Two constraints impede our understanding of the biology of marine animals: First, because we routinely probe only a tiny fraction of the ocean’s biosphere, the diversity and abundance of animals in much of the world’s oceans are extremely undersampled. Second, in many cases even familiar species are observed only during portions of their life cycles. Intertidal barnacles and sea urchins do not seem very mysterious, yet a crucial portion of their life histories—when millions of fertilized eggs develop and disperse in the water column—remains poorly understood. Eggs and larvae of even common species are frequently difficult or impossible to identify on the basis of morphology. These facts make already difficult questions all the more intractable. What habitats are critical to the complete life cycle of a given species? What is the geographic range of individual planktonic larva? Do apparently isolated populations exchange migrants? We can answer these questions for remarkably few marine organisms.

This lack of knowledge cripples efforts to manage ocean resources. Fisheries typically target adult populations of organisms, often without knowledge of the source of recruits to the harvested stocks. Offshore fish stocks can be destroyed if management policies for coastal wetlands unwittingly permit unidentified key nursery areas to be disturbed. Without knowledge of spawning grounds and dispersal patterns within and between populations, the role of many habitats in fisheries production remains unknown, and, consequently, at risk.

Exciting new tools are continually providing novel insights into previously inaccessible aspects of the natural history of marine organisms. Examples include the use of trace elements, isotopes, and satellite tracking to monitor the movements of animals in order to better understand their geographic ranges for feeding and reproduction (Thorrold et al. 2002, Levin 2006, Teo et al. 2007). Although the capabilities of these methodologies continue to expand, the focus here will be limited to DNA- and RNA-based molecular genetic technologies. The goal of this article is to provide an overview of the rapidly expanding applications of molecular genetic approaches for understanding the ecology of marine animals.

The application of molecular tools in marine ecology spans a much broader range of topics than considered here. For example, Hofmann and Place (2007) discussed the use of microarrays in assessing the role of gene expression in adaptation to environmental change, and Edwards and Dinsdale (2007) discussed the use of metagenomics in analyses of marine microbial diversity and biogeochemical processes. Here the focus will be conceptually quite simple—DNA (and RNA) analyses permit estimates of genetic similarity among organisms. At the coarsest level, we can identify species. At a finer level, we can examine differences among populations within species. Levels of diversity within and between populations can also provide insights into historical population sizes and demography. Each level of inquiry presents challenges, and the molecular and analytical tools available for meeting these challenges are evolving rapidly.
Identification of species

By definition, species differ in their DNA sequences. Evolutionary analyses have clearly shown that the extent of sequence differentiation is correlated with the time since the taxa of interest shared a common ancestor. Consequently, DNA sequences can be used both to identify organisms and to reveal their evolutionary relationships. Although conceptually simple, the use of DNA sequencing for species identification has had a great impact in marine ecology. A large proportion of marine organisms have complex life histories, including larval forms that occupy different habitats and neither resemble nor behave like their respective adult forms (figure 1). In the absence of laborious (and frequently unsuccessful) efforts to culture organisms through their respective life cycles, we cannot identify larval forms to the species level for a vast majority of marine organisms. This limitation not only hinders our understanding of natural histories but also may result in erroneous assessments of commercial fisheries stocks, where larval surveys are used to predict future stock abundance.

The technology involved in obtaining DNA sequence data from even the smallest eggs and larval forms using the polymerase chain reaction (PCR) is now well developed (box 1). The challenge in determining the species identity of a specimen lies in matching its DNA sequence to a conspecific adult. To match larvae with adults requires a database with appropriate DNA sequences from correctly identified adult specimens. A fingerprint at a crime scene cannot be matched to the FBI fingerprint database if the perpetrator has never been fingerprinted. However, unlike the case with fingerprints, evolutionary relationships among DNA sequences mean that even when there is no match on file, DNA

