Highly accurate species identification of eastern Pacific rockfishes (*Sebastes* spp.) with high-throughput DNA sequencing

Diana S. Baetscher[1,2,#,*], Hayley M. Nuetzel[1,2,^], John Carlos Garza[1,2,3]

1. Department of Ocean Sciences, University of California, Santa Cruz, CA USA
2. Southwest Fisheries Science Center, National Marine Fisheries Service, Santa Cruz, CA USA
3. Institute of Marine Sciences, University of California, Santa Cruz, CA USA

#. Current affiliation: Alaska Fisheries Science Center, National Marine Fisheries Service, Juneau, AK USA

^. Current affiliation: Columbia River Intertribal Fisheries Commission, Portland, OR USA.

* Correspondence: 17109 Point Lena Loop Rd, Juneau, AK 99801;
diana.baetscher@noaa.gov

Keywords

Sebastes, species identification, microhaplotype, genetic stock identification, phylogeny, ascertainment bias

Abstract

Genetic species identification is often necessary for species flocks, such as rockfishes in the genus *Sebastes* (Teleostei, Scorpaenidae). Traditional visual identification methods are challenged by the presence of many sympatric rockfish species with morphologically similar juveniles. Here we present a straightforward approach for species identification in rockfishes using 96 nuclear microhaplotype loci that can be efficiently genotyped using high-throughput DNA sequencing. Self-assignment of nearly 1 000 samples from 54 species resulted in >99% accurate species identification at a 95% confidence threshold. Phylogenetic relationships of *Sebastes* uncovered with these same loci were highly concordant with relationships previously derived primarily with mitochondrial DNA. We also assessed ascertainment bias and consequent reduced nucleotide diversity and heterozygosity in non-ascertainment species to understand the potential utility of these markers for those species. The data and protocol presented here will be useful for research and management of rockfishes in the northeastern Pacific Ocean.

Introduction

Species identification is necessary when taxa that are the subject of study have closely related and morphologically similar congeners. Generally, visual identification is the first priority, as it is typically low-cost and rapid. However, it can be inaccurate, particularly for juvenile life stages, which often lack the morphological characteristics - especially color and pattern - used to identify adults. Genetic identification has emerged as a compelling alternative, by exploiting fixed nucleotide differences in specific gene regions between potential species of interest. Genetic identification approaches are commonly used both for ecological studies and for management of fisheries and wildlife. Some approaches further identify sampled individuals to population or family group. Genetic analyses provide such identifications through methods such as categorical assignment tests, mixed stock analysis, and parentage analysis (Pella and Milner 1987; Pella and Masuda 2001; Jones et al. 2003). These methods utilize nucleotide variation within species, employing differences in allele frequencies among distinct populations or stocks, and through segregating polymorphisms within populations (e.g., Abadía-Cardoso et al. 2013; Clemento et al. 2014).

In genetic studies of wild populations, it is possible to select a set of markers with the capacity to both distinguish the species identity of an individual, as well as its population or family of origin. In this scenario, the same genetic markers need to simultaneously contain fixed, or large frequency, differences between species and variation within a species, thereby providing inference at both levels of identification (e.g., Baetscher et al. 2019). However, ascertainment bias impacts the ability to use

57 genetic markers for this type of multi-level identification. Ascertainment bias occurs
58 because the initial assessment of genetic variation within a small number of samples
59 means that more common SNPs are more likely - and rare SNPs are less likely - to be
60 identified (Nielsen et al. 2004; Clark et al. 2005). When the ascertainment samples
61 consist of a small subset of the total number of species analyzed, ascertainment bias
62 suggests that only some of the variation from the initial SNP discovery samples will be
63 shared across species due to different demographic histories and rates of mutation (Li
64 and Kimmel 2013). One manifestation of this bias can be reduced variation in the
65 genetic markers commensurate with the evolutionary genetic distance between the taxa
66 used for marker discovery and other species of interest (Wakeley et al. 2001; Vowles
67 and Amos 2006). Another outcome is that markers may not amplify phylogenetically
68 distant species because of uncharacterized variation in the primer sites.

