doi: 10.1111/j.1420-9101.2009.01760.x

Demographic history, geographical distribution and reproductive isolation of distinct lineages of blue rockfish (*Sebastes mystinus*), a marine fish with a high dispersal potential

M. O. BURFORD

Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA, USA

Keywords:

admixture;
assignment test;
microsatellite loci;
mismatch distribution;
mtDNA control region;
population expansion;
reproductive isolation;
speciation.

Abstract

Understanding the barriers to genetic exchange in taxonomic groups that have a high dispersal potential will provide critical information on speciation in general. Blue rockfish (*Sebastes mystinus*) are good taxa to examine speciation because they are nonmigratory inhabitants of shallow rocky reef habitats along the eastern North Pacific with a pelagic larval stage lasting 3–5 months. The goal of this study was to analyse the evolutionary history and distribution patterns of different lineages within *S. mystinus* described previously and use this information to understand the speciation process in this group of high dispersal fish. The molecular data derived from specimens sampled over approximately 1650 km of the *S. mystinus* range revealed a northerly and southerly distribution for the two lineages. Almost equal frequencies of both lineages occurred at centrally located sample locations, with evidence of reproductive isolation between the lineages. A demographic analysis showed that the two lineages diverged and experienced sudden expansion prior to the last glacial maximum, which affected the observed pattern of genetic structure. The spatial distribution, demographic history and degree of genetic distinctiveness found from the genetic analysis, despite the high potential for dispersal in *S. mystinus*, suggest both lineages diverged in allopatry.

Introduction

Contemporary and evolutionary time-scale factors such as current or wind patterns, geographical barriers, demographic characteristics, physical changes in the environment, or a combination of these factors can act to reduce gene flow among populations throughout a species' range (Avise, 1994). In marine systems, reduction or elimination in gene flow may lead to genetic differentiation or speciation even within species with long pelagic stages, especially if populations are separated into different ecological regions. Selective forces can also lead to genetic structure or speciation in marine taxa with a long pelagic stage only if it is strong enough to counteract the homogenizing effect of gene flow. The evaluation of the genetic relationship among populations or distinct lineages in areas where divergent lineages overlap and where only one lineage is found can improve our understanding of the evolutionary forces and events causing speciation in the marine environment. For example, high potential dispersal may not translate to high levels of realized gene flow. We may also gain insight into why there is high species diversity in taxonomic groups with high dispersal potential.

Although the rockfishes (Scorpaenidae: *Sebastes*) are considered an ancient species flock (Johns & Avise, 1998) with a high potential for dispersal (through a pelagic stage lasting 2–5 months; Love *et al.*, 2002), there is evidence of recent speciation within the genus. Rocha-Olivares *et al.* (1999a) found recent speciation between sister-taxa in the subgenus *Sebastomus*. Two investigations within rockfish species found cryptic speciation within vermilion rockfish (*Sebastes miniatus*) that was caused by difference in depth distribution (Hyde *et al.*, 2008) and a similar pattern of sympatric distribution of distinct genetic populations of Pacific ocean perch

Correspondence: Martha O. Burford, 208 Fernow Hall, Cornell University, Ithaca, NY 14853, USA.
Tel.: +1 831 419 7586; fax: +1 831 459 3383; e-mail: mob8@cornell.edu

© 2009 THE AUTHOR. *J. EVOL. BIOL.* **22** (2009) 1471–1486
JOURNAL COMPILATION © 2009 EUROPEAN SOCIETY FOR EVOLUTIONARY BIOLOGY

(*Sebastes alutus*) in the Queen Charlotte Islands (Withler *et al.*, 2001). Collectively these findings of recent and cryptic speciation, combined with the suggestion of ecological speciation between congeners (Larson, 1980; Alesandrini & Bernardi, 1999; Narum *et al.*, 2004), provide substantial evidence of contemporary speciation within the genus.

In addition to evidence of recent speciation within the genus, there is evidence of geographical genetic structure and incipient speciation with the blue rockfish, *Sebastes mystinus* (Cope, 2004; Burford & Larson, 2007; Burford & Bernardi, 2008). This marine fish exhibits high potential for gene flow due to a protracted pelagic larval phase (3–5 months) compared to other marine taxa (Love *et al.*, 2002). Using *F*-statistics from mitochondrial DNA, Cope (2004) reported a population genetic break between two sample locations in the north (Washington and Oregon) and four locations in the central/southern range (California), but little evidence of geographical structuring at smaller spatial scales or support for this break in phylogenetic analyses. He proposed a dispersal barrier between these regions at Cape Mendocino in northern California. Burford & Larson (2007), using five microsatellite markers, found genetic differentiation between northern and central California samples of juvenile *S. mystinus* in the 2000 year-class. Because these juvenile sample locations were south of the proposed break at Cape Mendocino, this study provided evidence of smaller-spatial scale structuring during year-class formation that was not evident in Cope's (2004) study. These two studies, using different molecular markers, suggested that both historic and recent events promoted population genetic structure in this species despite the predicted homogenizing effect of a long pelagic phase.

Burford & Bernardi (2008), in a recent study of rockfish in the subgenus *Sebastosomus*, demonstrated the existence of two species (Type 1 and Type 2) within *S. mystinus* using mtDNA sequence and nuclear microsatellite data. In their analysis of three sample locations, they found only Type 1 individuals at Neah Bay in Washington State, all but one Type 2 individuals at Gaviota in southern California, and equal frequencies of the two lineages at Fort Bragg in northern California, near the centre of the *S. mystinus*' range. The admixed lineages at Fort Bragg were south of Cape Mendocino and this pattern was not found in previous adult samples from Fort Bragg (Cope, 2004; Burford & Larson, 2007). The body of work on *S. mystinus* reveals a complex pattern of genetic structure, which provides an opportunity to investigate factors that promote speciation in an organism with a high dispersal potential and to examine historical causes of this genetic structure.

There are two main hypotheses that address the phylogeographical distribution and stability of historic ranges of marine taxa during the Pleistocene with a similar distribution and nearshore habitat requirements to *S. mystinus*: (i) northern persistence during the last glacial maximum (LGM: 19 ka) in nearshore refugia (Marko, 2004; Sotka *et al.*, 2004; Hickerson & Cunningham, 2005) and (ii) southern population contraction and subsequent range expansion to the north after the LGM (Burton, 1998; Edmands, 2001; Hickerson & Cunningham, 2005). Both of these hypotheses assume the effects of the LGM on distribution patterns of nearshore species in the eastern North Pacific. The recent study on *S. mystinus* (Burford & Bernardi, 2008) found the support for the northern persistence hypothesis, but tested a very limited number of locations. If there was population persistence in the northern region via refugia in either one or several nearshore areas, the physical separation of the two populations may have resulted in allopatric speciation. Given previous evidence of northern persistence in *S. mystinus*, the question is whether the LGM was the cause of allopatric speciation in this group. The northern persistence hypothesis predicts a coalescence time beyond the LGM, an effective population size large enough to allow population persistence during the LGM, and subsequent population growth (range expansion) of contracted populations from north to south after the LGM. The southern contraction hypothesis, predicts lower average genetic diversity in northern populations due to a range expansion of individuals from south to north. Evidence of within lineage population structure or genetic isolation with geographical distance could reveal whether there was rapid or recent expansion of either lineage north or south after the LGM. Understanding the demographic history, difference in genetic diversity, and the shape and characteristic of the distribution of the two lineages (e.g. genetic cline vs. patchy distributions) will provide evidence of whether historic events promoted the divergence and speciation of these two groups.

Building on the previous work identifying a genetic break or distinct lineages (Cope, 2004; Burford & Bernardi, 2008 respectively), the first goal of this study was to define the distribution pattern and geographical limit of each lineage of *S. mystinus* throughout its range, both of which were unknown. The second goal was to investigate the following questions that have not been previously answered: (i) Is there patchiness in the distribution or a smooth genetic cline in areas of overlap between the lineages suggestive of ecological segregation or introgressive hybridization respectively? (ii) Is there genetic differentiation within lineages and is there a geographical pattern within lineages suggestive of recent range expansions? (iii) Is there evidence of northern persistence or southern contraction in the demographic and evolutionary history of the lineages using samples throughout the range and can this explain the speciation event? These questions were addressed in an attempt to elucidate both large-scale genetic structure between distinct lineages and fine-scale genetic structure within a lineage using two independent markers, and, importantly, provide information on potential mechanisms that promote speciation in a species with a high potential

for dispersal. Because *S. mystinus* is a member of a genus with over 60 species present in the Eastern Pacific, this study also contributes to our understanding of speciation within the highly diverse *Sebastes*.

Materials and methods

Study species

Sebastes mystinus resides in rocky subtidal habitats from northern Baja California, Mexico to Canada. The centre of distribution (i.e. an area with the largest number of individuals), is along the central and northern California coast from the northern Channel Islands to Fort Bragg (approximately 680 km) (Leet *et al.*, 2001; Love *et al.*, 2002). After the protracted pelagic larval stage, juveniles settle approximately from April through July (Wyllie-Echeverria, 1987) and adults tend to show high site fidelity (Miller & Geibel, 1973; Jorgensen *et al.*, 2006).

