Comparative Population Genetic Analysis of Bocaccio Rockfish Sebastes paucispinis Using Anonymous and Gene-Associated Simple Sequence Repeat Loci

VINCENT P. BUONACCORSI, CAROL A. KIMBRELL, ERIC A. LYNN, AND JOHN R. HYDE

From the Department of Biology, Juniata College, Huntingdon, PA 16652 (Buonaccorsi); and the Southwest Fisheries Science Center, National Marine Fisheries Service, La Jolla, CA (Kimbrell, Lynn, and Hyde).

Address correspondence to Vincent P. Buonaccorsi at the address above, or e-mail: buonaccorsi@juniata.edu.

Abstract

Comparative population genetic analyses of traditional and emergent molecular markers aid in determining appropriate use of new technologies. The bocaccio rockfish Sebastes paucispinis is a high gene-flow marine species off the west coast of North America that experienced strong population decline over the past 3 decades. We used 18 anonymous and 13 gene-associated simple sequence repeat (SSR) loci (expressed sequence tag [EST]-SSRs) to characterize range-wide population structure with temporal replicates. No FST- outliers were detected using the LOSITAN program, suggesting that neither balancing nor divergent selection affected the loci surveyed. Consistent hierarchical structuring of populations by geography or year class was not detected regardless of marker class. The EST-SSRs were less variable than the anonymous SSRs, but no correlation between FST and variation or marker class was observed. General linear model analysis showed that low EST-SSR variation was attributable to low mean repeat number. Comparative genomic analysis with Gasterosteus aculeatus, Takifugu rubripes, and Oryzias latipes showed consistently lower repeat number in EST-SSRs than SSR loci that were not in ESTs. Purifying selection likely imposed functional constraints on EST-SSRs resulting in low repeat numbers that affected diversity estimates but did not affect the observed pattern of population structure.

Key words: comparative genomics, marine fisheries, microsatellite, stock structure, threatened species

Discerning geographic patterns of population connectivity in high gene-flow marine species is difficult due to low signal to noise ratios and genetic disequilibrium (Waples 1998; Bradbury and Bentzen 2007). Genetic assays including loci involved in local adaptation may reveal dispersal barriers that would otherwise be obscured at neutrally evolving markers. Polymorphic markers developed from expressed sequence tags (ESTs) (partial sequence of reverse transcribed mRNAs) have the advantage of being variable and closely associated with polymorphisms that may reflect local adaptation along environmental gradients such as sea surface temperature (Ellis and Burke 2007; Yatabe et al. 2007; Nielsen, Hemmer-Hansen, Paulsen, et al. 2009, Nielsen, Hemmer-Hansen, Larsen, et al. 2009). Gene flow may be high enough to homogenize neutral polymorphisms but not so high that formation of genetic clines is precluded (Schmidt et al. 2008; Nielsen, Hemmer-Hansen, Larsen, et al. 2009). Only a few comparative studies have been performed to date regarding performance of gene-associated simple sequence repeats (SSRs) (i.e., ESTs containing SSRs or EST-SSRs) and anonymous (i.e., unannotated) SSRs for discernment of population genetic structure (Woodhead et al. 2005; Yatabe et al. 2007; Kim et al. 2008). Divergent (Nielsen, Hemmer-Hansen, Paulsen, et al. 2009; Nielsen, Hemmer-Hansen, Larsen, et al. 2009), balancing (Charlesworth 2006), or purifying selection (Charlesworth et al. 1993; Hahn 2008) can act on gene-associated polymorphisms, and repeat variation is functionally constrained in genes (Metzgar et al. 2000; Li et al. 2004). Therefore, it is still unclear how informative SSRs associated with genes will typically be in studies of population genetic structure compared with anonymously distributed SSRs.
The bocaccio rockfish (*Sebastes paucispinis*) is a large, long-lived late-maturing species that has been a mainstay of the California commercial and recreational fishery for several decades. Bocaccio adults can be found in or near kelp forests and rocky reefs, occupying midwater or bottom areas (Love et al. 2002). The species ranges from Kodiak Island, Alaska to central Baja California, Mexico (Miller and Lea 1972) but is common only in the waters south of British Columbia. Individuals have a pelagic larval and juvenile phase followed by a demersal juvenile and adult existence where they inhabit the continental slope and shelf regions from 50 to 250 m. Bocaccio were assessed as overfished in 1999 (MacCall et al. 1999) and have been listed as a species of concern for National Oceanic and Atmospheric Administration since 2002. The species is susceptible to overfishing given the low intrinsic rate of population increase characteristic of *Sebastes* rockfishes in general (Musick et al. 2000), due largely to their late age at maturity and highly variable recruitment. During the El Niño dominated warm oceanographic regime of the 1980s and 1990s, there were consistently poor year classes followed by a major recruitment event in 1999, with subsequently strong year classes in 2003, 2005, and 2009 (Field et al. 2010).