69 In marine fishes, few groups are as speciose as the rockfishes of the genus
70 *Sebastes*, which includes over 100 species globally, almost all of which are found
71 exclusively in the North Pacific Ocean (Love et al. 2002). Nearshore species are
72 abundant in kelp forests and are prominent in studies of ecology and community
73 structure along the west coast of North America (Carr 1991). Rockfishes also support
74 important commercial and recreational fisheries throughout their Northeastern Pacific
75 range from Alaska to Mexico, where some regulations do not differentiate among
76 species and others apply to species complexes, both because many species co-occur
77 and also to alleviate the need to identify each fish (e.g., the ”Other Rockfish” stock
78 complex in the Gulf of Alaska; Tribuzio et al. 2017). Despite this inconsistent regulatory
79 framework, adult rockfishes can be accurately identified in most cases based on

morphometric characteristics; however, juveniles and cryptic species are frequently misidentified (Butler et al. 2012).

Mitochondrial DNA (mtDNA) data suggest that *Sebastes* arose during the middle Miocene in the Northwest Pacific and quickly diversified and dispersed into habitats produced by high-latitude cooling and upwelling systems throughout the North Pacific (Hyde and Vetter 2007). Originally, phylogenetic relationships among rockfishes were defined by morphologic and meristic characters, with genetic data - specifically mitochondrial DNA - incorporated by the early 2000s (see Kendall 2000 for a comprehensive review). Closely related species have been the subject of recent genetic studies, which have identified cryptic species where adult specimens are morphologically similar and sometimes indistinguishable (Orr and Blackburn 2004; Gharrett et al. 2005; Burford and Bernardi 2008, Orr and Hawkins 2008; Hyde and Vetter 2009, Hess et al. 2013; Frable et al. 2015). These discoveries of cryptic species have coincided with increased genetic monitoring of rockfish populations for commercial and recreational groundfish fisheries and population assessments (Orr and Blackburn 2004; Berntson and Moran 2009). Previous research used mitochondrial and nuclear loci to genetically identify rockfish species (Hyde and Vetter 2007; Pearse et al. 2007). In this study, we describe a new protocol for genetic species identification of rockfishes, including almost all lineages found in the California Current and Gulf of Alaska Large Marine Ecosystems. We efficiently assay 96 nuclear microhaplotype loci using high-throughput DNA sequencing of amplicons and provide almost perfect species identification in this group of fishes. The genetic markers (described in Baetscher et al. 2018; 2019) were discovered using double-digest restriction-site associated DNA

sequencing (ddRADseq; Peterson et al. 2012) of *S. atrovirens* (kelp rockfish) samples. Initially, these markers were selected for the high heterozygosity necessary for identifying family relationships within *S. atrovirens* and its sympatric close relatives *S. chrysomelas* (black-and-yellow rockfish) and *S. carnatus* (gopher rockfish). Additionally, the markers were designed for multiplexed analysis using next-generation DNA sequencing of amplicons, which allows researchers to generate genotype data for hundreds-to-thousands of fish in a single sequencing reaction. Given that some collection techniques employed to sample juvenile rockfishes can capture hundreds of fish in a single sampling event (Ammann 2004), a high-throughput method is particularly useful." In the approach we describe here, we conducted species identification using genetic assignment tests. Such assignment tests are employed to determine the likelihood that a sample originates from one or more populations based on allele frequencies derived from reference samples taken from those populations (Paetkau et al. 1995; Rannala and Mountain 1997). Self-assignment provides a metric of how well a particular set of genetic markers can differentiate among taxa when the identity of the true taxon is known. Intuitively, the accuracy of assignment tests is limited by the biology and life history of the organism – species with high gene flow have populations that are more difficult to differentiate, and require high-resolution genetic data, whereas species with almost no gene flow have populations that are typically easily discriminated using a sufficient number of polymorphic loci. Given that our study involved classifying species rather than populations, we anticipated identifying samples to true species with high accuracy, assuming little-to-no ongoing gene flow among species.

Genotype data generated for testing species assignment allowed us to estimate phylogenetic relationships of more than 50 rockfish taxa using nuclear DNA markers and compare these results with a previously published phylogeny for *Sebastes* based on seven mtDNA and two nuclear genes (Hyde and Vetter 2007). Depending on the evolutionary history of the organism, nuclear and mtDNA genes can produce discrepant signals of diversification (Shaw 2002; Chan and Levin 2005) and, thus, comparing the nuclear phylogeny against patterns derived in large part from mtDNA highlights areas where the two marker types depict inconsistent relationships. A recent phylogenetic study of six rockfish species using nuclear markers provides us with a comparison for the subgenus Sebastosomus (Wallace et al. 2022). Furthermore, we describe phylogenetic relationships for a recently described cryptic species relevant to our geographic study region.