Figure 1. Adult barnacles and urchins are familiar sights, but they are less so when they start life as planktonic larvae, as many marine animals do. (a) Adult barnacles on an exposed intertidal rock and (b) the planktonic nauplius stage of a barnacle. (c) The pluteus-stage larva of a sea urchin and (d) adult urchins on the sea floor. Photographs: Ronald S. Burton (a, b), Elisa Maldonato (c), and Dave Rudie (d).
Molecular tools have long played a role in marine ecological studies; protein polymorphisms were used to study fish populations more than 50 years ago (e.g., Frydenberg et al. 1965; see Burton 1983). But protein polymorphisms scored on electrophoresis gels (allozymes) all derive from amino acid substitutions in nuclear genes. This greatly limits the availability of genetic markers for ecological studies. Much of the nuclear genome does not encode proteins, and even in coding regions, many mutations do not change amino acid sequences. Allozyme studies also do not include any markers in the mitochondrial genome (mtDNA); since mtDNA is typically maternally inherited without recombination, it can provide a unique, sex-biased window into the genetic structure of natural populations (Avise 2004).

The single technological breakthrough that allowed widespread application of DNA-level studies to marine ecology is the polymerase chain reaction (PCR). The PCR makes it possible to amplify specific DNA fragments from complex genomes and from vanishingly small and often poorly preserved tissue samples; analyses of PCR-amplified fragments are the primary data for most recent population-level genetic analyses. Products from PCR amplification from different individuals are analyzed by determining either their sizes or their sequences. The most popular methodology for the examination of intraspecific population structure focuses on microsatellite loci, or simple sequence repeats. Microsatellite polymorphisms are variations in the number of simple di-, tri-, or tetranucleotide repeats found between two PCR primer sites (see panel a in the figure). However, the PCR primers required to amplify microsatellite loci must be derived separately for each target species, a laborious task of obtaining sequences of DNA that flank each microsatellite locus; this startup investment can be substantial and constrains the use of microsatellites. A contrasting approach (amplified fragment length polymorphisms, or AFLPs) permits rapid fingerprinting of entire genomes without any prior sequence data. In brief, this PCR-based method amplifies fragments from the DNA of a single individual on the basis of the presence or absence of short sequences (recognized by restriction endonucleases) at hundreds of sites across the genome (see panel b; Vos et al. 1995). Variation in the presence or absence of a particular band in the PCR products is scored across individuals. Studies using AFLP do not require species-specific primer sets.

Several approaches are available for directly assessing DNA sequence variation among individuals within and between species (see panel c). Direct sequencing of a DNA fragment amplified by PCR is the most straightforward. Sequence variation of mtDNA has become particularly accessible as “universal primers” have become more available (e.g., Folmer et al. 1994); some similarly universal primers are now available for nuclear genes as well (Jarman et al. 2002).
Box 1. (continued)

If the goal is to distinguish among alternative DNA sequences (as in species identification using barcode information), DNA hybridization methods are more efficient than direct sequencing. Short DNA probes can be synthesized that match species-specific sequences in the barcode database. Sample DNA (or RNA) is fluorescently labeled; that which complements the probe sequence will hybridize to the probe and can then be detected. For microbes, a phylochip containing thousands of taxon-specific probes can determine which species are present in a field-collected sample (DeSantis et al. 2007); a suspension array approach, with probes attached to color-coded microspheres, has been used to monitor marine phytoplankton (Ellison and Burton 2005). In a dramatic demonstration of hybridization technology, a robotic adaptation of this approach (using only a small number of probes) has already been deployed in Monterey Bay, California, to monitor marine invertebrate larvae (Jones et al. 2008).

sequence data can typically provide at least some taxonomic information about the specimen (e.g., identification of a related species).

