Rethinking the mechanisms that shape marine decapod population structure

BREE K. YEDNOCK & JOSEPH E. NEIGEL

Department of Biology, University of Louisiana at Lafayette, Lafayette, LA 70504, U.S.A.

ABSTRACT

In research on the population genetics of marine decapod crustaceans it is often assumed that planktonic larvae disperse widely and genetic markers are not strongly influenced by natural selection. Although population genetic surveys often seem to corroborate these assumptions, some have revealed genetic differentiation among populations that lack discernable barriers to dispersal. These cases, along with insights from related areas of investigation, suggest it may be time to revise our thinking. We now know that dispersal of meroplanktonic larvae can be strongly limited by oceanographic processes and larval behavior. There is also evidence suggesting that directional selection may be stronger in marine populations than previously thought. These new perspectives have important implications for decapod population genetics and phylogeography. This paper will review the findings that have led to a revised view of larval dispersal and will examine natural selection as a potential modifier of dispersal and influence on genetic population structure in decapod crustaceans.

1 INTRODUCTION

Research on the population genetics of marine decapod crustaceans was long grounded in two widely held assumptions. First was the assumption that a planktonic larval phase ensured widespread dispersal (Thorson 1950, Strathmann 1993). This was based on early observations of the occurrence of larvae in oceanic plankton far from any suitable adult habitats (Scheltema 1971). Such observations also contributed to the idea that the planktonic larval phase is itself an adaptation to promote long-distance dispersal to new habitats. Planktonic larvae were viewed as passive drifters in oceanic currents whose success in ultimately reaching a suitable habitat was largely a matter of chance. Local populations of benthic adults were assumed to be open and dependent on upstream larval sources (Caley et al. 1996). The second assumption, which followed from the first, was that planktonic dispersal results in such high gene flow that it overwhelms natural selection and prevents local adaptation (discussed by Hedgecock 1986). This assumption was bolstered by surveys of genetic markers in marine species in which little population differentiation
was found (e.g. Shulman and Bermingham 1995, Brown et al. 2001). Exceptions to this pattern of geographic homogeneity, such as sharp breaks or gradual clines in allele frequencies, were interpreted as imprints of past disruptions of gene flow that were somehow maintained by present day barriers (Hellberg et al. 2002). Heterogeneity on smaller spatial scales as well as temporal shifts in allele frequencies were also explained as consequences of planktonic dispersal. Because chance was thought to play a major role in the success of planktonic larval recruitment, it followed that individual pulses of recruitment might consist of the progeny of just a few lucky individuals that happened to spawn at just the right time and place to deliver their progeny into favorable currents. The sampling error associated with the small number of winners in this reproductive sweepstakes would cause the genetic makeup of the successful recruits to differ from the population as a whole (Hedgecock 1994). Despite scant direct evidence, this model has gained wide acceptance and the term “sweepstakes reproduction” is often used as synonymous with any occurrence of fine scale temporal or spatial heterogeneity in genetic markers.

Over the past fifteen years, a revised view of planktonic larval dispersal has emerged, which places greater emphasis on mechanisms that retain larvae near natal habitats or return them to those habitats (Warner and Cowen 2002). The impressive mobility and complex behaviors of fish and crustacean larvae in particular were found to be incompatible with their characterization as passive drifters. Estimates of larval dispersal distances based on a variety of approaches were consistently much less than would be expected for passive drifters (Shanks et al. 2003, Leis 2006), and tagging studies showed the larvae of reef fish to possess an uncanny ability to return to their natal reefs after weeks spent in the plankton (Jones et al. 1999). This revised view of planktonic larval dispersal has important implications for our understanding of the population genetics of meroplanktonic decapod crustaceans. It implies that we cannot assume that gene flow is always an overwhelming force that homogenizes populations or that strong physical barriers are required to isolate populations. We are now faced with a much richer variety of possibilities, in which gene flow, genetic drift and natural selection can all play significant roles. In this chapter we will review the evidence that supports a revised view of larval dispersal, and we shall look at natural selection in particular as a potential modifier of dispersal and as a cause of genetic population structure in decapod crustaceans. We will also consider the implications of a revised view of decapod population genetics for phylogeographic studies of decapods. Phylogeography was originally conceived as the application of phylogenetic and biogeographic approaches from systematics to the genealogical structure of mitochondrial DNA (mtDNA) variation within animal species (Avise 2000). As such, it was viewed as “the mitochondrial DNA bridge between population genetics and systematics” (Avise et al. 1987). A central assumption in phylogeography has been that natural selection doesn’t play a significant role in shaping phylogeographic structure, which can thus be interpreted solely in terms of historical patterns of genetic drift and gene flow. However, there is growing evidence to suggest that selection can mimic what are assumed to be historical patterns, especially if mtDNA is the sole genetic marker used. If, as we suggest, selection is an important factor in decapod population genetics, it is likely to be important in decapod phylogeography as well.
2 LARVAL DISPERSAL

