14

Identifying units to conserve using genetic data

Barbara L. Taylor, Karen Martien, and Phillip Morin

14.1 The biology of structure and the role of genetics

Genetics has provided new tools to understand units to conserve (UTC) for marine mammals. Direct observations of processes important to the management and understanding of marine mammal biology are impossible or costly for most species. Although insights to marine mammal movements have been gained using photographic identification and tagging (see Chapter 2), particularly for pinnipeds, these tools are impractical for many species, particularly those with more pelagic ranges and/or higher abundance. The long generation times of marine mammals also present challenges because studies conducted over just a few generations may not be informative about questions involving longer time-scales. Marine mammal taxonomy lags behind work on terrestrial species because of issues with inaccessibility coupled with practical difficulties in maintaining collections of large skulls. Genetics, therefore, becomes a much more important tool for marine mammal scientists than for those studying terrestrial species, since genetic data often provide the only means to address pressing questions in a timely fashion.

Nevertheless, genetics are not always the right tool for the job. It is important to discriminate when genetic studies are most useful and when genetic studies have been applied properly. Like other fields, genetics has special terms that we indicate in bold defined in Box 14.1 at the end of this chapter. Although genetic technologies and analytical methods will continue to change, the foundation of the relationship between biology and genetic patterns and the basic principles of project design will not. Project design generally is optimized through the following steps: (1) frame the question; (2) choose the marker(s) with consideration of how life history affects the markers; (3) if possible, do power analysis to aid in estimating sample size and the number of markers required to address the question; (4) obtain/choose the samples; (5) conduct the laboratory analysis;

(6) rigorously assess data quality and return to step 5 as needed; (7) employ appropriate analytical methods; and (8) interpret and write-up results.

While each step is necessary, none is more critical than carefully framing the research question. This careful reflection on project design is especially important because samples often result as by-products of other research rather than as the primary goal of the research. We begin, therefore, by describing general properties of marine mammal biology and genetics that will provide the foundation for good genetic project design. Additional examples and interpretation of population genetic structure are given in Hoelzel *et al.* (2002).

Although genetic markers are covered in detail below, some basics are needed to understand how genetic patterns relate to marine mammal biology. There are two types of DNA that yield very different insights: mitochondrial DNA (mtDNA) and nuclear DNA (nDNA). mtDNA is uniparentally inherited from the mother and has some unique properties useful in marine mammal research. Marine mammals are all relatively long-lived with high survival rates and relatively low reproductive rates. Because a female produces few offspring, her fitness is increased by investing in maternal care. This care may extend over several years, allowing learned behaviours, ranging from feeding strategies to migratory routes, to be transferred to offspring (both male and female). These learned behaviours are reflected in the patterns of mtDNA. These mtDNA patterns are important for identifying some types of unit to conserve discussed in more detail in Section 14.2.3, Demographically independent populations.

Since nDNA is biparentally inherited, it therefore provides data on male and female **gene flow**, and thereby indicates the structure of the breeding population. Most traits subject to selection result from the expression of nDNA, making these markers appropriate for questions relating to population structure that is most important to evolutionary significance.

Inferences drawn from genetic data are strengthened by good sampling design, which in turn depends on considering the biology of the species in question. For example, obtaining samples during the mating season can greatly simplify interpretation of data for answering questions about population structure. Without such 'pure state' data, interpretation of data from samples obtained on a feeding ground, where individuals may be present from multiple breeding populations, may prove impossible. Even at a breeding site, careful consideration should be given to obtaining a random sample if strong site fidelity is present. For example, some otariid females may return to the beach of their birth. If the question is to compare population structure from one island to the next, then sampling should avoid obtaining all samples from a single beach at each island, as the level of relatedness on a single beach may be higher than between beaches within a single island. Because estimates of differences between islands are based on comparing within to between island frequencies, inappropriate sampling could produce biased results indicating stronger separation than is actually present.

14.2 Scale—units to conserve

Are the mammal-eating killer whales part of a single **lineage** or have they originated multiple times? Is the form of the Irrawaddy dolphin found off Australia a new species? How many demographically independent populations of coastal spotted dolphins are there along the central American coasts? Do all humpback whales in the North Atlantic go to a single breeding ground? Is it likely that declining Steller sea lion rookeries in the western Aleutians will be rescued by rookeries in Russia or the eastern Aleutians? All these questions have been addressed using genetic data. Most of these questions involve identifying units to conserve. However, an accurate understanding of population structure is critical to correctly estimate trends in abundance, to assess the impact of a particular mortality source like fisheries by-catch or whaling, or to model the role of marine mammals in an ecosystem. Framing the research question involves first identifying the scale, either spatial or temporal, that is of interest. We begin at the longest time and greatest spatial scale with taxonomic units and work toward the shortest time and smallest spatial scale with demographically independent populations.

14.2.1 Taxonomy

Laws and treaties, including the Convention on the International Trade of Endangered Species (CITES) and the red list—developed by the International Union for the Conservation of Nature (IUCN) and the US Endangered Species Act (ESA)—and similar domestic legislation, make an implicit assumption that scientists have already properly defined species and subspecies. However, new species of cetaceans are being named nearly every year, including the North Pacific right whale (Rosenbaum *et al.* 2000), Perrin's beaked whale (Dalebout *et al.* 2002), and the Australian snubfin dolphin (Beasley *et al.* 2005). These species' descriptions were based largely or solely on genetic data. This reliance on genetic data departs from previous practice in taxonomy (the naming of species and subspecies), which has primarily used morphology, especially skull measurements, to designate species.