Phylogenetic relationships help to contextualize the low levels of heterozygosity and nucleotide diversity for species not included in our marker ascertainment process and allow us to assess this ascertainment bias based on evolutionary genetic distance. Reduced heterozygosity diminishes the utility of these markers for intraspecific genetic analyses, including population structure and pedigree inference, even for species within the same subgenus as the ascertainment species. This work describes a valuable analysis tool for research of rockfishes when confident species identity is required, an examination of phylogenetic relationships across the genus, and insight into how nucleotide diversity rapidly declines in species not included in the marker discovery process.

Methods

Samples

Samples from adults of 54 species of rockfishes (*Sebastes*) and cabezon (*Scorpaenichthys marmoratus*), the sister species of the genus *Sebastes*, were obtained by trawl and hook-and-line fishing. Rockfishes were identified morphologically by experts from the NOAA Southwest Fisheries Science Center or researchers at the University of California Santa Cruz. For the majority of samples, DNA was extracted from fin tissue using DNeasy 96 Blood & Tissue kits on a BioRobot 3000 (Qiagen, Inc.), eluted into 200 µL, with extracts stored at 4° C. For species with few adult samples available, DNA was extracted from juvenile samples as described. A small number of samples were received as previously extracted DNA and stored at 4° C prior to sequencing library preparation.

Genotyping and analysis

Samples were genotyped with a set of 96 microhaplotype markers ascertained in *S. atrovirens, S. carnatus* and *S. chrysomelas* using the Genotyping-in-Thousands by sequencing (GT-seq; Campbell et al. 2015) protocol, as modified by Baetscher et al. (2018). The amplicon-sequencing library preparation includes an initial multiplex PCR step to amplify target loci and a second PCR to add sequencing adapters and barcodes for identifying samples. Normalized libraries were sequenced using 2 x 75 bp paired-end approach on a MiSeq instrument (Illumina, Inc.). Raw sequence reads were sorted by individual barcode using the MiSeq Analysis Software (Illumina), and then paired

168 reads were combined and mapped to a reference using the bioinformatic workflow in
169 Baetscher et al. (2018). Variants were called across samples using FREEBAYES
170 (Garrison and Marth 2012) and the output variant call format (VCF) files were filtered for
171 quality (minQ = 30; minDP = 10) and merged using VCFTOOLS (Danecek et al. 2011).
172 In microhaplotypes, multiple single nucleotide polymorphisms (SNPs) segregate
173 together within a single sequencing read and do not require statistical phasing
174 (Stephens and Donnelly 2003), which makes it relatively straightforward to call
175 individual haplotypes from mapped data files and the combined VCF file using the
176 software program MICROHAPLOT (Ng and Anderson 2016). Resulting genotypes were
177 filtered in R (R Core Team 2016) using a minimum threshold of 20 reads per
178 individual/locus and a minimum read depth ratio of 0.4, which applies to heterozygotes
179 and is a measure of the number of reads of the second most common allele divided by
180 the read depth of the most common allele. Loci with high rates of missing data or
181 deviations from Hardy-Weinberg equilibrium (HWE) were removed and then samples
182 with missing data at more than 25 of the remaining loci were dropped from further
183 analysis. This missing data threshold was intentionally liberal to avoid removing
184 samples of species in which a larger proportion of loci failed to amplify due to
185 uncharacterized variation in the primer sites (Fig. S1). Such variation is more common
186 in genetic markers applied to species that are phylogenetically distant from the
187 ascertainment species due to different demographic histories and rates or
188 directionalities of mutation.

189 Since juvenile rockfishes are commonly misidentified, only genotypes for adults
190 were included, except for species in which we had fewer than five adult samples and

191 samples from juveniles were available. The veracity of the species identity for these
192 juvenile samples was evaluated by the self-assignment analysis (see below). A
193 maximum of 32 individuals per species was included, when available, to generate a
194 dataset with a representative estimate of assignment accuracy across the genus. The
195 data set was tested for deviations from HWE using the R package PEGAS (Paradis
196 2010) and pairwise F_{ST} was calculated with heterozygosity weighted by group size, also
197 in R using HIERFSTAT (Goudet 2005).