The stock structure of the species remains a significant source of uncertainty in the management of the bocaccio fishery. A zone of low adult abundance between Cape Mendocino and the Columbia River mouth was evident in both fishery-independent data from the 1990s (Ralston et al. 1996) and fishery-dependent data from Russian trawls taken between 1963 and 1978 that predate the population crash of the early 1980s (MacCall et al. 1999). A lack of pelagic juveniles in this geographic region was also reported (Sakuma et al. 2006). Because gene flow is the product of effective population size and migration rates, a disruption in range-wide connectivity is plausible. However, larvae and juveniles are pelagic for 2–4 months (Moser and Boehlert 1991; Woodbury and Ralston 1991) and have been detected as far as 300 nautical miles from the coast (Moser et al. 1993). Therefore, dispersal at these life history phases in particular may be extensive. Results of prior genetic stock identification efforts revealed no trace (Wishard et al. 1980) to weak (Matala et al. 2004) geographic partitioning of genetic variance using ad hoc statistical methods. Development of additional informative polymorphic markers may help clarify patterns of population structure.

In the present study, we used anonymous SSRs and EST-SSRs in *S. paucispinis* to examine its population genetic structure. To help prevent confounding effects on geographic structure due to sweeps recruitment (chance variation among cohorts) or unknown sampling biases, we aimed to sample the range of the species with 2 temporal replicate collections per location. Since the large 1999 year class dominated the adult population starting in 2003, we focused year class analysis on whether adult allele frequency patterns were consistent before and after 2003. The EST-SSR loci are all closely associated with functional areas of the genome. Directional or balancing selection at these loci could increase genetic heterogeneity among collections if selective pressures vary along the species range or conversely reduce divergence if selective pressures are similar along the species range. Purifying selection against deleterious mutations may reduce variation or repeat number predominantly at EST-SSR loci, but conversely, diversifying selection could increase diversity.