Designation of a species without corroborating morphological data, as in the right whale case, remains controversial among taxonomists. This case demonstrates why traditional taxonomy lags behind and may not be appropriate for species' designations involving marine mammals: adequate skull collections are not, and often will not become, available. The precise definition of what constitutes a 'species' or 'subspecies' has long been a subject of debate. Two dozen 'species concepts' exist (Mayden 1997) that can result in widely varying numbers of species when applied to any given dataset (Sites Jr and Marshall 2003). This debate reflects the underlying reality that species (and subspecies) boundaries are inherently 'fuzzy' and are not amenable to a one-size-fits-all definition (Hey *et al.* 2003).

To promote more timely progress in cetacean taxonomy, a workshop was held in 2004 to develop and agree on new definitions and criteria for species and

subspecies (Reeves *et al.* 2004). Workshop participants agreed that: (1) multiple species concepts should be acknowledged; (2) two congruent lines of evidence should be required to define a species, which would ideally incorporate both morphological and genetic data but could be based on two independent genetic markers; (3) only a single line of appropriate evidence should be required to define a subspecies and this may be a genetic marker; and (4) subspecies could be either a geographic form or an incipient species. Because these guidelines validate the use of genetic data in taxonomic designations, they pave the way for a more timely resolution of taxonomic uncertainty in many marine species for which substantial skull morphometric datasets are unlikely to be obtained in the foreseeable future.

14.2.2 Evolutionary significance

There are several types of units that are below the species level but are still likely to be following independent evolutionary paths: subspecies and units representing an important evolutionary component of the species. The latter units go by several names including evolutionarily significant units (ESUs) and distinct population segments (see USFWS 1996 for a full definition of DPS) under the US ESA and 'regional populations' by the IUCN (Gärdenfors *et al.* 2001). In this text we use the ESU term to broadly represent units needed to conserve the essential genetic variability for future evolutionary potential. Thus, the vision is to maintain sufficient evolutionary potential that species in the future can respond to environmental challenges through adaptation as well as they can today. The division between subspecies and the ESU-type unit is unclear given the new criteria for cetacean subspecies, as described above, but it seems likely that they differ by a matter of time elapsed since separation. For subspecies, the most closely related units have been separated long enough that morphological differences may accrue. ESUs, in contrast, are experiencing sufficiently low gene flow that local adaptation may occur. A rule of thumb often used to distinguish ESUs is that the level of gene flow between them is less than one disperser per generation (Gärdenfors *et al.* 2001).

14.2.3 Demographically independent populations (DIPs)

Demographically independent populations (DIPs) are units for which the internal dynamics within each group are more important to the group's maintenance than immigration from neighbouring groups. Units at this scale have been referred to by many names, including 'management units' (Moritz 1994), 'stocks', and 'population stocks' (Angliss and Wade 1997). DIPs are important for many management issues focused on shorter temporal scales, on the order of a human lifespan. Many marine mammals are highly social animals that form associations to exploit a marine habitat that varies over short time-scales of months to ones of decades. It is highly unlikely that individuals move about randomly like molecules and simply diffuse across space. Instead, a far more likely structure is that marine mammals operate like their terrestrial cousins and form herds that move according to their prey, predators,

and needs to rear their young in a safe environment. For some species, maximizing fitness has resulted in concentrated breeding grounds thousands of miles from their feeding grounds, as is the case for North Atlantic humpback whales (Palsbøll *et al.* 1997). For other species, prey specialization has resulted in localized resident populations, as is the case for bottlenose dolphins in the Gulf of Mexico (Sellas *et al.* 2005).

Because structure at this level exists on a continuum, framing questions about DIPs is difficult for researchers. There is no one level of discreteness that applies to all situations. DIPs are relevant to both applied questions (e.g. are fishery kills too high for a local population (DIP) to sustain?) and academic questions (e.g. are Steller sea lion populations declining because of shifts in prey distribution?). It is rarely the case that researchers are interested in the question 'Is there population structure or not?', yet this is the question posed by many analytical methods. There are some general principles of how genetic differentiation relates to gene flow that assist researchers in framing questions relating to DIPs and assessing whether a genetics approach will provide useful insights.

DIPs are typified by a genetic pattern of differences in allele frequencies. The magnitude of genetic differentiation between two groups (F_{ST}) is often represented by the formula

$$F_{ST} = 1/(4N_e m + 1) = 1/(4N_e dT + 1) \text{ for nuclear makers, or:} \quad (14.1)$$

$$F_{ST} = 1/(2N_e m + 1) = 1/(2N_e dT + 1) \text{ for mitochondrial markers} \quad (14.2)$$

where N_e is the **effective population size**, d is the annual **dispersal** rate, T is **generation time**, and $m = d*T$. The value of F_{ST} ranges between zero, with no differentiation, to one, when there is no overlap between the strata compared.

This formula demonstrates how the large abundances and long generation times of many marine mammal species work to reduce the level of differentiation expected for a given dispersal rate. Consider, for example, comparing two strata (putative populations) with moderate abundances of 1000, an even sex ratio, and a generation time of 10 years with one individual dispersing per year. For nDNA, (eqn 14.1) $F_{ST} = 0.000025$ and for mtDNA (eqn 14.2) $F_{ST} = 0.0001$. The researcher who wants to detect population structure at this level needs to distinguish between $F_{ST} = 0$ (no population structure) and at best $F_{ST} = 0.0001$ (worthy of designation as a DIP). Both geneticists and managers need to be wary of studies attempting to detect population structure with small sample sizes because low statistical power may make interpretation difficult. In fact, if increasing sample sizes to appropriate levels are not possible, genetic studies may be inappropriate (see Section 14.4, Analytical methods).