The great value of DNA sequences for species identification has been widely embraced and is the basis of recent global efforts for establishing databases of DNA “barcodes.” These efforts focus on obtaining a short diagnostic DNA sequence of the same gene from all species and placing the sequences into a publicly accessible database; anyone who wishes to identify a specimen can sequence its same gene and compare the results with those in the database. Thanks to the efforts of Paul Hebert at the University of Guelph in Canada, a broad consensus has been reached regarding the gene that is the primary focus of this effort: the cytochrome c oxidase subunit1 (CO1) gene encoded in the mitochondrial DNA. This gene has several attributes that make it an excellent choice for species identification, including some highly conserved regions that support the development of “universal” PCR primers, and sufficient between-species variation to permit its use for discrimination of all but the most closely related taxa (see www.barcoding.si.edu). As discussed below, barcode databases are a great resource for marine ecology (Ratnasingham and Hebert 2007, Stoeckle and Hebert 2008); one such database is found at the Web site for the Fish Barcodes of Life, known as FISH-BOL (www.fishbol.org).

Identification of the early life stages of fish is of significant value to management. For example, Robertson and colleagues (2007) examined small juvenile scombrids in northwestern Australia to determine which of nine morphologically similar species were breeding locally; mitochondrial DNA (mtDNA) sequences of only six species were observed, suggesting that three species were spawning elsewhere. In cases like this, in which one needs to discriminate among only a small number of likely candidate species, DNA sequencing is often unnecessary; methods such as multiplex PCR, in which a single PCR includes multiple pairs of species-specific primers, can cheaply and efficiently discriminate among alternative identifications. Such methods can be adapted to shipboard use, giving the rapid results needed for more adaptive field sampling programs. Recently, Hyde and colleagues (2005, 2006) applied this approach to billfish eggs and larvae, which are difficult to distinguish morphologically. Shipboard PCR and electrophoresis identified seven striped marlin larvae collected off the coast of Kona, Hawaii. Subsequent sequencing found that each larva had a unique mtDNA sequence. Because eggs receive the mtDNA of their mother, this means that each larva had a different mother; hence, numerous female marlin had spawned locally, suggesting that the region is a spawning ground for that species.

Species identification is also important as a forensic tool for conservation and enforcement. Using DNA sequences, the source of marketed seafood or other traded commodities can be determined to the species level. In probably the most publicized example, Baker and Palumbi (1994) sequenced DNA from samples of whale meat purchased in Japanese markets and identified several samples of whale meats for sale that were derived from protected species. This work has ultimately led to a recent suggestion that DNA samples can be used to continuously monitor marketed cetacean products; sample sequences could then be used to identify the species composition of the products as well as to estimate the number of individuals taken for each species appearing in the market by using a clever adaptation of “mark-recapture” analysis (Baker et al. 2007).

DNA identification is playing an ever-growing role in the conservation and management of marine resources. Strong demand in Asian markets for shark fins has led to the practice of cutting fins from captured sharks and discarding the carcasses at sea. Shivji and colleagues (2002) devised a multiplex PCR approach to identify the species composition of marketed fins. Another obvious application is checking retail fish markets for proper labeling. Marko and colleagues (2004) found that 75% of the fish sold as red snapper in the eastern United States were other species. Consumers are not getting what they intended to purchase, and widespread mislabeling of seafood may also lead to the perception that red snapper (or other species) is abundant when, in fact, stocks are depleted.

Gut contents and marine food webs

Given the difficulty of directly observing the behavior of many species, researchers glean important ecological insights and information about food-web relationships by examining the gut contents of rare specimens. Blankenship and Yamanos (2005) extracted DNA from the guts of deep-sea amphipods and used universal PCR primers to amplify, clone, and sequence mtDNA (CO1) and 18S rRNA genes from the extract. The results suggested that the amphipods have a broad diet; however, as might be expected, few sequences could be identified since the DNA database for deep-sea organisms—
including possible amphipod prey—is quite limited. Still, the general approach is quite straightforward and has been applied in diverse settings. For example, the gut contents of a rarely collected giant squid (Architeuthis sp.) were found to include a common, commercially harvested fish (the blue grenadier, Macruronus novazelandiae), as well as conspecific DNA, which was interpreted as evidence for cannibalism (Deagle et al. 2005).

