MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 3692–3707 doi: 10.1111/j.1365-294X.2010.04694.x

Seascape genetics along a steep cline: using genetic patterns to test predictions of marine larval dispersal

HEATHER M. GALINDO,*[1] ANNA S. PFEIFFER-HERBERT,† MARGARET A. McMANUS,‡ YI CHAO,§ FEI CHAI¶ and STEPHEN R. PALUMBI*

*Department of Biology, Stanford University, Hopkins Marine Station, 120 Ocean View Blvd., Pacific Grove, CA 93950, USA, †Graduate School of Oceanography, University of Rhode Island, South Ferry Road, Narragansett, RI 02882, USA, ‡Department of Oceanography, University of Hawaii at Manoa, Marine Sciences Building, 1000 Pope Road, Honolulu, HI 96822, USA, §Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA, ¶School of Marine Sciences, 453 Aubert Hall, University of Maine, Orono, ME 04469, USA

Abstract

Coupled biological and physical oceanographic models are powerful tools for studying connectivity among marine populations because they simulate the movement of larvae based on ocean currents and larval characteristics. However, while the models themselves have been parameterized and verified with physical empirical data, the simulated patterns of connectivity have rarely been compared to field observations. We demonstrate a framework for testing biological-physical oceanographic models by using them to generate simulated spatial genetic patterns through a simple population genetic model, and then testing these predictions with empirical genetic data. Both agreement and mismatches between predicted and observed genetic patterns can provide insights into mechanisms influencing larval connectivity in the coastal ocean. We use a high-resolution ROMS-CoSINE biological-physical model for Monterey Bay, California specifically modified to simulate dispersal of the acorn barnacle, *Balanus glandula*. Predicted spatial genetic patterns generated from both seasonal and annual connectivity matrices did not match an observed genetic cline in this species at either a mitochondrial or nuclear gene. However, information from this mismatch generated hypotheses testable with our modelling framework that including natural selection, larval input from a southern direction and/or increased nearshore larval retention might provide a better fit between predicted and observed patterns. Indeed, moderate selection and a range of combined larval retention and southern input values dramatically improve the fit between simulated and observed spatial genetic patterns. Our results suggest that integrating population genetic models with coupled biological-physical oceanographic models can provide new insights and a new means of verifying model predictions.

Keywords: genetics, marine connectivity, numerical modelling, oceanography

Received 19 November 2009; revision received 2 March 2010; accepted 15 March 2010

Introduction

Landscape genetics takes into account the impact of geographic and population features that affect the way organisms and their genes move among localities (Manel *et al.* 2003). Dispersal barriers are not as immediately obvious in the marine environment (Galindo *et al.* 2006), and research interest has focused on the way ocean currents provide dispersal conveyer belts from place to place. Larval connections among spatially separate marine populations can have significant impacts on the local demography and genetic diversity of these populations, thereby affecting their evolutionary adaptability to environmental change (Kinlan & Gaines 2003; Gaines *et al.* 2007). Historically, the

Correspondence: Stephen R. Palumbi, Fax: +1 831 655 6215; E-mail: spalumbi@stanford.edu

[1]Present address: School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA.

probability that populations were connected by larval dispersal was estimated based on knowledge of average seasonal circulation patterns and the amount of time it took larvae to physiologically develop in the lab (reviewed in Shanks *et al.* 2003). More recently, physical oceanographic circulation models verified by empirical data allow for more precise simulation of larval movement among locations (Cowen *et al.* 2006; Pfeiffer-Herbert *et al.* 2007; Werner *et al.* 2007; Treml *et al.* 2008). These models often take into account larval biology in the form of pelagic larval duration, behaviour (usually changes in vertical depth over time), growth rates, mortality, and habitat preferences. Data on larval behaviour and growth are often derived from laboratory studies (see Pfeiffer-Herbert *et al.* 2007 for an example). The result is a series of larval movement predictions over short spatial and temporal scales that can be combined across many simulations to produce an estimate of the probability that a larva produced at one place disperses and settles at another. Other methods such as graph theory (Treml *et al.* 2008) produce similar results: a connectivity matrix that estimates the relative directional dispersal among locations within a complex population.

These connectivity matrices can be compared to empirical data in multiple ways: (i) The connectivity matrix generated from a coupled bio-physical oceanographic model may be compared to a connectivity matrix developed from direct observations of dispersal; or (ii) the connectivity matrix generated from a coupled bio-physical oceanographic model may be used to generate predictions of empirical patterns that can then be compared to field-based observational data. The first approach is more appropriate for studies where actual counts of the relative number of individuals moving among locations are known. This is the case when larger individuals are physically tagged with marker, archival, or satellite tags (Schaefer *et al.* 2007; Weng *et al.* 2007). The use of otolith or statolith microchemistry and genetic assignment tests to identify larval origins also falls into this category (Thorrold *et al.* 2001, 2007; Rooker *et al.* 2008). However, the most abundant sources of readily available empirical connectivity data are population genetic studies. Because genetic data integrate the signal of realized connectivity over time, and because gene flow can occur across generations via a stepping stone pattern of movement, it is not straightforward to use genetic data to create a connectivity matrix. Instead, it is more feasible to use a connectivity matrix to predict long-term average genetic patterns that develop over many generations (Galindo *et al.* 2006). This comparison of physical predictions of migration to empirical measures based on gene frequency differences has defined the field of landscape genetics, and its marine counterpart, seascape genetics (Galindo *et al.* 2006; Selkoe *et al.* 2008).

The prediction of spatial genetic patterns has often been based on simple models of population structure such as the Island model (Wright 1931) or Isolation-by-Distance model (Rousset 1997). Though Island model analyses have been extremely useful in population genetics, few natural populations completely meet the assumptions of the model, and as a result, the calculations derived from this model have limited practical uses (Whitlock & McCauley 1999). Likewise, Isolation-by-Distance models have provided important alternatives to a strictly Island model (Rousset 1997), and have allowed new insights into the ecology and conservation of natural populations (Palumbi 2003). However, interpretations depend heavily on ecological and reproductive data, such as local effective densities, that are rarely available (Pinsky *et al.* 2010). One solution, the recently developed Isolation-by-Resistance models, allows the prediction of spatial and temporal genetic patterns in a heterogeneous landscape, although these models rely on spatially explicit habitat maps that are difficult to generate for the temporally unstable marine environment (McRae 2006). In parallel, the development of coupled spatially explicit dispersal and population genetic models have a more extensive history in terrestrial systems (as reviewed by Epperson *et al.* 2010). Instead, in marine systems, we can use the connectivity matrices produced by empirically-based biological-physical oceanographic models to make genetic predictions for dispersal patterns in a complex seascape.

For example, a previous study of broadcast spawning corals in the Caribbean Sea used this approach and found correspondence between simulated and observed genetic patterns at a regional scale (Galindo *et al.* 2006). This was the first example of using connectivity matrices produced by a coupled biological-physical oceanographic model to directly simulate genetic patterns that are then compared to empirical data, but the recent release of a new genetic model also applied to the Caribbean suggests this approach is on the rise (Kool *et al.* 2009). Baums *et al.* (2006) found that simulated coral larvae did not cross an ocean passage between the eastern and western Caribbean, a dispersal break that matched a significant genetic difference between populations east and west of this divide. Other studies have compared spatial genetic patterns to observed oceanographic features. Selkoe *et al.* (2006) were able to correlate the genetic signature of recruiting cohorts of kelp bass (*Paralabrax clathratus*) to changes in current flows in the Santa Barbara Channel that were likely coming from different larval sources. Veliz *et al.* (2006) used a genetic model balancing fitness values and dispersal

rates of two allozyme loci under selection in the acorn barnacle *Semibalanus balanoides* to estimate asymmetrical larval exchange in the Gulf of St Lawrence. The resulting predictions were consistent with previous data on larval dispersal patterns and oceanographic circulation in the region. Gilg & Hilbish (2003) were able to find similar patterns of connectivity for blue mussels (*Mytilus edulis* and *M. galloprovincialis*) in southwest England using both a circulation model and empirical genetic data collect from recently settled recruits. In this Special Issue, Selkoe *et al.* (2010) use correlations among genetic patterns in three marine species in the Southern California Bight and environmental data to show that kelp cover and ocean circulation patterns affect the spatial genetic structuring of these populations.

Successful implementation of this seascape genetics approach requires a physical oceanographic model driven by empirical data, the ability to simulate biologically-relevant larval characteristics, and a spatially explicit empirical genetic dataset to which the predicted patterns can be compared. Although all of these elements are not likely to be available for all species at all locations, we can use existing opportunities to investigate the identity and magnitude of factors likely to affect population connectivity for species having a similar spatial distribution and life history. To that end, initial applications of this modelling framework should focus on exploring the sensitivity of predicted and observed data matches to changing a variety of parameters. Further, where there is a mismatch, the nature of the mismatch should be used to inform the choice of how and which factors of the model should be modified.

