Asymmetric dispersal allows an upstream region to control population structure throughout a species’ range

James M. Pringle1, April M. H. Blakeslee2, James E. Byers6, and Joe Roman4

1Ocean Process Analysis Laboratory, University of New Hampshire, Durham, NH 03824; 2Marine Invasions Laboratory, Smithsonian Environmental Research Center, Edgewater, MD 21037; 3Odum School of Ecology, University of Georgia, Athens, GA 30602; and 4Gund Institute for Ecological Economics, University of Vermont, Burlington, VT 05405

Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved August 3, 2011 (received for review January 10, 2011)

In a single well-mixed population, equally abundant neutral alleles are equally likely to persist. However, in spatially complex populations structured by an asymmetric dispersal mechanism, such as a coastal population where larvae are predominantly moved downstream by currents, the eventual frequency of neutral haplotypes will depend on their initial spatial location. In our study of the progression of two spatially separate, genetically distinct introductions of the European green crab (Carcinus maenas) along the coast of eastern North America, we captured this process in action. We documented the shift of the genetic cline in this species over 8 y, and here we detail how the upstream haplotypes are beginning to dominate the system. This quantification of an evolving genetic boundary in a coastal system demonstrates that novel genetic alleles or haplotypes that arise or are introduced into upstream retention zones (regions whose export of larvae is not balanced by import from elsewhere) will increase in frequency in the entire system. This phenomenon should be widespread when there is asymmetrical dispersal, in the oceans or on land, suggesting that the upstream edge of a species’ range can influence genetic diversity throughout its distribution. Efforts to protect the upstream edge of an asymmetrically dispersing species’ range are vital to conserving genetic diversity in the species.

N ovel genetic material can appear in a population through mutation, migration, or long-distance dispersal events which may be human-mediated. In a single well-mixed population, the evolution of the frequency of novel neutral alleles will be governed by random genetic drift, not their initial spatial distribution (1). However, spatial structure and complexity can alter this expectation. In a metapopulation linked by migration, alleles introduced into source populations are more likely to persist than those that are introduced into sinks (2–4). Many meta-populations are embedded in complex spatial systems with a preferential direction of migration (“asymmetric dispersal”). In these systems, little is known about the equilibrium frequency of novel alleles or how this frequency depends on the location where these new lineages appear.

Asymmetric dispersal is common where propagules are carried long distances by wind or water. In atmospheric, riverine, and oceanic flows, there is usually a predominant flow direction (downstream or downwind) that biases dispersal, and eddies or weather systems that slow or reverse such flow add a stochastic (and potentially upstream) component of migration. For example, many terrestrial plant species have propagules that can be dispersed by the winds (5), and a recent study of the spatial patterns of diversity in moss, liverwort, and lichen flora in the Southern Hemisphere were found to be best explained by the predominantly downwind dispersal (6). Further evidence that asymmetric dispersal can structure a species’ genetic patterns in ecologically important ways can also be found in fungal rust plant pathogens that are dispersed by winds: long-distance dispersal events can lead to the repeated introduction of genetic lineages from upwind regions into downwind ones, where they cause new disease outbreaks (7, 8).

Asymmetric dispersal is common in coastal marine systems for species with planktonic larvae. Mean currents transport most larvae downstream even as some larvae are dispersed upstream by eddies (9). In such systems, a study of the spatiotemporal evolution of the distribution of newly introduced, or locally evolved, genotypes as a function of the location of their origin will help us understand how asymmetric dispersal structures the population. It will also demonstrate which regions are most important to the genetic diversity of the overall population and how changes in these areas can impact the genetic structure of the species over the rest of its range. Theoretical work argues that novel genetic information introduced into the upstream edge of a population is likely to persist (10–13). Neutral genotypes that are introduced or arise elsewhere in the range will tend to be lost as deceased adults are replaced by migrants from farther upstream. The upstream edge of the species domain can retain new genotypes because there is little or no external immigration from farther upstream (14, 15), and so the population must be maintained by local production. This upstream region will export migrants containing these genotypes downstream, which will, in time, spread the genotype throughout the entire species’ range.

Species introductions provide a unique opportunity to examine the genetic consequences of asymmetric dispersal. Because introductions in their early stages are typically not in equilibrium, they provide an opportunity to observe the evolution of gene frequencies. From an applied perspective, understanding the spread of invasive genotypes is also essential for identifying source populations and potential harbors for future invasions.