198 Genetic assignment

199 Genetic self-assignment was conducted in the R package RUBIAS (Moran and
200 Anderson 2018) using the leave-one-out self-assignment function with default allele
201 frequency prior. Leave-one-out procedures remove the gene copies for each sample
202 from the allele counts of its known population/taxon of origin before calculating the
203 likelihood that the sample came from that population, in order to avoid overestimating
204 assignment accuracy. RUBIAS provides a likelihood for each sample assigning to every
205 reference population and a z-statistic for each sample assignment. The z-statistic is the
206 difference (in number of standard deviations) between the observed log-probability of an
207 individual's genotype given it came from a specific population, and the log-probability
208 expected for an individual from that population. The mean and standard deviation of the
209 expected log-probability values are computed by RUBIAS using the locus-specific allele
210 frequencies and the assumption of HWE. When the probability of assignment is high for
211 a given reference population but the z-statistic is outside the expected range (<-3 or
212 >3), this can be an indication that the sample belongs to a population that is not

included in the reference dataset. In an effort to ensure that only samples that were confidently identified to true species were included, any samples that were assigned to a reference taxon with a z-statistic <-3 or >3 were excluded from the final dataset.

Phylogenetic analyses

Samples verified by self-assignment were used to construct phylogenetic trees. To generate consensus sequence data for building trees, species-specific VCF files were produced by FREEBAYES and then a consensus FASTA file for each species was created using VCFTOOLS (Danecek et al. 2011). A member of the sister genus to *Sebastes*, *Scorpaenichthys marmoratus* (cabezon) was used to root the phylogenetic trees. Loci in each species-consensus FASTA file were concatenated with the GENEIOUS software program (v 7.1.7; Kearse et al. 2012) before export to MUSCLE (Edgar 2004) with alignments output in ClustalW format. These were then used as input for MEGA (v. 7.0.26; Kumar et al. 2016) to generate maximum-likelihood trees using the General Time Reversible (GTR) model (Nei and Kumar 2000) with 1 000 bootstrap replicates, which was consistent with the model used by Hyde and Vetter (2007) for their *Sebastes* phylogeny. A similar analysis was performed to generate an unrooted maximum-likelihood tree, without cabezon, also using the GTR model and 1 000 bootstrap replicates.

For the Bayesian analysis, FASTA alignments were converted to Nexus format using PGDSpider (v. 2.1.1.5; Lischer and Excoffier 2012), and then used as input for MRBAYES (v. 3.2; Huelsenbeck and Ronquist 2001). Parameters included a GTR substitution model and one million generations, where generation time was increased

experimentally until the standard deviation of split frequencies dipped below 0.01 and the Potential Scale Reduction Factor (PSRF) converged to 1. This included a uniform Dirichlet prior (1,1,1,1) and 25% burn-in with sampling from the posterior every 5000 generations. Phylogenetic trees generated by this analysis were visualized using FigTree (v 1.4.3; Rambaut 2016).

Because the marker set was designed using data from *S. atrovirens*, *S. chrysomelas*, and *S. carnatus* based on the variability in those species, the amount of variation in other species was expected to be affected due to ascertainment bias. This bias was quantified as the decrease in mean internal heterozygosity and nucleotide diversity for each species with increasing genetic distance from *S. atrovirens*. Genetic distance was calculated in MEGA using a variety of model settings to determine the extent to which estimates of genetic distance in these data are sensitive to model choice (Fig. S2). Nucleotide diversity was calculated per variant site for each species in VCFTOOLS and then the sum of all sites within a species was divided by the total number of bases in the 96 loci to account for invariant sites.

Results

Genotyping and data analysis

A total of 997 rockfish samples from 54 species were genotyped and analyzed with a VCF file that had previously been generated from 1 690 rockfish samples and contained 4,322 variant sites from all species (Baetscher 2019; Table S1). Five loci (Sat_914,

Sat_934, Sat_1399, Sat_1871, Sat_2513) with large amounts of missing data across > 35% of species and one locus (Sat_1166) with three or more haplotypes per individual in some species, suggestive of a paralogous locus, were removed. Only genotypes that passed filtering thresholds for read depth, allelic ratio, and missing data were retained for analyses. In three species (*S. reedi, S. wilsoni*, and *S. crameri*) with fewer than two adult samples available, genotypes from juveniles were included. The number of samples per species ranged from two (*S. rufinanus*) to 32 (*S. atrovirens*; Table 1).