In addition to demonstrating links in the food web, new approaches promise to elucidate the actual rates of consumption of specific prey. Durbin and colleagues (2008) used quantitative PCR to estimate rates of predation by adult copepods (Centropages typicus females) on nauplii of another copepod species (Calanidae) (Acartia tonsa). The general approach was to determine the mitochondrial COI gene copy number of an individual prey organism in the stomachs of predatory copepods collected in the field. By measuring DNA disappearance rates in the stomachs under lab conditions, researchers could estimate field ingestion rates.

Cryptic species
The use of DNA in the identification of species has had rather surprising consequences. Many marine taxa appear to have global distributions, and given the apparent lack of barriers to dispersal in the open ocean and the dispersal capacities of planktonic larval stages, this is not unexpected. However, molecular analyses of many broadly distributed species have revealed that they are, in fact, groups of morphologically indistinguishable species (although criteria defining “species” are inconsistent across studies). Knowlton (2000) suggested that there has been an excessive tendency to lump taxa together as single species. For example, Goetze (2003) examined a large number of populations of 20 broadly distributed calanoid copepods in the family Eucalanidae. Not only were populations within what was identified as a single species sharply divergent genetically (and most likely cryptic species), but neighboring populations were frequently not even the most closely related taxa. Unfortunately, it is difficult to predict the existence of cryptic taxa; Provan and colleagues (2009) found no evidence for population differentiation across the entire North Atlantic range of populations of a related copepod, Calanus finmarchicus.

Identification of cryptic species is not just an academic exercise. Geller (1999) showed that the invasive European blue mussel Mytilus galloprovincialis replaced the native Mytilus trossulus in southern California over the past century essentially unnoticed because of the morphological similarity of the two species. Since Mytilus is used as an indicator species in marine toxicology studies, inadvertent use of cryptic species with potentially different physiologies can jeopardize the validity of monitoring programs. Similarly, taxonomic ambiguity could result in misguided management decisions. Recent genetic studies have shown that the globally distributed bonefish (Albula vulpes), a prized tropical sport fish, consists of eight or more genetic lineages, with sympatric cryptic species occurring in several parts of the world, including the Florida Keys (Colborn et al. 2001). Surprisingly, the juvenile bonefish observed off the Keys belong not to the A. vulpes species caught in the recreational fishery but to a different one, Albula sp. B. The juvenile habitat of A. vulpes remains unknown (Adams et al. 2008).

The genetic structure of natural populations—a starting point
For most organisms, habitat is patchy. Stretches of sandy beach are interrupted irregularly by rocky headlands, and the fauna on rocky intertidal habitat differs from that on adjacent sandy beaches. But many of the rocky habitat organisms have larval stages that are planktonic and may be transported among the rocky habitat patches. How frequently are the larvae dispersed between rocky patches? Why is it important? Consider a harvestable resource, such as a sea urchin. If one harvests the urchins at one rocky site, will the population be reseeded by larvae drifting in from neighboring sites? This is a question of demographic connectivity: Is interpopulation dispersal sufficiently frequent that it contributes significantly to local population sizes? We might also be interested in another level of connectivity—evolutionary connectivity—in which low levels of dispersal lead to just enough interbreeding (gene flow) to prevent the isolated populations from achieving genetic differentiation. Although they do not directly affect population sizes, low levels of gene flow determine the extent to which populations can adapt to local environments; such adaptation may limit the success of migrants between populations (Hauser and Carvalho 2008).

Gametes and early larval stages are frequently microscopic in size and not easily tracked. But connectivity can be inferred from patterns of genetic variation across populations (e.g., Hellberg et al. 2002). An example reveals the power of the general approach. The copepod Tigriopus californicus lives in high, intertidal rock pools along the Pacific coast of North America. Because it is a tiny, free-swimming organism, one would expect that T. californicus could be frequently transported by coastal currents among neighboring rocky sites. But when protein polymorphism and later DNA sequence analyses were used to assess the connectivity of T. californicus populations, the results were startling—populations 500 meters apart were sharply differentiated. In fact, levels of DNA sequence divergence between some populations indicate that they have been isolated for at least many thousands of generations (Burton 1997, 1998). Surprisingly high levels of population differentiation have been observed in a diversity of marine species, and it has become clear that life-history characteristics, such as the length of planktonic larval life, are not always good predictors of genetic structure (Hellberg et al. 2002).