Sampling

Sampling of the adults included fine-scale sampling within the centre of distribution and sites extending throughout the range from Santa Barbara to Washington (approximately 1650 km; Fig. 1). The originally reported range of *S. mystinus* (Baja to Aleutian Islands) does not reflect the contemporary range, due to extirpations and a range contraction of the southern population in 1977 (Stephens *et al.*, 1994) and false identifications in the north (Love *et al.*, 2002). Although the current range of *S. mystinus* extends slightly beyond the northern (to Vancouver Island) and southern-most samples (to border of California and Mexico; Love *et al.*, 2002), these extremes did not provide adequate sample sizes, as the populations in these regions were either too small or not observed (Burford, personal observation).

I sampled approximately 50 individuals from most sites for the microsatellite and 20 individuals from all sites for the sequence analyses (Fig. 1). At one location, Brookings, Oregon, I was unable to conduct microsatellite analyses. I used individuals collected in 1999 from San Miguel Island for the sequence analyses to understand whether individuals sampled in the Burford & Larson (2007) study classified as a particular type. I obtained individual specimens from port samplers with the California Department of Fish and Game and Oregon and Washington Department of Fish and Wildlife, from Central California spear fishing meets, and from samplers using both scuba and hook and line. I preserved all finclips and tissue in 95% ethanol for genetic analyses, measured the standard length of individuals sampled at each site to the nearest millimetre, and used these length data to confirm that individuals were reproductive size (over 200 mm standard length) and to indicate age. In addition, I analysed different sampling years within sites or regions for significant genetic differences to evaluate cohort effects on the genetic structure. I only pooled individuals within a location if there was no significant genetic differentiation among the sample dates.

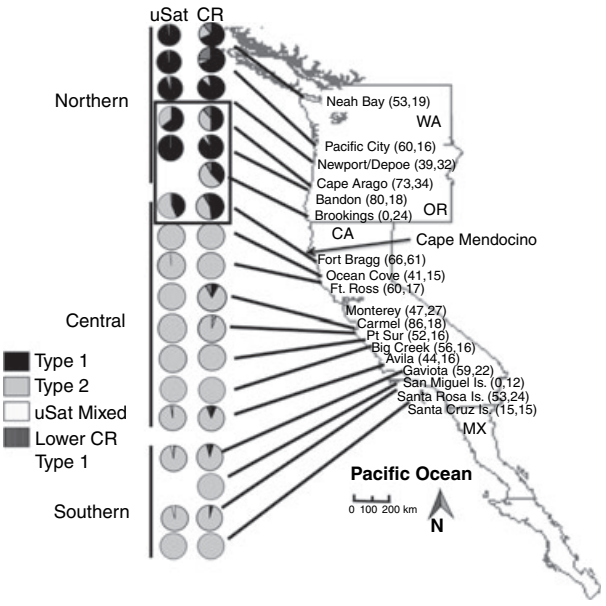

Fig. 1 Distribution of sample locations and geographical pattern of genetic structure of *Sebastes mystinus*. Sampling locations and sample sizes (uSat, CR) depicted in bold. Pie charts are the summary of structure assignment using six microsatellite loci and including 16 sampling locations (uSat), and results based on phylogenetic analysis using mtDNA CR sequence data including 18 sampling locations (CR). Bars on the left-hand side of the map depict regions as described in the text (northern, central and southern). The boxed sample locations depict the region of overlap between the two types. The legend depicts Type 1 (black), Type 2 (grey), mixed (white) that were mixtures of both types in the microsatellite assignment test, and lower CR (barred) individuals that were found in a lower group in the phylogenetic analysis, but were considered part of the Type 1 assemblage.

DNA extraction and amplification

For both sequence and microsatellite analyses, I extracted genomic DNA from the caudal fin of each specimen using a Qiagen DNA Extraction Kit (Valencia, CA, USA) following the manufacturer's protocol. From extracted genomic DNA, I amplified six microsatellite loci designed from *Sebastes rastrelliger* (Buonaccorsi *et al.*, 2004; Westerman *et al.*, 2005) using the polymerase chain reaction (PCR). I followed the protocol outlined in Burford & Larson (2007) for PCR reactions, fragment amplification, and scoring of the microsatellite loci for a total of 869 adult *S. mystinus* across 16 different sample locations (Fig. 1; Appendix S1).

I sequenced 309 bp of the 5′-end of the mitochondrial control region (CR) from the extracted DNA using the L15725rv (Sorenson *et al.*, 1999) and CR-E primers (Lee

et al., 1995). The cycling conditions were identical to those outlined in Burford & Bernardi (2008). I generated CR sequences from 402 individual *S. mystinus* from 18 sampling locations (Fig. 1), two individual *Sebastes melanops*, and two individual *Sebastes serranoides*, which were the outgroup because they belong to the same subgenus, *Sebastosomus*, and were closely related to *S. mystinus* (Rocha-Olivares *et al.*, 1999b; Hyde & Vetter, 2007; Burford & Bernardi, 2008).

Statistical analyses

Microsatellite loci

General microsatellite analysis of *S. mystinus* collections included: (i) an estimate of genetic diversity (Nei, 1987; Eq. 7.39) and allelic richness to measure diversity at equivalent sample sizes using FSTAT version 2.9.3.2 (Goudet, 1995), (ii) an estimate of expected and observed heterozygosity (H_E and H_O) using the software package ARLEQUIN v3.01 (Schneider *et al.*, 2000), and (iii) deviations from Hardy–Weinberg (HW) genotypic expectations and independence of microsatellite loci (linkage equilibrium) using an exact test (Guo & Thompson, 1992) as implemented in GENEPOP version 3.2 (Raymond & Rousset, 1995a). For all exact tests, I generated significance probabilities using the Markov chain method as described in Guo & Thompson (1992) (10 000 iterations) and a sequential Bonferonni correction (Rice, 1989) for multiple comparisons, and used Fisher's method of combining probabilities (Raymond & Rousset, 1995a, b; Sokal & Rohlf, 1995). For comparisons of genetic diversity, I employed a one-tailed paired *t*-test (Sokal & Rohlf, 1995) to check for decreases in genetic diversity in the northern vs. southern regions, a test of the southern expansion hypothesis. To conduct this *t*-test, each individual locus was the sampling unit. To analyse differences in genetic diversity measures between northern vs. southern regions using individual loci, I employed a two-tailed, two-sample *t*-test with unequal variances (Sokal & Rohlf, 1995). The sampling unit was individual sample locations within regions, which tested whether individual loci had lower than average genetic diversity combined with higher than average genetic differentiation indicating local adaptation.

Phylogenetic analyses

To addresses the congruence between two independent molecular markers' grouping of the two lineages, I conducted a phylogenetic analysis using both mtDNA and microsatellite nuclear DNA. I aligned sequences of the 5′-end of the CR using the program CLUSTALV (Applied Biosystems) in Sequence Navigator. I generated measures of the number of haplotypes (H#) and diversity (Hd) using ARLEQUIN (Appendix S2) and found the appropriate substitution model with MODELTEST v3.7 (Posada & Crandall, 1998; hLRT best-fit model =TVM+I+G; gamma shape = 0.4994 and pinvar = 0.3661). I assessed phylogeographical relationships among samples using Neighbour-joining (NJ) method implemented in PAUP* v.4.0b10 (Swofford, 1998), used 1000 bootstrap replicates to test the topology of the tree, and used the uncorrected p-distance model following Nei & Kumar (2000) because this model is more appropriate for closely related species with small genetic distances and did not vary significantly from the model provided by MODELTEST. I also used Bayesian inference methods generated in MrBayes v3.1 (Huelsenbeck & Ronquist, 2001). Using GDA v1.1 (Lewis & Zaykin, 2001) and Nei's (1972) distance measure derived from allele frequencies for the six microsatellite loci, I used the NJ method to generate a phylogenetic tree for sample locations and types.

Individual assignment and cluster analysis

To test the clustering of genotypes among individuals or samples and to test for individuals of mixed ancestry with the goal of analysing the degree of hybridization or introgression between the two lineages, I conducted a Bayesian assignment test in STRUCTURE v2.1 (Pritchard *et al.*, 2000). In this test, I used an admixture model and grouping priors ($K = 1$ to $K = 5$), and allowed for correlation among allele frequencies (Falush *et al.*, 2003). I generated posterior probabilities for each K using 100 000 iterations of the MCMC method after an initial burn-in period of 40 000. I verified the grouping priors and confirmed stability of the other model parameters by running the model with five replicates for each value of K, analysed the data estimate (Ln[Pr(X/K)]; Pritchard *et al.*, 2000), and measured ΔK (Evanno *et al.*, 2005). Evanno *et al.* (2005) found that the *ad hoc* statistic, ΔK, provided a superior prediction of the highest grouping level. For subsequent comparisons using individuals divided among or between types, I assigned individuals with a posterior probability of 70% or greater to that type and considered all others as a group of mixed ancestry individuals.

Genetic structure

To address whether there was population genetic structure between or within lineages using both markers, I analysed the overall genetic structure for *S. mystinus*. I used *F*-statistics, including unbiased estimates of F_{ST} and F_{IS} (Weir & Cockerham, 1984), and analysed pairwise differences of F_{ST}-values in FSTAT using 10 000 permutations to generate *P*-values for both markers. I used an *a posteriori* analysis of molecular variance (AMOVA) as implemented in ARLEQUIN (Excoffier *et al.*, 1992) to test the significance of the grouping of sample locations or *S. mystinus* lineages generated by either the sequence or microsatellite data and based on the phylogenetic and assignment test analyses. I generated significance values by analysing the expected Φ_{ST} and Va (among group variance). To test genetic structure between the two lineages using CR data, I grouped individuals by either

grouping sample locations based on the AMOVA, or by the combined results of microsatellite and mtDNA analysis. To test for recent expansion or dispersal limitation by an increase in genetic divergence with increasing geographical distance among locations within a region or within a lineage, I conducted a regression analysis of isolation-by-distance (IBD) using linear pairwise F_{ST} [$F_{ST}/(1 - F_{ST})$] values and pairwise distance (km) and conducted a Mantel test (10 000 permutations; GENEPOP) using microsatellite data.