Here, we describe one such introduction of novel genotypes into a coastal marine system and document how the location and prevailing circulation patterns that govern dispersal allowed newly introduced and quite rare haplotypes to grow in frequency in the population. The European green crab, Carcinus maenas (hereafter Carcinus), is a widespread invasive species that was first observed in the eastern United States in 1817 (16) (Fig. 1). It took ~80 y for the crab to expand north of Cape Cod to the Gulf of Maine and then another 50 before it became established in...
Halifax Harbor. For several decades, ephemeral populations were witnessed northeast of Halifax, Nova Scotia, but they did not persist. In the 1990s, the crab displayed what appeared to be a rapid range expansion throughout the Canadian Maritimes (17). Genetic analyses revealed that these new populations of Carcinus were genetically distinct from downstream populations, suggesting a recent, novel introduction to northern Nova Scotia in the late 1980s or early 1990s, likely originating from northern populations in the green crab’s native European range (18). The repeated introductions of Carcinus therefore provide an excellent natural experiment to test theories of population genetics and asymmetric dispersal.

In an 8-y study of the mitochondrial haplotype distribution of Carcinus in the Gulf of Maine and along the Scotian Shelf of northeast North America, we examined the evolution of a cline that has formed between the population established in the 19th century and the newly introduced lineages that have been established for less than 20 y. Young of the year from 2002 and 2007 were collected, sequenced, and compared with a baseline collection from 1999 to 2000 (18) (Fig. 2). The haplotypes are presented in two major classes in our analysis, one for the preexisting 19th-century lineages (southern haplotypes), and a second for the recently introduced haplotypes (northern haplotypes). Details on these groupings can be found in ref. 18 and Materials and Methods.

Over our 8-y timeframe, there has been a widespread increase in the recently introduced lineages. The spatial averaged frequency of the northern haplotypes between Louisbourg, Nova Scotia, and Cape Cod increased by 25% from 1999 to 2000 to 2007 ($P < 0.01$, Table 1). This average is not equivalent to the change of the haplotype frequency in the population because it fails to account for spatial variation in population density (Materials and Methods). However, the northern haplotype frequency increases from 1999 to 2007 at nearly all sampling locations except for those at the extreme northern edge of the range (Fig. 2), suggesting that the haplotype frequency in the entire population is also increasing.

Coincident with the overall increase in the frequency of the northern genotypes, the haplotype frequencies across the sampling area have become more similar to those at the upstream edge of the domain. A comparison of the level of population differentiation ($F_{st}$) value from Louisbourg, at the upstream edge...
of the crab’s range, to all downstream sampling points shows a decrease in pairwise differentiation from 1999 to 2000 to 2007 for all but one of the sampling locations (Fig. 3) (the change from 1999 to 2002 to 2007 is not significant (P < 0.10), but the differences between 2002 and 2000 and 2007 are (P < 0.01)).

To examine whether the observed pattern can reasonably be ascribed to larval dispersal driven by coastal circulation under the assumption of neutral genetic markers, we created a numerical model of the evolution of haplotype frequencies based on our estimates of the green crab’s demographic parameters. The model uses the observed haplotype frequency in 1999–2000, 2002, and 2007 to estimate dispersal, produce optimally smoothed estimates of haplotype frequency in 2002 and 2007, and predict haplotype frequency in the years after 2007 (Materials and Methods). The estimates of larval dispersal are then compared in the discussion to estimates derived from the observed regional ocean circulation to determine whether the evolution of the haplotype frequency is consistent with circulation-driven asymmetric dispersal. Assuming allelic neutrality, the model estimates that the mean distance the larvae transported downstream before recruiting, $L_{adv}$ is $67 \pm 17$ km, whereas the SD of the larval transport distance, $L_{diff}$ is $234 \pm 19$ km. $L_{diff}$ is greater than $L_{adv}$ suggesting significant upstream transport of larvae even as most move downstream. Interannual variability does not influence these estimates greatly, for we get similar estimates of $L_{adv}$ and $L_{diff}$ if the model is fit to only the data from 2002 or 2007 (Table 2). These estimates are robust to errors in the demographic parameters; changing the lifespan by 2 y or changing the age/ fecundity relation biases the estimates of $L_{adv}$ by $\sim 20\%$ (Table 3; the origin of the sensitivity is described in Materials and Methods).

Table 1. The along-shore distance-weighted mean allele frequency of the northern type allele calculated from Cape Cod (Barnstable, MA) to Louisbourg, Nova Scotia, along with the SE of the estimates

<table>
<thead>
<tr>
<th>Year</th>
<th>Frequency (F)</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999–2000</td>
<td>0.36 ± 0.014</td>
<td>Mixed ages</td>
</tr>
<tr>
<td>2002</td>
<td>0.39 ± 0.012</td>
<td>Young of the year</td>
</tr>
<tr>
<td>2007</td>
<td>0.45 ± 0.017</td>
<td>Young of the year</td>
</tr>
</tbody>
</table>

The difference in frequency between 1999–2000 and 2002 is not significant (P < 0.10), but the differences between 2007 and 2002 and 2000 are (P < 0.01).