Although F_{ST} is a useful parameter because it gives the researcher an ideal magnitude of differentiation, there are some important caveats. All models make assumptions that simplify from biological reality. One assumption that is particularly

important to be aware of for marine mammals is the assumption that populations are in equilibrium. While all populations naturally fluctuate through time, most marine mammals have recently been hunted to small fractions of their normal range of abundances. There are two effects for which researchers should carry out sensitivity tests to guard against false conclusions. The first effect is a case for a population that is typically at high abundance, but has recently been severely reduced in number. This population will retain the genetic diversity from times of high abundance, and estimating statistical power to detect structure should use the historical effective population size. The second effect is a case for a population with a long generation time and sampling that is not random with respect to age. Rapidly growing populations will have cohorts that are more related to one another than would be the case in a stable population. The interpretation of genetic data when populations are out of equilibrium is difficult and analytical methods are still in their infancy. Since analytical methods are being rapidly developed, researchers should make a practice of seeking out new analytical methods that may be more appropriate for the case-specific conditions of their question.

In studies of marine mammals, researchers are often better able to detect fine-scale population structure with mitochondrial sequence data than with nuclear markers. There are several possible explanations for this pattern. First, because the mitochondrial **genome** is **haploid**, the rate of **genetic drift** in mitochondrial markers is double that of the **diploid** nuclear genome. This difference is reflected in the factor-of-two difference in the denominators of equations 14.1 and 14.2. Second, because only females pass their mitochondrial genome on to their offspring, the effective population size for the mitochondrial genome is half that of the nuclear genome (assuming an even sex ratio and random mating). Thus, the haploid, uniparentally inherited nature of the mitochondrial genome means that, for a given level of dispersal, the expected degree of differentiation for mitochondrial markers is four times greater than for nuclear markers.

Finally, male-biased dispersal or male-mediated gene flow can result in stronger differentiation at mitochondrial markers than nuclear markers. Due to its strictly maternal inheritance, the mitochondrial genome only reflects the population structure in the female component of the population. Thus, if females exhibit a high degree of site fidelity while males are more cosmopolitan, genetic structuring may be present in the mitochondrial genome but not in the nuclear genome. Similarly, the maternal transmission of mtDNA will affect patterns of differentiation in some highly **migratory** species. For instance, humpback whales exhibit strong, maternally transmitted fidelity to high-latitude feeding grounds, resulting in strong mitochondrial differentiation among those feeding grounds (Palsbøll *et al.* 1997). However, because whales from multiple feeding grounds converge on the same breeding ground, females are likely to breed with males from different feeding grounds, resulting in mixing of the nuclear genome (Palsbøll *et al.* 1997).

The North Atlantic humpback whale example nicely illustrates the utility of different markers to questions at different scales for units to conserve (UTCs). Consider a case where whales in the Gulf of Maine are being killed by entanglement in fishing gear and by ship strikes. In addressing the question of whether the local 'population' can sustain this mortality, it would be inappropriate to count whales from other discrete feeding grounds (like off Greenland) into mortality rate calculations because whales from other feeding grounds do not recruit into the Gulf of Maine feeding aggregation at demographically significant levels. Thus, mtDNA is both necessary and sufficient to address this question. Using nDNA would be neither necessary nor sufficient; and in this case would give an incorrect picture of structure, as whales from the demographically discrete feeding grounds could not be distinguished because of mixing of the nuclear genome on the breeding grounds. Thus, based on genetic criteria alone, whales from Greenland and the Gulf of Maine belong to different DIPs but the same ESU. The ESU question requires nDNA markers.

Some questions can only be addressed with nDNA. For example, consider a case where whaling is occurring on a feeding ground that has whales from multiple breeding grounds. If one breeding population is small, and hence their recovery could be delayed by excessive whaling, managers would need to know which breeding population the harvested whales were coming from. Such a determination would require the use of an assignment test, in which animals of unknown origin are assigned to populations on the basis of their genetic make-up. Because mtDNA is a single non-recombining unit with limited variability, most breeding populations have many shared mtDNA **haplotypes**. Consequently, it is generally not possible to identify the breeding population origin of samples using mtDNA. With nDNA, there is the possibility of using many independent markers to increase variability, thereby increasing the precision of population assignments.

14.3 Genetic markers

The toolbox of genetic markers continues to grow with the advances in genomic technologies and the availability of whole genome sequences. Some methods and markers (such as **allozymes** and **DNA fingerprinting**) have lost their appeal because of technical limitations or replacement by higher quality or more broadly applicable methods. We will focus on the markers that are currently most commonly used, or show the greatest promise for application, to detect genetic variation from the individual to the species level, all of which can be useful in defining units to conserve. Other reviews of genetic markers and applications contain additional information for application to molecular ecology questions that are not addressed here (e.g. Beaumont and Nichols 1996; Sunnucks 2000; Luikart *et al.* 2003). In the following sections, we will discuss the benefits and limitations of markers for each level of variation of interest for defining UTCs. Analytical methods to identify structure at various levels assume that markers are 'neutral'.

Thus, we do not discuss markers that code for actual proteins and therefore are subject to selection and focus on non-coding markers (though this does not ensure that they are selectively neutral). A summary of the applicability of different markers to each level of investigation is given in Table 14.1.

14.3.1 Mitochondrial DNA sequencing (mtDNA)

Mitochondrial DNA has been the primary molecular marker for **phylogenetics** and phylogeography for over 20 years. Several characteristics of mtDNA make it highly suitable for these types of studies:

- High copy number in cells, so that it can be easily amplified by **PCR** from even very poor samples.
- Maternal haploid inheritance with no recombination. Haploid status means that the effective population size of mtDNA is 25% the effective population size of nDNA, resulting in faster **lineage sorting** of haplotypes in populations. The smaller effective population size allows mtDNA to more rapidly resolve to **monophyly** after divergence of populations. Lineage sorting generally resolves to monophyly in $4N_e$ generations (Avise *et al.* 1987; Avise 1989), where N_e is the effective population size.
- Availability of PCR primers that are widely applicable across taxa for amplification of well-characterized portions of the mtDNA genome, including the non-coding control region (Ross *et al.* 2003), cytochrome b (Kocher *et al.* 1989), and the 'DNA Barcode' region of cytochrome c oxidase subunit I (CO-I) (Glaubitz *et al.* 2003; Waugh 2007).