**Materials and Methods**

**Samples and Laboratory Methods**

Collections of 47–49 bocaccio were obtained from different years at 3 locations spanning the species’ range, including the Southern California Bight, Central California, and British Columbia (Table 1). Fish from 2003 and 2004 were comprised predominantly of smaller and younger fish, consistent with recruitment stemming from the 1999 spawning year class. Development, amplification, and scoring procedures for 13 novel EST-SSR markers for this study were described in Buonaccorsi et al. (2011). In addition to those 13 loci, we tested 44 primers for amplification and polymorphism from a compilation of 67 anonymous SSR loci that had previously been previously characterized in other species (Bernston et al. 2007). Here we also introduce an additional locus, *Sr5-l* 6-52, (forward: 5’-ATCGGGTGTTCCCTACGTAC-3’, reverse: 5’-CGCTTTAATTTCCCGTGA-3’; GenBank number AF269057). We scored all samples for 18 anonymous SSR loci. PCR primers were modified by addition of 5’ M13 tags following Schuelle (2000) to allow for generic fluorescent labeling of SSR loci and a GTTT tail on the 5’ end of the reverse primer to complete nontemplated adenylation of Taq polymerase (Brownstein et al. 1996). PCR amplifications were performed in 5 μl reactions. For single locus reactions, 0.04 μM of forward, and 0.16 μM of reverse, and M13 labeled primers were used, along with 2.5 μl of 2x Qiagen Hotstar Master Mix and 20 ng of template DNA. Duplexed reactions were run as above except the M13 labeled primer was run at 0.32 μM. Annealing temperatures of 55 and 60 °C were used. For PCR amplification with an annealing temperature of 55 °C, cycling conditions included a single cycle of 94 °C for 5 min, 36 cycles of 94 °C for 15 s, 55 °C for 15 s, 72 °C for 30 s, and finally extension of 72 °C for 10 min. For PCR amplification with an annealing temperature of 60 °C, cycling conditions included one cycle of 94 °C for 5 min, 28 cycles of 94 °C for 15 s, 60 °C for 15 s, and 72 °C for 30 s, and 8 cycles of 94 °C for 15 s, 55 °C for 15 s, 72 °C for 30 s. Again, a final extension of 72 °C for 10 min was performed. We genotyped individuals using the ABI 3730 Genetic Analyzer. Electropherograms produced by the genetic analyzer were analyzed and scored using GeneMapper v4.0 Software (Applied Biosystems) with visual confirmation of all automated allele calls. For analyses below involving repeat number (MICROCHECKER, LOSITAN, general linear models [GLMs], aka GLMs), *Sr15-8, Srme10, Ru2a, and LBH* were removed because they
had off-repeat alleles (i.e., allele sizes deviated from a "perfect" repeat pattern).

**Statistical Analysis**

Presence of null alleles and scoring errors were evaluated using MICROCHECKER v.2.2.3 (van Oosterhout et al. 2004). Exact-significance testing methods were used to evaluate probability of deviation from Hardy–Weinberg equilibrium (HWE) and linkage equilibrium (LE), as implemented in GENEPOP ver. 4 (Raymond and Rousset 1995) using Markov Chains with 10 000 dememorization steps, 500 batches, and 5000 iterations per batch. Probabilities over loci were combined using Fisher’s method (Sokal and Rohlf 1995). Significant differences in allele frequencies were considered over multiple tests was evaluated using sequential Bonferroni corrections (Rice 1989).

To investigate whether divergent or balancing selection may have affected any loci, we examined whether genetic divergence at any locus was greater or lesser than expected given the average divergence level. Island-model coalescent simulations of mutation-drift-equilibrium were performed to generate the sampling distribution of single locus \( F_{ST} \) values using LOSITAN (Antao et al. 2008). Simulations (50 000) were performed on 3 populations (i.e., the number of geographic locations). Quantiles (central 99%) on single locus \( F_{ST} \) values given the mean \( F_{ST} \) (iteratively adjusted for possible outliers) were calculated. Simulations were run under both stepwise (SMM) and infinite allele (IAM) mutation models.

ARLEQUIN v. 3.1 (Excoffier et al. 2005) was used to perform hierarchical AMOVA examining significance of allele frequency variation among regions and among temporal collections within regions using 10 000 iterations. If dispersal is limited, we expect variation among regions to be stronger than among temporal collections within regions. In order to determine if collections clustered significantly by year class, we also performed hierarchical AMOVA examining significance of clustering before and after 2003, as well as among collections within year classes. We also performed exact \( G \)-tests for allele frequency homogeneity among the 6 collections using GENEPOP with run parameters as described above. For comparative purposes, EST and anonymous SSRs were considered both separately and combined in the analyses above. Factorial correspondence analysis by populations, as implemented in GENETIX version 4.05.2 (Belkhir 1996–2004), was used to visualize relationships among collections.