Fig. 3. (A) Pairwise $F_{st}$ values calculated between the northernmost sample site (Louisbourg, Nova Scotia) and all other sites for the years 1999–2000, 2002, and 2007. Higher values of pairwise $F_{st}$ indicate greater differentiation of haplotype frequencies between Louisbourg and downstream populations (*$P > 0.05$). (B) The frequency of the southern haplotype class. Markers represent observed frequencies of the southern haplotype class in 1999–2000 (black circles), 2002 (blue triangles), and 2007 (red squares). Colored lines are the model’s best estimate of haplotype frequency for these years for the age classes observed; the model and data match in 1999–2000 because this data are used to initialize the model. The two gray dashed lines represent the model’s predictions of haplotype frequencies in 2014 and 2021.

Discussion of the Carcinus System

In our 8-y investigation, we observed several noteworthy patterns that illuminate the effects of asymmetric dispersal on genetic diversity patterns in advective systems: (i) We discovered a downstream shift in the position of the center of the genetic cline between established haplotypes and those introduced within the past several decades, (ii) we observed an increase in the overall frequencies of the newly introduced haplotypes throughout the crab’s range, and (iii) we witnessed a decrease in pairwise differentiation ($F_{st}$) between sites across the crab’s range. Along with our estimates of mean dispersal given above, these patterns are consistent with recent theories of genotype spread and clinal evolution in advective coastal ocean systems (10). The mean downstream transport of larvae, $L_{adv}$, drives the dispersal parameters and selection is not strong enough to change the conclusions presented below.

The observed cline between the northern and southern haplotype assemblages shifts downstream (to the southwest) and broadens with time (Figs. 2 and 3). The modeling predicts that this evolution will continue after 2007, with the frequency of the southern haplotype continuing to decrease in the south (Fig. 3) and increasing slightly in northern populations. The location of the midpoint of the cline should shift south by $\sim 300$ km from 1999 to 2000 to 2021. This prediction is robust to errors in the demographic parameters and selection; the model compensates for any error in these parameters by lowering the estimates of larval dispersal parameters $L_{adv}$ and $L_{diff}$ to maintain the fit to the observations. The model’s robustness is quantified by the error of the fit between model and data (the sum of the squared difference between the modeled and observed haplotype frequencies; “Error” in Tables 2 and 3). This error only changes slightly, and the predictions of the future cline positions through 2021 do not change, when we include moderate selection ($s \sim \pm 0.1$) or when we change demographic parameters (Materials and Methods). In the long run, however, a favored southern haplotype would begin to grow in frequency again if, as observed, it is present at the upstream edge of the species’ range.
Table 2. Model estimates of mean larval dispersal parameters obtained by fitting to the observed haplotype data and using the best estimates of lifespan, age/fecundity ratios, and selection

<table>
<thead>
<tr>
<th>Selection for northern assemblage</th>
<th>Fit to 2002 data</th>
<th>Fit to 2007 data</th>
<th>Fit to 2002 and 2007 data</th>
<th>Fit to 2002 and 2007 data, no outlier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L_{adv}, km</td>
<td>L_{diff}, km</td>
<td>Error</td>
<td>L_{adv}, km</td>
</tr>
<tr>
<td>-0.1</td>
<td>78</td>
<td>160</td>
<td>1.12</td>
<td>91</td>
</tr>
<tr>
<td>0</td>
<td>62</td>
<td>157</td>
<td>1.12</td>
<td>69</td>
</tr>
<tr>
<td>0.1</td>
<td>45</td>
<td>153</td>
<td>1.13</td>
<td>44</td>
</tr>
</tbody>
</table>

To show the sensitivity to selection, we estimate L_{adv} and L_{diff} for different values of selection, and to estimate interannual variability in the system, we present estimates from fits to data from different years. Negative values of s indicate selection against the northern haplotype assemblage, and for the southern assemblage, vice versa. Entries in boldface indicate the base parameter set presented in the article. Error is the sum of the squared difference between the observed and modeled haplotype frequencies in both 2002 and 2007. The calculations are made by fitting the model to just the 2002 data, just the 2007 data, both the 2002 and 2007 data, and both the 2002 and 2007 data with the exception of the outlier in Liverpool, Nova Scotia, in 2007. The SEs of L_{adv} and L_{diff} do not vary greatly between the various estimates, and for the base parameter set are 17 km for L_{adv} and 20 km for L_{diff}.

downstream shift in the cline between the two haplotype classes and leads to the increase in the introduced haplotypes’ abundances in the population. Fewer than 20 y after the upstream establishment of the new invasion front, the dominant southwestward coastal currents (Fig. 1) have increased the frequency of the introduced haplotypes throughout the Scotian Shelf, the Gulf of Maine, and even as far as Long Island Sound, 1,800 km to the south. These population-level changes in haplotype frequencies illustrate the sensitivity of the upstream border of a species’ range to anthropogenic change, whether in the form of new introductions or habitat changes that create new retention zones or allow the species range to expand upstream.