2.1 The complexity of oceanographic processes

A central concept in dispersal biology is a species’ dispersal kernel, a profile of the probabilities of different dispersal distances. Dispersal kernels have been estimated for many terrestrial species, but none are available for decapods, or for that matter, any meroplanktonic species. Even estimates of mean dispersal distances for planktonic larvae have been difficult to obtain, and are either limited to ecologically atypical circumstances, such as colonization of empty habitats, or based on assumption-laden indirect methods (such as population genetic methods). As a substitute for direct measurements, estimates of larval dispersal kernels and distances are often based on models of planktonic dispersal in oceanographic flows (e.g. Siegel et al. 2003). The accuracy of these models is difficult to test, for the very reason that they are needed in the first place: a lack of empirical data.

The simplest model of planktonic dispersal treats larvae as passive particles advected in a uniform flow; dispersal distance is then the product of the length of the planktonic period and the flow’s velocity (e.g. Roberts 1997). However, this model has been found to greatly overestimate actual dispersal distances (Shanks et al. 2003, Shanks 2009). Complex current patterns, eddies, tides, and winds can reduce the transport of planktonic larvae to far below the model’s simple predictions (Sponaugle et al. 2002). Larvae spawned in shallow waters do not immediately enter offshore currents, but may remain in the relatively sluggish coastal boundary layer for days or longer (Largier 2003). In some cases, transport in one direction is followed by transport in the opposite direction. For example, the prevailing current in a well-studied upwelling zone along the coast of California transports larvae tens of km. However, during relaxations of upwelling the current is reversed and larvae are carried back towards their natal locations (Wing et al. 1995). Similarly, regions of counter-current flow have been identified as areas where larvae can be retained in natal habitats (Sponaugle et al. 2002). Tidal fluxes can export larvae away from shore on ebb tides and return them on flood tides, and the larvae of some decapods vertically migrate to take advantage of tidal transport shoreward and into estuaries (Hughes 1969, Cronin and Forward 1986, Shanks 1986). High concentrations of brachyuran larvae in tidal fronts at the mouths of estuaries (Epifanio 1987) provide further evidence that decapod larvae can undergo development in coastal waters without being dispersed by offshore currents.

2.2 The complexity of larval behavior

The old supposition that a planktonic larval phase is an adaptation to promote dispersal has not been supported by studies of larval morphology or behavior. Instead, it is now suggested a planktonic phase may allow larvae to exploit particular resources, reduce competition with other life stages, avoid predation (Johannes 1978) or escape from parasites (Strathmann et al. 2002). For example, the projecting spines of crab larvae, although superficially resembling the structures that carry plant seeds on the wind, are unlikely to provide buoyancy or facilitate dispersal. The spines instead deter predators, and are thus an adaptation for survival rather than dispersal (Morgan 1989, 1990). Some behaviors of decapod larvae are directed to reduce, rather than promote dispersal. Despite their small size, decapod larvae are capable of swimming over 10 cm s\(^{-1}\) (Cobb et al. 1989, Luckenbach and Orth 1992, Fernandez et al. 1994, Chiswell and Booth 1999), which can...
be enough to make headway against ocean currents. Decapod larvae also possess sensory abilities that enable them to recognize physical and biological cues for orientation and navigation (reviewed in Kingsford et al. 2002). The combination of mobility and the ability to respond to sensory cues allows decapod larvae to direct their movements in ways that increase their chances of reaching suitable habitats for settlement.

2.3 Pelagic duration and dispersal distance

While the pelagic duration (PD) of most fish larvae can be directly estimated by counting otolith growth rings (e.g. Victor 1986, Wellington and Victor 1989), decapod PDs are nearly impossible to measure in the field. As a result, measurements of decapod PDs are generally made in the laboratory (e.g. Costlow and Bookhout 1959, Gore and Scotto 1982, Anger 1991, Nates et al. 1997, Strasser and Felder 1999). These laboratory studies have demonstrated the potential for larval PDs to vary with environmental parameters such as salinity (Anger 1991, Field and Butler 1994) temperature (Anger 1991, Field and Butler 1994, Sulkin et al. 1996) and food availability (Shirley and Zhou 1997). This plasticity could allow decapod larvae to respond to variable field conditions, for example, by delaying settlement until favorable cues indicate a suitable habitat (Christy 1989, Strasser and Felder 1999).