In order to better understand variation among loci and collections in parameter estimates, heterozygosity was modeled as a linear function of region, replicate nested within region, marker type (EST or anonymous SSR), and locus nested within marker type, using GLM analysis in Minitab v. 16 [Heterozygosity = Region + Replicate(Region) + MarkerType + Locus(MarkerType); State College, PA]. Repeat number is known to influence SSR mutation rate and therefore diversity levels. Because we found consistently low diversity within EST markers (see Results below), we wished to further discern whether variation in mean repeat number alone was sufficient to explain heterozygosity differences, by modeling mean heterozygosity (over all collections) against mean repeat number (over all collections) and marker type, with interaction (MeanHeterozygosity = MeanRepeatNumber + MarkerType + MeanRepeatNumber \times MarkerType). Low repeat number itself has been suggested to result from selective constraints on polymorphism within genes. If this were the case in general, one would expect a lower repeat number within coding regions than outside in the model fishes sequenced so far. To test this hypothesis, we performed a comparative genomic analysis on Gasterosteus aculeatus, Takifugu rubripes, and Oryzias latipes. Repeat number and position of all di-, tri-, or tetranucleotide repeats and locations of all ESTs from each genome were downloaded from the UCSC table browser (simple repeats track). EST-SSRs were identified as SSRs whose genome coordinates intersected with those of EST blocks using Galaxy (Goecks et al. 2010). Any SSRs that overlapped both EST and non-EST regions were removed. For each species, standard 95% confidence intervals (CIs) on the true mean difference in repeat number between EST and non-EST SSR markers were assessed in Minitab v. 16 (State College, PA).

In order to examine whether variation in genetic divergence depended on marker type and/or diversity, \( F_{ST} \) was modeled against mean locus heterozygosity, marker type, and their interaction (\( F_{ST} = \text{MeanHeterozygosity} + \text{MarkerType} + \text{MeanHeterozygosity} \times \text{MarkerType} \)). A difference in divergence among marker classes after accounting for variation in diversity could indicate that divergent or balancing selection is acting on the marker

---

**Table 1** *Sebastes paucispinis* collection information

<table>
<thead>
<tr>
<th>Collection</th>
<th>Location</th>
<th>Lat. °N (mean)</th>
<th>Long. °W (mean)</th>
<th>Total length (mean [mm] ± SD)</th>
<th>N</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>QC99</td>
<td>Queen Charlotte Islands</td>
<td>52.58</td>
<td>130.25</td>
<td>677 ± 54</td>
<td>48</td>
<td>May 1999</td>
</tr>
<tr>
<td>QC04</td>
<td>Queen Charlotte Islands</td>
<td>53.29</td>
<td>132.47</td>
<td>303 ± 26</td>
<td>48</td>
<td>May–July 2004</td>
</tr>
<tr>
<td>MB98</td>
<td>Monterey Bay</td>
<td>36.74</td>
<td>121.98</td>
<td>Not available</td>
<td>49</td>
<td>June 1998</td>
</tr>
<tr>
<td>CB00</td>
<td>Cordell Bank</td>
<td>38.00</td>
<td>123.04</td>
<td>632 ± 79</td>
<td>47</td>
<td>May–June 2000</td>
</tr>
<tr>
<td>SCB94</td>
<td>Southern California Bight</td>
<td>32.58</td>
<td>118.89</td>
<td>505 ± 68</td>
<td>48</td>
<td>March 1994–August 1995</td>
</tr>
<tr>
<td>SCB03</td>
<td>Southern California Bight</td>
<td>32.65</td>
<td>117.97</td>
<td>373 ± 151</td>
<td>48</td>
<td>November 2002–March 2003</td>
</tr>
</tbody>
</table>

SD, standard deviation.

* Capture method: hook and line, trawl, or spear.
class. For each GLM, normality of residuals was examined using an Anderson–Darling test, and heteroscedasticity evaluated by visual examination of residual scatter against fitted values for no striking trend, as implemented in Minitab v. 16.

Results

Over the 13 EST-SSR and 18 anonymous SSR loci, no evidence of null alleles or scoring error was detected using MICROCHECKER. Deviations from HWE were not apparent at any locus over collections (Table 2). Deviations from linkage equilibrium were not significant after corrections for multiple tests. Over all 465 pairwise tests of LE among the 31 loci (combined over populations), all P values were greater than 0.002 and, none were significant after corrections for multiple tests (initial alpha = 0.0001). LOSITAN analysis revealed that regardless of whether SMM or IAM, mutation models were run (IAM includes all loci including those with off-repeat alleles), no $F_{ST}$ values were outside expected 99% quantiles.