In order to demonstrate this approach, we present a comparison between simulated connectivity patterns for the intertidal barnacle *Balanus glandula* in Central California with empirical genetic data from the same species. Biologically relevant models including growth rates based on food availability and ontogenetic vertical migration have previously been developed for this species (Pfeiffer-Hoyt & McManus 2005; Pfeiffer-Herbert *et al.* 2007). These models were then integrated into a nested-grid Regional Ocean Modelling System (ROMS-CoSINE) framework for a region centered on Monterey Bay, CA. The *B. glandula* populations in this region have also been previously shown to exhibit a clinal pattern in allele frequencies at two loci, the mitochondrial gene cytochrome oxidase I and the nuclear gene elongation factor 1 alpha (Sotka *et al.* 2004). Together, these studies lay the foundation to test the ability of the coupled biological-physical oceanographic model to predict the observed clinal pattern. Our goal will be to use the modelling framework to investigate the factors influencing dispersal in an intertidal marine species living in an eastern boundary current system with seasonal upwelling.

Materials and methods

Genetic data

Approximately 50 adult individuals were collected at each of the 13 sampling locations: Pillar Point (PP)*, Pescadero State Beach (PSB), Waddell Creek (WC), Sand Hill Bluff (SHB), Natural Bridges/Terrace Point (TP), Capitola Beach (CB), Moss Landing Jetty (MLJ), Pacific Grove (PG)*, Asilomar/Point Piños (APP), Stillwater Cove (SC), Carmel Point (CP), Sobranes Point (SP), and Andrew Molera Beach (AM) (details in Table 1 and Fig. 1; * A subset of these samples were included in Sotka *et al.* 2004). These samples are concentrated in the center of the steep genetic cline extending between Cape Mendocino and Point Conception in *B. glandula* populations (Sotka *et al.* 2004). For each individual, a small amount of cirral tissue was dissected and placed in 100 μL 10% Chelex-100 resin (Bio-Rad Laboratories) solution for DNA extraction (Sotka *et al.* 2004). A mitochondrial locus (cytochrome oxidase I, COI) and a nuclear locus (elongation factor 1 alpha, EF1a) were sequenced in forward and in both directions, respectively, in all individuals. Individual haplotypes for COI were assigned to haplotype clades (i.e. A, B, or C) based on 386 basepairs of DNA sequence data according to the criteria in Sotka *et al.* (2004). Haplotype clade frequences for EF1a were determined by DNA sequence at basepairs 105–108 relative to the 156bp fragment in Sotka *et al.* (2004). (Clade A=TCCA, Clade B=TCAT, Clade C=CCCA.)

Larval transport model

To simulate pathways of larval transport among the 13 populations from which genetic data were collected, we used a series of numerical models. An ocean circulation model (ROMS) coupled with an ecosystem model (CoSINE) produced fields of horizontal current velocities, temperature and chlorophyll. A larval development model predicted larval growth rates from temperature and chlorophyll levels. Larvae also undergo ontogenetic vertical migration. Finally, a particle-tracking model used current velocities to advance the position of larvae over time.

Circulation model

We used the Regional Ocean Modelling System (ROMS; Shchepetkin & McWilliams 2005) to generate predictions of ocean currents along the central California coastline. For the Monterey Bay region, ROMS

Table 1 Frequency of COI-C haplotypes and EF1a-C haplotypes in populations of *Balanus glandula* sampled along the coast of central California. Clade designations are as in Sotka *et al.* (2004). Sample sizes (*N*) are listed separately for each locus. (* Simulated larvae were released 2 km offshore from these locations. †Our collecting permits for barnacles forbid collection of animals within the borders of PG, California except for at the Hopkins Marine Station (PG). For the Asilomar/Point Pinos oceanographic station, we interpolated gene frequencies based on Hopkins (PG) and SC data, from approximately 4 and 15 km away, respectively)

Locality	Latitude, Longitude*	Year	COI (C)	EF1 (C)	COI (*n*)	EF1 (*n*)
PP	37° 30′ N, 122° 28′ W	2002, 2004	0.61	0.40	54	54
Pescadero (PSB)	37° 16′ N, 122° 24′ W	2004	0.71	0.40	45	42
WC	37° 05′ N, 122° 16′ W	2004	0.83	0.53	41	48
SHB	36° 58′ N, 122° 09′ W	2004	0.80	0.71	48	45.5
Natural Bridges/TP	36° 57′ N, 122° 04′ W	2004	0.87	0.72	46	37
CB	36° 58′ N, 121° 57′ W	2004	0.87	0.65	39	48
MLJ	36° 49′ N, 121° 47′ W	2004	0.86	0.63	45	45
PG	36° 37′ N, 121° 54′ W	2003	0.90	0.68	58	58
Asilomar/Pt. Piños (APP)	36° 38′ N, 121° 56′ W	2007	0.94	0.67	–†	–†
SWC	36° 34′ N, 121° 56′ W	2005	0.98	0.66	43	48
CP	36° 32′ N, 121° 56′ W	2004	0.91	0.70	45	45
SP	36° 27′ N, 121° 56′ W	2004	0.95	0.70	43	44
AM	36° 18′ N, 121° 53′ W	2005	0.98	0.78	44	43

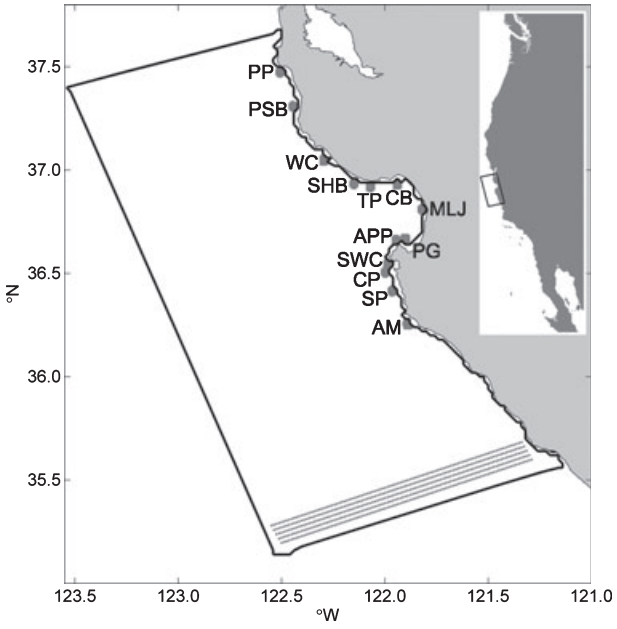

Fig. 1 Map of model domain in central California and location of larval release sites. Sites with genetic data are PP, PSB, WC, SHB, TP, CB, MLJ, PG, APP, SWC, CP, SP and AM Beach. Additional larvae were released along the southern boundary (SBY) of the model. Inset map shows location of model domain along the US west coast.

configurations consist of three nested domains: the entire US West coastal ocean at 15 km resolution, the central California coastal ocean at 5 km resolution, and the Monterey Bay region (Fig. 1) at 1.5 km resolution (Chao *et al.* 2009; Wang *et al.* 2009). All three ROMS configurations have 20 vertical layers and a minimum depth of 50 m. In the Monterey Bay region, the shoreward model boundary is on average 2 km offshore of the actual shoreline due to imprecision in the process of generating the model grid. The 15, 5 and 1.5 km regional ROMS are nested on-line as a single system and run simultaneously, exchanging boundary conditions at every time step of the parent grid. See Chao *et al.* (2009) and Wang *et al.* (2009) for more details about the ocean model.

The ROMS configuration described above has been integrated for many seasonal cycles using empirically measured climatological conditions and forced by known climatological air-sea fluxes. This manuscript describes simulations that used data from August 2003 (1 month) and from the period between August 2006 and July 2007 (~1 year). During the August 2003 integration, the air-sea fluxes were derived from the Coupled Ocean-Atmosphere Mesoscale Prediction System (Doyle *et al.* 2009) at 3 km resolution. Ocean temperature and salinity measurements collected from both remote sensing (e.g. satellites and aircraft) and in situ platforms (e.g. ships, gliders, Autonomous Underwater Vehicles) during the Autonomous Ocean Sampling Network II project of August 2003 were assimilated into ROMS using a three dimensional variational method (Li *et al.* 2008, 2009). During the August 2006–July 2007 integration, we assimilated both the *in situ* and remotely sensed observations every 6 h (Chao *et al.* 2008). The *in situ* data are derived from gliders, AUVs, moorings, and ship measurements, while the remotely sensed data are derived from satellite sea surface temperature measurements and high frequency (HF) radar surface current measurements. Chao *et al.* (2009)

compared ROMS model output to salinity data collected by gliders over depths of 10–50 m at five locations within the model domain. Correlations between the model and data were greater than 0.75 and RMS differences were less than 0.1 psu throughout the depth range (Chao *et al.* 2009). The velocity fields predicted by the model also compare well, qualitatively, with both vertical profiles of velocity and horizontal maps of surface currents (Chao *et al.* 2009). The model output was saved hourly, and the daily snapshots (at 1500 GMT) are used here. The result is a set of predicted ocean current patterns on grid points spaced 1.5 km apart throughout the Monterey Bay region.