In a compilation of empirical estimates of dispersal distances for marine propagules (spores, eggs as well as larvae), PD was a poor predictor of dispersal distance. Although they were correlated, dispersal distances varied by several orders of magnitude for species with similar PDs (Shanks et al. 2003). However, despite the study’s conclusion that dispersal distance is strongly dependent on propagule behavior, Shanks et al. (2003) has been most frequently cited as showing that dispersal distances are determined by PDs. Shanks (2009) has repudiated this interpretation and offered further evidence against it.

2.4 Pelagic duration and genetic population structure

The use of genetic population structure, and in particular $F_{ST}$, to estimate gene flow has had a long and troubled history. Wright originally defined $F_{ST}$ as the probability of alleles that are identical-by-descent from an ancestral population being combined in zygotes within a subpopulation (Wright 1951). Wright also showed that in a simple island model $F_{ST}$ was determined by the relative strengths of genetic drift (inversely proportional to effective population size, $N$) and gene flow (the proportion of a subpopulation’s gene pool originating from other subpopulations, $m$) as follows:

$$F_{ST} \approx \frac{1}{(4Nm + 1)}$$

An important but generally overlooked aspect of this equation is that $F_{ST}$ is just as dependent on effective population size as it is on the rate of gene flow. Later, with the introduction of allozymes and other genetic markers to population genetics, estimators of $F_{ST}$ based on the frequencies of genetic markers came to replace Wright’s parametric definition, for example:
This estimator, along with all others based on allele frequencies, is only valid when other forces such as selection and mutation are not acting on allele frequencies. Theoretical work by Slatkin and others showed that in small populations (with effective population sizes on the order of 100) $F_{ST}$ would reach equilibrium quickly and estimates of gene flow based upon it would be relatively insensitive to selection or mutation (Slatkin 1987, Slatkin and Barton 1989). However, the critical assumptions behind this conclusion were often overlooked and $F_{ST}$ came to be viewed simply as an estimator of gene flow that could be applied on any scale with any genetic marker. Eventually, the flaws in this approach became painfully obvious and led to a series of papers critiquing the use of $F_{ST}$ as an estimator of gene flow (Neigel 1997, Bossart and Prowell 1998, Whitlock and McCauley 1999).

For marine decapods, it makes little sense to view $F_{ST}$ as an estimator of gene flow because the effect of gene flow is confounded with the effect of genetic drift, and we may know less about genetic drift in marine populations than about gene flow. Although populations of marine benthic species can be very large, high fecundities and other life history characteristics suggest that effective population sizes could be much smaller than census sizes (Orive 1993). Just how much smaller is open to speculation. For oysters, which have extremely high fecundities, estimates of effective population sizes from temporal variation in allozyme frequencies have been as low as 10-100, many orders of magnitude below census sizes (Hedgecock et al. 1992). These remarkably low values led to the hypothesis of “sweepstakes reproduction”, an extreme reduction in effective population size that occurs when, by chance, all of the successful progeny in a cohort are produced by a tiny fraction of the entire population (Hedgecock 1994). Sweepstakes reproduction can explain not only temporal changes in allele frequencies, but also microspatial variation in allele frequencies, known as “chaotic patchiness” (Johnson and Black 1982). Genetically distinct patches could represent different cohorts of sweepstakes progeny. However, there is little hard evidence for sweepstakes reproduction apart from the phenomena it is intended to explain. Alternative hypotheses, such as variability in larval sources or selection are not always given full consideration. Decapods generally don’t have high enough fecundities for the reproductive output of a few individuals to account for all of the larval settlement in a large population (Corey and Reid 1991, Reid and Corey 1991). Nevertheless, the range of plausible effective population sizes for most decapods is so wide that nothing should be assumed about the relative strengths of genetic drift and gene flow.

There is an important distinction between dispersal and gene flow. Dispersal is an ecological process that increases the distances between organisms or gametes (Baker 1978), whereas gene flow is a population genetic process that changes the genetic composition of populations through the movement of organisms or gametes among populations (Slatkin 1987). Gene flow does not occur when organisms or gametes disperse but fail to contribute to a population’s gene pool. Considering the weakness of the relationship between PD and dispersal distance, the possibility that dispersal does not result in gene flow, and the dependency of $F_{ST}$ on factors other than gene flow, it may be surprising that some studies have found a clear relationship between PD and $F_{ST}$ (reviewed in Bohonak 1999).
However, in a large meta-analysis based on a random sample of published studies (Weersing and Toonen 2009) no significant relationship was found between PD and $F_{ST}$ for species with a pelagic phase. Similarly, Shanks (2009) found only a weak relationship between PD and genetic differentiation for species with PDs over 10 h. This dichotomy between the findings of narrow and broad comparative studies suggests that the relationship between PD and $F_{ST}$ may hold for similar species with minimal variation in other factors, but that in a broad comparative setting PD isn’t a major factor. $F_{ST}$ continues to be a very useful quantity in population genetics, but we should avoid thinking of it as a measure of gene flow or population connectivity (Neigel 2002).