Table 14.1 *Appropriate use, difficulty, and cost of molecular techniques. The number of Xs in each cell indicates the level of appropriateness.*

	mtDNA sequence	Microsatellite	SNP	AFLP	Nuclear locus sequence
Taxonomy	XXX	X	XX	XXX	XXX
ESUs	XXX	XXX	XXX	XXX	X
DIPs	XXX	XXX	XX	XXX	NA
Assignment tests	X	XXX	XX	XXX	NA
Individual ID, relatedness	NA	XXX	XX	XX	NA
Historical DNA	XXX	X	XXX	NA	NA
Difficulty	Low	Moderate	Low	*High	Low–moderate
Cost	Low	Moderate	**Low–moderate	Low	Moderate–high

Key: XXX, highly appropriate; XX, appropriate; NA, not appropriate.

* Initial methods development and optimization can be very labour-intensive, but subsequent typing can be rapid and easy.

** Cost depends on the number of SNPs used, **ascertainment,** and genotyping methods.

Mitochondrial DNA is especially useful for population structure at the species, ESU, and DIP levels, as the rate of mutation fixation and allele frequency differences, due to complete or partial reproductive isolation of populations resulting from genetic drift, are greater for mtDNA than for nuclear loci. For many mammalian species, sex-biased dispersal also favours differentiation of mtDNA among populations, when females are the non-dispersing sex.

Despite its wide use and utility, mtDNA also has some limitations. The relatively high mutation rate, especially in the non-coding control region, can lead to **homoplasy** that decreases the signal among more divergent taxa (Galtier *et al.* 2006). In addition, fragments of the mitochondrial genome frequently (in evolutionary time) become incorporated into the nuclear genome, creating copies that evolve at a different rate than the mtDNA, and which can confound analysis (Sorenson and Fleischer 1996; Bensasson *et al.* 2001). Finally, the fact that the whole mitochondrial genome still only represents a single non-recombining locus means that it represents only a single gene lineage, or **gene tree**, and may not represent the true phylogeny of the taxa. In some cases, lineage sorting of mtDNA can be at odds with the true taxonomy, especially with more widely divergent taxa (Murphy *et al.* 2001).

14.3.2 Microsatellites

Microsatellites are a class of highly repetitive and variable nDNA loci that are widely distributed in almost all genomes. They consist of tandemly repeated copies of one to six nucleotides, with the number of copies varying between alleles in the population. Most microsatellite loci used for population studies are di-, tri-, or tetra-nucleotide repeats, as these are the most variable and easiest to score (genotype). Several review papers describe the characteristics, strengths, and weaknesses of microsatellites (Jarne and Lagoda 1996; Schlötterer and Pemberton 1998).

The great benefit of microsatellites as nuclear markers is that they are typically highly **polymorphic** in populations, providing a high level of statistical power for individual identification (Herraeza *et al.* 2005; Seddon *et al.* 2005), paternity (Morin *et al.* 1994; Fung 2003), and population structure analyses (Luikart and England 1999) with relatively few loci, typically between 10 and 20. Microsatellite loci are often also conserved across taxa, so PCR primers designed to amplify a locus in one species will often work in closely related species (and sometimes across widely divergent taxa, e.g. S.S. Moore *et al.* 1991).

While by far the most widely used nuclear marker for analyses from the individual to the species level, microsatellites have several limitations that can be significant, and careful consideration should be made prior to starting a project. Microsatellite **genotypes** are obtained by PCR amplification followed by **electrophoresis** and **allele** size inference. Scoring of alleles can have many pitfalls, potentially resulting in population biases that can affect results (van Oosterhout *et al.* 2004; Hoffman and Amos 2005; Morin *et al.* 2009a). Comparison of genotypes across laboratories, technologies, or even just time can be difficult, as allele

size inferences can vary with the physical and chemical conditions during electrophoresis (LaHood *et al.* 2002; Davison and Chiba 2003). Finally, the mutation rates and patterns vary greatly among microsatellites, making application of an appropriate model of evolution difficult (Luikart and England 1999).

14.3.3 Single nucleotide polymorphisms (SNP)

Single nucleotide polymorphisms (SNPs) have only very recently been introduced to the fields of **systematics**, population genetics, and molecular ecology, but show great promise for accurate, efficient nuclear genotype data that provide some technical and analytical benefits over microsatellites (Brumfield *et al.* 2003; Morin *et al.* 2004). SNPs are simply mutations at individual nucleotide positions in the DNA sequence, and are the most abundant type of variation in the genome, being found typically every few hundred nucleotides, on average, in mammalian genomes (Aitken *et al.* 2004; Morin *et al.* 2004, 2007). This makes SNP discovery relatively easy (Aitken *et al.* 2004; Morin *et al.* 2004). Higher numbers of SNPs are needed relative to microsatellites because statistical power is related to the number of independent alleles in the whole set of markers, and SNPs are most often bi-allelic (Chakraborty *et al.* 1999, Krawczak 1999; Kalinowski 2002). For this reason, estimating the number and types of SNPs needed is an important first step in any study (Ryman *et al.* 2006). Recent applications of SNPs have shown that they can be used for individual identification, parentage, and population assignment (Herraeza *et al.* 2005; Seddon *et al.* 2005; Rohrer *et al.* 2007; Morin *et al.* 2009b), population structure (Turakulov and Easteal 2003; Seddon *et al.* 2005; C.T. Smith *et al.* 2005, 2007; Rosenblum and Novembre 2007; Narum *et al.* 2008), phylogenetics (Shaffer and Thomson 2007), and studies of historical samples such as museum collections of bone and baleen (Morin and McCarthy 2007).