The variability in the transport of individual larvae L_{diff} increases the width of the cline (Fig. 3), leading to genetic admixture, with the southern haplotype becoming present in the northernmost sampling sites and vice versa. The magnitude of L_{diff} is greater than that of L_{adv}, indicating significant upstream larval transport, which drives the spread of the southern haplotypes into the upstream-most portion of the range; this spread is significant because the appearance of the southern haplotype class at the northern edge of the range should allow those haplotypes to persist in the population (10). Both the downstream shift in the cline and the incursion of the southern haplotypes into northern populations has reduced the pairwise differentiation observed between the northern sites and the downstream locations. Left unanswered in this analysis is why stochastic larval transport had not moved the crabs with the southern haplotypes northward before the second introduction. The area just to the west of Halifax may have lacked sufficient retention zones for Carcinus to persist and spread to the Strait of Canso and Chebucto Bay, currently an important retention zone. It is also possible that Allee effects before the establishment of crabs in northern regions after the second introduction prevented sporadic larval recruitments from establishing populations upstream on their own; however, our current observations cannot confirm or refute this hypothesis.

The observed downstream transport of the larvae that drives the movement of the cline is consistent with local oceanographic conditions and our understanding of coastal larval dynamics. Estimates of the regional currents (summarized in Materials and Methods) suggest a mean downstream transport of 280 km for a surface drifter in the core of the coastal current for the planktonic duration of the larvae; this would be the mean downstream transport of larvae that moved immediately to the core of the coastal current after spawning (~30–50 km off shore), remained at the surface, shifted in the core of that current for their entire larval duration, and then moved immediately back to shore to settle. It is implausible to expect the Carcinus larvae to make this exact journey with this exact timing because, at least until their later stages, the planktonic larvae cannot swim a significant distance across the shelf. Also, the larvae do not spend all of their time at the surface (19, 20). Shanks and colleagues have found the observed mean dispersal distance of larvae is usually three to five times less than what one would predict from the larval duration and the mean surface currents in the central part of the continental shelf (21, 22). Our estimate of L_{adv} is consistent with this finding.

The SD of the larval transport distance (L_{diff} in Table 2), which describes the variability of dispersal experienced by individual larvae, is higher than would be expected based on the regional circulation statistics. L_{diff} is driven by the variability in the currents experienced by larvae spawned at different times or entrained into different coastal eddies (9, 14); it could also include the effects of other larval transport mechanisms, such as shipping or fisheries. There is strong weather system-driven (23, 24) and interannual variability (25, 26) in the coastal current; its SD is comparable to the mean, suggesting that the upper limit on the stochastic component of larval transport should be the magnitude of the mean larval transport L_{adv} (14). Despite these expectations, the variability in larval transport diagnosed by the model from the evolution of the cline is ~3.5 times greater than the mean larval transport. It does not seem likely that this discrepancy is caused by larval behavior because behavior that shelters larvae from the mean larval transport. It does not seem likely that this discrepancy is caused by larval behavior because behavior that shelters larvae from the mean larval transport. It does not seem likely that this discrepancy is caused by larval behavior because behavior that shelters larvae from the mean larval transport. It does not seem likely that this discrepancy is caused by larval behavior because behavior that shelters larvae from the mean larval transport. It does not seem likely that this discrepancy is caused by larval behavior because behavior that shelters larvae from the mean larval transport. It does not seem likely that this discrepancy is caused by larval behavior because behavior that shelters larvae from the mean larval transport. It does not seem likely that this discrepancy is caused by larval behavior because behavior that shelters larvae from the mean larval transport. It does not seem likely that this discrepancy is caused by larval behavior because behavior that shelters larvae from the mean larval transport. It does not seem likely that this discrepancy is caused by larval behavior because behavior that shelters larvae from the mean larval transport.

We cannot determine from the data whether the large stochastic long-distance dispersal (i.e., the high value of L_{diff}) is because of the transport of larvae by unusual and extreme oceanographic conditions or by some other (perhaps anthropogenic) transport mechanism. The model is sensitive to any larval transport mechanism, and even a small number of extreme events can significantly alter the SD. For example, there is a clear
How it all changes in a species range to modify its genetics throughout its range. Ultimately, our observations suggest that asymmetric migration allows any processes that affect the population genetics of a species at the upstream edge of its range to alter the population genetics of the species everywhere.