3 NATURAL SELECTION

3.1 The potential for selection on early life stages

Planktonic larvae experience high rates of mortality, caused primarily by starvation and predation, although variation in environmental conditions such as temperature, salinity, dissolved oxygen, and pollutants can also be important (Thorson 1950, Morgan 1995). Where rates of mortality at early life stages are high there is the opportunity for strong natural selection (Williams 1975). Pre-settlement selection during the pelagic larval phase could alter the genetic composition of potential recruits before they reach suitable habitats. Alternatively, post-settlement selection could eliminate individuals that are ill-equipped for the conditions they encounter. If post-settlement selection results from phenotype-environment mismatches rather than intrinsic fitness differences, long-distance dispersal of larvae may not necessarily result in effective gene flow (DeWitt et al. 1998, Marshall et al. 2009). Most studies of decapod populations sample only adults or juveniles, so that what appear to be barriers to dispersal could actually be selective filters on larval survival.

Definitive evidence of post-settlement selection has been documented for some marine invertebrates. For example, Koehn et al. (1976) and Hilbish (1985) observed a cline in allele frequencies of the Lap allozyme locus in a population of mussels, *Mytilus edulis*, in Long Island Sound, New York. The cline was found to correspond with a salinity gradient that separated mussels with genotypes specific to oceanic habitats from those with genotypes specific to brackish waters. Although oceanic genotypes were found among recruitment cohorts of mussels in brackish locations of the Sound, the oceanic genotypes failed to persist through subsequent life stages. Therefore, despite larval dispersal between oceanic and brackish water habitats, selection for favorable alleles within recruiting cohorts at brackish locations prevented gene flow from genetically homogenizing these local populations. Biochemical analysis of the allozymes produced by the different Lap alleles provided definitive support of selection. The allozymes were found to be functionally different (Koehn and Siebenaller 1981), resulting in lower fitness for mussels with oceanic genotypes in brackish water and consequently higher mortalities.

3.2 The appropriate scale for decapod population genetics

Most studies of genetic population structure in decapods (as well as other marine benthos) have been based on surveys conducted over large distances (100’s of km) to detect pronounced shifts in genetic marker frequencies, typically referred to as “genetic breaks” (Hellberg et al. 2002). This bias probably reflects both the old assumption that planktonic
dispersal results in gene flow over large distances and practical considerations about the sample sizes needed to detect small differences in genetic marker frequencies. However, the emphasis on large-scale geographic patterns may be misplaced. Genetic breaks imply deep historical divisions that have been maintained by strong barriers to gene flow. In many cases these phylogeographic divisions might be more appropriately viewed as cryptic taxa than as units of population structure (e.g. Barber et al. 2000). In light of the many factors that can limit the planktonic dispersal of decapod larvae, there is reason to investigate more subtle patterns of population structure over smaller distances. We might expect at these finer scales to identify ecologically relevant subpopulations that are structured by oceanographic and behavioral mechanisms. In these investigations, we should not discount the potential importance of small, but statistically significant differences in genetic marker frequencies. Any significant frequency difference indicates a restriction in gene flow; a small difference could mean either that the rate of migration is only slightly reduced or that large effective populations have made the diversifying force of genetic drift relatively weak.

For decapods, plausible limits for such fundamental population genetic parameters as effective population size and gene flow span at least several orders of magnitudes. These limits encompass radically different possibilities for how decapod populations function. At one extreme we can envision a scenario in which gene flow is strong as a consequence of planktonic dispersal and genetic drift is strong because a few highly fecund individuals are responsible for most of the successful reproduction. In this scenario, temporary reductions in gene flow allow genetic drift to produce localized fluctuations in allele frequencies but these are quickly removed when gene flow resumes. Selection is overwhelmed by both gene flow and genetic drift, so local adaptation is prevented. At the other extreme is a scenario in which gene flow is weak because oceanographic and behavioral mechanisms limit dispersal and genetic drift is weak because local populations consist of large numbers of individuals with limited fecundity. Even without any gene flow, differences in allele frequencies develop slowly because drift is weak and the differences never become great because there are occasional episodes of moderate gene flow. However, selection is strong enough to overcome both gene flow and genetic drift, so that populations become locally adapted to their environments. These two scenarios, both plausible, obviously have very different implications for the evolution, ecology and management of decapod populations.