Ecosystem model

An ecosystem model (Carbon, Silicate, and Nitrogen Ecosystem—CoSINE), developed by Chai *et al.* (2002), was embedded into the 3-level nested ROMS. CoSINE consists of 12 compartments: two classes each of phytoplankton (P1, P2) and zooplankton (Z1, Z2), dissolved inorganic nitrogen in the form of nitrate (NO_3) and ammonium (NH_4), detritus nitrogen (DN), detritus silicate (DSi), and total CO_2. Nitrate and ammonium are treated as separate nutrients, thus dividing primary production into new production and regenerated production. Below the euphotic zone, sinking particulate organic matter is converted to inorganic nutrients by a regeneration process similar to the one used by Chai *et al.* (1996), in which organic matter decays to ammonium and then is nitrified to NO_3. The modeled chlorophyll concentration is derived from the phytoplankton biomass (mmol N m^{-3}), converted to mg m^{-3} using a nominal gram chlorophyll-to-molar nitrogen ratio of 1.64, which corresponding to a chlorophyll-to-carbon mass ratio of 1:50 and a carbon-to-nitrogen molar ratio of 7.3.

The CoSINE model has been coupled with the ROMS for the Pacific basin and evaluated against available observations for different regions in the Pacific Ocean (Chai *et al.* 2009). The modeled nutrients and phytoplankton dynamics and carbon fluxes have been analysed in response to seasonal and interannual climate variation (Bidigare *et al.* 2009; Fujii *et al.* 2009; Liu & Chai 2009). For the Monterey Bay ROMS-CoSINE model simulations, all the biogeochemical fields (e.g. nutrients, planktons, and TCO_2) were initialized from the Pacific ROMS-CoSINE model simulation. Boundary conditions for the biogeochemical components were also used from the Pacific ROMS-CoSINE outputs.

The patterns of surface chlorophyll predicted by CoSINE reproduce patterns observed in remotely sensed chlorophyll data obtained from the MODIS satellite (http://oceancolor.gsfc.nasa.gov/cgi/l3) reasonably well. Mean chlorophyll levels, however, were lower in the model than in the satellite data. *In situ* data were available from nine vertical profiles collected in August 2006 between 36.6°N and 37°N (R. Kudela, Univ. of California, Santa Cruz, unpublished data). Comparisons of CoSINE output and the *in situ* chlorophyll data also revealed an offset in absolute levels of chlorophyll below the sea surface. To correct for this under-prediction of chlorophyll concentration in the model, we augmented the modeled chlorophyll by a scaling factor of 3.5, which was the median difference between the *in situ* data and the model. Although the duration of the larval period is sensitive to variations in chlorophyll (Pfeiffer-Hoyt & McManus 2005), the scaling factor did not have a large impact on predicted larval transport patterns. For example, when the CoSINE output was increased or decreased by the root mean square difference between modeled and observed chlorophyll, 96% of larvae settled within 1 km of their original settlement location (Pfeiffer-Herbert *et al.* 2007).

Larval development model

The development rate of *B. glandula* larvae is calculated as in Pfeiffer-Hoyt & McManus (2005). *B. glandula* larvae develop through six feeding naupliar stages that are not competent to settle, followed by a non-feeding cyprid stage that is competent to settle (Brown & Roughgarden 1985). The length of time that barnacle larvae spend as nauplii varies with temperature and food concentration (e.g. Anil & Kurian 1996). The results of laboratory experiments (Brown & Roughgarden 1995; Hentschel & Emlet 2000; R.B. Emlet, Oregon Institute of Marine Biology, unpublished data) that reared *B. glandula* larvae under varying temperature and food concentration levels and measured the duration of the naupliar stages were compiled into a lookup table. The duration of each naupliar stage as a fraction of total naupliar duration was obtained from laboratory experiments by Brown & Roughgarden (1985). The larval development subroutine uses temperature from ROMS and chlorophyll (as a proxy for food concentration) from CoSINE to determine the fraction of naupliar development that is completed during each model time step. The larval development subroutine tracks the naupliar stage achieved by each individual. Temperature and chlorophyll are stored at each 30 min time step of the particle tracking routine and averaged at the end of the day. Larval maturation is updated once per day.

Larval depth distribution

Field observations suggest that the depth distributions of many species of barnacle larvae change through the naupliar and cyprid stages. In field observations, the

larvae of many barnacle species, including one in the *Balanus* genus, appear to have a similar average depth range (B. Grantham, personal communication). For example, the average depth of *Balanus improvisus* larvae increases as naupliar stage increases (Bousfield 1955). Existing observations of *B. glandula* larvae are not adequate to determine stage-specific depth distributions for this modelling study, however detailed observations of larvae of the barnacle *Chthamalus* spp. were collected in the Monterey Bay region in 1993–1994 (B. A. Grantham, unpublished data). Thus, *Chthamalus* spp. was chosen as a substitute for *B. glandula* larval depth distributions based on our general knowledge of barnacle depth-distributions, and because the recruitment patterns of *Chthamalus* spp. and *B. glandula* are similar in the study area (Connolly *et al.* 2001).

The depth distribution of *B. glandula* larvae was parameterized after the average depth distribution of *Chthamalus* spp., which increases from 0 to 25 m as the larvae mature (after Pfeiffer-Herbert *et al.* (2007)). The cyprid stage returns to the surface layer of the model, as observed in the field (Grosberg 1982). Larvae are assumed to be capable of maintaining their preferred depth, as suggested by comparing averaged swimming speeds and typical vertical current velocities (Hardy & Bainbridge 1954; Franks 1992); thus all larvae of a particular larval stage maintain a prescribed depth in the model. It should be noted that the output of the larval transport model is not very sensitive to variations in the stage-specific depths within this depth range (Pfeiffer-Herbert *et al.* 2007).

Particle-tracking model

In the particle-tracking model, larvae are treated as particles that are advected by the horizontal current velocities predicted by ROMS, as described in Pfeiffer-Herbert *et al.* (2007). Simulated larval trajectories are calculated by a 4th-order Runge–Kutta integration scheme with a time step of 30 min. Larvae that cross a land boundary are returned to their previous water position. Larvae that cross an open sea boundary are considered 'lost' from the simulation.

Once a larva is competent to settle into its adult habitat, it is tracked until it arrives at a designated habitat zone, is lost through a sea boundary, or it reaches the 8-week time limit for larval stage duration (Strathmann *et al.* 1981). Habitat zones are 10 km^2 boxes (5 km in the along-shore direction and 2 km in the across-shore direction) placed along the model shoreline boundary. The 2 km width of habitat zones was selected to include larval trajectories that approach but never reach the model shoreline due to boundary effects on the predicted currents. Competent larvae that arrive within a habitat zone before the maximum larval period duration are called 'settlers'.

In order to estimate the amount of time it might take for a larva to be transported between the shoreline and the 50 m isobath (or ~2 km offshore), we used 7.5-year time series of current velocity data collected on the 20 m isobaths in both northern and southern Monterey Bay by the Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO) program. Over the whole time series, the near-surface currents are onshore 46% of the time at both current measuring stations. Our calculations show that, on average, it would take 0.6 and 2.8 days for a larva that hatches into the surface waters to be transported from the shoreline to the centre of the habitat boxes in northern and southern Monterey Bay, respectively. This time frame is similar to the average duration of naupliar stage I in laboratory studies (Brown & Roughgarden 1985), which suggests that it is reasonable for the larvae to begin naupliar stage II at the model boundary. The pre-settlement cyprid larval stage is also found near the surface. The observed near-surface currents are onshore 37% of the time series at the northern station and 51% of the time at the southern station. We estimate transit times of 0.5 and 3 days from the center of the habitat box to the shoreline during periods of onshore surface flow in northern and southern Monterey Bay, respectively. Because these estimated time frames for transit of larvae through the nearshore zone is short relative to the entire larval duration, we make the assumption that the larvae that arrive at a habitat zone are subsequently able to reach the intertidal zone directly adjacent to the modeled habitat zone. In the discussion section, we discuss implications of assuming that larval connectivity patterns do not depend on the nearshore flow patterns that are not included in the circulation model (please see 'Local retention').

Larval transport model simulations

The 'summer base run' of the larval transport model (simulation 1 [L1], Table 2) simulated larval dispersal between 30 January 2007 and 10 July 2007. A cluster of 450 larvae was released every 7 days, at each of the 13 sites at which we have empirical genetic data (Fig. 1). In total 157,950 larvae were released. Larvae were counted as settlers if they matured to the cyprid stage and if they reached a habitat zone around one of the larval release sites within 56 days. The 'annual base run' of the larval transport model (simulation 2 [L2], Table 2) simulated larval dispersal between 15 August 2006 and 10 July 2007. A cluster of 450 larvae was released every 7 days, at each of the 13 sites at which we have empirical genetic data (Fig. 1). In total 274,950 larvae were released.