Historical demography

To test whether the two lineages experienced a sudden population expansion after the LGM, I analysed sequence variation by a mismatch distribution (Rogers & Harpending, 1992) and a maximum likelihood coalescent approach (Kuhner *et al.*, 1998). Theoretical studies provide evidence that stable populations in demographic equilibrium have chaotic mismatch distributions whereas populations experiencing rapid expansions or bottlenecks have unimodal (Poisson) distributions (Rogers & Harpending, 1992). First I tested for departures from mutation-drift equilibrium with Fu's (1997) *F*s and Romos-Onsins & Rozas (2002) R_2 using ARLEQUIN and DnaSP v.4.10.7 (Rozas *et al.*, 2003) and using 10 000 iterations or coalescent simulations for each assemblage respectively. An analysis of new neutrality tests for detecting population growth showed that both of these tests were powerful at detecting constant size vs. growth for both large and small sample sizes respectively (Romos-Onsins & Rozas, 2002). Second, I analysed the mismatch distribution using the sum-of-squared-difference statistic to determine significant departures from a model of sudden population expansion, as implemented in ARLEQUIN (Schneider & Excoffier, 1999) using 10 000 iterations. Using parameters generated by this analysis, I calculated the time of expansion (t) through the relationship of $\tau = 2ut$ (Rogers & Harpending, 1992). To estimate the time of coalescence for *S. mystinus* lineages, I assumed that coalescence was reached when estimated population size was 1% of the present estimated value (Wares & Cunningham, 2001; Burford & Bernardi, 2008). I based coalescence time on two parameters: mutation rate per generation year (μ) and population growth (g). I estimated the maximum likelihood of parameters θ ($\theta = 2N_f\mu$) and population growth (g) using unconstrained exponential growth (exponential growth parameter in units in μ^{-1}) in FLUCTUATE v1.4 (Kuhner *et al.*, 1998). To check the stability of the parameter estimates, I used 10 short Monte Carlo chains of 200 steps each and 10 long chains of length 20 000, with a sample increment of 20, and ran the estimates 10 times per dataset with randomly generated seeds. I calculated the mean frequency and standard deviation for all analyses based on these 10 runs and provided a range of estimates given both divergence dates and incorporating the respective standard deviations. I calculated an estimate of the population effective size, the idealized genetic population as opposed to the actual, and an estimate of the relative effective population size for five thousand generations ago (kga) (approximately time of LGM) to confirm population persistence during the LGM or colonization and expansion. I compared both distinct lineages and regions (north vs. south) for these analyses, used a previously reported *Sebastes* specific μ per generation year for the CR for two different divergence dates (3.0 and 3.5 Ma; Burford & Bernardi, 2008), and to calculate timing used the generation time of 4 years (Love *et al.*, 2002).

Results

Sampling

With the exception of Cape Arago, there was no significant genetic structure between sample dates at a given sample location (Appendix S1). Therefore I pooled samples from different dates at these locations. Within in the coexistence region, each sample date at Cape Arago comprised two different geographical sites, with a high frequency of Type 1 in the first and Type 2 in the second sample date. The remaining locations in this region were all Type 1 individuals (Bandon) or had no pattern of length or genetic signature associated with sample date or length frequency (Brookings and Fort Bragg).

Phylogeny and phylogeographical pattern

Using sequence data, the NJ and Bayesian analyses revealed two distinct assemblages (types 1 and 2; Fig. 2) within *S. mystinus*, which correspond genetically to the types 1 and 2 reported by Burford & Bernardi (2008). There were two statistically equivalent trees using the uncorrected p-distance model, one with a constrained Type 2 assemblage (Fig. 2) and one with an unconstrained Type 2 assemblage (Tree not shown; Shimodaira–Hasegawa criterion $P = 0.464$). The lower assemblage in the constrained tree was part of the Type 1 lineage in the unconstrained tree and Type 1 microsatellite group (Fig. 2). Statistically equivalent trees and the low-level resolution due to ties encountered during the NJ algorithm were expected given the lack of fixed differences between the two assemblages. Within assemblages, approximately 11 of 251 individuals sampled in the south were found in the Type 1 assemblage (4%) and nine of 151 individuals sampled in the north were found in the Type 2 assemblage (6%). At three sampling locations at the boundary of the northern and southern regions, Cape Arago, Brookings and Fort Bragg, individuals were dispersed throughout the NJ tree and were not found in their own assemblage (Fig. 2).

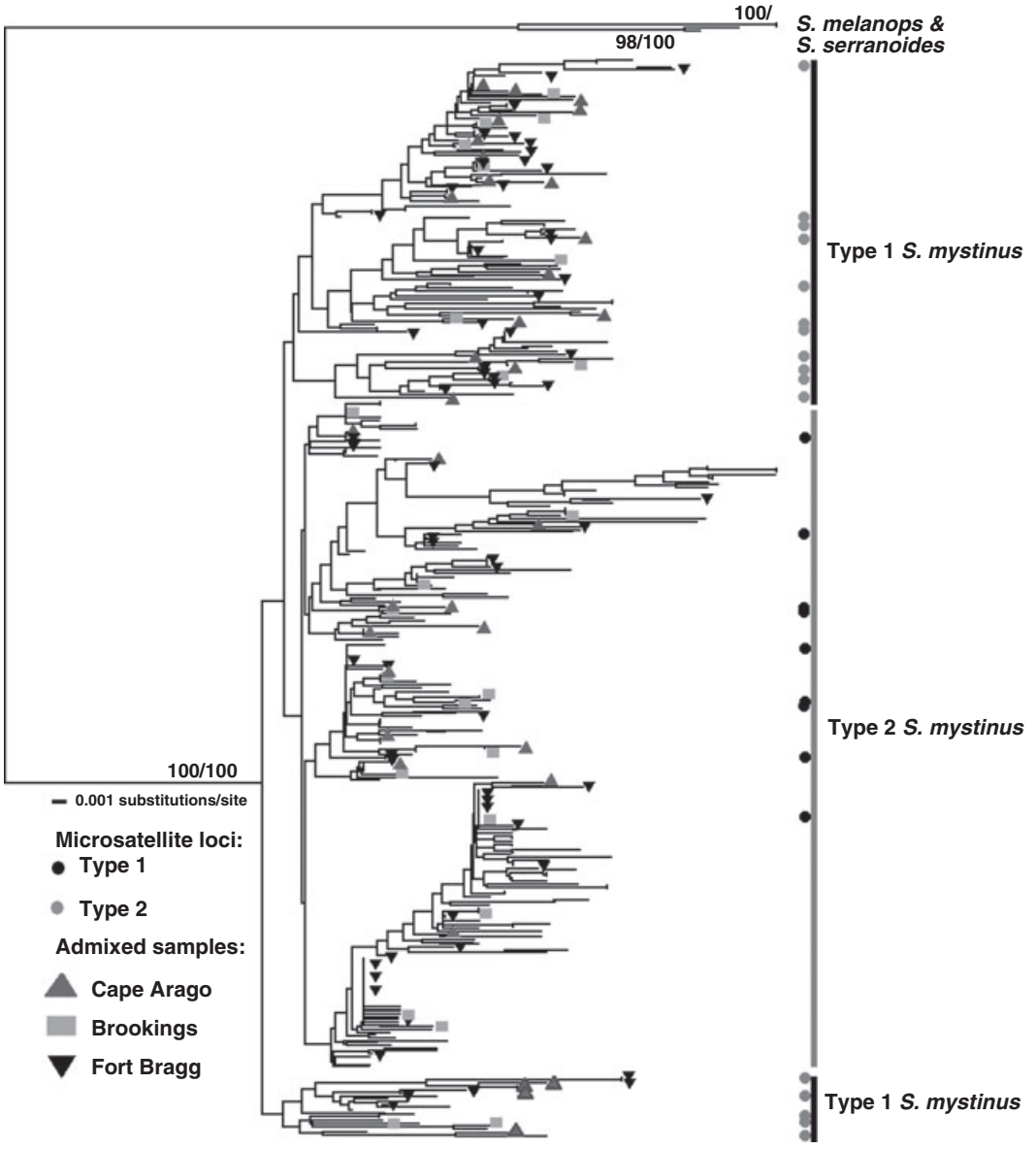

Fig. 2 Phylogenetic relationships within *Sebastes mystinus* using 309 bp of the mtDNA control region and the Neighbour-joining method, the uncorrected (p) substitution model, and *Sebastes melanops* and *Sebastes serranoides* as the outgroup. The left-hand panel is a list of three populations found throughout the tree, Cape Arago, Brookings and Fort Bragg. Colour-coded bars on the right-hand side represent microsatellite assignment to type based on the STRUCTURE analysis (black for Type 1 and grey for Type 2). The colour-coded circles found next to individuals within these assemblages indicate individuals that were exceptions to the assignment bars. Numbers on the left of the major nodes indicated bootstrap support (1000 replicates) using the Neighbour-joining method and numbers on the right of the major nodes are posterior probabilities based on Bayesian inference. Bootstrap values or posterior probabilities of 50% or greater were reported. Major assemblages are indicated on the right-hand side of the tree.