Identification of stocks for management and conservation
Management of fish stocks requires understanding connectivity patterns, and efforts to characterize population structure using genetic markers have been extensive. One major
conservation challenge involves species that reproduce as discrete populations, but migrate to common feeding grounds where populations of mixed origin may be subject to significant mortality. Sea turtles exhibit such a life history. For example, hawksbill turtles are endangered in every ocean basin as a result of exploitation of “tortoiseshell” scutes for artisan products. To determine the potential impact of harvest on nesting populations, Bowen and colleagues (2007) used mtDNA sequence variation to determine the relationships between nesting and feeding populations. Most nesting populations were found to be strongly differentiated; in the most extreme example, a sample of 69 turtles from the Yucatan included 64 individuals with a haplotype that was not observed at any other sampled nesting site. Yet that haplotype was recovered in all the feeding areas, suggesting that the Yucatan nesting population is an important contributor to the entire Caribbean hawksbill metapopulation. All feeding grounds are mixtures of multiple nesting populations, so any local harvesting on the feeding grounds will affect nesting populations in multiple countries. These data clearly indicate the essential need for international cooperation if conservation efforts for this endangered species are to be successful.

The sharp differentiation of hawksbills enabled a simple genetic analysis of population mixing on the feeding grounds. Other species of great importance for conservation and fisheries present greater challenges. The evolutionary recent origin of populations or large population sizes often result in limited genetic divergence among marine stocks. Recent work has shown that approaches relying on large numbers of loci with low divergence are sometimes capable of teasing apart subtle genetic signals. Nikolic and colleagues (2009) nicely demonstrated how increasing the number of microsatellite markers improves the genetic resolution of populations of Atlantic salmon, *Salmo salar* (figure 2). Of course the ability of salmon to return to their natal streams to spawn results in the gradual genetic divergence of local populations, but the relatively recent origin of many populations reduces observed levels of genetic divergence. Efforts to use increasing numbers of loci to discriminate relatively young (post-glaciation) populations of chinook salmon (*Oncorhynchus tshawytscha*) require careful cross-calibration of methods among laboratories. Seeb and colleagues (2007) reported on a large international collaboration to standardize microsatellite DNA data needed to identify the stock composition of chinook salmon caught during their oceanic migrations. Nine laboratories identified 13 microsatellite loci and produced a database of more than 16,000 salmon representing 110 putative populations, ranging from southeastern Alaska to the Sacramento River in California. This data set now provides a key tool for genetic stock identification of samples collected from mixed fisheries.

Genetic identification of stocks has proven important for stock enhancement through fish hatchery activities. Banks and colleagues (2000) noted that broodstock used in a hatchery supplementation program for winter-run chinook salmon were unwittingly hybridized with spring-run salmon; such hybridization could have detrimental effects on the very population targeted for conservation efforts. Appropriate broodstock are now identified using a panel of 10 microsatellite loci that distinguish winter and spring runs; only fish classified as having a high probability of belonging to the desired population (see below) are used as broodstock.

The great commercial value of salmon stocks has made the large investment in salmon genetics feasible. But the considerable effort required for the startup of microsatellite studies has made this method less accessible to investigators lacking the resources to construct and screen genomic libraries. In recent years, the AFLP (amplified fragment length polymorphism) approach (box 1, figure 2) has seen increasing application. This method produces a genetic fingerprint that can be used to assess population connectivity. Amplified fragment length polymorphisms have recently proven useful in a variety of ecological genetic studies (Campbell et al. 2003, Brazeau et al. 2005, Murray and Hare 2006). For example, using AFLPs, Shank and Halanych (2007) found significant genetic divergence among populations of the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*; these results contrasted with earlier allozyme and mtDNA studies. Similarly, AFLP studies revealed population differentiation in red abalone that was not evident in allozyme, mtDNA, or microsatellite studies (Gruenthal et al. 2007). These and other studies suggest that AFLPs may provide enhanced sensitivity, presumably due to the large number of markers scored. In addition to adding sensitivity to studies of population structure, the large number of markers available in AFLP studies makes whole new types of investigations possible. Campbell and Bernatchez (2004) showed that AFLPs provide a “genomic scan” that permits analyses of associations between specific markers and habitat parameters—associations that very likely reflect the role of natural selection in population divergence.