Table 2 Larval transport model simulations

Description	Simulation	Dates	Larval release locations	Number of larvae
Summer Base Run	L1	30 January 2007–10 July 2007	Genetic sampling sites	157,950
Annual Base Run	L2	15 August 2006–10 July 2007	Genetic sampling sites	274,950
Southern sources	L3	15 August 2006–10 July 2007	Along southern boundary	23,500
Interannual comparison	L4	1 August 2003–31 August 2003	Genetic sampling sites	23,400

Two additional larval transport model simulations are reported in this paper. The 'southern sources run' of the larval transport model (simulation 3 [L3], Table 2) was designed to assess the effect of input of larvae from sources south of the model domain. In this simulation, larvae were distributed in a cross-shore rectangle beginning 11 km inside the southern boundary of the model and extending 11 km north (Fig. 1). Larvae were released weekly and tracked to the 13 habitat boxes, or to within 2 km of the model shoreline boundary. The 'interannual variability run' of the larval transport model (simulation 4 [L4], Table 2), was designed to assess differences in connectivity patterns for the same season in two different years. As in the base runs, but using ROMS-CoSINE output from August 2003, larvae were released weekly from the 13 genetic sampling sites and tracked for 56 days (Table 2). Because only 1 month of output from ROMS-CoSINE was available in 2003, the velocity, temperature and chlorophyll fields were used in a loop (i.e. 31 August was followed by 1 August). Results from the August 2003 simulation are compared to results from 15 August–15 September 2006 from the yearlong base run (simulation 2 [L2], Table 2).

Connectivity matrices

At the end of each larval transport simulation, the number of larvae that settled at each of the 13 genetic sampling sites was summed. The contribution of each of the 13 initial populations to the total settlement at a single population can be determined. Likewise, we can determine the final arrival place of all larvae released from a particular population. The simulation results were placed in a connectivity matrix, a two-dimensional grid of larval settlement separated by release location versus settlement location' (Fig. 2).

Population genetics model

The connectivity matrices based on movement of simulated larvae among the empirical sampling locations were used to estimate genetic structure. To simulate mtDNA, we used the empirical frequencies of the southern haplotype clade C at AM and PP as the input into the population genetic model. To simulate EF1, we used the frequencies of the southern allele C (see Sotka *et al.* 2004) as the input into the population genetic model. Because barnacles are hermaphrodites, all individuals add both mtDNA and nuclear genes to the population. For each generation and each of the 13 sampling locations in Fig. 1, we estimated the gene frequency that results from the combination of local retention, and larval input from elsewhere. The fraction of settling larvae at *j* derived from each location *i* along

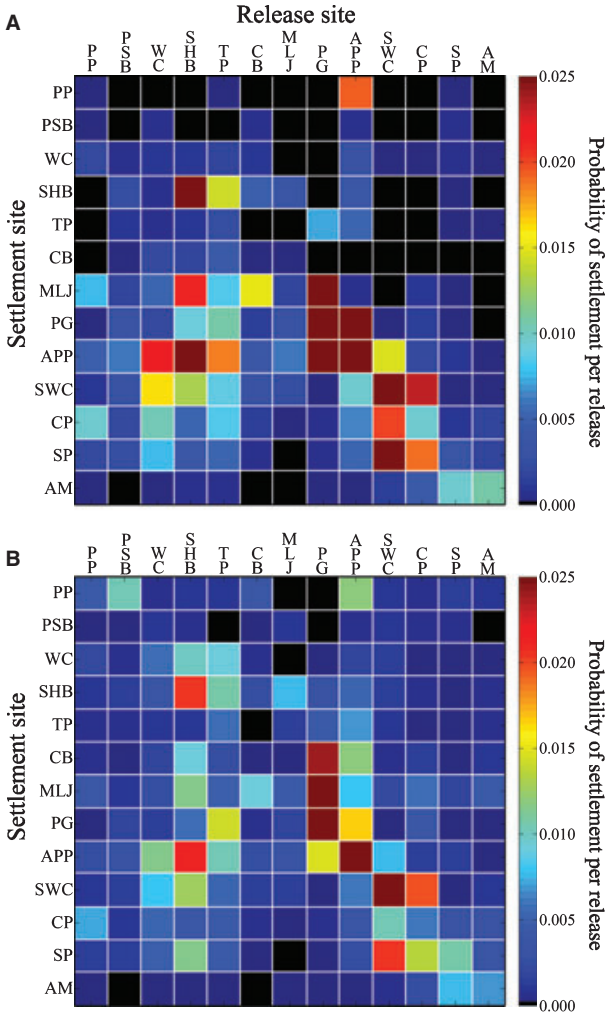

Fig. 2 Population connectivity matrices for (A) 30 January–10 July 2007 and (B) 15 August 2006–10 July 2007. Columns are sites of larval release and rows are settlement sites.

the coast is $p_{ij}/p_{j,\text{total}}$, where p_{ij} is the fraction of particles leaving deme i that land at deme j, and $p_{j,\text{total}}$ is the sum of these fractions for deme j. The gene frequency at location i in generation t ($f_i(t)$) was used to estimate the new gene frequency at each location j ($f_j(t+1)$) as the average of gene frequencies at the different demes weighted by their fractional contribution to settlement at deme j

$$f_j(t+1) = \sum [p_{ij}/p_j, total \times f_i(t)].$$

These deterministic equations do not include changes due to genetic drift as in previous models (Galindo *et al.* 2006) because an initial sensitivity analysis suggested that gene flow in this area was a much more powerful force than drift. As a result, we are using these matrices to model the clinal transitions between northern and southern gene frequencies rather than using them to model equilibrium levels of genetic differentiation among populations.

In order to model clinal structure, we fixed northern and southern gene frequencies at both loci according to the frequencies observed at AM and PP. We added a small amount of dispersal from these fixed gene pools into AM and PP. In general, the geographic patterns of the results we present here had low sensitivity to changes in the amount of this gene flow into the edges of the model. We ran the genetic population model for all sampled populations for 200 generations. Inspection of the data typically showed that the cline reached a stable state by 100–150 generations. Here we report gene frequencies at generation 200, although it should be noted that in no case do our results change if we run the model for more generations.

Simulating spatial genetic patterns

The population genetic model was used to predict spatial genetic patterns under six different scenarios (Table 3). First, we simulated the genetic patterns resulting from the summer (30 January 2007–2010 July 2007, L1, G1) base run of the coupled biological-physical model (Table 2). Second, we simulated patterns based on the annual (15 August 2006–10 July 2007, L2, G2) run. Next, we added three features to the summer base run connectivity matrix (L1) to test whether the augmented matrices provided a better match between the predicted and observed genetic patterns.

The first added feature was additional input of larvae from the southern boundary of the model based on the relative amounts of settlement of these larvae predicted by the annual connectivity matrix generated from the southern sources simulation (L3, G3) of the biological-physical oceanographic model. The distribution of settlement of these southern larvae, $S(i)$, was estimated as the number larvae settled at location i divided by the total southern larvae settled. For each location j in the genetic model, we increased settlement into j from the south by $S(j) \times SI \times p_{j,\text{total}}$, where SI is the southern input as a fraction of the total larval settlement that occurs during the summer months. When SI = 0.1, for example, the settlement into the model from southern sources is equal to 0.10 times the total settlement that occurs during summer. During our genetic model simulations, we altered southern input (SI) from 0.0 to 1.0 to investigate the sensitivity of our model to input of larvae from the southern boundary.

The second feature added to the summer connectivity matrix (L1), which we term a retention multiplier (RM), was used to represent varying rates of retention of larvae in their parent population (G4). For all entries in the diagonal of our connectivity matrix (e.g. the p_{jj} terms), we multiplied these values by a constant, *RM*. We used *RM* to increase the amount of recruitment from deme j to j from 1 to 200. Then, we used combinations of both the southern larval input and larval retention features to augment the summer connectivity matrix (L1, G5).

The third added feature was selection in favor of individuals carrying alleles characteristic of the southern populations in the summer connectivity matrix (L1,

Table 3 Population genetic model simulations

Description	Genetic simulation	Dates	Larval transport simulation
Spring/summer connectivity	G1	30 January 2007–10 July 2007	L1
All year connectivity	G2	15 August 2006–10 July 2007	L2
Spring/summer connectivity + 15% Southern input	G3	30 January 2007–10 July 2007	L1 + L3
Spring/summer connectivity + 150x local retention	G4	30 January 2007–10 July 2007	L1
Spring/summer connectivity + 15% Southern input + 150x local retention	G5	30 January 2007–10 July 2007	L1 + L3
Spring/summer connectivity + selection	G6	30 January 2007–10 July 2007	L1

G6). This was done by allowing selection to act on allele frequencies. We changed the relative frequency of southern alleles in each population each generation by $p(1 - p)s/(1 + ps)$, where p is the frequency of the southern allele in the model, and s is the selection coefficient on the southern allele (i.e. relative fitness (W) of the southern allele is $1 + s$ (see Hartl and Clark 1997, eq. 6.6)). We make the simple assumption that s is constant over space within the model, except for the northern boundary.

Methods for comparing genetic model predictions and empirical data

We graphed the gene frequencies predicted by the genetic model for each deme for each generation and compared them to the observed gene frequencies. To compare the fit of different hypothetical models to the data, we used a simple sum of squares, $\Sigma(f_i(t) - \text{obs}_i)^2$, where obs_i is the observed gene frequency at deme i.