Results revealed little evidence for population subdivision by geography or year class. When collections were arranged in a Northern/Central/Southern geographical hierarchy using AMOVA, neither single locus nor overall $F_{CT}$ or $F_{SC}$ values were significant after corrections for multiple tests, regardless of whether loci were combined within marker class or over all 31 markers (Table 2). Similar results occurred when collections were arranged temporally into a pre-2003/post-2003 hierarchy (Table 2). Visualization of population relationships showed little clustering by geography or year class for any marker class (Figure 1). Using exact G-tests, allele frequency divergence among the 6 collections was not significant when $P$ values were combined over the 18 anonymous loci ($F_{ST} = -0.0002; P = 0.536$). Although $F_{ST}$ values were negative, overall allele frequency divergence using the exact G-tests was significant.

Table 2 Locus summary statistics

<table>
<thead>
<tr>
<th>Locus name</th>
<th>N</th>
<th>AT</th>
<th>A</th>
<th>$H_e$</th>
<th>$F_{IS}$</th>
<th>$P$</th>
<th>Geographic structure</th>
<th>Temporal structure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$F_{SC}$</td>
<td>$P$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$F_{CT}$</td>
<td>$P$</td>
</tr>
</tbody>
</table>

Average (over 6 collections) sample size (N), annealing temperature (AT), number of alleles (A), expected heterozygosity ($H_e$), local inbreeding coefficient ($F_{IS}$) with significance value combined over collections, allele frequency divergence among collections within groups ($F_{SC}$) or among groups ($F_{CT}$), under geographic or temporal hierarchical grouping schemes.
over the 13 EST loci ($F_{ST} = -0.0003; P = 0.006$) and over all 31 loci ($F_{ST} = -0.0002; P = 0.0045$). No pairwise comparisons among collections over all loci were significant after corrections for multiple tests ($P < 0.014$), but over EST loci, SCB94 differed from both SCB03 ($P = 0.0003$) and MB98 ($P = 0.00008$). These collections were visually separated by Factor 1 in factorial correspondence analysis for ESTs (Figure 1). Divergence over loci was mainly attributable to a single locus, LBH (limb–bud–heart) ($F_{ST} = 0.003; P = 0.0006$). For this locus, only pairwise comparisons between the 2 southern collections against MB98 were significant after corrections for multiple tests (Table 3). Population relatedness was not geographically coherent at this locus (Figure 1; Table 2), with the only pattern being that the 2 southern California populations were most similar, in contrast to the overall EST pattern described above. With this locus removed, allele frequency divergence over EST ($P = 0.105$) or all other loci ($P = 0.24$) were not significant.

Variation in diversity among loci was related to marker class and mean repeat number rather than collection. GLM analysis showed that mean heterozygosity did not vary significantly among the 3 geographic regions ($F_{2,150} = 0.08; P = 0.924$; Figure 2) or among replicate collections within geographic regions ($F_{3,150} = 1.18; P = 0.32$). However, there was clearly a relatively low diversity at the EST-SSRs ($F_{1,150} = 8.29; P = 0.007$), with significant variances among loci within each marker type ($F_{29,150} = 171.5; P < 0.001$). Very similar results were found when groups were defined

![Figure 1. Factorial correspondence cluster analysis of *Sebastes paucispinis* collections. Northern collections were from Queen Charlotte Island (QC99, QC04). Central collections were from Monterey Bay (MB98) and Cordell Bank (CB00). Southern collections were from the southern California Bight (SCB94, SCB03).](image1.png)

![Figure 2. Boxplots of expected heterozygosities over loci by collection and marker type. Collection abbreviations follow Figure 1. Marker type is designated as either anonymous microsatellites (Anonymous) or gene-associated microsatellites (EST). Median is shown within boxes. Box limits indicate interquartile range among loci. Whiskers indicate last non-outlier datapoint on either side of median. Stars indicate outlier loci.](image2.png)