**Materials and Methods**

**Genetic Data.** J.R. collected adult and juvenile *Carcinus* in 1999 and 2000 from 23 sites throughout the species’ range in the Northwest Atlantic (18). In late summer 2002 and 2007, young-of-the-year juvenile crabs were collected approximately every 50 km along the northern Mid-Atlantic Bight, the Gulf of Maine, and the Scotian Shelf. Thirty locations each were sampled between Rye, NY, and Louisbourg, Nova Scotia, in both 2002 and 2007 (see SI Materials and Methods for details of the collection and sequencing of the mitochondrial CO1 gene).

**Data Analysis.** Haplotypes were grouped into two classes as described in ref. 18. The southern haplotype class, shown in white in Fig. 2, includes haplotypes that were part of the original founding population and is found in the southern part of the range and in museum samples collected in southern Nova Scotia before 1976. The newly introduced northern haplotypes, shown in black in Fig. 2, were introduced on the northern edge of the range in the late 20th century. The frequency of the haplotype classes is given for each year and location in Fig. 2. Summary statistics for genetic variability are available in Supporting Information, and details of the mapping of this data to alongshore distance and the Monte Carlo error estimation of allele frequency are given in SI Materials and Methods.

The fraction of the southern haplotype is shown as a function of alongshore distance in Fig. 3. To examine temporal change in the genetic cline recorded in 1999–2000, 2002, and 2007, we calculated pairwise differences with Arlequin version 2.0 (30) to assess change in population structure at all locations. Pairwise Fst values were calculated with the complete (non-grouped) haplotype data. Confidence limits in Fst were obtained by the default error estimation in Arlequin.

**Population Modeling.** The numerical model is based on the work of Pringle and Wares (10, 11) and simulates the downstream motion of the cline between the two haplotype assemblages. This spatially explicit model allowed for an age-structured population with age-dependent fecundity and density dependence caused by habitat limitation. It can include selection in favor of either haplotype assemblage. The model parameterized the larval dispersal in terms of the mean distance the larvae that successfully recruited are dispersed downstream (Ladv) and the SD of their dispersal distance (Ladv, SD). Because of the great uncertainty in the alongshore variability of the population density, especially the subtidal population (31), we assume that the population density is uniform. The model runs the dispersal kernel is assumed to be Gaussian (9); qualitatively identical results were obtained with a Laplacian dispersal kernel. The model is initialized with the distribution of the allele classes observed in 1999–2000; these observations mixed individuals from all year classes, and so the model is initialized with the same allele frequency for all ages and the population at its carrying capacity. The details of the demographic and life-history parameters used in the model are given in SI Materials and Methods. The demographic parameter with the largest impact on the models prediction of the cline’s evolution is the maximum age of the adults. The origin of this error is described in SI Materials and Methods, and a doubling of lifespan roughly halve the speed of downstream advance of a cline (cf. refs. 32 and 33).

Any selection for one of the haplotype assemblages would also change the speed the cline moves. As discussed in SI Materials and Methods, if the northern haplotype assemblage (or those genes associated with it) is favored everywhere in the domain, it will cause the cline to move southwestward faster, and if they are selected against, it will tend to cause the cline to move downstream more slowly, or even upstream.

The model is then fit to the observed haplotype frequencies to estimate the larval dispersal parameters that produce the best match between the modeled and observed haplotype frequencies. Powell’s method is used to find the values of mean larval transport Ladv and the stochastic component of larval transport Ladv, sth that minimizes the sum of the square of the difference between the modeled and observed haplotype distributions in 2002 and 2007 (34). Because the 2002 and 2007 sampling focused on the young of the year, we compare modeled young-of-the-year haplotype distributions to the data. Results are shown in Tables 2 and 3. Random errors in Ladv and Ladv, sth attributable to sampling statistics are estimated with a Monte Carlo
resampling of each location 1,000 times from a binomial distribution with the same expected value and number of samples as the observations. We cannot precisely quantify the errors in our estimate of \( L_{\text{obs}} \) and \( L_{\text{dist}} \) because of errors in our understanding of demographics or selection, for we do not have an a priori estimate of the uncertainty in our knowledge of these parameters. However, because the model finds the \( L_{\text{obs}} \) and \( L_{\text{dist}} \) that provide the best fit to the observed haplotype data, any error in our understanding of demographic parameters or selection will lead to errors in our estimate of \( L_{\text{obs}} \) and \( L_{\text{dist}} \) while keeping the evolution of the haplotype frequency consistent with the observations. Errors in parameters that tend to slow the downstream motion of the cline, e.g., a longer lifespan or selection for the southern haplotype assemblage, would lead to an increase in the model’s estimate of \( L_{\text{obs}} \) to compensate, whereas the converse would happen if we underestimated the lifespan of Carcinus or if there were selection for the northern haplotype assemblage. The sensitivity of the model’s estimates of \( L_{\text{obs}} \) and \( L_{\text{dist}} \) to these errors is shown in Tables 2 and 3. However, even if we err in our estimate of the demographic parameters of Carcinus, the model still produces estimates of the past and future distribution of the haplotypes that do not change. This lack of change is seen in the sum of the squared difference between the modeled and observed haplotype distributions, the error in Tables 2 and 3. This error does not increase significantly, and the predictions of the future cline positions do not change, when we include selection (Table 2) or when we change demographic parameters (Table 3).