**Sex-biased dispersal**

Differences in inheritance patterns of different genetic markers have permitted some remarkable inferences regarding the ecology of marine animals. The logic is simple. Mitochondrial DNA is maternally inherited, so the relationships revealed by marker similarity or identity reflect only the dispersal of females, whereas markers in the nuclear DNA, with biparental inheritance, reflect migration of both sexes. Comparison of apparent gene flow using different markers can therefore reveal sex-biased dispersal.

For example, analyses of mtDNA variation have revealed significant population differentiation in sperm whales across ocean basins; in contrast, microsatellite polymorphisms (nuclear markers) did not reveal such differentiation. These contrasting patterns suggest that males probably disperse across ocean basins but females do not, an inference that is consistent with observations of females being the philopatric sex and having a more limited latitudinal distribution than males (Lyrholm et al. 1999). The genetic composition of
white shark populations in South Africa and Australia led to a similar result; mtDNA variation clearly distinguished the populations and microsatellite variation did not. Again, the simplest explanation is that male white sharks disperse across ocean basins significantly more frequently than do females (Pardini et al. 2001).

**Equilibrium-based estimation of population structure versus assignment methods**

Historically, estimation of interpopulation migration has relied almost exclusively on methods assuming that (a) differences accumulate between populations as a result of mutation and genetic drift, (b) any migration between the populations will homogenize their gene pools, and (c) these processes jointly result in the observed level of population divergence (presumed to be at equilibrium). The estimation of migration typically focuses on estimates of \( F_{ST} \), the standardized variance in allelic frequencies among populations; \( F_{ST} \) varies from 0 (no variation among populations) to 1 (all variation among populations). At equilibrium, \( F_{ST} \) can be related to gene flow through the approximation \( F_{ST} \approx \frac{1}{(4N_e m + 1)} \), where the product \( N_e m \) (effective population size \( \times \) proportion of migrants to a population per generation) is the number of migrants exchanged per generation. It must be remembered that \( N_e m \) is an equilibrium estimate of genetic exchange; a high \( N_e m \) (low \( F_{ST} \)) estimate can reflect a low level of ongoing exchange or a high level of exchange in the distant past, followed by a period of no exchange (but large population sizes). From an evolutionary point of view, the difference may not be important, since both scenarios prevent population differentiation. But from a management point of view, low \( F_{ST} \) values (i.e., low levels of genetic differentiation, as commonly observed in marine systems) are difficult to interpret.

To make inferences of contemporary levels of connectivity among populations from genetic data, a variety of new methods of analysis has been developed. Manel and colleagues (2005) discussed the use of "assignment tests" that take advantage of the multilocus genotypes of individuals to, for example, identify individual migrants within population samples. In the simplest approach, potential source populations are sampled, and expected multilocus genotypic probabilities are calculated using allele frequencies. Then, for any individual multilocus genotype, the probability that it came from a particular source can be calculated, and the individual is assigned to the source with the greatest probability. Of course, there is also the assumption that all potential source populations have been sampled. Alternatively, it is sometimes easier to use an exclusion approach: On the basis of sample allele frequencies, one can generate the probability that a certain genotype would be observed in that sample. If the probability is very low for a particular individual, the individual is probably a migrant. Such analyses are used to discriminate between the winter and spring runs of chinook salmon discussed above.