Table 3  Pairwise comparisons of allele frequency divergence at the $LBH$ locus

<table>
<thead>
<tr>
<th></th>
<th>SCB94</th>
<th>SCB03</th>
<th>MB98</th>
<th>CB00</th>
<th>QC99</th>
<th>QC04</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCB94</td>
<td>—</td>
<td>0.0081</td>
<td>0.0238</td>
<td>0.0038</td>
<td>0.0048</td>
<td>−0.0067</td>
</tr>
<tr>
<td>SCB03</td>
<td>—</td>
<td>—</td>
<td>0.0141</td>
<td>0.0052</td>
<td>0.0008</td>
<td>−0.003</td>
</tr>
<tr>
<td>MB98</td>
<td>***</td>
<td>**</td>
<td>—</td>
<td>0.0007</td>
<td>−0.0021</td>
<td>0.0079</td>
</tr>
<tr>
<td>CB00</td>
<td>*</td>
<td>*</td>
<td>—</td>
<td>—</td>
<td>0.0028</td>
<td>−0.0074</td>
</tr>
<tr>
<td>QC99</td>
<td>*</td>
<td>*</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.0066</td>
</tr>
<tr>
<td>QC04</td>
<td>*</td>
<td></td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

$F_{ST}$ values are shown above the diagonal and $P$ values for exact $G$-tests of allele frequency divergence below diagonal. Values with greater than one star were significant after corrections for 15 pairwise tests. Shaded cells indicate comparison of temporal replicates.

*P < 0.05, **P < 0.001, ***P < 0.001.
by year class (not shown). In these analyses, heterozygosity values were transformed via arcsine square root in order for model residuals to be approximately normal and to reduce heteroscedasticity. GLM analysis of heterozygosity against marker type and mean repeat number over collections revealed no interaction between type and repeat number \((F_{1,23} = 0.16; P = 0.691)\), a significant relationship between heterozygosity and mean repeat number \((F_{1,23} = 28.43; P < 0.001)\), and no relationship between marker type and heterozygosity after accounting for these effects \((F_{1,23} = 0.70; P = 0.41)\; (\text{Figure 3}).\) GLM analysis also showed that \(F_{\text{ST}}\) was not significantly predicted by locus mean heterozygosity \((F_{1,27} = 0.57; P = 0.46)\), marker type \((F_{1,27} = 0.06; P = 0.804)\), or their interaction \((F_{1,27} = 0.48; P = 0.493)\).

EST-SSRs in \(G.\) aculeatus, \(T.\) rubripes, and \(O.\) latipes all had lower repeat numbers than SSRs outside of coding regions (Table 4). CIs (95%) on the true difference in mean repeat number ranged from 4.9 to 11.3 lower in EST-SSRs over the 3 species. The magnitude of difference in the 3 species was therefore sufficient to explain the difference detected in \(S.\) paucispinis (3.5 lower mean in EST-SSRs; Table 4).

## Discussion

Results from the Lositan \(F_{\text{ST}}\) outlier analysis suggested that neither divergent nor balancing selection occurred on any locus, in that there were no unexpectedly high or low \(F_{\text{ST}}\) values given the mean divergence level. In addition, GLM analysis showed that there were no differences in mean divergence among marker classes with both classes essentially showing no genetic divergence. In Atlantic cod, divergent selection at loci sensitive to salinity and temperature was widespread in the North Atlantic, suggesting that populations are following their own evolutionary trajectories at relatively small spatial scales (e.g., Nielsen, Hemmer-Hansen, Paulsen, et al. 2009). Similar results were also recently detected between Baltic and North Sea populations of European Flounder (Hemmer-Hansen et al. 2007). The authors found a weak cline at neutral SSR loci and a strong cline associated with temperature and salinity differences between collections from the 2 seas at a gene coding for heat shock protein (\(Hs70\)). In the present study, although no locus followed a pattern consistent with divergent selection, the number of loci was small relative to the total number of metabolic genes. Given the cost and labor of genotyping SSR loci, fewer loci are used in such studies as compared with high-throughput SNP assays. Only a small fraction of EST-SSR loci are expected to be under divergent selection (Ellis and Burke 2007), and comparative studies have shown that most EST-SSRs have been shown to behave as effectively neutral (Woodhead et al. 2005; 9 anonymous SSR vs. 10 EST-SSR loci) Kim et al. 2008 [17 anonymous SSR vs. 17 EST-SSR loci]. However, Yatake et al. (2007) did find increased prevalence of \(F_{\text{ST}}\) divergence outliers among EST-SSR (5 of 64 screened) than anonymous SSR loci (44 screened) and suggested that selection was working directly on the EST-SSR loci. Lazrek et al. (2009) found results suggestive of balancing selection at 7 of 18 predominantly EST-SSRs in a study on legumes. In the present study, although only a few outliers would be expected of the 13, the markers as a whole still serve to provide the most robust evaluation of population structure within bocaccio rockfish to date and allow exploration of trends among marker classes.