Expected Larval Dispersal Parameters from the Regional Ocean Circulation. The expected mean and SD of the Carcinus larval transport distance can be roughly estimated for the Scotian Shelf and Gulf of Maine under rather idealized models of larval behavior. Although these estimates are very rough, because of the uncertainty in our knowledge of the larval behavior, they form a comparison point to the estimates found by fitting the haplotype evolution model to the observed haplotype data. The details of the regional oceanography and larval planktonic durations used to estimate these distances are given in Si Materials and Methods.

The mean downstream transport of larvae \( L_{\text{dist}} \) would be 280 km if the larvae moved immediately to the core of the coastal current after spawning (~30–50 km off shore, over the 60- to 100-m isobaths), drifted in the core of the coastal current for their entire larval duration, then moved immediately back to shore on settle. However, it is quite implausible to expect the planktonic larvae to be able to move immediately to the core of a current tens of kilometers off shore and then to return promptly when they are ready to recruit. Thus, we must treat 280 km as an upper bound on their transport distance and expect the actual larval transport to be smaller, perhaps quite significantly so. Comparisons of mean larval transport estimated from mean currents to indirect estimates of larval transport suggest that the estimates from midshelf currents tend to overestimate the mean transport distance by a factor of 3–5 (22).

The SD of the larval dispersal distance \( L_{\text{dist}} \) of the larvae released by an adult is driven by two distinct phenomena: (i) the dispersal of larvae released in a single dispersal event by eddies and fluctuations in the mean currents (9) and (ii) the varying dispersal distance of larvae released by a single adult during different seasons or different years (9, 14). The larger of these two will dominate the total \( L_{\text{dist}} \). Details of the estimate of \( L_{\text{dist}} \) from the variable circulation in the Gulf of Maine and Scotian Shelf are given in Si Materials and Methods; they show that the estimate of \( L_{\text{dist}} \) is dominated by interannual and intraseasonal variability in \( L_{\text{obs}} \) and should be of the same order as \( L_{\text{obs}} \).

ACKNOWLEDGMENTS. The original data collection and preliminary analyses were performed by J.R. in the laboratory of S. Palumbi, who was essential in catalyzing this research. We thank Carrie Keogh for help with data collection, Caitlin Hamer for field and laboratory assistance, and Michael Lesser for use of equipment and facilities. A.M.H.B. thanks Greg Ruiz, Whitman Miller, the Invasions laboratory, and the Plant Ecology laboratory at the Smithsonian Environmental Research Center for support and use of facilities. The 2007 data collection was supported by the Sloan Foundation’s Census of Marine Life. J.M.P. and J.E.B. were supported by National Science Foundation Grant OCE0961344.
Supporting Information

Pringle et al. 10.1073/pnas.1100473108

SI Materials and Methods

Gathering Genetic Data. In late summer 2002 and 2007, young-of-the-year juvenile crabs were collected approximately every 50 km along the northern Mid-Atlantic Bight, the Gulf of Maine, and the Scotian Shelf. Thirty locations each were sampled between Rye, NY, and Louisbourg, Nova Scotia, in both 2002 and 2007 (Dataset S1). Crabs were gathered in the upper intertidal zone, among shell hash, rockweed (Fucus vesiculosus), and wrack (Ascophyllum nodosum), beneath cobbles, or among rip-rap. The 20 smallest crabs located at each site (determined by carapace width) were collected and stored in 95% ethanol, except for 13 crabs collected during an author’s vacation near Rye, NY, which were stored in 80-proof vodka (40% alcohol; Absolut).

On the basis of Berrill’s study in the Gulf of Maine (1), young-of-the-year green crabs are defined as individuals ≤10-mm carapace width. In 2002, only crabs with a carapace width of ≤12 mm were analyzed, ensuring that there was no overlap in generations between 2000 and 2002 and that most crabs were in the same larval cohort. In 2007, there were several locations (Belfast, Lubec, Chance Harbor, West Quaco, Cape Sable, Kingsport, Liverpool, Irish Cove, and Louisbourg) where the average size of crabs at a site was >12 mm (ranging from 14 to 26 mm). Some of these crabs may have been 1 y old and in rare instances 2 y old; no large adults were analyzed to avoid overlap in generations. When possible, we gathered crabs from the same areas sampled in 1999–2000 or the nearest location possible (always within 5 km of the original location). In total, 19 collection sites were repeat sampled in 1999–2000 and 2002, 25 sites were repeat sampled in 2002 and 2007, and 18 sites were sampled across all three generations. Full details on sampling locations are provided in Dataset S1.