The majority of species-by-locus combinations conformed to HWE; however, the six species with the greatest number of deviations (> 10 loci out of HWE), were *S. rosaceus* (18 loci), *S. carnatus* (15 loci), *S. chrysomelas* (13 loci), *S. ensifer* (13 loci), *S. diaconus* (11 loci), and *S. mystinus* (10 loci). Thirty percent of the loci were out of HWE in four of the 54 species, and three loci, Sat_770, Sat_875, and Sat_2178, were out of HWE in 8-13 species. Pairwise F_{ST} ranged from 0.015 between *S. carnatus* and *S. chrysomelas* to 0.746 between *S. levis* and *S. entomelas* (mean F_{ST} = 0.45, s.d. = 0.13).

Self-assignment

Self-assignment resulted in 98.3% accuracy at a scaled-likelihood value of 0.95, and all mis-assigned individuals at > 50% likelihood were either *S. carnatus* assigning to *S. chrysomelas*, or vice versa. These assignment results indicated that this set of genetic markers cannot consistently distinguish between *S. carnatus* and *S. chrysomelas* and that a single genetic reporting group would be appropriate for assignment.

The self-assignment analysis was performed again after creating a single *S. carnatus/chrysomelas* reporting group and 100% of samples were correctly assigned at a 50% scaled-likelihood threshold. At the 95% confidence level, assignment accuracy was 99.2% and all lower confidence assignments were *S. carnatus* or *S. chrysomelas* samples that assigned to the joint reporting group, but at a scaled-likelihood below 95%.

Phylogenetic trees

Species relationships were elucidated with maximum-likelihood and Bayesian trees. Both rooted trees (Fig. 1, Fig. S3) and an unrooted tree (Fig. S4) recovered very similar phylogenetic relationships. Branch support on the Bayesian tree was generally higher than for the maximum-likelihood trees, which had consistent bootstrap values, but with slight differences at some of the deeper nodes. Some of the most confident relationships in the Bayesian tree included the position of *S. atrovirens* clustered with members of the *Pteropodus* subgenus, as well as that *S. saxicola* and *S. semicinctus* appeared proximate to *Pteropodus* and distant from other members of the subgenus *Allosebastes* (Fig. 1). Monophyletic relationships among taxa within the subgenus *Sebastomus* garnered strong support with the exception of *S. rufus*, which groups with the subgenus *Acutomentum* (Fig. 1). While the branch support for these phylogenetic positions varied between the maximum-likelihood and Bayesian analyses, the overall pattern among these subgenera appeared consistent.

Ascertainment bias

Observed heterozygosity in most species declined substantially when compared to *S. atrovirens*, with a smaller decrease in *S. chrysomelas* and *S. carnatus* (mean for *S. atrovirens, chrysomelas, carnatus* = 0.423, overall mean = 0.130; range = 0.012-0.458; Fig. 2). Nucleotide diversity sharply declined with genetic distance from *S. atrovirens* (Fig. 3), with low levels of diversity even in species in the same subgenus as *S. atrovirens*. Genetic distance was calculated as pairwise differences since a comparison indicated that nucleotide substitution model does not substantially alter distance estimates for this dataset (Fig. S2). Based on these results, over 80% of species analyzed in this study contained less than half of the nucleotide diversity of *S. atrovirens* over a genetic distance of fewer than 0.04 base differences per site for 10 695 total sites, excluding gaps and missing data.

Discussion

Here we demonstrate the high accuracy (>99% correct assignment) of a set of short haplotypic markers for identifying 54 species of the genus *Sebastes*, including all of the species commonly found in the California Current Large Marine Ecosystem along the Pacific coast of North America. Using these loci, we distinguish between closely related and recently described cryptic species, describe phylogenetic relationships, and quantify a decrease in the heterozygosity and nucleotide diversity of these genetic markers in species with increasing evolutionary genetic distance from the ascertainment species.

317 Ecological studies and management of fisheries require efficient methods to
318 conclusively identify sympatric marine species, particularly at the larval and juvenile
319 stages. In rockfishes, planktonic larvae from many species coexist during their pelagic
320 phase and remain challenging to identify morphologically as they recruit to settlement
321 habitats (Butler et al. 2012). Even as adults, the number of species present in
322 overlapping habitats, the presence of cryptic species (e.g., Frable et al. 2015), and
323 subtle differences in coloration or morphology (Ingram and Kai 2014) underscore the
324 need for genetic species identification. Previous marker types have been used for this
325 task; one such study included 33 species with 97.4% assignment accuracy (Pearse et
326 al. 2007), and the other, a much more complete survey of the genus, genotyped 103
327 individuals from 101 species at seven mitochondrial and two nuclear genes, but did not
328 test these data for genetic assignment accuracy (Hyde and Vetter 2007). Our method of
329 genotyping fewer than 100 multiplexed microhaplotype loci with high-throughput DNA
330 sequencing is highly accurate, efficient for large sample sizes and can be coupled with
331 a reproducible analysis workflow based on the reference database for species
332 assignment generated by this study.