EST-SSRs displayed much lower variation relative to anonymous SSRs. Although relatively high variation in EST-SSRs compared with anonymous SSRs has been reported in several fishes (e.g., Yue et al. 2004; Coulibaly et al. 2005), low variation in EST-SSRs versus anonymous SSRs has been more commonly reported (e.g., Ellis and Burke 2007; Kim et al. 2008; Simko 2009). Evolutionary conservation of coding regions is typically invoked to explain the low variation because EST-SSR loci are expected to be under greater functional constraint than loci outside coding regions (Metzgar et al. 2000; Li et al. 2004). Purifying selection against deleterious polymorphisms may reduce

---

**Table 4** Mean repeat number (sample size) for SSR loci within EST or non-EST genomic regions, the difference in mean repeat number, 95% CI on true difference in mean repeat number, and \(P\) value associate with significance of the difference in mean repeat number

<table>
<thead>
<tr>
<th>Species</th>
<th>EST</th>
<th>Non-EST</th>
<th>Difference</th>
<th>95% CI true difference</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasterosteus aculeatus</td>
<td>18.4</td>
<td>24.0</td>
<td>-5.6</td>
<td>-6.0, -5.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Takifugu rubripes</td>
<td>18.2</td>
<td>27.3</td>
<td>-9.1</td>
<td>-11.3, -6.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Oryzias latipes</td>
<td>16.5</td>
<td>22.4</td>
<td>-5.9</td>
<td>-6.9, -4.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sebastes paucispinis</td>
<td>12.9</td>
<td>16.4</td>
<td>-3.5</td>
<td>7.0, 0.9</td>
<td>0.12</td>
</tr>
</tbody>
</table>
variation in regions linked to the polymorphism under selection (Charlesworth et al. 1993; Begun et al. 2007). In this report, we showed that variation in repeat number of the locus, which is known to affect mutation rates of SSR loci (Lai and Sun, 2003), was a sufficient explanation of the heterozygosity difference between marker classes. Low diversity was also accompanied by a low repeat number of EST-SSRs relative to anonymous SSRs in studies of sunflower (Yatabe et al. 2007) and fern (Woodhead et al. 2005). However, these studies did not attempt to directly account for the effect of repeat number on variation through GLM analysis. Low repeat numbers within EST-SSRs are likely a general phenomenon in fishes as EST-SSRs in G. aculeatus, T. rubripes, and O. latipes all had a lower repeat number than anonymous SSRs. It is likely that low diversity in EST-SSRs resulted from directional selection having reduced EST-SSR repeat number. The selective mechanism for how this occurred is unclear. However, given that EST-SSR loci from the present study produced levels of heterozygosity comparable to non-EST SSR loci with the same repeat numbers, recent selection against high variation is not likely.

No association between diversity and divergence levels was detected in the present study, but extremely high diversity levels are known to depress $F_{ST}$ (e.g., Buonaccorsi, McDowell, et al. 2001), and this effect has been implicated in a comparative study of EST-SSR to anonymous SSRs (Yatabe et al. 2007). The increasing availability of large pools of ESTs will allow researchers to select markers with repeat numbers that will produce markers of the desired variability. For population genetic studies, statistics such as Hedrick’s $G_{ST}$ (Hedrick 2005) that are robust to polymorphism levels can also be used to address high polymorphism bias.