Results

Empirical genetic data

Haplotype clade frequencies of COI (NCBI accession numbers HM028681–HM029141) and EF1a (see Table S1, Supporting information) shift dramatically from PP to AM on the Big Sur coast (AM). The sharpest breaks occur from PSB to WC and Natural Bridges/TP in Santa Cruz. Across this distance of approximately 60 km, the C haplotype of COI increases from 71% to 87%, and the C haplotype of elongation factor 1 alpha increases from 40% to 72% (Table 1). At the southern end of the sampling range, COI-C is nearly fixed, and EF1-C occurs at 78% frequency.

Larval transport model

Summer connectivity patterns (simulation L1; Fig. 2A): In 30 January–10 July, the major spawning season for *B. glandula* in this region (Gaines *et al.* 1985; Broitman *et al.* 2008), population connectivity was dominated by southward movement of larvae. In this season, a majority of the larvae from all but two of the populations settled to the south of their release location. Only PG and APP had more than 50% of their larvae settle to the north.

Winter connectivity patterns (simulation L2; Fig. S1, Supporting information): Simulated larval settlement rates were lower overall in the winter. In contrast to the summer period, connectivity patterns were more strongly influenced by northward movement of larvae. Larvae released from nine of the populations (all but WC, SHB, CP, and SP) were more likely to settle to the north of their release location.

Transport from southern sources (simulation L3; Fig. S2, Supporting information): Most of the larvae released near the southern boundary were rapidly lost from the model domain. Few larvae successfully moved as far north as AM (the most southerly population modelled in simulations L1 and L2). Larvae from the southern boundary did settle near the southern populations in the early fall and late winter. Movement of southern larvae to populations in the northern half of the domain occurred only in August and January.

Interannual variability (simulation L4): A comparison between larval transport simulations of the early fall of 2003 and 2006 revealed that, aside from a difference in the mean settlement rate, connectivity patterns were quite similar. Seven of the populations had more than 50% of their larvae settle to the south in both years. More than 50% of the larvae from three other populations settled to the north in both years. The remaining three populations switched between predominantly southward or northward movement of larvae.

Comparing population genetic model simulations and observed genetic spatial patterns

Summer connectivity base simulation (G1): During simulations for the summer season, when observed larval settlement is generally highest (Gaines *et al.* 1985; Broitman *et al.* 2008), gene flow is predominately north to south. There is so little gene flow from the south in this simulation that gene frequencies quickly shift to be similar to the frequency of the northern gene pool (Fig. 3A). Only the population at the southernmost site (AM) retains a higher frequency of southern alleles, in this case entirely due to the input imposed from the southern gene pool.

Annual connectivity base simulation (G2): Using the cumulative connectivity patterns across an entire year (L2 matrix) causes only minor changes to the predicted genetic spatial patterns compared to *G1*. Gene frequencies remain nearly identical along the study area (Fig. 3B). However, the south to north flowing currents in the winter months increase the frequency of the southern alleles in the analysis. This is an unrealistic model in that it assumes that settlement is equivalent in summer and winter, but even with this assumption, the pattern of southern alleles does not match the empirical data. The main result of this and other simple models is that the expected gene frequencies are more biased towards the northern alleles than seen in the genetic data. These results suggest that there is an additional factor increasing the relative amount of southern alleles in these populations.

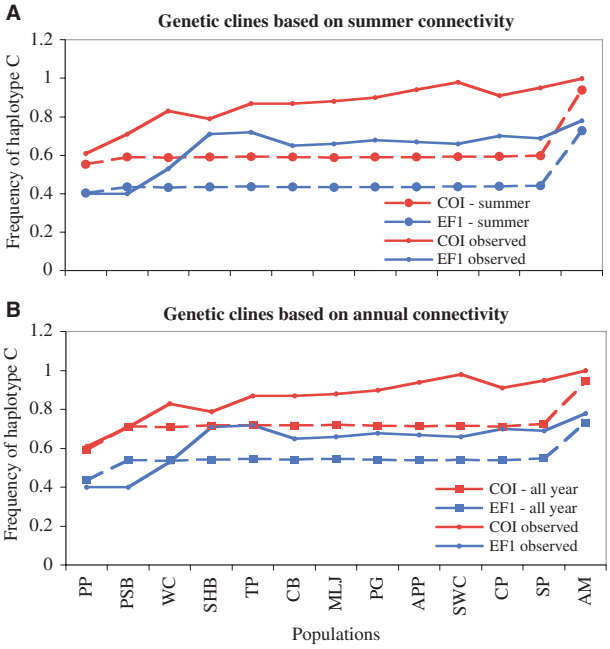

Fig. 3 Comparison of the seascape genetic model (dashed lines) to observed data (solid lines) for mtDNA and nuclear genes (red and blue, respectively). Results based on the connectivity matrix (A) from the main barnacle settlement season 30 January–10 July, and (B) averaged over the whole year.

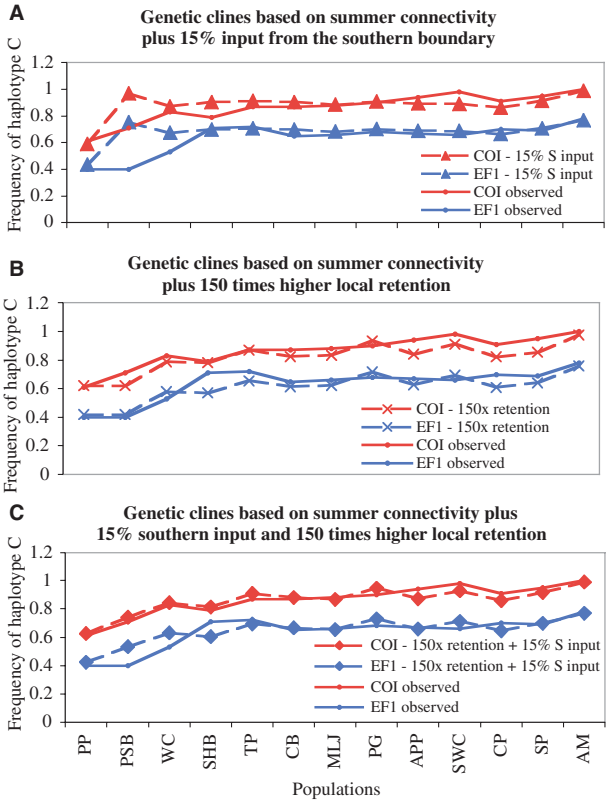

Fig. 4 (A) A comparison of observed and predicted genetic patterns if 15% of the settlers come from beyond the southern boundary of the ocean model, (B) a comparison of observed and predicted genetic data based on 150-fold higher retention than allowed in the simple ocean model, and (C) a combination of higher retention and southern input.

Summer connectivity plus southern larval input (G3): Input of larvae from beyond the southern boundary of the model can substantially improve the fit between the genetic model's predictions and empirical observations. We conducted simulations with southern input equal to 5–50% of the total settlement recorded in the annual settlement model. If southern input equals 10–20% of the input seen during summer, then the resulting genetic clines are similar to the observed clines except at PSB where the model predicts gene frequencies that are much more southern than those observed (Fig. 4A shows southern input at 15%). Artificially increasing the direct gene flow from PP to PSB has little effect. However, adding input into PSB directly from beyond the northern edge of the model can help resolve this difference (not shown).

Summer connectivity plus larval retention (G4): Because the ROMS model only considers ocean movement beyond approximately 2 km from shore, it cannot be used to simulate larvae that stay closer than this to the shoreline. If such larvae also tend to disperse low distances, our model may underestimate the proportion of larvae that settle close to home. To address this possibility, we added a multiplier (*RM*) to all p_{jj} values in order to increase the proportion of larvae that settle at the same locality at which they are released. We also adjusted $p_{j,\text{total}}$ values to take the higher p_{jj} values into account. Increasing local retention by an order of magnitude (e.g. *RM* = 10) had little effect on the results. Increasing it by up to 150-fold brought the observed and expected genetic data into close agreement (Fig. 4B).

Summer connectivity plus southern input and larval retention (G5): The above analyses suggest that additional input of larvae from the south and higher rates of local retention can substantially improve the fit of our current model (Fig. 4A,B). To explore what combinations of these factors may best improve fit, we ran our genetics model with various combinations of southern input from 0% to 20% of the total, and various levels of our RM from 1 to 300 (e.g. Fig. 4C). We evaluated each model by comparing the predicted genetic cline with the observed cline using least-squares criteria. We summed the squared difference between observed and predicted gene frequencies at each location, and compared the sums across different models (Fig. 5). Our original model, with no extra retention or southern input, produced the largest difference between mod-

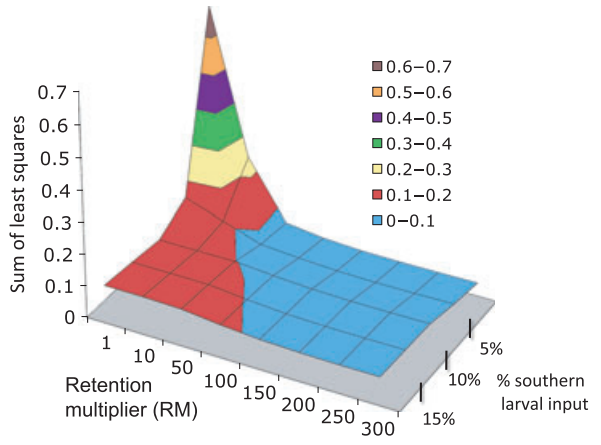

Fig. 5 The fit of the observed and predicted genetic data (estimated as the sum of least squares) for various combinations of the RM and the fraction of southern input. Higher values represent a poorer fit of predicted and observed genetic patterns. Even moderate amounts of southern input and extra retention dramatically improve the data fit.

elled and observed data. Models with 0 < RM < 50 and southern input over 10% have a much better fit (red panels, Fig. 5). Models with RM > 50 and southern input between 0% and 20% have a slightly better fit to the data (light blue panels in Fig. 5), though the improvement drops the sum of squares only marginally from < 0.2 to < 0.1 (see legend in Fig. 5).