The phylogenetic analysis of the two genetic types using the microsatellite data showed two distinct groups, a Type 1 with northern sample locations and a Type 2 with central/southern sample locations, and two locations, Fort Bragg and Cape Arago, that partitioned into the two types and group accordingly (Fig. 3). The shape of the tree was not altered when I included the six individuals that were from the opposite type to their region or when all individuals were grouped together at Fort Bragg and at Cape Arago, other than these two locations were intermediate to the two major groups (data not shown).

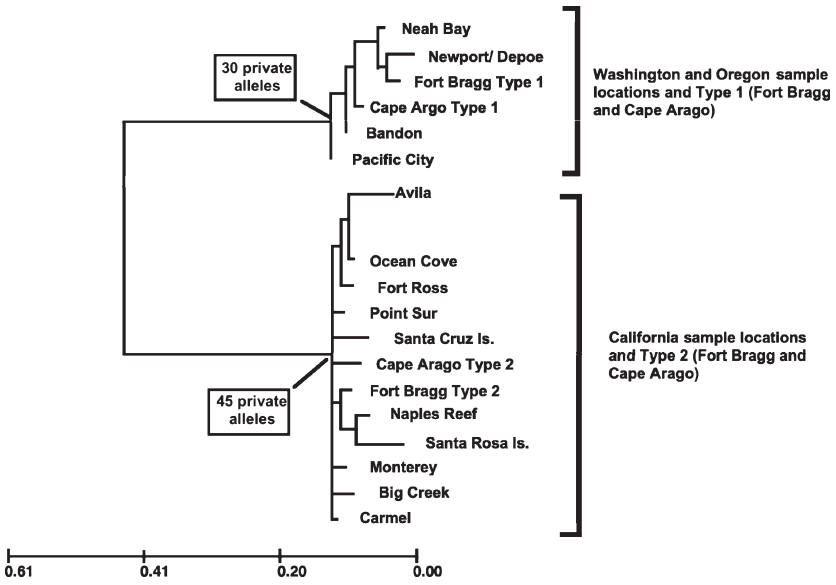

Fig. 3 Phylogenetic relationships of *Sebastes mystinus*. Neighbour-joining dendogram of 16 sampling locations of Type 1 and Type 2 lineages (six and 10 locations respectively) and Cape Arago and Fort Bragg divided into two types grouping based on Nei's (1972) distances derived from the allele frequencies of six microsatellite loci.

Microsatellite analysis

The most common allele differed at five of six microsatellite loci between the Type 1 and Type 2 lineages, including a common allele within Type 2 (46%) that was not present in any individuals within Type 1, and at two of six loci between the central and southern regions. There were 30 vs. 45 alleles unique (private alleles) to Type 1 and Type 2 lineages respectively (Fig. 3). Within types, a percentage of the private alleles were unique to one population within the Type 1 (47%) and Type 2 (38%) regions whereas the remaining private alleles were found at multiple locations. At three individual loci (*Sra.7-2*, *Sra.7-25* and *Sra.6-52*), there were significant differences in both genetic diversity and allelic richness between locations in the north and locations in the combined central and southern regions excluding admixed sample locations (two-sample, unequal-variance *t*-test, d.f. = 11, $P < 0.001$ for both diversity measures). There was also a significant decrease in allelic richness between the north and combined southern locations at the locus *Sra.7-7* (two-sample, unequal-variance *t*-test, d.f. = 11, $P = 0.003$). Genetic diversity for all measures using microsatellite loci was not significantly different between the regions or between types (for regions and types, respectively, one-tailed paired *t*-tests, d.f. = 10, genetic diversity $P = 0.44$ and 0.43; allelic richness $P = 0.38$ and 0.49).

Microsatellite genotype frequencies for all loci, with samples and types pooled, significantly violated HW expectations due to a deficiency of heterozygotes at the following four sample locations: Neah Bay, Pacific City, Cape Arago and Fort Bragg ($P < 0.05$, corrected for multiple comparisons; Table 1). With the two types separated, the samples at Neah Bay, Cape Arago and Fort Bragg were in HW equilibrium, but the sample of Type 1 individuals at Pacific City remained out of HW equilibrium. There was also significant linkage disequilibrium (LDE) at Pacific City, Fort Bragg and Santa Rosa Island (Table 1). As with the evaluation of HW equilibrium, individual types at Fort Bragg were in linkage equilibrium. However, although there was a reduction in the number of linked loci, individual types at both Pacific City and Santa Rosa Island remained in LDE.

Individual assignment and cluster analysis

The results from the STRUCTURE assignment test supported two groups based on higher ΔK (Appendix S3). Overall, there was geographical differentiation between the genotype clusters, with a higher frequency of Type 1 in the northern region and lower in the central and southern regions (97% and 3% respectively; Fig. 1). Analysing the clustering of individuals within a sample location, two locations had approximately equal frequencies of both types (Cape Arago 64% and Fort Bragg 44% Type 1; Figs 1 and 4). At other locations, there were four individuals that assigned at 70% or greater to the opposite type from the regional type (two at Depoe Bay, one at Pacific City and Gaviota; Fig. 4). In total, only seven of 882 of the individuals sampled (0.79%) were mixed ancestry (one at Fort Bragg, Fort Ross, and Avila and two at Gaviota and Santa Rosa Island). Comparing the results of the assignment test to the CR sequence data, there was congruence between the two molecular markers when I overlaid the assignment test results on the NJ tree generated from sequence data (Fig. 2). However, there were a few exceptions, including individuals that assigned to the opposite microsatellite type within a mtDNA assemblage (e.g. individual assigned to the microsatellite Type 1 group and Type 2 CR assemblage; Fig. 2).

Table 1 Summary statistics of (a) 15 populations of *Sebastes mystinus* and (b) two lineages, using six microsatellite loci.

	All loci combined						*Sra.7-2* A# (all samples) 27							*Sra.7-7* 15						
	A#	AR	F_{IS}	H_O	H_E	PA#	*N*	A#	AR	F_{IS}	H_O	H_E	PA#	*N*	A#	AR	F_{IS}	H_O	H_E	PA#
Sample location																				
Neah Bay (NB)	16	13.980	0.020	0.740*	0.789	4	53	8	6.890	0.173	**0.491***	0.592	2	53	7	6.608	0.008	0.736	0.751	0
Pacific City (PC)	14	12.258	0.003	0.757*	0.789	1	60	7	6.033	0.050	0.57	0.597	0	59	7	6.807	−0.052	0.780	0.741	0
Newport/Depoe (ND)	14	13.750	0.006	0.753	0.786	0	38	5	4.947	0.090	0.500	0.563	0	39	6	6.000	0.037	0.692	0.718	0
Cape Arago (CA)	18	15.151	0.087	0.761*	0.846	3	73	15	12.403	0.099	**0.712***	0.791	0	73	9	7.886	0.102	**0.699***	0.778	2
Bandon (BA)	17	13.312	0.059	0.722	0.775	5	80	5	4.479	0.172	0.450	0.543	0	80	7	5.902	0.030	0.688	0.708	0
Ft. Bragg (FB)	18	15.051	0.076	0.764*	0.839	3	66	13	11.427	0.227	**0.636***	0.823	1	64	10	8.864	0.068	0.719	0.771	0
Ocean Cv. (OC)	13	12.893	0.015	0.768	0.790	1	41	12	11.804	0.042	0.829	0.867	0	41	8	7.707	0.198	**0.610***	0.768	1
Ft. Ross (FR)	15	13.109	−0.004	0.811	0.807	3	60	13	12.818	−0.065	0.967	0.908	0	60	8	7.088	−0.025	0.767	0.755	1
Monterey (MB)	14	12.837	0.051	0.752	0.785	3	46	13	12.363	0.046	0.826	0.866	1	47	8	7.735	0.034	0.745	0.779	0
Carmel (CB)	16	12.875	−0.004	0.793	0.801	1	85	17	13.589	0.028	0.871	0.897	0	86	8	7.617	−0.024	0.814	0.795	0
Pt. Sur (PS)	14	12.494	0.022	0.784	0.801	2	52	13	12.046	0.019	0.865	0.883	0	52	7	6.896	−0.085	**0.808***	0.746	0
Big Creek (BC)	14	12.402	0.057	0.761	0.785	2	56	12	11.305	−0.038	0.929	0.895	0	56	8	7.208	0.152	0.643	0.759	0
Avila (AV)	13	12.268	−0.009	0.798	0.815	0	44	12	11.814	0.014	0.886	0.903	0	44	7	6.952	0.001	0.750	0.751	0
Gaviota (GA)	14	12.392	0.007	0.799	0.792	1	59	14	12.587	−0.083	0.966	0.893	1	59	9	8.541	0.047	**0.729***	0.765	0
Santa Rosa Is. (SRI)	13	11.669	−0.023	0.799	0.787	0	53	12	11.281	−0.004	**0.887***	0.883	0	53	9	8.598	0.023	0.755	0.775	0
Average	15	14.436	0.024	3.004	0.799	1.93		23	10.386	0.051	0.759	0.794	0.33		15	8.37	0.034	0.723	0.757	0.27
Lineage																				
Type 1 (T1)	21	20.472	0.050	0.675	0.78	30	302	6	5.930	0.109	0.5	0.56	2	302	9	8.970	0.031	0.71	0.73	2
Type 2 (T2)	23	20.800	0.012	0.720	0.8	45	567	20	18.43	0.007	0.889*	0.9	16	567	13	11.4	0.035	0.739*	0.77	6