SNPs are still in their infancy relative to mtDNA and microsatellites, especially for applications to natural populations, so there is much to learn about their benefits and limitations. We know that individual SNPs provide far less information than individual microsatellite markers for most applications, and this may limit their applicability for fine-scale population structure analysis (Morin *et al.* 2009b) and relatedness.

Methods are still being developed for more rapid and efficient SNP discovery, genotyping, and data analysis. SNPs have become the marker of choice for studies of humans, medical genetics, and economically important or model organisms, so adaptation for their use in conservation research of natural populations will no doubt follow quickly.

14.3.4 Amplified fragment length polymorphisms (AFLP)

AFLP methods have been around for many years, and are widely used for genetic diversity studies of plants and agriculturally important animals. The AFLP method makes use of **restriction endonucleases** to digest genomic DNA into small

fragments, and then preferentially amplifies a small portion of those fragments, some of which are variable within and among populations (Bensch and Akesson 2005). The great benefit of this method is that it is universally applicable to DNA from any organism without the need to design species-specific PCR **primers**. AFLPs are also able to generate a substantial number of variable loci with relatively few primer combinations. These characteristics have made AFLPs attractive for analysis of relatedness (J. Wang 2004), population structure, and phylogenetics (Giannasi *et al.* 2001; Kingston and Rosel 2004).

Despite the apparent benefits of AFLPs, they have not been widely used for animal studies (Bensch and Akesson 2005). One reason is the type of data generated; AFLP 'alleles' are **dominant**, so that the allelic state is unknown (i.e. each **locus** is scored as present or absent, so a **homozygote** for the present allele cannot be distinguished from a **heterozygote**). This means that, although the method produces data from many loci, the loci are less informative than **co-dominant** markers. The data are also nearly impossible to compare across studies, as all the data essentially need to be generated at the same time and with the same conditions to allow the many alleles to be compared and correctly sized. The quality of the data can also be highly dependent on DNA quality and strict adherence to laboratory and allele scoring protocols. While the methods are, in principle, straightforward, the robust application of them has often proven difficult.

14.3.5 Nuclear locus sequencing

To date, direct sequencing of multiple nuclear loci has been used primarily for phylogenetic studies (Palumbi and Baker 1994; Murphy *et al.* 2001). Sequencing of multiple individual loci from multiple individuals remains an expensive technology, both in terms of the laboratory work and the time it takes to analyse the data. Researchers must also know enough about the genes and genomes of the organisms studied to design PCR primers for each locus. One way to do this in the absence of whole genome projects for the target species is to use conserved primers designed based on the alignment of genes from multiple divergent species. This has been done for cetaceans and other mammals using the method of '**exon** priming, **intron** crossing' (EPIC; Palumbi and Baker 1994), also called 'comparative anchor tagged sequences' (CATS: Lyons *et al.* 1997; Aitken *et al.* 2004). The benefit of direct sequencing is that all variant sites can be detected for the phylogenetic information content of each locus (as opposed to genotyping one or a few SNPs in a locus). Most researchers have opted for less information from each locus (e.g. SNPs, AFLPs) while sampling more loci.

As with most genetic methods, however, sequencing technologies are changing rapidly. New 'next generation' methods are extremely high throughput, generating over a million base pairs of sequence from a single experiment (Margulies *et al.* 2005). Initial applications of this technology have been for rapid sequencing of whole genomes or transcriptomes (the transcribed genes), but it is also amenable to

highly parallel sequencing of smaller sets of genes (Meyer *et al.* 2008). While it may be a while before the costs come down sufficiently to make this approach practical for population level studies, it is likely that phylogenetics will benefit quickly. The question will be, how many genes are needed to obtain the true phylogeny (Gatesy *et al.* 2007)?

14.4 Analytical methods

As outlined above, there are many types of units to conserve that are of interest to conservation geneticists. Because these different units differ greatly in the expected magnitude of genetic differentiation, it is important to choose an analytical method appropriate to the unit that you need to define. In this section, we review the genetic patterns that are expected for each of the UTCs described above, and the analytical methods that are useful for detecting those patterns. Genetic analytical methods are rapidly evolving, with new approaches and critiques of existing methods published frequently. Consequently, we focus on the general types of analyses that are most useful in identifying UTCs at different levels rather than attempting to review specific methods. We conclude by discussing the importance of assessing the strength of an analysis for inferring population structure.

14.4.1 Choosing methods to match questions

Taxonomy

Analytical approaches to identifying taxonomic units are reviewed in Sites and Marshall (2003). Approaches can generally be categorized as tree-based and non-tree-based. Tree-based approaches consist of using a dataset to construct a phylogenetic tree and making inferences based on the topology of the resulting tree. Non-tree-based approaches rely on patterns in the genetic distance within and among hypothesized taxa and the identification of diagnostic characters. Kingston and Rosel (2004) used a combination of tree-based and non-tree-based approaches to examine genetic differentiation in pairs of sister taxa in the genera *Delphinus* and *Tursiops*.