Such approaches classify individuals to likely source populations. But the rather continuous distributions of some marine organisms make population delineation itself a problem. This problem has been addressed by recent statistical methods that assess the probabilities that the samples fall into some unknown number of different clusters (or populations). These approaches rely on detecting the nonrandom association of alleles across loci (termed "linkage disequilibrium") generated when individuals from different populations are mixed in a sample. Clustering methods break the sample into groups of individuals (populations) by minimizing disequilibrium within the constructed groups (Pritchard et al. 2000). This method has been applied in a number of recent studies; for example, Moller and colleagues (2007) applied the methods to analyze microsatellite variation among samples of bottlenose dolphins from southeastern Australia. Cluster analysis verified that the five geographic samples are best sorted into only two biologically significant populations, an embayment population and a coastal population, representing recent differentiation between the two habitat types.

**Estimation of effective population size**

Genetic data are increasingly being employed to give insight into the demography of populations of marine organisms.
Several methods have been proposed for the estimation of present-day effective population sizes. For example, a temporal approach is based on changes in allelic frequencies in samples taken one or more generations apart; the magnitude of genetic drift (random sampling error resulting in changes in allelic frequencies) is inversely dependent on effective population size. Hence, if large changes are observed, \( N_e \) must be small (assuming there is no natural selection favoring specific alleles). Although this method can estimate \( N_e \), the relationship between \( N_e \) and census population size in many marine organisms is difficult to infer. In fact, the level of genetic variation maintained in many marine species is orders of magnitude below what census population sizes would lead one to expect (Hauser and Carvalho 2008). Several factors may contribute to this observation. Depending on the vagaries of the environment, some females in highly fecund species may be wildly successful, whereas others, by chance, may fail to leave any offspring. This sweepstakes pattern of recruitment may account at least in part for the discrepancy between census and effective population sizes (Flowers et al. 2002, Hedgcock et al. 2007, Palstra and Ruzzante 2008).

Alternative methods are currently being developed to estimate \( N_e \) from single samples. Tallmon and colleagues (2008) estimated \( N_e \) from microsatellite data from a single sample; the calculation was based on simulations of many populations with the same number of individuals sampled, and comparable microsatellite data (number of loci and repeat motif). Ultimately, the \( N_e \) estimate is determined by comparing the observed data to the simulated data for a series of statistics relating genetic variation to \( N_e \), including expected heterozygosity, number of alleles per locus, mean and variance of multilocus homozygosity, and others. Although the idea of calculating \( N_e \) from observed within-population genetic variation (or temporal change in allele frequencies) is attractive, empirical results can be erratic; a recent study employed two different methods for estimating \( N_e \) in the spiny spider crab and obtained values that varied from 100 to 1 million for the same population (Sotelo et al. 2008). Palstra and Ruzzante (2008) have discussed some of the important complications in estimating \( N_e \), including importance of both proper sampling and consideration of population demographic factors.

**Historical population sizes**

A recurrent problem in marine conservation is that we frequently lack baseline information on species abundances prior to historic exploitation. The problem of shifting baselines was first cited by Pauly (1995), who noted that successive generations of fishers have quite different personal experiences and expectations; in many parts of the world, a big catch today is probably substantially smaller than a big catch was one or two generations ago. How do we know when a depleted stock has recovered? If we have no knowledge of the size of the original, unexploited stock, how can we set appropriate management goals? A 50% increase in the number of whales might seem a great management success, but if our frame of reference is one in which stocks have been reduced to less than 10% of their preexploitation levels, the increase (to 15% of preexploitation levels) is not so impressive.

Although it would seem that we cannot turn back the clock to genetically sample the populations of the past, several studies have been able to use DNA extracted from shells or bones recovered from archaeological middens and museum samples to investigate genetic changes over ecologically important time frames. For example, studies of mtDNA extracted from ancient bones of the now endangered seabird, the Chatham Island taiko, show that despite precipitous population declines (from millions to less than 200 birds over the past several hundred years), there has been no significant loss of genetic diversity (Lawrence et al. 2008). Analyses of ancient DNA promise to provide some important insights into temporal changes in genetic composition in response to human activities in the past several centuries.