We included sequences from refs. 2 and 3 for the 1999–2000 and 2002 collection years and followed the published PCR and sequencing protocols (4) for the 2007 collection year, which consisted of amplifying and sequencing a 502-bp fragment of the mitochondrial CO1 gene. DNA was extracted with a 10% Chelex (Bio-Rad) solution (5) or a standard hexadecyltrimethylammonium bromide (CTAB) protocol (6) from gill or muscle tissue in periodods. PCR amplifications (30 cycles: 1 min at 94 °C, 1 min at 50 °C, and 1 min at 72 °C) were conducted with 2.5 μL of 10x buffer, 1.5 mM MgCl2, 0.5 mM dNTPs, 1 μM of each primer [5′-GCT TGA GCT GGC ATA GTA GG-3′, 5′-GAA TGA GGT GTT TAG ATT TCG-3′ (4)], and 0.4 unit Taq DNA polymerase. PCR fragments were sequenced on an ABI 3100 using BigDye (Perkin-Elmer) terminator chemistry. Forward and reverse sequences were aligned by eye using the program Sequencher 4.8 (Gene Codes).

Data Analysis. To model larval dispersal, it is necessary to linearize the coastline by mapping the observations onto a measure of alongshore distance that is parallel to the dominant currents. The along-shore distance is defined by the great-circle distance measured between the observation and waypoints marked by the black line in Fig. S1. The mean currents in the area are shown in Fig. 1 (7–10). In the Bay of Fundy, Manning et al.’s analysis of drifter data (9) shows that most drifters (and thus presumably larvae) cut northwestward across the bay around the location of Port George (see drifter tracks in Fig. S1). As modeled, our path cuts across the Bay of Fundy at this location. The zero distance is located on Cape Cod at Barnstable, MA, our southwestern-most collection locale in the Gulf of Maine; distance increases to the northeast.

The frequencies of the two haplotype classes are linearly interpolated between the observation points to provide a continuous estimate of haplotype frequency from Connecticut to Louisbourg (northeast Cape Breton). Several closely spaced sites are combined together for the analysis. These sites are shown as red dots in Fig. S1.

Errors in our data were dominated by sampling statistics: at each location and time, we have between 10 and 58 individuals collected, with ~20 individuals in most samples. Error analysis of the frequency data were performed with a Monte Carlo resampling of each location from a binomial distribution with the same expected value and number of samples as the observations. The linearly interpolated data are then averaged between Barnstable, MA, and Louisbourg, Nova Scotia, to obtain a space-weighted haplotype frequency. The means and SEs are given in Table 1. This frequency is equivalent to the frequency of the haplotype classes in the population only if there is no alongshore variation in population density.

Demographic and Life-History Parameters of Carcinus. The demographics and fertility of Carcinus are parameterized from observations and laboratory data (11, 12). The adults are assumed to live 5 y, and they become reproductively mature in the second year. Fecundity of the adults is assumed to increase linearly with age, doubling from the second year to the fifth. (Fecundity grows linearly with size is from ref. 13. Linear increase in size after sexual maturity is from ref. 14.) Because the model as used here assumes the total population density is in steady state, the absolute levels of fecundity do not matter, only the relative change with age.

Sources of Error in Modeling Cline Evolution. The demographic parameter with the largest impact on the models prediction of the cline’s evolution is the maximum age of the adults. Lifespan impacts the speed at which a cline drifts downstream because the allele mix of downstream populations in the next generation is determined not only by the larvae that arrive from upstream but also by the adults at the downstream location that have not died yet, relieving competition for habitat. For a species that lives for five generations, larvae from upstream replace only 1/5 of the population at any point each generation, but in the species that only lives one generation, all of the adults at a location are descended from larvae that came from upstream in the last generation. Thus, the longer the adults live, the longer it will take for the population at a downstream location to be entirely replaced by migrants from upstream. The persistence of adults carrying the downstream haplotypes in turn reduces the export of larvae carrying the upstream haplotype from that location to locations farther downstream. Both of these mechanism slow downstream cline motion and are illustrated in Fig. S2. A doubling of lifespan roughly halves the speed of downstream advance of a cline (cf. refs. 15 and 16).