In the past decade, considerable attention has been devoted to the development of **Bayesian** approaches to the reconstruction of phylogenetic trees (reviewed in Ronquist 2004). These powerful methods allow explicit incorporation of uncertainty into phylogenetic analyses. The availability of increasingly large datasets from ever-increasing numbers of taxa has driven the development of new analytical methods capable of dealing with the size and complexity of datasets now available. The 'supertree' (Sanderson *et al.* 1998; Bininda-Emonds 2004) and 'supermatrix' (Gatesy *et al.* 2004; de Queiroz and Gatesy 2007) approaches allow diverse datasets to be integrated into a single phylogenetic analysis. These approaches may eventually be able to provide resolution in groups such as the *Stenella–Tursiops–Delphinus* complex, which has long been problematic (LeDuc *et al.* 1999).

ESUs

The requirement that ESUs be evolutionarily distinct means that they will be strongly genetically differentiated. Therefore, many of the tree-based methods that are used for identifying taxonomic units are often applied to the identification of ESUs. ESUs often display a **phylogeographic signal**. However, species with very high genetic diversity, and those in which population structure has changed in the recent evolutionary past, may not exhibit a strong phylogeographic signal even when multiple ESUs are contained in a sample set.

Moritz (1994) suggested that in order for two groups to be considered separate ESUs, their mitochondrial sequences should exhibit a pattern of **reciprocal monophyly**. Again, while the presence of reciprocal monophyly between geographically, morphologically, or ecologically distinct groups can provide strong support for the designation of ESUs, the lack of this pattern does not mean that management as separate ESUs is unwarranted. Reciprocal monophyly develops as a result of lineage-sorting between two reproductively isolated groups. The rate at which lineage-sorting occurs is affected by the effective population size and generation time (Avise *et al.* 1987; Avise 1989). Because many marine mammal populations have very large abundances and some, particularly the large whales, have long generation times, complete reciprocal monophyly may take millions of years to develop, even between groups that are completely reproductively isolated. In fact, some currently recognized species of cetaceans are not reciprocally monophyletic (e.g. *Delphinus delphis* and *D. capensis*; LeDuc *et al.* 1999; Kingston and Rosel 2004). Of course, complete reproductive isolation is not expected between ESUs. Even the very low levels of gene flow expected between these units can be sufficient to prevent reciprocal monophyly from developing.

Bayesian clustering algorithms can be very powerful tools for identifying ESUs. These algorithms typically use multilocus genotypes to cluster samples into groups. Because they work at the level of the individual, these methods do not require the a priori stratification of samples nor any a priori hypotheses regarding the locations of population boundaries. Thus, these methods can be particularly valuable in cases where a species is continuously distributed over a large geographical area with few obvious barriers to movement, as is the case for many marine mammals. The development of Bayesian clustering algorithms is a rapidly developing research area, with new methods being introduced every year. While older methods made very limited use of spatial data (e.g. 'BAPS' (Corander *et al.* 2003, 2004), 'PARTITION' (Dawson and Belkhir 2001), 'STRUCTURE' (Pritchard *et al.* 2000; Falush *et al.* 2003)), many newer 'geographically constrained' methods use information regarding the spatial location of samples to constrain the clustering algorithm so that the final groups exhibit some degree of geographic cohesion (e.g. 'TESS' (Chen *et al.* 2007), 'GENECLUST' (Francois *et al.* 2006), 'GENELAND' (Guillot *et al.* 2005)). Because these geographically constrained methods are more likely to produce groups that are geographically contiguous, they may be more easily applied to management questions.

Bayesian clustering algorithms represent a subset of methods from a rapidly evolving discipline known as 'landscape genetics' (Manel *et al.* 2003). One of the primary aims of landscape genetics is the detection of genetic discontinuities, making many of its methods well suited to the definition of ESUs (and, to a lesser extent, DIPs). Although new methods and software are being developed every year, many of the available landscape genetic methods are included in the software package Alleles in Space (Miller 2005).

DIPs

The lower levels of genetic differentiation expected for DIPs make them far more challenging to detect using currently available analytical methods. Bayesian clustering algorithms hold some appeal, but performance analyses indicate that they will often be unable to detect the low levels of genetic differentiation expected for some DIPs (Latch *et al.* 2006; Waples and Gaggiotti 2006; Chen *et al.* 2007). Other landscape genetic methods for detecting genetic discontinuities are also well-suited to the task of defining DIPs (e.g. Monmonier's algorithm (Monmonier 1973), 'Wombling' (Crida and Manel 2007)), but have not been adequately tested to see how they will perform in this context.

To date, the most common analytical approach to defining DIPs is traditional null hypothesis testing. Hypothesis testing has been subject to growing criticism in the last decade. It requires the researcher to define hypothesized units in advance, a step that is often highly subjective and can greatly influence results (Taylor and Dizon 1999; Martien and Taylor 2003). Statistical power to detect genetic differentiation is often low, but is rarely estimated (Avise 1995; Taylor and Dizon 1996; Taylor *et al.* 1997; Martien and Taylor 2003). There also remains considerable uncertainty and disagreement regarding how and when to account for the impact of conducting multiple hypothesis tests (Perneger 1998; Ryman and Jorde, 2001; Moran 2003; Nakagawa 2004). The choice of α (the value at which a p-value is considered statistically significant) is arbitrary, and represents a hidden policy decision regarding the relative importance of Type I errors (concluding that there is population structure when there really is not) and Type II errors (failing to detect population structure that really is present) (Taylor and Dizon 1999).