4 CASE STUDIES

In this section we review two studies of genetic population structure in marine decapods with benthic adult phases and pelagic larval stages. We choose these examples because they revealed unexpected genetic structure that would have been missed in routine broad range surveys and defy simple interpretation.

4.1 The burrowing ghost shrimp, Callichirus islagrande

The ghost shrimp, *Callichirus islagrande*, is restricted to the quartzite beaches and barrier islands of the northern and western Gulf of Mexico. This species’ range stretches from Pariso, Tobasco, Mexico, to northwest Florida, with a gap corresponding to the muddy sediments of the Chenier Plain in the northwestern Gulf of Mexico (Staton and Felder 1995). *C. islagrande* excavates burrows that can extend as far as 2 m below the surface and
occur in high densities (up to 100 m\(^{-2}\)), which implies large, continuous, and genetically well-mixed local populations. Females can brood several thousand eggs, and the pelagic larval phase is estimated to last from 16 and 20 days, although laboratory experiments indicate some plasticity in PD in response to settlement cues (Strasser and Felder 2009).

A pronounced genetic break has been documented for \textit{C. islagrande} at the Chenier Plain for allozymes (Staton and Felder 1995), mitochondrial DNA (mtDNA) and microsatellites (Bilodeau et al. 2005). Alongshore currents are fast enough to carry the larvae of \textit{C. islagrande} across the Chenier Plain before they complete their pelagic phase, and the location of the break differs from those found for other species in the Gulf of Mexico, which are generally further to the east (Neigel 2009). There is thus nothing to suggest an oceanographic barrier that would prevent gene flow across the Chenier Plain. An alternative possibility is suggested by recent theoretical work that predicts a genetic break will occur where an environmental gradient in selection crosses a region of reduced dispersal (Pringle and Wares 2007). The division within \textit{C. islagrande} could thus be maintained by selection but localized by reduced gene flow across the Chenier Plain.

Intensive sampling on the barrier islands off the coast of Louisiana revealed another unexpected level of population structure in \textit{C. islagrande}. Slight but statistically significant differences in microsatellite allele frequencies were found between locales separated by as little as 10 km. Coalescent analysis indicated that this was not due to small effective population size, but rather to limited gene flow. However, this does not necessarily imply that local populations are nearly isolated. A Bayesian analysis indicated that most of the sampled locales contained mixtures of several different source populations. Thus the differences in allele frequencies could reflect differences in the proportions of larvae they receive from upstream sources (Bilodeau et al. 2005).

4.2 \textit{The blue crab, Callinectes sapidus}

The blue crab, \textit{Callinectes sapidus}, inhabits a wide geographic range spanning temperate, sub-tropical and tropical regions of the western Atlantic Ocean. The life history of \textit{C. sapidus} suggests the potential for widespread dispersal and mixing of larvae. Females can produce multiple clutches of over 3 million eggs (Hsueh et al. 1993). Ovigerous females release well-developed zoea in coastal waters that are then transported offshore by currents where they go through 7 or 8 zoeal stages before molting into a megalopal stage (Costlow and Bookhout 1959) and recruiting to coastal waters. Their entire pelagic development is estimated to take between 37 and 69 days (Costlow and Bookhout 1959). Early assumptions of passive larval dispersal for this species were supported by multiple genetic surveys that found very little or no genetic variation over broad geographic scales (McMillen-Jackson et al. 1994, Berthelemy-Okazaki and Okazaki 1997, McMillen-Jackson and Bert 2004). However, in stark contrast to these findings, Kordos and Burton (1993) uncovered significant heterogeneity in allozyme allele frequencies of blue crab adults, juveniles and megalopae collected along a 600 km stretch of coastline in the Gulf of Mexico. Adult blue crabs were found to be genetically differentiated among nearby bays, indicating reduced gene flow over surprisingly short distances. At three different locations, allele frequencies among megalopae recruits also varied spatially and often differed from those of nearby adults. Even more striking, however, were the extreme temporal shifts in allele frequencies among groups of megalopae recruiting to an area and the significant loss
of alleles between megalopal and adult life stages. The temporal genetic variation seen among recruits could be explained by seasonal spawning differences in source populations and changes in current patterns supplying recruits to an area (Kordos and Burton 1993). However, this scenario fails to explain why genotypes common in recruits were underrepresented in later life stages. *C. sapidus* megalopae typically recruit to the mouths of bays and estuaries, molt to an early crab stage within a few days of settling and move further into marshes and bays as they grow (Morgan et al. 1996). If the megalopae recruiting to the beaches in Texas follow the typical pattern for *C. sapidus*, it is reasonable to expect they would become the juveniles and adults in the nearest bay. The fact that Kordos and Burton (1993) did not find similar allele frequencies between megalopal recruits and nearby juveniles and adults indicates a differential loss of alleles within each cohort, which suggests the possibility of post-settlement selection.