333 Self-assignment using genotype data from 90 retained microhaplotype markers
334 accurately identified the true species identity of every sample for all 54 species, with the
335 exception of two extremely closely related species. At a stringent likelihood threshold (>
336 95%), eight samples of *S. carnatus* and *S. chrysomelas* assigned to the combined *S.*
337 *carnatus/chrysomelas* group at a lower level of confidence, but still above a 50%
338 scaled-likelihood. Notably, these sister species have been the subject of ongoing
339 research (Narum et al. 2004; Buonaccorsi et al. 2011) and our results from the self-

assignment demonstrate the challenge of separating the two groups with existing genetic markers and call into question their taxonomic status as two distinct species.

Coincidentally, *S. carnatus/chrysomelas* are also the most phylogenetically proximate to the primary ascertainment species (*S. atrovirens*; Fig. 1; Fig. S3, Fig. S4), and with nearly as much variation in these loci (Fig. 2, Fig. 3). And while these genetic markers easily differentiate juvenile-stage cryptic species (e.g., *S. mystinus/diaconus, S. aleutianus/melanostictus*) and those commonly misidentified even as adults (e.g., *S. flavidus/serranoides*), they underperform for *S. carnatus/chrysomelas*. This indicates that these taxa are more genetically similar than every other pair of sister species included in our dataset, at least in the portion of the genome surveyed with these loci, consistent with the lowest pairwise F_{ST} value (0.015) in the study. Previous work on *S. carnatus* and *S. chrysomelas* identified a single, highly diverged locus and concluded that the pair is likely in the final stages of speciation, but with ongoing gene flow (Narum et al. 2004, Buonaccorsi et al. 2011). A more recent investigation using reduced-representation and whole genome resequencing found three distinct genomic regions with elevated divergence and variation in genes pointing to the importance of coloration and vision (Behrens et al. 2021). Results from these studies are consistent with the general idea that speciation mechanisms in rockfishes likely involve both allopatric and sympatric processes, including habitat differentiation associated with depth gradients (Ingram 2011) and mate choice reinforced by internal fertilization (Buonaccorsi et al. 2011).

Previously described rockfish species relationships relied heavily on mitochondrial DNA data (Kai et al. 2003; Li et al. 2006; Hyde and Vetter 2007; Li et al. 2007), providing an opportunity to apply the nuclear markers from this study to estimate phylogenetic relationships for comparison (Fig. 1, Fig. S3, Fig. S4). Rooted and unrooted maximum-likelihood trees produced consistent topologies with very similar branch support, although some deeper nodes in the unrooted tree garnered higher support, while other nodes were better supported in the rooted tree (Fig. 1, Fig. S4). High confidence nodes in the Bayesian tree were generally well supported in the maximum-likelihood tree, with most differences occurring at nodes with lower support, such as the position of either *S. alutus* or *S. borealis* in a clade with *S. melanostictus* and *S. aleutianus* (Fig. 1, Fig. S3). Few instances of well-supported Bayesian relationships deviate from the maximum-likelihood tree, although *S. polyspinis* presents one such case. The Bayesian tree topology from our data is the most appropriate for comparison with the phylogeny in Hyde and Vetter (2007) since the analyses are equivalent and, although Bayesian methods can overestimate node support, bootstrapped maximum-likelihood values may be overly conservative (Douady et al. 2003).

Most relationships remain consistent between the microhaplotype tree topologies and the more complete *Sebastes* tree from Hyde and Vetter (2007). Although Hyde and Vetteranalyze species that are absent from our dataset, primarily from the northwest Pacific and North Atlantic, we analyze representatives from each major clade with the exception of the subgenera *Sebastocles* and *Mebarus*, whose constituents are exclusively in the northwest Pacific, with the exception of *S. atrovirens* which should