We detected little evidence of population structure across the range of bocaccio rockfish. Hierarchical AMOVA showed that there was not a significant proportion of variation among geographic or year class groupings, regardless of which combinations of markers were tested. In contrast, allele frequency heterogeneity among the 6 collections was detected over all EST markers and over all markers using exact G-tests. Heterogeneity among individual collections combined with a lack of hierarchical structure suggests that a sizable fraction of the allele frequency heterogeneity detected was attributable to the difference among replicates within groups and does not represent stable genetic structure (Waples 1998). Significance of overall EST-SSR divergence was also largely a function of divergence found at a single locus, LBH. LBH is a conserved protein among vertebrates involved in early development of limbs and heart (Briegel and Joyner 2001; Ai et al. 2008). At this locus, the 2 southern California collections clustered together but showed higher similarity to northern than central collections. Two alleles were detected at this locus that were a single base-pair divergent from a perfectly repeating trinucleotide pattern. The region of LBH amplified contains both a trinucleotide and mononucleotide repeat, so the alleles appear valid. Because 1) divergence was not strong enough at this locus for a consistent geographic or year class pattern to emerge, 2) the repeat pattern was not straightforward, and 3) this was the sole source of population heterogeneity, we maintain that there is not enough evidence to reject the single homogeneous gene pool hypothesis for bocaccio rockfish. It may be advisable in future genetic studies of this species to include LBH and redesign primers to target its individual repeat motifs to reduce the possibility of homoplasy. The lack of genetic divergence in this study does not support subdivision of the species’ range into separate management units, but does not necessarily refute division as moderate levels of population exchange may be sufficient to homogenize allele frequencies (e.g., Buonaccorsi, Starkey, et al. 2001), yet may not be sufficient for populations to quickly recolonize extirpated areas. In particular, the region of low abundance off the Oregon coastline is reasonable evidence for demographic discontinuity, albeit not at levels low enough to be detected by the panel of markers used in the current study. The findings of this study are concordant with an early allozyme study of bocaccio stock structure from southern to northern California, which detected no significant allele frequency differences at 2 polymorphic loci (Wishard et al. 1980). A population structure analysis of bocaccio from Canada to southern California based on 7 SSR loci (including Sra.7-2 and Sra.7-25 from this study) detected very weak structure that was best partitioned into 3 or 4 geographical groupings. The most significant signal of population heterogeneity from that study was at the Sra.7-25 locus. Using our hierarchical approach, allele frequency heterogeneity among geographic groups at Sra.7-25 or among all loci were not close to significant in the present study, suggesting that the prior study may have reported a Type I error at this locus and overall.

**Conclusion**

We found striking differences between marker classes in diversity levels in *S. panzophilus* that are likely explained by functional constraints against high repeat number. Although differences in repeat number affected diversity, geographic divergence estimates were equivalent, and we were able to rule out action of divergent or balancing selection on EST-SSR polymorphism. A single locus showed significant heterogeneity among collections, but consistent hierarchical structuring of populations by geography or year class was not detected regardless of the large panel of SSR markers, large samples sizes, and broad geographic spread of collections in the present study. The fact that systematically lower heterozygosity in the EST-SSRs marker class (mean 0.5) did not bias divergence estimates should be encouraging to researchers targeting EST-SSRs with similar repeat numbers because moderately polymorphic loci are easier to score and multiplex, have fewer null alleles, and cross amplify across a broader spectrum of species than anonymous SSRs (Ellis and Burke 2007).
Funding

Internal funds from Juniata College and the Southwest Fisheries Science Center, National Marine Fisheries Service, La Jolla, CA, were used to complete this work.

References


Briegel KJ, Joyner AL. 2001. Identification and characterization of Lbh, implications for dispersal estimation in anadromous and marine fish populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171; Montpellier (France): Université de Montpellier II.


Received November 2, 2011; Revised December 21, 2011; Accepted January 22, 2012

Corresponding Editor: Jennifer Jackson