Summer connectivity plus selection in favour of southern alleles (G6): An additional factor that could improve the fit between the observed and modelled data is natural selection in favour of southern alleles ($s > 0$). When selection at each locus is increased from $s = 0$ to 0.02 (for EF1) or $s = 0.05$ (for COI) the fit between the observed and modelled data gradually improves over the results from simulations G1 and G2 (Fig. 6). The best selection regime explains the data more poorly

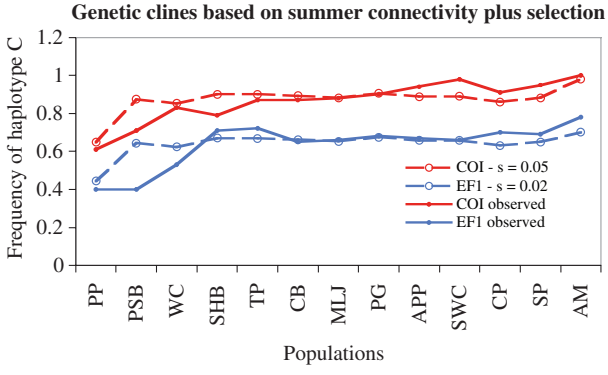

Fig. 6 A comparison of observed and predicted genetic patterns if southern C alleles have a relative fitness advantage in all populations for both COI ($s = 5\%$) and EF1a ($s = 2\%$).

than the best hypothetical retention regime (Mean squared difference = 0.08 rather than 0.04), mainly because the selection regime is assumed to be identical across Monterey Bay. A more complex selection regime may perform better. However, it is clear than hypothetical selection and hypothetical retention are both powerful solutions to the difference between the basic model and the observed data.

Discussion

There is an increasing need to evaluate population genetic data in the light of realistic, local meta-population models (Manel *et al.* 2003). By integrating biological-physical oceanographic and genetic models, we are able to gain new insights into larval connectivity patterns relevant to real world scenarios. In this study on acorn barnacle dispersal in central California, the larval transport model shows dispersal from north to south across Monterey Bay, generating enough gene flow to completely prevent the persistence of the observed genetic cline. Adding parameters to the population genetic model dramatically improves the fit to observed data. If a small fraction of the settling larvae come from far to the south, or if local retention of larvae is higher than suggested by the model, then the models predict the observed cline with greater accuracy (Figs 4 and 5). Natural selection in favour of southern genotypes also improves the accuracy of the model predictions (Fig. 6).

Insights gained from simulated and empirical data mismatches

Using coupled biological–physical oceanographic models to advect larvae as particles with growth and behaviour in three dimensional circulation fields is one method of predicting the dispersal pathways of larvae. Through this type of modelling, environmental parameters can be prescribed, the location of each simulated larva is known over time, and predictions can be made through repeated model simulations. However, this modelling has inherent uncertainties: first the biological models may imperfectly represent important biological attributes of the system, and second, the physical models may under-resolve features of water flow that are important to population dynamics. The ability to use demographic predictions of ocean models to simulate gene flow allows us to test the validity of coupled biological-physical models using genetic data and to use the resulting comparisons to refine model predictions (Galindo *et al.* 2006). Our results along the central coast of California suggest that the available connectivity model is incomplete, and might be improved by further consideration of at least two factors.

Local retention

The model simulations presented in this paper include biological assumptions and geographic limitations that may underestimate the tendency of larvae to remain close to their parents. Fine scale foraging and aggregation behaviours (as reviewed by Woodson & McManus 2007) have not yet been parameterized in this, or other oceanographic models. These fine scale behaviours would help to maintain the larvae in the vicinity of persistent fronts where food concentrations are often elevated, thus increasing fitness of the larvae and reducing dispersal distances by hundreds of kilometres. Woodson *et al.* (unpublished data, C.B. Woodson, Stanford University, Palo Alto, CA, USA) used 7 years of high-resolution satellite sea surface temperature to identify regions of high coastal front activity and compared these to corresponding time series of marine invertebrate recruitment across the California Current Large Marine Ecosystem (CCLME). These investigators found that front probability explained a high percentage of the variance in recruitment for each species. In northern Monterey Bay, a persistent front was identified from satellite data where the coastal upwelling jet separates from the coastline. An inclusion of fine scale foraging and aggregation behaviours resulting in increased numbers of larvae at persistent fronts would most likely increase retention of larvae within the Monterey Bay region.

In addition, our ability to assess the effects of nearshore circulation on local retention of larvae is limited by the ocean model configuration. The physical circulation fields obtained from ROMS do not include any current patterns inshore of the 50 m isobath ($\sim$ 2 km from shore). This means that any larvae that might stay nearshore never enter the model, and are unaccounted for. Shanks & Shearman (2009) recently concluded from larval sampling that about half the *B. glandula* larvae along the Oregon coast were within 2 km of shore. Such larvae are more likely to disperse only short distances (Largier 2003), and so ocean models might be biased towards higher dispersal than actually exists. Nearshore circulation patterns will need to be understood and modelled in greater detail before we can accurately predict the delivery of larvae directly into the intertidal zone.

Long-distance dispersal

Our simulations may miss input of southern alleles that travel from beyond the southern boundary of the model. This input could be due to long distance gene flow from the south that is beyond the model grid. In a simulation in which larvae were released along the southern boundary of the model, few larvae successfully moved north, and these only occurred in the fall-winter. No larvae moved into the model from the southern boundary during the spring-summer period. Of the larvae that entered the model, most arrived near the southern demes, with a dramatic drop off of arrivals in the middle and northern reaches (Fig. S2, Supporting information). One exception was an increase in movement of larvae from the southern boundary to the vicinity of PSB, which occurred in one month of the simulation. There is also a slight increase in southern delivery to the Monterey Peninsula.

A targeted field study, combining observations of physical oceanographic structure and processes, as well as the plankton and barnacle recruitment in north Monterey Bay, indicates that during upwelling relaxation events persistent fronts can move in a poleward direction resulting in transport of larvae from south to north (Woodson *et al.* 2009). Other researchers have suggested that coastally trapped waves, interacting with topographic features can produce strong secondary flows in a poleward direction (Auad & Hendershott 1997; Beckenbach & Washburn 2004). It has been suggested that these coastally trapped waves, which travel at speeds of 75 km day^{-1} (86 cm s^{-1}) have the potential to transport barnacle larvae alongshore and poleward (Pineda & Lopez 2002).

Two other predictions can be made about the genetics of barnacles in this system if there is long distance gene flow from both the north and south. First, it should be possible to detect the inflow of larvae with multi-locus genotypes characteristic of distant adult populations (e.g. Brazeau *et al.* 2005). Second, mixing of larvae from different source should generate a signal of multi-locus linkage disequilibrium. The second prediction is met in these barnacle populations, especially in Monterey Bay (Sotka & Palumbi 2006).

Identifying realistic combinations of retention and southern input

By itself, additional retention must be very strong (e.g. a multiplier > 50) to bring observed and predicted frequencies substantially into alignment (Fig. 5). Likewise, by itself a large amount of southern input is required (Fig. 4A). However, when both features are included together, we see that modest additional retention and low levels of southern input greatly improve the model (Fig. 5). In this case, southern input delivers southern alleles and extra retention reduces the rate at which they are advected away. For example, a RM of 50 with a southern input of 5% provides good model fit. However, at this multiplier, settlement into localities is up to 99% from local sources (range 50–99%, mean 75%), a value higher than recently measured for short-dispersal

coral reef fish (Almany *et al.* 2007). At a RM of 10, and southern input of 10–15% also provides a good fit to the data (Fig. 5). When the RM 10, average retention is about 57% of larvae, which is similar to the data in Almany *et al.* (2007). Further field observations are needed to understand whether such rates of retention are seen in central California marine systems. In particular, there is a need for biological observations, both in the water column and on the shoreline, combined with observations of nearshore currents.