	Sra.7-25 58							*Sra.16-5* 41							*Sra.15-8* 18							*Sra.6-52* 15						
	N	A#	AR	F_{IS}	H_O	H_E	PA#	*N*	A#	AR	F_{IS}	H_O	H_E	PA#	*N*	A#	AR	F_{IS}	H_O	H_E	PA#	*N*	A#	AR	F_{IS}	H_O	H_E	PA#
NB	52	36	31.440	0.09	**0.865***	0.955	2	50	28	24.044	0.11	0.840	0.945	0	53	9	7.916	−0.09	0.774	0.709	0	53	7	6.984	0.06	**0.736***	0.779	0
PC	59	35	28.574	0.14	**0.814***	0.94	0	60	21	19.028	0.07	0.867	0.938	0	59	7	6.111	−0.1	0.780	0.710	1	60	7	6.992	0.090	0.733	0.805	0
ND	37	31	31.000	0.13	**0.838***	0.959	0	38	23	22.867	0.010	0.947	0.958	0	39	11	10.690	0.03	0.718	0.742	0	39	7	6.998	−0.06	0.821	0.778	0
CA	71	38	28.992	0.12	**0.831***	0.945	1	72	24	20.918	0.15	0.806	0.946	0	73	12	10.405	−0.11	0.836	0.754	0	73	11	10.3	0.21	**0.685***	0.862	0
BA	79	40	30.408	0.02	0.937	0.953	1	80	29	23.432	0.12	**0.838***	0.947	3	80	11	7.869	0.03	0.700	0.728	1	79	8	7.781	0.06	0.722	0.774	0
FB	64	36	27.198	0.04	**0.860***	0.882	0	65	27	23.101	−0.01	0.954	0.946	2	66	12	10.827	0.09	0.697	0.763	0	64	9	8.886	0.16	**0.719***	0.849	0
OC	41	19	17.909	0.01	0.756	0.761	0	41	21	20.558	0.04	0.902	0.939	0	41	11	10.601	−0.06	0.732	0.705	0	41	9	8.780	−0.12	0.780	0.697	0
FR	60	27	20.536	0.05	0.767	0.809	1	59	23	20.219	−0.01	0.949	0.937	0	60	11	9.631	0.03	0.683	0.706	1	60	9	8.362	−0.010	0.733	0.726	0
MB	47	21	17.952	0.02	0.745	0.758	1	47	23	21.655	−0.04	0.979	0.940	0	46	9	8.600	0.05	0.609	0.641	1	46	9	8.718	0.16	**0.609***	0.724	0
CB	86	23	16.427	0.03	0.733	0.757	0	83	27	22.429	0.06	0.892	0.944	1	84	10	9.225	−0	0.714	0.713	0	86	10	7.964	−0.05	0.733	0.700	0
PS	52	20	17.190	0.01	0.750	0.763	0	51	19	18.147	−0.01	0.941	0.931	0	51	11	10.66	0.11	0.647	0.734	0	52	11	10.03	0.08	0.692	0.749	2
BC	55	23	18.954	−0.02	0.745	0.732	0	56	19	17.725	0.08	0.857	0.928	0	56	11	9.844	0.050	0.625	0.668	1	56	11	9.378	−0.05	0.768	0.731	1
AV	43	18	16.839	0.12	0.721	0.818	0	42	19	18.587	−0.03	0.952	0.930	0	44	9	8.972	−0.08	0.795	0.746	0	44	11	10.45	0.07	0.682	0.742	0

© 2009 THE AUTHOR. J. EVOL. BIOL. 22 (2009) 1471–1486

Table 1 *(Continued).*

	Sra.7-25							*Sra.16-5*							*Sra.15-8*							*Sra.6-52*						
	58							41							18							15						
	N	A#	AR	F_{IS}	H_O	H_E	PA#	*N*	A#	AR	F_{IS}	H_O	H_E	PA#	*N*	A#	AR	F_{IS}	H_O	H_E	PA#	*N*	A#	AR	F_{IS}	H_O	H_E	PA#
GA	59	20	16.360	−0.01	0.763	0.758	0	59	23	20.841	0.05	0.898	0.942	0	59	9	8.422	−0.02	0.763	0.745	0	59	8	7.599	−0.04	0.678	0.651	0
SRI	53	16	13.888	−0.01	0.698	0.693	0	53	21	19.397	−0.03	0.943	0.915	0	53	9	8.394	−0.13	0.811	0.717	0	50	9	8.455	0.05	0.700	0.737	0
Avg		58	26.536	0.049	0.795	0.832	0.40		40	21.636	0.036	0.904	0.939	0.40		18	9.583	−0.014	0.726	0.719	0.33		15	10.101	0.040	0.719	0.754	0.20
T1	297	49	49.00	0.06	**0.892***	0.95	17	298	36	35.980	0.09	**0.862***	0.95	7	302	15	14.950	−0.03	0.75	0.730	2	302	8	8.000	0.04	0.75	0.78	0
T2	565	41	38.11	0.01	0.740	0.75	9	561	33	29.690	0.02	**0.918***	0.94	4	565	16	13.770	−0.01	0.72	0.71	3	562	12	13.360	0	0.71	0.71	7

The table includes sample size (*N*), number of distinct alleles (A#), allelic richness (AR) at $N = 37$, observed and expected heterozygosities (H_O and H_E) and private alleles (PA#).
*Significant departures from Hardy–Weinberg expectation using an exact test after sequential Bonferroni corrections are shown in bold ($P < 0.05$).

Genetic structure

As with the assignment test, geographical partitioning between the two lineages was evident in the frequency-based analyses of the microsatellite and CR data, with a high frequency of the Type 1 lineage in the northern (Oregon and Washington) and Type 2 in the central and southern part of the distribution (Global F_{ST} microsatellite = 0.051 and CR = 0.056; $P < 0.05$, Table 2; Fig. 1). Each of the six independent microsatellite loci (Table 2) and the pairwise F_{ST}-values (Appendix S4) supported this overall divergence. As in the results of the F_{ST}-analysis, the AMOVA for the microsatellite and sequence data supported a break between the northern and the central/southern regions with Fort Bragg, Brookings and Cape Arago locations, which did not differ consistently from either the more northerly or southerly locations supporting an overlap zone (Fort Bragg grouped with south AMOVA microsatellite percentage $Va_1 = 8.33$, $Vb_{13} = 0.87$, $P < 0.001$; CR percentage $Va_1 = 7.68$, $Vb_{16} = 1.60$, $P < 0.001$). Although three locations sampled in this region, Cape Arago, Brookings and Fort Bragg, were mixtures of both lineages, Bandon was exclusively Type 1 in the microsatellite analysis and with the exception of one individual in the Type 1 assemblage in the NJ tree. Global estimate of F_{ST} showed no evidence of structure within the Type 1 lineage locations, however, there was higher global F_{ST} within the Type 2 lineage (Table 2). This smaller but nonetheless significant genetic divergence in the Type 2 lineage was supported by the microsatellite data and occurred between two of the southern locations (Gaviota and Santa Rosa Island), and most other locations for pairwise F_{ST} measures and AMOVA results (Gaviota and Santa Rosa Island vs. central and northern groups AMOVA microsatellite $Va_2 = 6.83$, $Vb_{12} = 0.71$, $P < 0.001$). Using the microsatellite data, there was no significant signature of IBD (Mantel test, $P > 0.05$) within regions (north and central/south, excluding Fort Bragg) or within lineages (Type 1 and Type 2).

Historical demography

The mismatch-distribution analysed using the CR sequence data indicated that both the Type 1 and Type 2 lineages were expanding, as neither lineage was significantly different from that predicted by the sudden expansion model and both lineages violated neutrality assumptions (Fig. 5). Using the τ generated in ARLEQUIN for Type 1 and Type 2 lineage (Table 3a), the date of sudden expansion was 72–84 and 29–34 kga, respectively, which was beyond the estimated timing of the LGM (20 ka or 5 kga). The coalescence time for each lineage was over 100 kga, which was over an order of magnitude greater than the LGM with a slightly larger coalescence time for the Type 1 than for the Type 2 lineage (Table 3b). The population growth based on the

Fig. 4 Results of the assignment model using for the STRUCTURE for *Sebastes mystinus* using microsatellite data and populations are depicted on the *x*-axis and posterior probabilities along the *y*-axis. Adult sample locations include, Neah Bay (NB), Pacific City (PC), Newport/Depoe (NDP), Cape Arago (CA), Bandon (BA), Fort Bragg (FB), Ocean Cove (OC), Fort Ross (FR), Monterey (MB), Carmel (CB), Point Sur (PS), Big Creek (BC), Avila (AV), Gaviota (GA), Santa Rosa Island (SR), Santa Cruz Island (SC).

Table 2 Global F_{ST} of *Sebastes mystinus* with all sample locations, all sample locations divided into two lineages, or all individuals grouped into two lineages for (a) microsatellite analysis using all six loci combined and 309 bp of CR sequence data including number of groups (*N*) for the microsatellite data and CR sequence data, and (b) for individual microsatellite loci

	N		F_{ST}	
	Microsat	CR	Microsat	CR
(a)				
Sample locations	16	18	0.051	0.056
Sample locations and 2 types	18	21	0.064	0.100
Individuals and 2 types	2	2	0.120	0.208
Type 1 and sample locations	6		0.000	
Type 2 and sample locations	10		0.001	

	F_{ST}					
	Sra.7-2	*Sra.7-7*	*Sra.7-25*	*Sra.16-5*	*Sra.15-8*	*Sra.6-52*
(b)						
Sample locations	0.085	0.024	0.061	0.005	0.027	0.104
Sample locations and 2 types	0.104	0.028	0.080	0.006	0.030	0.134
Individuals and 2 types	0.186	0.056	0.147	0.008	0.057	0.239

coalescence model corresponded with the sudden expansion model predicted by the mismatch-distribution analysis (Table 3). Converting the relative effective size from generation time to time in years, the Type 1 lineage at 20 ka was approximately 84–90% of the current estimated effective size and 77–83% for the Type 2 lineage and both lineages were approximately 50% of the current estimated effective size 80 ka.