Even in the absence of historic samples, genetic analyses can be used to estimate historic population sizes. Standing levels of neutral genetic diversity within populations are determined by a balance between loss of variation due to genetic drift (a function of \( N_e \)) and introduction of variation due to mutation. For maternally inherited mtDNA, for example, the formula \( q = 2 \left( \frac{1}{N_e} \right) \mu \), where \( q \) is a measure of genetic diversity, and the effective female population size is \( N_e (f) \), describes the expected relationship. Estimation of \( N_e \) requires an estimate of the mutation rate (\( \mu \)) in the studied sequence.

Roman and Palumbi (2003; also see Lubick 2003) used this approach to estimate long-term population sizes of three North Atlantic baleen whale species. They estimated genetic diversity from mtDNA sequences, and estimated the rates of mutation on the basis of fossil evidence suggesting that the humpback genus *Megaptera* is at least 6 million to 10 million years old. Once they had estimated \( N_e (f) \), they used three factors to convert this to estimates of the long-term census population sizes: (1) sex ratio, (2) proportion of breeding adults versus total adult numbers, and (3) proportion of adults versus juveniles in the population. Their results were surprising—the estimates of long-term population sizes of humpbacks in the North Atlantic were on the order of 240,000 animals (current population estimates are 10,000), which contrasts with historical estimates suggesting that only about 20,000 humpbacks existed in the North Atlantic before commercial exploitation began in the 19th century.

The conservation implications of such genetic estimates of past population sizes are considerable, given the concept of shifting baselines. The international moratorium on commercial whaling sets in place targets for recovery based on estimates of preexploitation population sizes. If the genetic data are correct, those targets might be far too low. However, the drop in abundances from those estimated to current levels could have happened rather far back in evolutionary time. Has the carrying capacity actually changed over time, or have anthropogenic actions been the primary cause of large-scale changes in whale numbers?
In another study, Jackson and colleagues (2008) have also used measures of extant mtDNA variation in southern right whale (Eubalaena australis) populations to estimate several parameters of historical demography of this formerly heavily exploited species. On the basis of whaling records and current abundance, the International Whaling Commission (IWC) has estimated current rates of population growth. By using the genetic data to inform historical reconstructions of population dynamics, Jackson and colleagues (2008) found that IWC estimates are likely to be inaccurate in important ways. Existing mtDNA variation in the population suggests that (a) pre-exploitation population sizes were probably substantially higher than the IWC estimates (as also suggested by Roman and Palumbi [2003] for other whale species), (b) the number of whales surviving through the bottleneck (when population size was at its minimum) was most likely larger than the IWC assessment, and (c) the current population growth rate estimated by the IWC is probably too high. The study again suggests the potentially key role genetics can play in informing management decisions.

Conclusions

Despite decades of study, two issues continue to obscure our understanding of populations of marine organisms: (1) We routinely visit or monitor biological activities in an infinitesimally small fraction of the world’s oceans, and (2) the complex life histories of many marine species make it difficult or impossible to track individuals from birth to death. Against this difficult backdrop, molecular genetics has made some remarkable contributions. Strong intuitions regarding population connectivity in species with planktonic larvae have been proven wrong. Geographic ranges of “cosmopolitan” species have been corrected as cryptic species have been uncovered. Larval forms have been matched to adults, and spawning sites in the open ocean have been elucidated. Gut contents analyzed by molecular tools provide new information about marine food webs. Sex-biased migration patterns have been inferred in megafauna that are very difficult to observe, including the great sharks and whales. Contributions of different populations to mixed population fisheries have been clarified. Estimates of contemporary and historical effective population sizes have been made. Even this partial list is impressive, but both the molecular genetic toolbox and the analytical approaches available to marine ecologists have advanced faster than their respective applications, suggesting that molecular marine ecology is still far from mature.

Acknowledgments

This work has been supported by National Science Foundation grants DEB-0717178 and OCE-0526979. Xavier Mayali, Jeffry Mitton, and two anonymous reviewers provided many helpful comments on an earlier version of the manuscript.

References cited


Ronald S. Burton (burton@ucsd.edu) is with the Marine Biology Research Division at Scripps Institution of Oceanography, University of California, San Diego.