clearly be included in the *Pteropodus* subgenus. Generally, we find very high concordance with Hyde and Vetter (2007) at the subgeneric level. Areas in which the microhaplotype tree (Fig. 1) deviates from their tree include clade ”D” nesting within *Pteropodus*, and members of *Eosebastes*, *S. aurora* and *S. diploproa*, nesting within *Sebastichthys*. At the species level, more variation exists. For example, both trees depict close phylogenetic relationships among *S. atrovirens, S. carnatus*, and *S. chrysomelas*, with the microhaplotype tree placing *S. maliger* as a closer relative of the three species than *S. caurinus*, as in the mitochondrial tree. Other small differences in the topologies include strong support that *S. melanops* is more closely related to *S. flavidus* than *S. serranoides (a relationship also identified by Wallace et al. 2022)*; and that *S. goodei* is more closely related to *S. paucispinis* than to *S. jordani*. We also show that *S. diaconus* and *S. mystinus* are easily distinguished and nearest neighbors in the phylogeny, which is unsurprising since these species were only recently described as separate taxa (Frable et al. 2015; Wallace et al. 2022).

Taxonomy of rockfishes, particularly of subgenera, has been and continues to be dynamic, as highlighted by multiple revisions of subgeneric classifications (Love et al. 2002). For example, *S. diploproa* is part of the subgenus *Sebastichthys* in Kendall (2000), who cites Eigenmann and Beeson (1894), but Li et al. (2006) designate *S. diploproa* as a member of *Allosebastes*, attributed to Gilbert (1890). Phylogenetic relationships described by the microhaplotype data are generally consistent with mitochondrial data and support polyphyly of generally accepted subgenera, including *Acutomentum*, *Allosebastes*, and *Sebastosomus* (Hyde and Vetter 2007; Li et al. 2007). A formal re-description of these subgenera would alleviate some of the taxonomic

407 confusion but comprehensive taxonomic revision would require data from more species
408 in the genus than are included in this study.

409 The set of nearly 100 microhaplotype loci target substantial variation in the
410 ascertainment species, *S. atrovirens*, *S. carnatus,* and *S. chrysomelas* (Baetscher et al.
411 2018; 2019) and contain a similar amount of variation in a closely related taxon (*S.*
412 *maliger*). However, variation declines rapidly with increasing genetic distance (Fig. 3),
413 even for members of the *Pteropodus* subgenus. Such reduced variation has been
414 documented in studies of ascertainment bias in microsatellite loci across multiple
415 genera (Vowles and Amos 2006). Even so, the ascertainment bias we observe here is
416 even more significant than previously observed, with dramatically decreased nucleotide
417 diversity over relatively small evolutionary genetic distances, with only the most closely
418 related species to those included in the marker discovery process found to have
419 substantial variation (Fig. 3). The surprising amount of variation in *S. rosaceus* and *S.*
420 *ensifer*, despite their evolutionary distance from *Pteropodus*, might be explained by
421 cryptic structure in those species, as indicated by the relatively high number of loci that
422 deviated from HWE. However, selectively removing loci for individual species would be
423 challenging with the > 50 species included in this analysis.

424 Although the relatively low observed heterozygosity found in this set of markers
425 for the majority of species analyzed here suggest limited utility for purposes other than
426 species identification (e.g., pedigree reconstruction), the amplicon library preparation
427 protocol is highly flexible and enables researchers to add additional loci or swap out
428 markers that would increase power for species of particular interest. Such an effort

could bolster this set of markers for population genetic structure or pedigree analyses in additional species, and previous research has shown that genotyping samples with a single set of genetic markers to both identify species and analyze pedigree relationships is an economical approach (Baetscher et al. 2019).

Here, we describe an efficient method for genotyping and analyzing genetic data to identify species of rockfishes, particularly for taxa commonly captured together as juveniles. The genetic markers we employ, and our subsequent analytical workflow, provide highly accurate species identification and estimates of phylogenetic relationships largely consistent with previous genetic data. In addition, we describe a flexible protocol for modifying the set of target loci and accounting for ascertainment bias to suit the specific needs of a variety of ecological studies and fisheries management objectives.

Acknowledgements

We thank C. Columbus, E. Campbell, E. Correa and E. Gilbert-Horvath for laboratory assistance. E. Anderson, A. Clemento, C. Edwards, and M. Carr contributed to discussions and provided helpful comments on the manuscript. This work was supported by a grant from the National Science Foundation (Award number 1260693; PIs: M.H. Carr, E.C. Anderson, C. Edwards and J.C. Garza).

References

Abadía-Cardoso, A., Anderson, E.C., Pearse, D.E., and Garza, J.C. 2013. Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (Oncorhynchus mykiss). Mol. Ecol. 22: 4733–4746. doi:10.1111/mec.12426.