Discussion

Distribution pattern and demographic history

The results with this expanded data set: (i) showed geographical differences in the distribution of genetic types, (ii) identified the break between two lineages of S. *mystinus*, which was south of Fort Bragg and the proposed break at Cape Mendocino for Type 1, and north of Cape Arago for Type 2, and (iii) revealed an expansive area of co-occurrence (≈430 km) of the two main lineages along the coast of southern Oregon and northern California. This region of overlap was not a gradual cline, as found with the acorn barnacle in the same region (Sotka *et al.*, 2004), but a 'chunky' step-cline characterized by a large region of overlap and precipitous declines in frequencies of both lineages within a short distance of their southern or northern limit. In addition, there was little evidence for introgressive hybridization in this step-cline.

Although the six independent microsatellite loci and sequence data supported differences between the lineages, and these differences were geographically-based, there was little genetic structure within lineages. Finding low genetic structure within lineages was not surprising given the extended pelagic duration of *S. mystinus*. However, there were exceptions at two locations in the southern region, where there was evidence of small-scale genetic structure within the Type 2 lineage. This finding of genetic differentiation with the Type 2 lineage was inconsistent between the two markers and there was no evidence of isolation by distance within lineages. Therefore, the main pattern of genetic structure revealed in this study was between the lineages and not within, which suggests that low gene flow or realized dispersal is not an inherent characteristic of this group nor is there strong evidence of recent extensions of lineage limits.

The demographic analysis of the genetic lineages showed a unimodal mismatch distribution representing a contraction and rapid expansion of both lineages at a time prior to the LGM and close to the time of coalescence. With extensive sampling in the northern and southern regions the coalescence times for both lineages, which were well beyond the LGM, support the previous finding of the northern persistence hypothesis for Pleistocene distributions of marine taxa (Burford & Bernardi, 2008). In contrast to other species showing northern persistence (e.g. acorn barnacle, Sotka *et al.*,

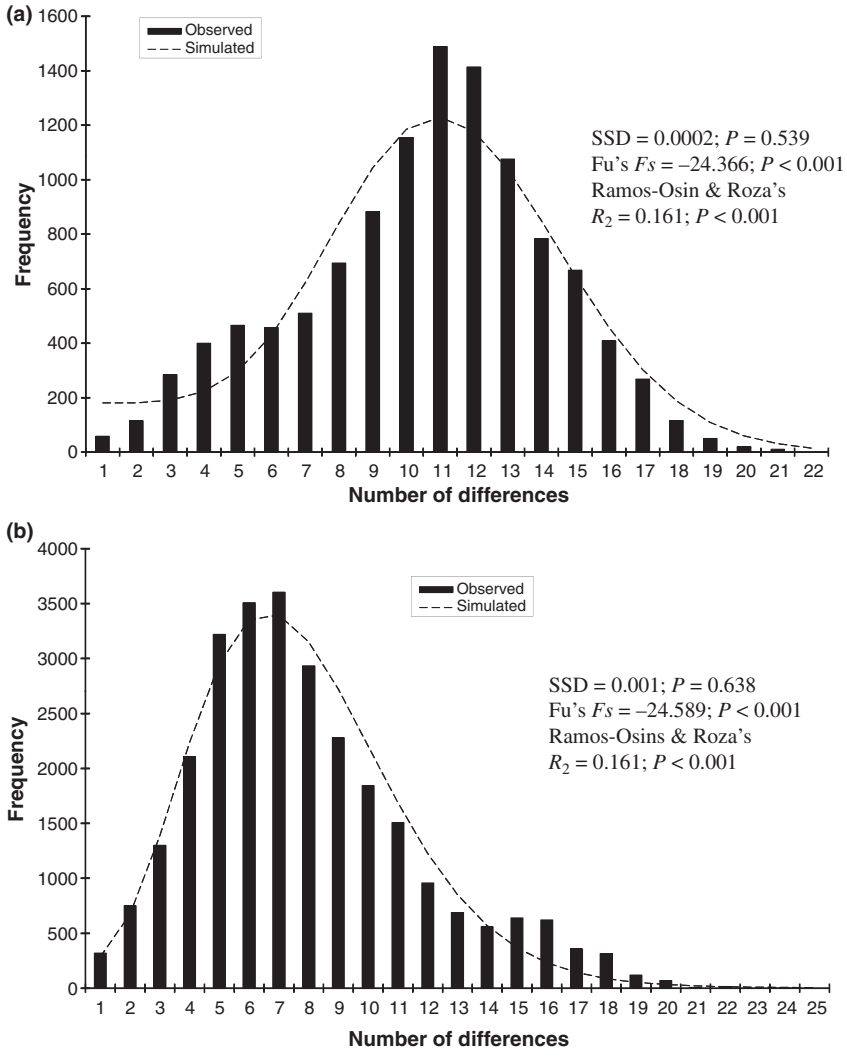

Fig. 5 Mismatch-distribution test of sudden expansion model and test of neutrality for lineages of *Sebastes mystinus* (a) Type 1 and (b) Type 2.

2004; black prickleback, Hickerson & Cunningham, 2005; marine gastropod *Nucella lamellose*, Marko, 2004), *S. mystinus* did not show evidence of a contraction during the LGM indicating the formation of refugium during this time. Due to a large estimated effective population sizes in both regions during the LGM, and no evidence of an expansion dating after this time, neither type was subjected to extensive periods of reduced populations in their respective regions. If there was an effect of population declines during the LGM, there should be a bimodal mismatch distribution supporting two contractions and expansions, one for the origin of the species and one during the LGM. Therefore, the demographic history of *S. mystinus* confirms the existence of the Type 1 lineages during the LGM, and refutes that this event was the cause of speciation in this group. Although environmental changes during the LGM may have differentially affected the two lineages, previous glaciations on the same order of magnitude as my estimates of coalescence time, such as one of the pre-Illinoian glaciations (3.8, 4.6 or 5.4×10^5 years ago), may be the source of distribution fragmentation and origin of the two lineages.

The contemporary distribution of the lineages, the contraction and sudden expansion after the two lineages diverged, and little evidence of introgression between the two lineages, suggest these two species diverged in allopatry. Given the high realized genetic exchange within lineages (i.e. no strong evidence of within lineage genetic structure), substantial physical or evolutionary forces must have countered the potential for genetic exchange between lineages. Before and after the LGM, the two groups may have expanded into the extensive region of overlap. Given the dispersal ability of *S. mystinus*, the lack of exchange between the two lineages supports the hypothesis that reproductive isolating mechanisms were in place prior to coexistence. Unravelling the potential mechanisms that supported speciation in *S. mystinus* requires an investigation of the level of

Table 3 Demographic analysis of *Sebastes mystinus* using (a) estimates of time to expansion, including τ pre- and post-expansion θ_0 and θ_1 populations sizes, respectively, estimate of *uhat* calculated using τ and time (generations) since expansion (t) for both mutation rates calculated with the 3.0 and 3.5 Ma divergence dates and confidence intervals (CI) generated using 10 000 simulations in ARLEQUIN, and (b) estimates of θ ($\theta = 2N_E\mu$) and g based on 309 bp of mtDNA generated in FLUCTUATE.

Lineage	Parameters	Estimated	Lower CI	Upper CI
(a)				
Type 1	τ	10.72	5.75	14.28
	θ_0	0.001	0.000	4.62
	θ_1	61.99	31.18	3960.74
	uhat (3.0 Ma)	6.35×10^{-5}		
	uhat (3.5 Ma)	7.41×10^{-5}		
	t Generations (3.0 Ma)	84 418.37	45 288.83	112 465.94
	t Generations (3.5 Ma)	72 368.07	38 824.07	96 411.98
Type 2	τ	4.28	2.57	13.75
	θ_0	2.77	0.00	10.00
	θ_1	153.73	22.73	5274.98
	uhat (3.0 Ma)	6.35×10^{-5}		
	uhat (3.5 Ma)	7.41×10^{-5}		
	t Generations (3.0 Ma)	33 671.26	20 234.26	108 275.74
	t Generations (3.5 Ma)	28 864.85	17 345.92	92 819.91

Lineage	Region	Sample size	θ (growth)	g	Coalescence time generation (years ago)	N_E (generations)	Proportion of N_E (20 kga)	Proportion of N_E (5 kga)
(b)								
Type 1 (pops)	Northern	67	2.48 (±0.18)	341.67 (±8.68)	129 620–158 605	11 301 518–15 201 301	0.49–0.56	0.84–0.87
Type 1 (mtDNA assemblage)	Northern	97	1.24 (±0.33)	303.80 (±28.55)	136 714–189 781	4 468 640–9 015 128	0.51–0.62	0.84–0.89
Type 2 (pops)	Central/ Southern	198	2.26 (±0.51)	434.97 (±19.85)	101 123–127 818	10 360 419–13 808 855	0.40–0.49	0.79–0.83
Type 2 (mtDNA assemblage)	Central/ Southern	216	2.68 (±0.50)	507.38 (±17.52)	85 958–107 558	10 661 029–18 191 601	0.34–0.43	0.77–0.81

Parameters generated by averaging over 10 runs of the model for each region and lineage including, a sample size, exponential growth, θ, coalescence time to 1% of present population size, effective population size (N_E), and relative N_E at 20 and 5 kga (LGM) compared to current estimate. The range of numbers for the estimated effective and relative effective sizes were generated by using both mutation rates (mutations per generation year) and the upper and lower standard deviation generated using 10 runs of the model.

reproductive isolation within the region of overlap and patterns associated with these levels and the characteristics of the two independent molecular markers throughout the range.

