehead H, Rendell L, Miller P, De Stefanis R, Canadas A, Airoldi S, Mignucci-Giannoni A (2009). Female philopatry in costal basins and male dispersion across the North Atlantic in a highly mobile marine species, the sperm whale (*Physeter macrocephalus*). *Molecular Ecology* 18: 4193-4205

Ferriere R, Michod RE (2011). Inclusive fitness in evolution. *Nature* 471: E6-E8

Fisher RA (1930). The Genetical Theory of Natural Selection. Oxford: Oxford University Press.

Freeland JR (2005). Molecular Ecology. West Sussex: John Wiley and Sons.

Garvin MR, Saitoh K, Gharrett AJ (2010). Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* 10: 915-934

Gero S, Engelhaupt D, Rendell L, Whitehead H (2009). Who cares? Between-group variation in alloparental caregiving in sperm whales. *Behavioural Ecology* 29: 838-843

Gero S, Engelhaupt D, Whitehead H (2008). Heterogenous social associations within a sperm whale, *Physeter macrocephalus*, unit reflect pairwise relatedness. *Behavioural Ecology and Sociobiology* 63: 143-151

Gilson A, Syvanen M, Levine K, Banks J (1998). Deer gender determination by polymerase chain reaction: validation study and application to tissues, bloodstains, and hair forensic samples from California. *California Fish and Game* 84: 159-169

Giordano M, Oefner PJ, Underhill PA, Sforza LLC, Tosi R, Richardi PM (1999). Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. *Genomics* 56: 247–253

Goldenberg EM, Bickel A, Weihls P (1996). Effect of highly fragmented DNA on PCR. *Nucleic Acids Research* 24: 5026-5033

Hamilton WD (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology* 7: 1-52

Herre EA, Wcislo WT (2011). In defence of inclusive fitness theory. *Nature* 471: E8-E9

Hinde RA (1976). Interactions, relationships, and social structure. *Man* 11: 1-17

Jefferson TA, Stacey PJ, Baird RW (1991). A review of killer whale interactions with other marine mammals: predation to co-existance. *Mammal Review* 21: 151-180

Kalinowski ST, Taper ML, Marshall TC (2007). Revising how the computer program CERVUS accomodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099-1106

Kawakami T (1980). A review of sperm whale food. *Scientific Reports of the Whales Research Institute* 32: 199-218

Li CC, Weeks DE, Chakravarti A (1993). Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity* 43: 45-52

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

Lynch M, Ritland K (1999). Estimation of pairwise relatedness with molecular markers. *Genetics* 152: 1753-1766

Lyrholm T, Leimar O, Johannson B, Gyllensten U (1999). Sex-biased dispersal in sperm whales: contrasting mitochondrial and nuclear genetic structure of global populations. *Proceedings of the Royal Society of London B* 266: 347-354

Mann J, Connor RC, Tyack PL, Whitehead H (2000). Cetacean Societies: Field Studies of Dolphins and Whales. Chicago: University of Chicago Press.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639-655

Mesnick SL, Taylor BL, Archer FI, Martien KK, Treviño GE, Hancock-Hanser BL, Medina SCM, Pease VL, Robertson KM, Straley JM, Baird RW, Calambokids J, Schorr GS, Wade P, Burkanov V, Lunsford CR, Rendell L, Morin PA (2011). Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites, and mitochondrial DNA. *Molecular Ecology Resources* 11: 278-298

Morin PA, Aitken NC, Rubio-Cisneros N, Dizon AE, Mesnick S (2007). Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*). *Molecular Ecology Notes* 7: 626-630

Morin PA, Luikart G, Wayne RK, the SNP workshop (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19: 208-216

Morin PA, McCarthy M (2007). Highly accurate SNP genotyping from historical and low-quality samples. *Molecular Ecology Notes* 7: 937-946

Morin PA, Nestler A, Rubio-Cisneros NT, Robertson KM, Mesnick SL (2005). Interfamilial characterization of a region of the ZFX and ZFY genes facilitates sex determination in cetaceans and other mammals. *Molecular Ecology* 14: 3275-3286

Myakishev MV, Khripin Y, Hu S, Hamer DH (2001). High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. *Genome Research* 11: 163-169

Noren DP, Mocklin JA (2012). Review of cetacean biopsy techniques: Factors contributing to successful sample collection and physiological and behavioural impacts. *Marine Mammal Science* 28: 154-199

Nowak MA, Tarnita CE, Wilson EO (2010). The evolution of eusociality. *Nature* 466: 1057-1062

Orr HA (2009). Fitness and its role in evolutionary genetics. *Nat Rev Genet* 10: 531-539

Ortega-Ortiz JG, Engelhaupt D, Winsor M, Mate BR, Rus Hoelzel A (2012). Kinship of long-term associates in the highly social sperm whale. *Molecular Ecology* 21: 732-744

Papastavrou V, Smith SC, Whitehead H (1989). Diving behavior of the sperm whale, *Physeter macrocephalus*, off the Galapagos Islands. *Canadian Journal of Zoology* 67: 839-846