Despite these drawbacks, hypothesis testing remains the primary tool for identifying DIPs. Statistics for estimating genetic differentiation in a hypothesis test generally fall into two categories—frequency-based statistics and distance-based statistics. Frequency-based statistics rely exclusively on allele or haplotype frequencies to estimate genetic differentiation. Distance-based statistics take into account both allele/haplotype frequencies and the evolutionary relationships between alleles/haplotypes. Thus, with a distance-based statistic, two strata whose haplotypes differ by an average of only one or two nucleotides will be considered less differentiated than two strata whose haplotypes differ by an average of three to five nucleotides, even if the differences in haplotype frequencies are comparable

between the two pairs of strata. Despite the fact that they seem to make better use of the available information (by taking account of both frequency data *and* evolutionary relationships), distance-based statistics tend to perform more poorly than frequency-based statistics for detecting low levels of genetic differentiation (Hudson *et al.* 1992; Goudet *et al.* 1996). An exception is when sample size is very small relative to genetic diversity (Chivers *et al.* 2002). In these cases, many alleles/haplotypes will be represented by a single individual and will therefore contribute no information to a frequency-based statistic. Reviews of the statistical power of different measures of differentiation under a variety of circumstances are provided by Hudson *et al.* (1992) for mtDNA and Goudet *et al.* (1996) for microsatellite markers.

The critical parameter for determining whether two putative populations are demographically independent is the annual rate of dispersal between them (d in equations 14.1 and 14.2 above). While a finding of statistically significant genetic differentiation can be used to infer a low dispersal rate, a more direct approach is to use genetic data to estimate dispersal rates. Estimates of dispersal rates, along with their uncertainties, are far more useful to managers and decision makers than the results of a hypothesis test (Taylor and Dizon 1999). They can easily be brought into a formal risk-analysis to evaluate the relative costs and benefits of various management actions. They can also be incorporated into population dynamics' and genetics' models, thereby informing other aspects of research on species that are of management concern.

Estimating dispersal rates remains an analytical challenge. Several Bayesian and maximum likelihood methods have been developed in recent years (Beerli and Felsenstein 1999, 2001; Fisher *et al.* 2002). However, they can be computationally prohibitive and have been found to perform poorly in some instances (Abdo *et al.* 2004). Nonetheless, as new methods are developed and computers continue to get faster, estimation of dispersal rates is likely to become the standard in defining DIPs.

14.4.2 Assessing the strength of inference

Attempts to define units to conserve are often plagued by a mis-communication between researchers and managers. All too often, the manager wants an answer to the question, 'Are there multiple units-to-conserve in this geographic region?', while the researcher sets off to address the question, 'Can I detect genetically distinct groups in this geographic region?' These two questions do not always have the same answer. There are many reasons why a geneticist might fail to accurately define UTCs. Awareness of these potential pitfalls can help researchers and managers alike to interpret the results of genetic studies better and use them to define UTCs.

Many studies of population structure are hampered by inadequate sample size or inadequate sample coverage. This can be particularly problematic in studies of

marine mammals, due to the difficulty of obtaining samples in a marine environment. Studies based on opportunistically collected samples often end up with sample distributions that better reflect the distribution of people (e.g. fishermen collecting samples from by-caught animals, beachcombers happening across stranded animals, etc.) than the distribution of the study species (Chivers *et al.* 2002). In species that exhibit social structure, inclusion of multiple samples collected from a single group could result in disproportionate representation of related individuals in some strata, which could bias the results of tests of genetic differentiation. In these cases, it may be necessary to exclude samples from potential relatives from some strata, further reducing the sample size. For pelagic species that rarely come into contact with people, collecting adequate samples for even phylogenetic studies may be extremely challenging.

The probability of correctly identifying population structure with traditional hypothesis tests can be assessed through power analysis (Martien and Taylor 2003; Ryman and Palm 2006; Morin *et al.* 2009b). Estimating statistical power is computationally burdensome. Faster computers and the development of new, user-friendly programs for estimating power will likely result in power analyses becoming more commonplace. However, in order to calculate statistical power, the researcher must specify a specific alternate hypothesis. In other words, the researcher must specify the level of differentiation s/he is trying to detect. This remains an important stumbling block to the useful application of power analysis (Morin *et al.* 2009b).

Many of the analytical methods outlined above require the a priori stratification of samples. If samples can be stratified in advance into groups that accurately reflect the underlying population structure, the likelihood of detecting genetic structure between those groups is much higher than if the analysis has to rely on methods that do not require stratification. If samples are stratified incorrectly, however, that error will propagate through and affect the entire analysis (Martien and Taylor 2003). Such an error could result in the definition of UTCs that does not accurately reflect underlying population structure or, worse, failure to define UTCs at all. The problem of a priori stratification can be partly dealt with by examining the sensitivity of results to stratification, or by starting with small strata and relying on careful consideration of patterns of significant and non-significant test results, to guide the grouping of strata into UTCs (e.g. O'Corry-Crowe *et al.* 2006). The sensitivity of results to the inclusion of data from stranded animals should also be examined, as the stranding location of an animal may be poorly correlated with its home range while alive.

DIPs remain the most challenging UTC to detect using genetic data. Most genetic analytical methods are developed by academic scientists for evolutionary studies. While these methods are likely to prove very useful for clarifying taxonomy and identifying ESUs, most are unlikely to detect the relatively high (from an evolutionary perspective) levels of gene flow typical of DIPs (Waples and Gaggiotti 2006). Because of its focus at the ecological rather than evolutionary scale, the

emerging field of landscape genetics holds promise for the development of analytical tools appropriate for the definition of DIPs. It is critical, however, that new methods be subject to rigorous performance testing prior to their application in a management setting. Although such performance testing is becoming more common (e.g. Abdo *et al*. 2004; Latch *et al*. 2006; Waples and Gaggiotti 2006; Chen *et al*. 2007), methods are rarely tested at the levels of differentiation typical of DIPs. A new project, known as TOSSM (testing of spatial structure methods) aims to fill this gap by testing the performance of methods in a management context (IWC, 2004; Martien *et al*. 2009). Results of the TOSSM project and other studies like it will be critical in the development of analytical tools that enable us to make full use of genetic data to define UTCs.