5 PHYLOGEOGRAPHY

Phylogeography became possible with the technology available in the early 1980’s because of the unique properties of animal mtDNA (Moritz et al. 1987). In contrast to nuclear DNA, mtDNA is circular and relatively small, which facilitates its isolation by ultracentrifugation. This allowed mtDNA sequence variation to be characterized by routine restriction fragment analysis (Brown 1980). And unlike allozymes, which were then the standard marker for population genetics, mtDNA could be used to infer intraspecific phylogenies (Avise et al. 1979). Although phylogeographic analysis is no longer tied to it, mtDNA has remained the marker of choice. Phylogeography’s origins in the 1980’s also coincided with the growing acceptance of the neutral theory of molecular evolution, which predicted that most of the DNA sequence variation within a population is likely to be selectively neutral (Kimura 1983). Acceptance of the neutral theory allowed phylogeographers to focus on historical biogeographic explanations of the patterns they found without the complications that selection would introduce (Avise et al. 1987). Although the issue of selection on mtDNA was raised periodically, it did not become a major concern until recently.

The central project of phylogeography, to assign historical causes to the biogeographical patterns revealed by mtDNA (Avise 2000), continues today although its methods have become more sophisticated. Where once mtDNA phylogenies were constructed by hand from restriction fragment data and phylogeographic patterns were interpreted by eye, large data sets are now generated by direct sequencing of PCR products and computers are tied up for weeks with their analysis. Population genetics and molecular evolution have also become more sophisticated. The data provided by PCR and large-scale sequencing of genes and genomes have focused new interest on detecting selection and led to more nuanced forms of the neutral theory of molecular evolution (Austin 2008). Coalescent models have linked the genealogical approaches of phylogeography to a broader theoretical framework that applies to nuclear sequence polymorphisms as well as mtDNA (Avise 2009). A blurring of the boundaries between phylogeography and population genetics developments has opened up new possibilities for phylogeography, but it has also led to sharp debates about the robustness of traditional phylogeographic approaches.
5.1 The debate over methods of phylogeographic analysis

One debate that appears to have lasted far too long concerns the interpretation of phylogeographic patterns as “signatures” of particular historical processes. This approach reached its extreme form in nested clade phylogeographic analysis (NCPA), which likens the problem of identifying historical processes that generate phylogeographic patterns to the problem of identifying plant or animal specimens with a dichotomous key (Templeton et al. 1995). The diagnostic characters used in the NCPA key are statistics that can be easily calculated from phylogeographic data and are intended to capture various predictions of population genetic theory. A major problem with NCPA and similar approaches is that population genetics theory does not predict unique and easily identifiable signatures for distinct historical processes, but rather overlapping ranges of possible outcomes with different probabilities (Nielsen and Wakeley 2001). Support for alternative historical scenarios from phylogeographic data are therefore best evaluated in terms of likelihoods or Bayesian posterior probabilities calculated from probabilistic models (Knowles 2004). The concept of distinct historical signatures can be useful as a heuristic aid, but it is not a sound statistical basis for hypothesis testing. Despite pointed efforts to explain the conceptual flaws of NCPA (Knowles and Maddison 2002, Petit and Grivet 2002, Beaumont and Panchal 2008), which have been backed up by multiple computer simulation studies demonstrating that it is usually wrong (Knowles and Maddison 2002, Panchal et al. 2007, Panchal and Beaumont 2010), NCPA is still used in phylogeographic studies of marine decapods and other taxa. It has been suggested that in spite of its proven inaccuracy NCPA continues to be used because it is the only method that promises to reconstruct detailed biogeographic histories from modest amounts of data (Knowles 2008).