Pinela AM, Quérouil S, Magalhães S, Silva MA, Prieto R, Matos JA, Santos RS (2009). Population genetics and social organization of the sperm whale (*Physeter macrocephalus*) in the Azores inferred by microsatellite analyses. *Canadian Journal of Zoology* 87: 802-813

Pitman RL, Ballance LT, Mesnick SI, Chivers SJ (2001). Killer whale predation on sperm whales: observations and implications. *Marine Mammal Science* 17: 494-507

R Development Core Team (2008). R: A language and environment for statistical computing. R Vienna, Austria: Foundation for Statistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reidman ML (1982). The evolution of alloparental care and adoption in mammals and birds. *The Quarterly Review of Biology* 57: 405-435

Rice DW (1989). Sperm whale *Physeter macrocephalus* Linnaeus, 1758. In *Handbook of Marine Mammals, vol. 4: River Dolphins and the Larger Toothed Whales* (eds. Ridgeway SH, Harrison R). London: Academic Press.

Richard KR, Dillon MC, Whitehead H, Wright JM (1996). Patterns of kinship in groups of free-living sperm whales (*Physeter macrocephalus*) revealed by multiple molecular genetic analyses. *Proceedings of the National Academy of Sciences of the United States of America* 93: 8792-8795

Ryman N, Palm S (2006). POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology* 6: 600-602

Sachs JL, Mueller UG, Wilcox TP, Bull JT (2004). The evolution of cooperation. *The Quarterly Review of Biology* 79: 135-160

Sambrook J, Russell DW (2001). Molecular Cloning: A Laboratory Manual. NY: Cold Spring Harbor Laboratory Press

Sanger F, Nicklen S, Coulson R (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74: 5463-5467

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011). Single nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol Ecol Res* 11: 1-8

Seutin G, White BN, Boag PT (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* 69: 82-90

Strassmann JE, Page Jr. RE, Robinson GE, Seeley TD (2011). Kin selection and eusociality. *Nature* 471: E5-E6

Sunnucks P (2000). Efficient genetic markers for population biology. *Trends in Ecology and Evolution* 15: 199-203

Sunstad PD, Simmons MJ (2012). Principles of Genetics (6th ed). Hoboken, New Jersey: John Wiley and Sons.

Taberlet P, Waits LP, Lulkart G (1999). Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution* 14: 323-327

Teloni V, Mark JP, Patrick MJO, Peter MT (2008). Shallow food for deep divers: Dynamic foraging behavior of male sperm whales in a high latitude habitat. *Journal of Experimental Marine Biology and Ecology* 354: 119-131.

Trivers RL (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology* 46: 35-57

Valsecchi E, Glockner-Ferrari D, Ferrari M, Amos W (1998). Molecular analysis of the efficiency of sloughed skin sampling in whale population genetics. *Molecular Ecology* 7: 1419-1422

Van de Casteele T, Galbusera P, Matthysen E (2001) A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology* 10: 1539-1549

Vignal A, Milan D, SanCristobal M, Eggen A (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution* 34: 275-305

Watkins WA, Daher MA, DiMarzio NA (2002). Sperm whale dives tracked by radio tag telemetry. *Marine Mammal Science* 18: 55-68

Watwood SL, Miller PJO, Johnson M, Madsen PT, Tyack PL (2006). Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). *Journal of Animal Ecology* 75: 814-825

Weilgart LS, Whitehead H (1988). Distinctive vocalizations from mature sperm whales (*Physeter macrocephalus*). *Canadian Journal of Zoology* 66: 1931-1937

Wey T, Blumstein DT, Shen W, Jordan F (2008). Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal behaviour* 75: 333-344

Whitehead H (1993). The behaviour of mature male sperm whales on the Galápagos Islands breeding grounds. *Canadian Journal of Zoology* 71: 689-699

Whitehead H (1996). Babysitting, dive synchrony, and indications of alloparental care in sperm whales. *Behavioural Ecology and Sociobiology* 38:237-244

Whitehead H (2003). Sperm Whales: Social Evolution in the Ocean. Chicago: University of Chicago Press

Whitehead H, Gordon J, Mathews EA, Richard KR (1990). Obtaining skin samples from living sperm whales. *Marine Mammal Science* 6: 316-326

Whitehead H, Waters S, Lyrholm T (1991). Social organization of female sperm whales and their offspring: constant companions and casual aquaintances. *Behavioural Ecology and Sociobiology* 29: 385-389

Whitehead H, Weilgart L (2000). The Sperm Whale. In: *Cetacean Societies: Field Studies of Dolphins and Whales* (eds. Mann J, Connor RC, Tyack PL, Whitehead H), 154-172. Chicago: University of Chicago Press

Wilson DS, Wilson EO (2007). Rethinking the theoretical foundation of sociobiology. *The Quarterly Review of Biology* 82: 327-348

Williams GC (1966). Adaptation and Natural Selection. Princeton, New Jersy: Princeton University Press.

Woodroffe R, Vincent A (1994). Mother's little helpers: patterns of male care in mammals. *Trends Ecol Evol* 9:294–297

Wynne-Edwards VC (1962). Animal Dispersion in Relation to Social Behaviour. Edinburgh: Oliver & Boyd