Box 14.1 Genetics' definitions

Allele: One of two copies of a gene or locus in a diploid genome. One allele is inherited from each parent.
Allozymes: Protein variants that cause them to migrate differently in gel electrophoresis, or to stain differently based on enzymatic activity.
Ascertainment: The process of finding and selecting polymorphic loci for use in population analysis. Ascertainment bias results from use of a limited number of genetically restricted populations or samples, and can bias the outcome of genetic studies.
Bayesian: A theory of statistics that estimates the probability of events. A genetics example would be using genetic data from putative populations to estimate the probability of different rates of dispersal (usually represented by a probability distribution over the range of plausible dispersal rates).
Co-dominant: Genetic markers for which both alleles in a sample can be detected, resulting in genotypes that may be homozygous for either allele, or heterozygous.
Diploid: Having two sets of chromosomes, one inherited from each parent.
Dispersal: Emigration and reproduction of individuals outside of the population in which they were born.
DNA fingerprinting: The use of combined genotypes from many loci to generate a multilocus genotype that is highly likely to be unique to an individual.
Dominant: Genetic markers for which only one allele can be determined, the other being recessive or undetected, though present. Genotyping of dominant markers results in only presence/absence genotypes.
Effective population size (N_e): For nuclear DNA, roughly the number of reproducing adults in the population, but is reduced if the sex ratio is unequal. For mtDNA, N_e is roughly the number of reproducing females in the population. Both are affected by variance in individual reproduction. A genetics text will give several formulae that may be applied depending on available data, including formulae to apply if temporal data are available.
Electrophoresis: The separation of DNA or proteins by size and electrical charge. The DNA or protein is placed at one end of a matrix or polymer and exposed to an

electrical current that moves it through the matrix, with smaller fragments moving faster than larger ones.
Exon: Any region of DNA within a gene that is transcribed to the final messenger RNA (mRNA), which is, in turn, translated into a protein. Exons are often, but not always, interspersed with introns within genes.
Gene flow: The movement of genetic material among populations through dispersal of individuals.
Gene tree: A tree-like figure consisting of branches that connect individuals, showing the evolutionary relationships or similarity among those individuals. Basal branches represent more ancestral relationships.
Generation time: In the context of genetic drift, generation time is the average age of parents.
Genetic drift: The process of changing allele frequencies due to random inheritance and loss of alleles in a finite population. The rate of genetic drift between populations is affected by the rate of mutation, the effective population size, and time since divergence.
Genome: The collective genetic material cells or mitochondrial organelles of an individual. The nuclear genome consists of the diploid chromosomes inherited from both parents in sexually reproducing organisms, and the mitochondrial genome consists of the circular chromosome found in the mitochondria within each cell.
Genotype: The designation of alleles for a locus. This usually refers to the combination of two alleles at each locus in an individual, but can also refer to the presence or absence of an allele for dominant loci.
Haploid: Having only one copy of a genome. The nuclear genome is only haploid in gametes, but the mitochondrial genome is always haploid.
Haplotypes: An identity given to the sequence or genotype of a haploid genome (an allele of a locus in a diploid genome).
Heterozygote: An individual locus containing copies of two different alleles.
Homoplasy: Repeated mutation at a site in DNA resulting in alleles that are identical in state but not identical by descent.
Homozygote: An individual locus containing two copies of the same allele.
Intron: Portions of genes that do not code for part of the gene product (protein), and that are not present in the final messenger RNA (mRNA) after the exon segments have been spliced together.
Lineage: For mtDNA, this refers to the lineal transmission of an mtDNA haplotype through a group of organisms. It can refer to any lineal descent of individuals or taxa.
Lineage sorting: The process by which reciprocal monophyly develops between two reproductively isolated groups. Bias in estimating dates of separation can occur when small populations are founded from a large diverse population and subsequently lose diversity and become fixed for the marker being used. Without knowing the population history, the genetic differences between the small populations will be falsely interpreted to represent the time taken to evolve such lineages.
Locus: A region of DNA defined by characteristics of the region (e.g. a gene), or by molecular markers (e.g. PCR primers for a microsatellite or SNP).

Migratory/migration: Seasonal movement between geographic areas used for breeding and feeding, or sometimes using different feeding grounds in different seasons. Geneticists sometimes incorrectly use 'migration' to mean 'dispersal' or 'gene flow' and therefore use the symbol 'm' in formulae (see equations 14.1 and 14.2).
Monophyly: All members of a taxa share a common ancestor to the exclusion of other taxa. This can apply to any character being used for phylogenetic analysis.
PCR: Polymerase chain reaction: a laboratory method for replicating a specific segment of DNA *in vitro* by the use of two PCR primers, DNA polymerase, and the building blocks of DNA, to cause an exponential increase in the number of copies of the DNA segment.
Phylogenetics: The study of the evolutionary relatedness among taxa, using genetic characters.
Phylogeographic signal: A pattern in which samples that are collected from the same geographical region also appear close to each other on a phylogenetic tree.
Polymorphic: Having two or more alleles in a population.
Primers: Synthetic short fragments of DNA used in PCR to specify the segment of a genome to be replicated.
Reciprocal monophyly: A group of taxa (species, subspecies, or populations) is considered reciprocally monophyletic when all members of each taxon share a most common recent ancestor that is not ancestral to any other taxa.
Restriction endonucleases: Enzymes commonly found in all organisms that function to splice or degrade DNA by cutting it into fragments at specific DNA sequence recognition sites. They are extracted from bacteria for use in laboratory methods that require DNA fragmentation.
Systematics: The study of the relationships among taxa through evolutionary time.