Natural selection

The distinction between our modelling results and genetic data suggests that southern larvae might make up 10–20% of settlers to balance the steady gene flow from the north. If this gene flow persisted for long periods of time, then the genetic cline would gradually decay. We do not see this in our genetic model simulations because we artificially hold northern and southern gene frequencies constant. Sotka & Palumbi (2006) suggested that natural selection against northern alleles in the south and southern alleles in the north must be acting, likely through post-settlement mortality, to prevent the decay of this genetic cline. Such selection, acting generation after generation might be low enough to be ecologically difficult to detect (e.g. $s < 0.05$), yet still be powerful enough to generate a cline with the width seen in our study (Sotka & Palumbi 2006). Hare *et al.* (2005) point out that unidirectional gene flow can alter the equilibrium equations often used to predict the balance between gene flow and selection.

The results of our simple selection model indicate that selection in favor of southern genotypes along the length of the cline would in fact help bring the model and observations into alignment. Such selection is in accord with predictions of cline theory that the broad differences in gene frequency we see along the west coast for *Balanus glandula* are probably due to a balance of selection and gene flow (Sotka & Palumbi 2006).

The selection levels required are not huge: selection coefficients favouring southern alleles ranging from 0.02 to 0.05 across loci are sufficient to substantially improve the model. However, current data show little hint of this level of selection. Among 869 barnacles sequenced at Hopkins Marine Station for COI, there was no increase in southern haplotype frequency with age. Cyprids had a COI-C frequency of 91.4% ($n = 347$) whereas juveniles had a lower frequency of 88.8% ($n = 563$) and adults showed a frequency of 89% ($n = 58$). The same is true for EF1a data; there is a drop of southern allele frequencies from 70.7% to 69.5% to 68.1% ($n = 694, 1126, 118$ respectively, unpublished data). These data do not preclude the action of mild selection because our models suggest that an increase in gene frequencies by only 0.5% per generation can rectify our model and data, and such tiny differences are difficult to refute. In addition, selection may occur unevenly at different times and places. Another possibility not visible before the present analysis is that a combination of local retention, southern input and selection may be enough to explain the persistence of the cline and the patterns of gene frequency changes in Monterey Bay. A model that includes all three parameters suggests that selection as low as $s = 0.01$ for COI and a RF factor of 10 can be a good fit to the data. Further work separating and independently testing these factors is needed.

Other potential influential factors

Although our analyses show a good fit between the augmented models and gene frequency patterns, this fit does not pinpoint the cause of the mismatch between modelled and observational data. Other possible causes of the mismatch include possible inaccuracies in the temperature and chlorophyll levels that control the planktonic larval duration, the fact that the numerical model was only run for 1 year, and assumptions in the population genetics models including the relative unimportance of genetic drift in populations of this size. Though we cannot rule out these other factors, our current analysis suggests that they play a smaller role than selection, retention or southern input.

Interannual variability

Interannual differences in current flow are not likely to erase the basic oceanographic patterns shown in our connectivity matrices. A paired *t*-test showed no significant difference ($p > 0.05$) between the connectivity matrix calculated for August 2003 and for August 2006. There is a trend, however, that localities in southern Monterey Bay tend to send larvae a little further north in 2003 than in 2006. This may slow the rate of north-to-south gene flow, but by itself will not tend to deliver more southern alleles into the model. Other years, especially those during El Niño events, which results in more frequent relaxation events in central California (Chavez *et al.* 2002) may provide more south-to-north gene flow than seen in 2006, but such rare years may not play a major role in genetic patterns. Our general conclusion is that the delivery of more southern alleles into central California is likely to be a regular feature of dispersal, either through a phenomenon like poleward moving coastally trapped waves, or frequent years in which northern dispersal dominates. Nevertheless, ocean dispersal patterns are expected to be somewhat stochastic from year to year (Siegel *et al.* 2008) and

future work on the role of annual variation in accurate depictions of ocean gene flow is greatly needed.

Oceanography as a predictor of dispersal

Using oceanographic dispersal models to functionally represent scales relevant to marine conservation efforts is a challenging problem; with each new study, our scientific community gets a step closer to the answer. It is important to remember that this is an iterative process, which requires ongoing discussions between geneticists, modellers and observationalists. Inadequate data sets are a major limitation for these important studies. There is a clear need for interdisciplinary programs, with multi-scale observations of both physical features of the environment and genetic patterns, to provide the basis for coupled physical-biological modelling studies. Specifically, empirical observations of nearshore physical oceanographic processes and *in situ* measurements of larval behaviour, development, and mortality rates are needed to better parameterize coupled biological–physical models. Increasing the domain of high-resolution oceanographic models and minimizing edge effects of model domains will also be important. Finally, empirical field studies should be designed in concert with the development of biological-physical models to produce comparable observed and predicted datasets. Together, these efforts will provide unprecedented insight into the larval connectivity among marine populations at a scale relevant to policy and management decisions.

Acknowledgements

We thank E. Jacobs-Palmer, E. Sotka, and V. Alla for assistance with sample and data collection. We also thank three anonymous reviewers for their comments, which helped us to significantly improve the manuscript. This is contribution 359 from PISCO, the Partnership for Interdisciplinary Studies of Coastal Oceans, funded primarily by the Gordon and Betty Moore Foundation and the David and Lucille Packard Foundation. Additional funding was provided by the Andrew Mellon Foundation, Stanford University, and a National Science Foundation Graduate Research Fellowship awarded to H.M. Galindo. The research for Y. Chao was carried out, in part, at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA).

References

Almany GR, Berumen ML, Thorrold SR, Planes S, Jones GP (2007) Local replenishment of coral reef fish populations in a marine reserve. *Science*, **316**, 742–744.

Anil AC, Kurian J (1996) Influence of food concentration, temperature and salinity on the larval development of *Balanus Amphitrite*. *Marine Biology*, **127**, 115–124.

Auad G, Hendershott MC (1997) The low-frequency transport in the Santa Barbara Channel: Description and forcing. *Continental Shelf Research*, **17**, 779–802.

Baums IB, Paris CB, Cherubin LM (2006) A bio-oceanographic filter to larval dispersal in a reef-building coral. *Limnology and Oceanography*, **51**, 1969–1981.

Beckenbach E, Washburn L (2004) Low-frequency waves in the Santa Barbara Channel observed by high-frequency radar. *Journal of Geophysical Research-Oceans*, **109**, C02010.

Bidigare RR, Chai F, Landry MR *et al.* (2009) Subtropical ocean ecosystem structure changes forced by North Pacific climate variations. *Journal of Plankton Research*, **31**, 1131–1139.

Bousfield EL (1955) Ecological control of the occurrence of barnacles in the Miramichi estuary. *National Museum of Canada Biological Series Bulletin*, **137**, 1–69.

Brazeau DA, Sammarco PW, Gleason DF (2005) A multi-locus genetic assignment technique to assess sources of *Agaricia agaricites* larvae on coral reefs. *Journal of Marine Biology*, **147**, 1141–1148.

Broitman BR, Blanchette CA, Menge BA *et al.* (2008) Spatial and temporal patterns of invertebrate recruitment along the west coast of the United States. *Ecological Monographs*, **78**, 403–421.

Brown SK, Roughgarden J (1985) Growth, morphology, and laboratory culture of larvae of *Balanus glandula* (Cirripedia: Thoracica). *Journal of Crustacean Biology*, **5**, 574–590.

Chai F, Barber RT, Lindley ST (1996) Origin and maintenance of high nutrient condition in the equatorial Pacific. *Deep-Sea Research II*, **42**, 1031–1064.

Chai F, Dugdale RC, Peng TH, Wilkerson FP, Barber RT (2002) One dimensional ecosystem model of the equatorial Pacific upwelling system. Part I: Model development and silicon and nitrogen cycle. *Deep-Sea Research II*, **49**, 2713–2747.

Chai F, Liu G, Xue H *et al.* (2009) Seasonal and interannual variability of carbon cycle in South China Sea: a three-dimensional physical-biogeochemical modeling study. *Journal of Oceanography*, **65**, 703–720.

Chao Y, Li Z, Farrara JD, Moline MA, Schofield OME, Majumdar SJ (2008) Synergistic applications of autonomous underwater vehicles and the regional ocean modeling system in coastal ocean forecasting. *Limnology and Oceanography*, **53**, 2251–2263.

Chao Y, Li Z, Farrara J *et al.* (2009) Development, implementation and evaluation of a data-assimilative ocean forecasting system off the central California coast. *Deep-Sea Research II*, **56**, 100–126.

Chavez FP, Pennington JT, Castro CG *et al.* (2002) Biological and chemical consequences of the 1997–1998 El Nino in central California waters. *Progress in Oceanography*, **54**, 205–232.

Connolly SR, Menge BA, Roughgarden J (2001) A latitudinal gradient in recruitment of intertidal invertebrates in the northeast Pacific Ocean. *Ecology*, **82**, 1799–1813.

Cowen RK, Paris CB, Srinivasan A (2006) Scaling of connectivity in marine populations. *Science*, **311**, 522–527.

Doyle JD, Jiang Q, Chao Y, Farrara J (2009) High-resolution real-time modeling of the marine atmospheric boundary layer in support of the AOSNII field campaign. *Deep-Sea Research II*, **56**, 87–99.

Epperson BK, McRae BH, Scribner K *et al.* (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology*, **19**, 3549–3564.