5.2 The debate over the importance of selection

A second debate that is still to be resolved concerns the role of natural selection in shaping phylogeographic structure. According to the neutral theory of molecular evolution, selection primarily clears deleterious mutations from populations (Kimura 1968) leaving behind selectively neutral polymorphisms (Kimura and Ohta 1971). This is a core assumption of phylogeographic analysis. However, the neutral theory also predicts that beneficial mutations also appear, although infrequently, and are swept to fixation while displacing the previously neutral polymorphisms (Kimura 1983). Sites in nuclear loci that undergo these selective sweeps will drag nearby tightly linked sites to fixation with them, creating small genomic regions of reduced polymorphism (Smith and Haigh 1975, Gillespie 2000). However, in the case of the animal mitochondrial genome the near absence of recombination means that selective sweeps will carry the entire genome to fixation, eliminating all mtDNA polymorphism. Without unlinked regions for comparisons, the reduction in mtDNA polymorphism from a selective sweep is indistinguishable from one caused by a population bottleneck or founder event. Furthermore, the larger the population, the more often favorable mutations will arise and undergo selective sweeps that eliminate polymorphism. This effect of selective sweeps is the opposite of that expected for purely neutral polymorphisms, which would reach higher levels in larger populations. It could explain why estimates of effective population size based on mtDNA polymorphism are often far below biologically reasonable values (Avise et al. 1988) and why levels of mtDNA polymorphism in different species may be similar despite what are likely to be
order of magnitude differences in effective population size (Bazin et al. 2006). The possibility of selective sweeps has led some to question estimates of historical effective population size based solely on mtDNA polymorphism (Galtier et al. 2009). However, without a reliable method to distinguish the effects of selective sweeps from population size effects this remains a heated subject of debate (Meiklejohn et al. 2007).

Selective sweeps do not necessarily begin with new, random mutations. Two other possibilities are selection of preexisting variants that become favored as a result of environmental change and introgression of selectively favored variants from other species. There is clear evidence of introgression in Drosophila and freshwater fish (reviewed by Ballard and Whitlock 2004). For mtDNA, a third possibility exists because mitochondria are maternally inherited in most animal taxa. Just as selection acting on any part of the mitochondrial genome drags along the entire genome because of linkage, so does selection acting on any maternally transmitted factor, such as vertically transmitted symbionts. Selection on symbionts has been well documented in two arthropod groups: insects and isopods (reviewed by Hurst and Jiggins 2005), and parasitic symbionts appear to be common in arthropods. For example, the alphaproteobacterium Wolbachia has been detected in over 20% of insect species, 50% of spiders, and 35% of isopods. Wolbachia alters its host’s reproductive system to favor its own transmission, typically by either a distortion of the sex-ratio or by cytoplasmic incompatibility, in which eggs of uninfected females are killed when fertilized by the sperm of infected males. The effects of indirect selection on mtDNA via symbionts are not limited to selective sweeps that reduce polymorphism within populations. They can also increase polymorphism (if there are multiple symbiont strains within a population), and either reduce or increase differentiation between populations (Hurst and Jiggins 2005). Although Wolbachia has not been detected in decapod crustaceans (Bouchon et al. 1998), inherited bacteria appear to be common in arthropods (Duron et al. 2008) and the potential for indirect selection on decapod mtDNA should not be ignored.

At present, it is unclear how often or to what extent phylogeographic inferences based on mtDNA alone are compromised by the effects of selection. However, the new view that planktonic dispersal does not necessarily imply overwhelming gene flow, along with a new appreciation for the potential of selection on mtDNA suggests the need for a broader view of the causes of phylogeographic structure in decapod crustaceans. Furthermore, we should avoid basing phylogeographic inferences solely on mtDNA data. Nuclear DNA markers are now relatively easy to assay and methods of phylogeographic analysis based on sound statistical principles can be applied to both mitochondrial and nuclear data.

6 DIRECTIONS FOR FUTURE RESEARCH

The hope that allozyme studies would uncover patterns of genetic differentiation caused by selection was based on the fact that allozymes correspond to protein polymorphisms that could conceivably cause differences in fitness (Lewontin 1974). However, the underlying sequence variation responsible for observed differences in electrophoretic mobility cannot be determined from traditional allozyme scoring methods. Revolutionary advances in DNA sequencing and other methods since the era of classic allozyme studies now make it relatively inexpensive and straightforward to uncover sequence variation. Single nucleotide polymorphism (SNP) markers offer an exciting opportunity to return to the intellectually
fertile ground of allozyme studies with the power of modern molecular methods. SNPs are bi-allelic codominant markers that correspond to single nucleotide substitutions in DNA sequences. They occur in high frequency across animal genomes, and are often considered preferable to other more commonly used frequency-based genetic markers such as microsatellites and AFLPs because their evolution corresponds to simple mutation models, they are comparatively easy to genotype and are not prone to null alleles (Ranade et al. 2001, Brumfield et al. 2003). These advantages have made SNPs a popular marker for medical and agricultural studies for a number of years, but they have only recently been applied to a broader range of taxa in molecular ecology, evolution and population genetics studies (Morin et al. 2004). Despite increasing acceptance of SNPs for marine conservation and fisheries research, to date, few decapod studies have utilized SNPs (Smith et al. 2005, Zeng et al. 2009).