Any selection for one of the haplotype assemblages would also change the speed the cline moves. Fisher, Cavalli-Sforza, and Moran (17–19) argue that selection will cause a cline to shift away from the initial location of a favored haplotype and that this shift in cline position is equivalent to the shift a mean downstream larval transport would produce (20). Thus, the speed of cline movement can be thought of as the sum of the “speed because of selection” and the “speed because of the mean downstream transport of larvae.” If the northern haplotype assemblage (or those genes associated with it) is favored every-
where in the domain, it will cause the cline to move southwestward faster, and if they are selected against, it will tend to cause the cline to move downstream more slowly, or even upstream.

Estimating Larval Transport Parameters from Oceanographic Conditions. The expected mean and SD of the Carcinus larval transport distance can be roughly estimated for the Scotian Shelf and Gulf of Maine under rather idealized models of larval behavior. The circulation in this area is dominated by a baroclinic current centered roughly over the 60- to 100-m isobath, with slower currents off shore and inshore of this depth (7, 9, 10). At the core of the current, the surface velocity is greatest, and the current decreases roughly linearly with depth to slower speeds near the bottom. We used the well-measured currents on the 65-m isobath in the Gulf of Maine (21) to parameterize larval transport. These currents are comparable to currents at similar depths in the Mid-Atlantic Bight (22) and along the Scotian Shelf (7). McMullin (21) found that the mean summer currents at the 65-m isobath in the Western Gulf of Maine was ~8 cm/s at the surface and 5.7 cm/s at middepth, with somewhat faster currents off shore and inshore of this depth (7, 9, 10). At the mean distance that the larvae at the 65-m isobath in the Western Gulf of Maine was ~17 °C when the larval duration is expected to be ~40 d; for the colder waters, ~10.5–12 °C near the Bay of Fundy, the larval duration is closer to 60 d. The waters at a depth of 25 m are more uniform in temperature and range from 10.5 to 12 °C, also suggesting a larval duration of ~60 d. [Temperature estimates are derived from a climatology of historical hydrographic data in the Bedford Institute of Oceanography Climate Database (8)]. It is unclear what the mean depth of the larvae is while they are in the plankton, although there is evidence of complex vertical behavior in Carcinus (24, 25). Using McMullin’s estimates of the flow speed and the larval durations given above, Ladv is ~280 km both for larvae that stay at the surface the entire time and those that stay at middepth because the decrease in current with depth is largely offset by the increase of larval planktonic duration (21).

The SD of the larval dispersal distance Ldiff of the larvae released by an adult is driven by two distinct phenomena: (i) the dispersal of larvae released in a single dispersal event by eddies and fluctuations in the mean currents (23) and (ii) the varying dispersal distance of larvae released by a single adult during different seasons or different years (23, 26). The larger of these two will dominate the total Ldiff (26). For larvae released in a single spawning event, Ldiff is given by the square root of the product of the larval duration, the decorrelation timescale of the currents, and the square of the SD of the currents. Observational data from the Gulf of Maine (8, 21, 26) give a SD of the subtidal alongshore currents of ~6 cm/s and a decorrelation timescale of ~2 d, leading to an estimate for Ldiff from a single release of ~50 km. The variability in larval dispersal driven by seasonal and interannual variability in the mean currents is set by the SD of the mean alongshore currents on these timescales; observations (10) suggest that this variability is of the same magnitude as the mean, which gives an Ldiff that is of the same size as the mean transport Ladv (26). Because our estimate of Ladv, from above is much greater than the 50 km variability expected in larval dispersal from a single release, the estimate of Ldiff is dominated by interannual and intraseasonal variability in Ladv and should be of the same order as Ladv.

Fig. S1. Map of the sampling locations. The green squares mark the C. maenas sampling locations used in the analysis, and the distance along the shelf is measured along the black lines connecting these points. The red dots are locations where haplotype data were gathered and lumped with nearby stations. Inset shows drifter paths in the Bay of Fundy region (data from ref. 1).


Fig. S2. A simple model of the spatial evolution of a cline in haplotype frequency in two organisms whose larvae move one habitat space to the right when they disperse. In A, the organism lives for five generations; in B, the organism lives for one generation. Haplotype frequency is represented by shading in the pie chart, and fecundity is constant with age. For organisms that live for 5 y, only 1/5 of habitat spaces open up each year. Thus, the larvae bearing novel haplotypes from upstream into a habitat take 5 y to replace the downstream haplotypes, and the downstream shift of the cline is slow. For the organisms that live for only 1 y, the invasion of the novel allele is much faster because the entire population at a location is replaced by larvae from upstream each year.

Dataset S1. Site locations and haplotype class frequencies for sites

Columns within the haplotype class section are frequencies of the southern haplotype class (represented by haplotypes 1, 2, and 3), the northern haplotype class (represented by haplotypes 4, 5, and 6), or other rare haplotypes (7, 8, 10, and 56) observed across sites. Haplotypes are classified as in ref. 1