Reproductive isolation

The region of overlap from southern Oregon to far northern California revealed admixture between the lineages and not introgression, which extended the previous finding of admixture at Fort Bragg (Burford & Bernardi, 2008) to southern Oregon and supported reproductive isolation between the two lineages. Reproductive or physical isolating mechanisms appear to be a characteristic within *Sebastes* and could be caused by either prezygotic or post-zygotic barriers. For example, Hyde *et al.* (2008) found depth differences between two cryptic species of *S. miniatus*, which suggested the source or maintenance of the cryptic species was due to a prezygotic barrier of physical segregation. Although there was no evidence of depth distribution differences in *S. mystinus*, two lineages sympatrically distributed over large geographical scales, but partitioned or patchily distributed at smaller spatial scales, may allow coexistence at high frequencies without introgressive hybridization. There was evidence of this patchiness in two northern locations (Cape Arago and Bandon). However, this pattern of geographical patchiness was not found at other locations in the coexistence region. Post-zygotic barriers, such as hybrid inviability or infertility, could develop after secondary contact and lead to reinforcement, increasing the degree of divergence between the lineages or completing the speciation process. Although the genes sampled in this study were not probably involved or closely connected to reinforcement genes, there was not greater genetic divergence between the two groups within mixed sample locations than at greater distances and no evidence of localized reinforcement. Due to the lack of strong evidence of current introgression between the two lineages and suggestion of physical or behavioural partitioning without reinforcement, it appears that the reproductive isolation is

prezygotic in nature and likely a by-product of the speciation process. A prezygotic barrier is plausible in this group of fishes, because their internal fertilization mating system probably provides more opportunities for individuals to develop a specific mate recognition system than in broadcast spawners. Although these data suggest prezygotic barriers, excluding post-zygotic barriers or reinforcement requires an investigation of the frequency of F1 hybrid juveniles in the wild and breeding experiments between adult lineages to confirm fertilization and gamete viability.

Congruence of molecular markers

Although the CR sequence data divided a majority of the Type 1 and Type 2 individuals into distinct assemblages in the NJ tree and showed higher F_{ST}-values than microsatellite markers, there was not reciprocal monophyly between the two lineages at this marker. The lack of fixed difference in the CR was not unique to this study as Cope (2004) showed significant genetic differences between Oregon and California samples only using frequency-based analyses. Previous studies on the black and yellow and gopher rockfish (*Sebastes chrysomelas* and *Sebastes carnatus* respectively) also showed a discrepancy between markers, as the microsatellite data demonstrated differentiation between these sibling species (Narum *et al.*, 2004) but the CR sequence data suggested incomplete lineage sorting (Alesandrini & Bernardi, 1999). As species progress from paraphyly to reciprocal monophyly or incipient species to highly diverged species, depending on the stage there should be discrepancies in the degree of lineage sorting between different genes or markers (Avise, 2000). Given the smaller effective size of mtDNA compared to nuclear markers it is curious that the mtDNA marker did not show more rapid sorting. The discrepancy between mtDNA and microsatellite markers suggests that the speciation process may be complicated by incomplete lineage sorting due to recent divergence or historic introgression between the two types. As explained above, there was little evidence of contemporary introgression, which would also slow lineage sorting. Alternatively, the discrepancy between the two markers could be due to selective forces influencing each marker at different rates during speciation causing greater divergence between the two lineages in the marker associated with genes influenced by adaptive selection.

Alternative scenarios for speciation in *Sebastes mystinus*

Given the pattern of reproductive isolation and the lack of reciprocal monophyly between the mtDNA assemblages, there are three different explanations for events involved in the speciation of *S. mystinus*. Either recent divergence or historic (not current) introgression could slow lineage sorting the mtDNA. Alternatively, ecological or adaptive selection could cause greater divergences in one set of markers. Although recent divergence may not allow enough time to complete the speciation process, the coalescence date and degree of distinctiveness at the microsatellite markers in this study do not support recent time to divergence. For example, nuclear markers with a higher effective population size and recombination should be less sorted than mtDNA markers, which was supported by the frequency-based measures but not by the lack of fixed differences between the two CR assemblages. However, low-level introgression of mtDNA and not nuclear genes between the two lineages could cause incomplete lineage sorting and could explain the slower sorting of the mtDNA compared to nuclear markers found in this study. Recombination at nuclear markers minimizing the contribution of opposite lineages as introgressed individuals backcross to the dominate regional type would cause lower introgression of nuclear compared to mtDNA genes. Although historic introgression, close to the time of speciation, was more likely the source of the discrepancy between the markers, both historic introgression and a short time since divergence result in incomplete lineage sorting of the mtDNA.

An alternative to incomplete lineage sorting is ecological or adaptive selection driving divergences at the microsatellite markers. Narrowing the focus from overall measures of divergence to individual microsatellite loci, comparing measures of microsatellites used here with the same microsatellites used in other *Sebastes* studies (Buonaccorsi *et al.*, 2004, 2005) both *Sra.7-2* and *Sra.7-25* showed higher genetic divergence, estimated by F_{ST}, in *S. mystinus* than in these other species. Significant divergence and significantly low genetic diversity at individual loci may suggest alleles that are closely linked or hitchhiking with genes that are involved in a selective sweep (Weihe *et al.*, 2007). Although the three main loci that showed the highest genetic divergence between the two lineages in this study also showed a significant decrease in both measures of genetic diversity in the northern Type 1 locations, this pattern could not be distinguished7 from that caused by signal noise. To assess whether ecological selection in allopatry drove the speciation process and the pattern was not due to signal noise requires sampling many loci to differentiate those above the average background levels and potentially linked to selected genes (Weihe *et al.*, 2007).

Overall, the genetic analysis of both lineages throughout their respective geographical ranges reveals a complex pattern of speciation. The current distribution pattern, different lineage centres and the step-cline, suggests allopatric speciation and subsequent expansion at an event prior to the LGM. The lack of substantial structure within lineages indicated adequate genetic exchange over large geographical distances, but the reproductive isolation indicated minimal genetic exchange between lineages

despite contact in regions of overlap. Given the lack of evidence of current reinforcement within this region, isolating mechanisms were likely in place prior to secondary contact and the lack of hybrids suggest prezygotic barriers. The discrepancy between the two markers reveals a pattern of either low-level historic introgression or recent divergence resulting in incomplete lineage sorting at the mtDNA. The focus of future investigations within the *Sebastes* group and other marine nearshore species should continue to explore divergence mechanisms, by analysing the genetic structure of lineages throughout their distributions and incorporating both the demographic history and genetic pattern of multiple markers. From this and subsequent studies, we can begin to unravel the complicated process of speciation in the marine environment despite a high potential to reproduce and disperse for many of these species.

Acknowledgments

I could not have completed this research without assistance in the field. For their assistance with sample collections I thank, R. Buckley, D. Bondemiller, K. Gordon, J. Headlee, D. VenTresca, P. Gundelfinger, R. Larson, T. Laidig, S. Parker, A. Chapelle, J. Barlow, M. O'Farrell, P. Tompkins, J. Fugurski, J. Grover, R. Nakamura, T. Olive, J. Hyde, M. Ramon, G. Bernardi, N. Crane, M. McCrea, J. Barr, the Oregon and Washington Department of Fish and Wildlife and the California Department of Fish and Game. I would like to thank Michael Reiskind, Vince Buonaccorsi, Giacomo Bernardi, Mark Carr, Ralph Larson, Stephen Palumbi, Grant Pogson and Peter Raimondi for comments on an earlier version of this manuscript. Anonymous reviewer's comments greatly improved this manuscript. The project was funded by grants to M.O. Burford from the PADI Foundation, Myers Trust Grant, The ASIH Raney Fund, Friends of the Long Marine Lab Fellowship, California Environmental Quality Initiative (CEQI) Graduate Fellowship, the Marilyn C. Davis Grant, and from the Packard Foundation's Partnership for the Interdisciplinary Study of Coastal Oceans (PISCO). This is contribution 334 from PISCO, funded primarily by the Gordon and Betty Moore Foundation and the David and Lucile Packard Foundation.

References

Alesandrini, S. & Bernardi, G. 1999. Ancient species flocks and recent speciation events: what can rockfish teach us about cichlids (and vice versa)? *J. Mol. Evol.* **49**: 814–818.

Avise, J.C. 1994. *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Avise, J.C. 2000. *Phylogeography, the History and Formation of Species*. Harvard University Press, Cambridge.

Buonaccorsi, V.P., Westerman, M., Stannard, J., Kimbrell, C., Lynn, E. & Vetter, R.D. 2004. Molecular genetic structure suggests limited larval dispersal in grass rockfish *Sebastes rastrelliger*. *Mar. Biol.* **145**: 779–788.