Within protein-coding genes, SNPs can either be synonymous, which means each allele codes for the same amino acid, or nonsynonymous, which means they correspond to amino acid differences that could potentially affect protein function. In this sense, SNPs can be viewed as highly informative allozyme markers, and comparisons between synonymous and nonsynonymous SNPs can be used to test for selection. For example, population parameter estimates generated from synonymous SNPs can be compared to those generated from nonsynonymous SNPs to test the hypothesis that selection on amino acid substitutions has altered those parameters. Such an approach can greatly extend our understanding of how decapods adapt to different environmental conditions and how selection influences population structure. In addition, recognizing and investigating the extent of local adaptation in decapod populations will better inform our conservation and management decisions.

Use of SNPs in investigations of selection in marine species has already produced some remarkable results. For example, a single amino acid residue substitution in the Na+ channel pore protein sequence in the softshell clam, *Mya arenaria*, is known to confer resistance to saxitoxin, a toxin associated with paralytic shellfish poisoning (Bricelj et al. 2005). Selection for alleles associated with saxitoxin resistance in regions with a history of toxic algal blooms provides a highly convincing explanation for the observed population level variation of *M. arenaria* along the coast of New England (Connell et al. 2007). However, a prior investigation into the population genetics of *M. arenaria* in the same region found no evidence of genetic population structure based on sequences from the ribosomal internal transcribed spacer (ITS) gene (Caporale et al. 1997). This case highlights the importance of including multiple loci in analyses that aim to describe population connectivity, but more importantly it also emphasizes the need to investigate genetic patterns based on loci from genes that are candidates for natural selection.

The unprecedented amount of genetic data currently being generated for model organisms, as well as an increasing number of non-model organisms, is quickly changing the field of population genetics to a genomic-based discipline. These data provide a wealth of opportunities to investigate sequence level variation in protein-coding genes. In particular, alignments of expressed sequence tags (ESTs) generated from cDNA libraries can yield large numbers of SNPs. Creating EST libraries for decapods (and mining those that are already available) will undoubtedly generate countless SNP markers that can be used to better infer population histories and estimate population parameters for species in this group.

Advances in statistical techniques that can be applied to SNP data also further our ability to detect selection. For example, outlier tests are a useful tool for identifying
individual loci that may be under selection (Beaumont and Nichols 1996, Antao et al. 2008). These tests utilize simulations to generate an expected neutral distribution based on $F_{ST}$ and heterozygosity values and classify loci that deviate from this distribution as potential candidates for selection. The importance of testing loci for selection in this manner can be illustrated by a study involving two morphotypes of the intertidal snail, *Littorina saxatilis*. Wilding et al. (2001) constructed phylogenies for *L. saxatilis* using AFLP markers and found two highly conflicting patterns. A phylogeny constructed from 290 AFLP loci showed snails grouped by morphotype and habitat despite being collected from locations separated by as much as 300 km of coastline. After removal of 15 outlier loci from the analysis, snails clustered by geographic location. This second scenario is more likely to reflect actual population history given this species’ direct development and exceptionally limited vagility as adults. The outlier pattern was observed across all geographic sampling locations and is consistent with a previously reported habitat-based selection gradient for *L. saxatilis* (Johannesson et al. 1995). Clearly, selection can produce complex patterns of population structure in marine species that are difficult to interpret under the assumption of marker neutrality. Therefore, *a priori* testing of loci for selection should be the first step of any population genetic analysis (Luikart et al. 2003).

7 CONCLUSIONS

A growing body of evidence provides overwhelming support for a new view of larval dispersal for meroplanktonic species, including decapods. Oceanographic processes and larval behavior can significantly alter dispersal distances from what would be expected in simplified models assuming passive particles traveling in unidirectional flows. Realized larval dispersal is further complicated by the potential for pre- and post-settlement selection acting on dispersing larvae. In this case, larvae may disperse, but they are not adapted for the conditions they encounter and gene flow does not occur. Several studies of meroplanktonic species suggest selection plays a significant role in structuring populations, and it is likely that investigations of selection in decapod populations will yield similar conclusions. The mounting genetic data currently being generated for decapods and advances in sophisticated statistical techniques offer the promise of numerous, highly informative SNP that can be used to test hypotheses of neutrality.

ACKNOWLEDGEMENTS

We would like to thank Christoph Schubart, Christoph Held and Stefan Köenemann for their invitation to contribute to this edition of *Crustacean Issues*, and for their excellent work in editing this volume. We would also like to thank two anonymous reviewers for detailed comments that have improved our chapter.

REFERENCES