Franks PJS (1992) Sink or swim: accumulation of biomass at fronts. *Marine Ecology Progress Series*, **82**, 1–12.

Fujii M., Chai F, Shi L, Inoue H, Ishii M (2009) Seasonal and interannual variability of carbon cycling in western and central tropical–subtropical Pacific: a physical-biogeochemical modeling study. *Journal of Oceanography*, **65**, 689–701.

Gaines S, Brown S, Roughgarden J (1985) Spatial variation in larval concentrations as a cause of spatial variation in settlement for the barnacle, *Balanus glandula*. *Oecologia*, **67**, 267–272.

Gaines S, Gaylord B, Gerber L, Hastings A, Kinlan B (2007) Connecting places: the ecological consequences of dispersal in the sea. *Oceanography*, **20**, 90–99.

Galindo H, Olson D, Palumbi S (2006) Seascape genetics: A coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Current Biology*, **16**, 1622–1626.

Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: Coupling genetics with fine-scale physical oceanography. *Ecology*, **84**, 2989–2998.

Grosberg RK (1982) Intertidal zonation of barnacles: the influence of planktonic zonation of larvae on vertical distribution of adults. *Ecology*, **63**, 894–899.

Hardy AC, Bainbridge R (1954) Experimental observations on the vertical migrations of plankton animals. *Journal of the Marine Biological Association of the United Kingdom*, **33**, 409–448.

Hare MP, Guenther C, Fagan WF (2005) Nonrandom larval dispersal can steepen marine clines. *Evolution*, **59**, 2509–2517.

Hartl DL, Clark AG (1997) *Principles of population genetics*, 3rd ed. Sinauer Associates Inc., Sunderland, Massachusetts.

Hentschel BT, Emlet RB (2000) Metamorphosis of barnacle nauplii: effects of food variability and a comparison with amphibian models. *Ecology*, **81**, 3495–3508.

Kinlan B, Gaines S (2003) Propagule dispersal in marine and terrestrial environments: A community perspective. *Ecology*, **84**, 2007–2020.

Kool JT, Paris-Limouzy CB, Andréfouët S, Cowen RK (2009) Complex migration and the development of genetic structure in subdivided populations: an example from Caribbean coral reef ecosystems. *Ecography*, **32**, 1–10.

Largier JL (2003) Considerations in estimating larval dispersal distances from oceanographic data. *Ecological Applications*, **13**, S71–S89.

Li Z, Chao Y, McWilliams JC, Ide K (2008) A three-dimensional variational data assimilation scheme for the Regional Ocean Modeling System: implementation and basic experiments. *Journal of Geophysical Research*, **113**, C05002.

Li Z, Chao Y, McWilliams JC, Ide K (2009) A three-dimensional variational data assimilation scheme for the regional Ocean modeling system. *Journal of Atmospheric and Oceanic Technology*, **25**, 2074–2090.

Liu G, Chai F (2009) Seasonal and interannual variability of primary and export production in the South China Sea: a three-dimensional physical-biogeochemical model study. *ICES Journal of Marine Science*, **66**, 420–431.

Manel S, Schwartz M, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.

McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, **13** (1 Supplement), S146–S158.

Pfeiffer-Herbert AS, McManus MA, Raimondi PT, Chao Y, Chai F (2007) Dispersal of barnacle larvae along the central California coast: a modeling study. *Limnology and Oceanography*, **52**, 1559–1569.

Pfeiffer-Hoyt AS, McManus MA (2005) Modeling the effects of environmental variability on *Balanus glandula* larval development. *Journal of Plankton Research*, **27**, 1211–1228.

Pineda J, Lopez M (2002) Temperature, stratification and barnacle larval settlement in two Californian sites. *Continental Shelf Research*, **22**, 1183–1198.

Pinsky ML, Montes Jr HR, Palumbi SR (2010) Using isolation by distance and effective density to estimate dispersal scales in anemonefish. *Evolution*, doi:10.1111/j.1558-5646.2010.01003.x.

Rooker JR, Secor DH, DeMetrio G, Kaufman AJ, Rios AB, Ticina V (2008) Evidence of trans-Atlantic movement and natal homing of bluefin tuna from stable isotopes in otoliths. *Marine Ecology Progress Series*, **368**, 231–239.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Schaefer KM, Fuller DW, Block BA (2007) Movements, behavior, and habitat utilization of yellowfin tuna (*Thunnus albacares*) in the northeastern Pacific Ocean, ascertained through archival tag data. *Marine Biology*, **152**, 503–525.

Selkoe KA, Gaines SD, Caselle JE, Warner RR (2006) Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology*, **87**, 3082–3094.

Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, **9**, 363–377.

Selkoe KA, Watson JR, White C *et al.* (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, **19**, 3708–3726.

Shanks AL, Shearman RK (2009) Paradigm lost? Cross-shelf distributions of intertidal invertebrate larvae are unaffected by upwelling or downwelling *Marine Ecology Progress Series*, **385**, 189–204.

Shanks AL, Grantham BA, Carr MH (2003) Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications*, **13**(1 Supplement), S159–S169.

Shchepetkin AF, McWilliams JC (2005) The Regional Oceanic Modeling System: A split-explicit, free-surface, topography-following-coordinate ocean model. *Ocean Modelling*, **9**, 347–404.

Siegel DA, Mitarai S, Costello CJ *et al.* (2008) The stochastic nature of larval connectivity among nearshore marine populations. *Proceedings of the National Academy of Sciences*, **105**, 8974–8979.

Sotka EE, Palumbi SR (2006) The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology*, **87**, 1094–1103.

Sotka EE, Wares JP, Barth JA, Grosberg RK, Palumbi SR (2004) Strong genetic clines and geographical variation in gene flow

in the rocky intertidal barnacle *Balanus glandula*. *Molecular Ecology*, **13**, 2143–2156.

Strathmann RR, Branscomb ES, Vedder K (1981) Fatal errors in set as a cost of dispersal and the influence of intertidal flora on set of barnacles. *Oecologia*, **48**, 13–18.

Thorrold SR, Latkoczy C, Swart PK, Jones CM (2001) Natal homing in a marine fish metapopulation. *Science*, **291**, 297–299.

Thorrold SR, Zacherl DC, Levin LA (2007) Population connectivity and larval dispersal using geochemical signatures in calcified structures. *Oceanography*, **20**, 80–89.

Treml EA, Halpin PN, Urban DL, Pratson LF (2008) Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecology*, **23**, 19–36.

Veliz D, Duchesne P, Bourget E, Bernatchez L (2006) Stable genetic polymorphism in heterogeneous environments: balance between asymmetrical dispersal and selection in the acorn barnacle. *Journal of Evolutionary Biology*, **19**, 589–599.

Wang X, Chao Y, Dong C *et al.* (2009) Modeling tides in Monterey Bay, California. *Deep-Sea Research II*, **56**, 219–231.

Weng KC, O'Sullivan JB, Lowe CG, Winkler CE, Dewar H, Block BA (2007) Movements, behavior and habitat preferences of juvenile white sharks *Carcharodon carcharias* in the eastern Pacific. *Marine Ecology Progress Series*, **338**, 211–224.

Werner FE, Cowen RK, Paris CB (2007) Coupled biological and physical models. *Oceanography*, **20**, 54–69.

Whitlock M, McCauley D (1999) Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). *Heredity*, **82**, 117–125.

Woodson CB, McManus MA (2007) Foraging behavior can influence dispersal of marine organisms. *Limnology and Oceanography*, **52**, 2701–2709.

Woodson CB, Washburn L, Barth JA *et al.* (2009) The northern Monterey Bay upwelling shadow front: observations of a coastally- and surface-trapped buoyant plume. *Journal of Geophysical Research*, doi:10.1029/2009JC005623.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

H.G. is a post-doc at the University of Washington using molecular ecology and coupled oceanographic-genetic models to study spatial connectivity among marine populations in a management context. A.P.H. is a PhD candidate at the University of Rhode Island's Graduate School of Oceanography whose research focuses on biological-physical interactions in coastal marine ecosystems. M.A.M. is an Associate Professor in the Department of Oceanography at the University of Hawaii at Manoa. She is recognized for her work assessing the effects of physical oceanography on the dynamics of marine ecosystems in the coastal environment. Y.C. is a physical oceanographer at the Jet Propulsion Laboratory at California Institute of Technology who uses numerical ocean models to study the 3D ocean circulation and variability including its response to weather and climate and impact on biogeochemistry and marine ecosystem. F.C. is an oceanographer at the University of Maine who develops and uses physical-biological models to investigate carbon cycle and marine ecosystem responses to climate change. S.P. works on the interface between population and evolutionary biology in marine systems using genetic tools.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Population connectivity matrix for 15 August 2006–29 January 2007. Columns are site of larval release and rows are settlement site.

Fig. S2 Distribution of settlers from the southern boundary of the ocean model based on the annual connectivity matrix.

Table S1 Haplotype clade frequences for EF1a determined by DNA sequence at basepairs 105–108 relative to the 156bp fragment in Sotka *et al.* (2004). (Clade A=TCCA, Clade B=TCAT, Clade C=CCCA.)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.