Buonaccorsi, V.P., Kimbrell, C.A., Lynn, E.A. & Vetter, R.D. 2005. Limited realized dispersal and introgressive hybridization influence genetic structure and conservation strategies for brown rockfish *Sebastes auriculatus*. *Conserv. Genet.* **6**: 697–713.

Burford, M.O. & Bernardi, G. 2008. Incipient speciation within a subgenus of rockfish (*Sebastosomus*) provides evidence of recent radiations within an ancient species flock. *Mar. Biol.* **154**: 701–717.

Burford, M.O. & Larson, R.J. 2007. Genetic heterogeneity in a single year-class from a panmictic population of adult blue rockfish (*Sebastes mystinus*). *Mar. Biol.* **151**: 451–465.

Burton, R.S. 1998. Intraspecific phylogeography across the Point Conception biogeographic boundary. *Evolution* **52**: 734–745.

Cope, J.M. 2004. Population genetics and phylogeography of the blue rockfish (*Sebastes mystinus*) from Washington to California. *Can. J. Fish. Aquat. Sci.* **61**: 332–342.

Edmands, S. 2001. Phylogeography of the intertidal copepod *Tigriopus californicus* reveals substantially reduced population differentiation at northern latitudes. *Mol. Ecol.* **10**: 1743–1750.

Evanno, G., Regnaut, S. & Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**: 2611–2620.

Excoffier, L., Smouse, P.E. & Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes – application to human mitochondrial-DNA restriction data. *Genetics* **131**: 479–491.

Falush, D., Stephens, M. & Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.

Goudet, J. 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. *J. Hered.* **86**: 485–486.

Guo, S.W. & Thompson, E.A. 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48**: 361–372.

Hickerson, M.J. & Cunningham, C.W. 2005. Contrasting quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (Xiphister) revealed by multilocus DNA analysis. *Evolution* **59**: 344–360.

Huelsenbeck, J.P. & Ronquist, F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Hyde, J.R. & Vetter, R.D. 2007. The origin, evolution, and diversification of rockfishes of the genus *Sebastes* (Cuvier). *Mol. Phylogenet. Evol.* **44**: 790–811.

Hyde, J.R., Kimbrell, C.A., Budrick, J.E., Lynn, E.A. & Vetter, R.D. 2008. Cryptic speciation in the vermilion rockfish (*Sebastes miniatus*) and the role of bathymetry in the speciation process. *Mol. Ecol.* **17**: 1122–1136.

Johns, G.C. & Avise, J.C. 1998. Tests for ancient species flocks based on molecular phylogenetic appraisals of *Sebastes* rockfishes and other marine fishes. *Evolution* **52**: 1135–1146.

Jorgensen, S.J., Kaplan, D.M., Klimley, A.P., Morgan, S.G., O'Farrell, M.R. & Botsford, L.W. 2006. Limited movement in blue rockfish *Sebastes mystinus*: internal structure of home range. *Mar. Ecol. Prog. Ser.* **327**: 157–170.

Kuhner, M.K., Yamato, J. & Felsenstein, J. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**: 429–434.

Larson, R.J. 1980. Territorial behavior of the black and yellow rockfish and gopher rockfish (Scorpaenidae, Sebastes). *Mar. Biol.* **58**: 111–122.

Lee, W.J., Conroy, J., Howell, W.H. & Kocher, T.D. 1995. Structure and evolution of teleost mitochondrial control regions. *J. Mol. Evol.* **41**: 54–66.

Leet, W.S., Dewees, C.M., Klingbeil, R. & Larson, E.J. (Eds) 2001. California's Living Marine Resources: Status Report. The Resource Agency, Department of Fish and Game, CA.

Lewis, P.O. & Zaykin, D. 2001. Genetic data analysis: computer program for the analysis of allelic data [www document]. URL http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php (Free program distributed by the authors over the internet).

Love, M.S., Yoklavich, M. & Thorsteinson, L. 2002. *The Rockfishes of the Northeast Pacific*. University of California Press, Berkeley, Los Angeles.

Marko, P.B. 2004. 'What's larvae got to do with it?' Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Mol. Ecol.* **13**: 597–611.

Miller, D.J. & Geibel, J.J. 1973. Summary of blue rockfish and lingcod life histories; a reef ecology study; and giant kelp, *Macrocystis pyrifera*, experiments in Monterey Bay, California. *Fish. Bull.* **158**: 1–137.

Narum, S.R., Buonaccorsi, V.P., Kimbrell, C.A. & Vetter, R.D. 2004. Genetic divergence between gopher rockfish (*Sebastes carnatus*) and black and yellow rockfish (*Sebastes chrysomelas*). *Copeia* **4**: 926–931.

Nei, M. 1972. Genetic distance between populations. *Am. Nat.* **106**: 283–292.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M. & Kumar, S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford.

Posada, D. & Crandall, K.A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Raymond, M. & Rousset, F. 1995a. GENEPOP (version-1.2) – population-genetics software for exact tests and ecumenicism. *J. Hered.* **86**: 248–249.

Raymond, M. & Rousset, F. 1995b. An exact test for population differentiation. *Evolution* **49**: 1280–1283.

Rice, W. 1989. Analyzing tables of statistical tests. *Evolution* **43**: 223–225.

Rocha-Olivares, A., Kimbrell, C.A., Eitner, B.J. & Vetter, R.D. 1999a. Evolution of a mitochondrial cytochrome b gene sequence in the species-rich genus *Sebastes* (Teleostei, Scorpaenidae) and its utility in testing the monophyly of the subgenus *Sebastomus*. *Mol. Phylogenet. Evol.* **11**: 426–440.

Rocha-Olivares, A., Rosenblatt, R.H. & Vetter, R.D. 1999b. Molecular evolution, systematics, and zoogeography of the rockfish subgenus Sebastomus (Sebastes, Scorpaenidae) based on mitochondrial cytochrome b and control region sequences. *Mol. Phylogenet. Evol.* **11**: 441–458.

Rogers, A.R. & Harpending, H. 1992. Population growth makes waves in the distribution pf pairwise genetic differences. *Mol. Biol. Evol.* **93**: 552–569.

Romos-Onsins, S.E. & Rozas, J. 2002. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**: 2092–2100.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X. & Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.

Schneider, S. & Excoffier, L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rate vay among sites: application to human mitochondrial DNA. *Genetics* **152**: 1079–1089.

Schneider, S., Roessli, D. & Excoffier, L. 2000. ARLEQUIN ver 3.01: A Software for Population Genetics Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Sokal, R.R. & Rohlf, F.J. 1995. *Biometry*. W.H. Freeman and Co., New York.

Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T. & Mindell, D.P. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phylogenet. Evol.* **12**: 105–114.

Sotka, E.E., Wares, J.P., Barth, J.A., Grosberg, R.K. & Palumbi, S.R. 2004. Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle Balanus glandula. *Mol. Ecol.* **13**: 2143–2156.

Stephens, J.S., Morris, P.A., Pondella, D.J., Koonce, T.A. & Jordan, G.A. 1994. Overview of the dynamics of an urban artificial reef fish assemblage at King Harbor, California, USA, 1974–1991: a recruitment driven system. *Bull. Mar. Sci.* **55**: 1224–1239.

Swofford, D.L. 1998. PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer, Sunderland, MA.

Wares, J.P. & Cunningham, C.W. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* **55**: 2455–2469.

Weihe, T., Nolte, V., Zivkovic, D. & Schlotterer, C. 2007. Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. *Genetics* **175**: 207–218.

Weir, B.S. & Cockerham, C.C. 1984. Estimating F-Statistics for the analysis of population-structure. *Evolution* **38**: 1358–1370.

Westerman, M.E., Buonaccorsi, V.P., Stannard, J.A., Galver, L., Taylor, C., Lynn, E.A., Kimbrell, C.A. & Vetter, R.D. 2005. Cloning and characterization of novel microsatellite DNA markers for the grass rockfish, *Sebastes rastrelliger*, and cross-species amplification in 10 related *Sebastes* spp. *Mol. Ecol. Notes* **5**: 74–76.

Withler, R.E., Beacham, T.D., Schulze, A.D., Richards, L.J. & Miller, K.M. 2001. Co-existing populations of Pacific ocean perch, *Sebastes alutus* in Queen Charlotte Sound, British Columbia. *Mar. Biol.* **139**: 1–12.

Wyllie-Echeverria, T. 1987. Thirty-four species of California rockfishes: maturity and seasonality of reproduction. *Fish. Bull.* **85**: 229–250.

Supporting information

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

Appendix S1 Sample data of *Sebastes mystinus*, including sample size (*N*), mean standard length in cm, and collection method.

Appendix S2 Summary statistics of *Sebastes mystinus* using 309 bp mtDNA control region sequence data including, number of sequences (*N*), number of segregating sites (*S*), number of haplotypes (H#), haplotype diversity (Hd) and nucleotide diversity (π).

Appendix S3 Results of parameter estimates for STRUCTURE assignment model using groupings (*K*) of 1–5, average and variance of the log-likelihood, data estimates for *K* (Pritchard *et al.*, 2000), and ΔK (Evanno *et al.*, 2005).

Appendix S4 Pairwise F_{ST}-values of *Sebastes mystinus* using both microsatellite data (below diagonal) and 309 bp of mtDNA CR sequence data (above diagonal), including pairwise comparisons of (a) all sample locations, and (b) all sample locations and the two types (Type 1 and Type 2).

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

Received 18 January 2009; revised 30 March 2009; accepted 1 April 2009

Copyright of Journal of Evolutionary Biology is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Copyright of Journal of Evolutionary Biology is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.