Ammann, A.J. 2004. SMURFs: Standard monitoring units for the recruitment of temperate reef fishes. J. Exp. Mar. Biol. Ecol. 299: 135–154. doi:10.1016/j.jembe.2003.08.014.

Baetscher, D.S. 2019. Larval dispersal of nearshore rockfishes. Doctoral dissertation, Department of Ocean Sciences, University of California, Santa Cruz, CA.

Baetscher, D.S., Anderson, E.C., Gilbert-Horvath, E.A., Malone, D.P., Saarman, E.T., Carr, M.H., and Garza, J.C. 2019. Dispersal of a nearshore marine fish connects marine reserves and adjacent fished areas along an open coast. Mol. Ecol. 28: 1611–1623. doi:10.1111/mec.15044.

Baetscher, D.S., Clemento, A.J., Ng T.C., Anderson, E.C., and Garza J.C. 2018. Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. Mol. Ecol. Resour. 18: 296–305. doi:10.1111/1755-0998.12737.

Behrens, K.A., Girasek, Q.L., Sickler, A., Hyde, J., & Buonaccorsi, V.P. 2021. Regions of genetic divergence in depth-separated Sebastes rockfish species pairs: Depth as a potential driver of speciation. Mol. Ecol. 30, 4259– 4275. https://doi.org/10.1111/mec.16046.

Berntson, E.A., and Moran, P. 2009. The utility and limitations of genetic data for stock identification and management of North Pacific rockfish (Sebastes spp.). Rev. Fish Biol. Fish. 19: 233–247. doi:10.1007/s11160-008-9101-2.

Buonaccorsi, V.P., Narum, S.R., Karkoska, K.A., Gregory, S., Deptola, T., and Weimer A.B. 2011. Characterization of a genomic divergence island between black-and-yellow and gopher Sebastes rockfishes. Mol. Ecol. 20: 2603–2618. doi:10.1111/j.1365-294X.2011.05119.x.

Burford, M.O., and 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. doi: 10.1007/s00227-008-0963-6

Butler, J.L., Love, M.S., and Laidig, T.E. 2012. A guide to the rockfishes, thornyheads, and scorpionfishes of the Northeast Pacific. University of California Press, Berkeley and Los Angeles, CA.

Campbell, N.R., Harmon, S.A., and Narum, S.R. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Resour. 15: 855–867. doi:10.1111/1755-0998.12357.

Carr, M.H. 1991. Habitat selection and recruitment of an assemblage of temperate zone reef fishes. J. Exp. Mar. Biol. Ecol. 146: 113–137. doi:10.1016/0022-0981(91)90257-W.

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., and Nielsen, R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 15: 1496–1502. doi:10.1101/gr.4107905.

Clemento, A.J., Crandall, E.D., Garza, J.C., and Anderson, E.C. 2014. Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Chinook salmon (Oncorhynchus tshawytscha) in the California Current Large Marine Ecosystem. Fish. Bull. 112: 112–130. doi:10.7755/FB.112.2-3.2.

Chan, K.M.A., and Levin, S.A. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution 59: 720–729. doi:10.1111/j.0014-3820.2005.tb01748.x

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156–2158. doi:10.1093/bioinformatics/btr330.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., and Douzery, E.J.P. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20: 248–54. doi:10.1093/molbev/msg042.

Edgar, R.C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797. doi:10.1093/nar/gkh340.

Frable, B.W., Wagman, D.W., Frierson, T.N., Aguilar, A., and Sidlauskas, B.L. 2015. A new species of Sebastes (Scorpaeniformes: Sebastidae) from the northeastern Pacific, with a redescription of the blue rockfish, S. mystinus (Jordan and Gilbert, 1881). Fish. Bull. 113: 355–377. doi:10.7755/FB.113.4.1,

Garrison, E., and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907v2, 9.

Gharrett, A.J., Matala, A.P., Peterson, E.L., Gray, A. K., Li, Z., and Heifetz, J. 2005. Two genetically distinct forms of rougheye rockfish are different species. Trans. Am. Fish. Soc. 134: 242–260. doi:10.1577/T04-055.1.

Goudet, J. 2005. HIERFSTAT: a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5: 184-186. doi:10.1111/j.1471-8286.2004.00828.x.

Hess., J.E., Chittaro, P., Elz, A., Gilbert-Horvath, E.A., Simon, V., and Garza, J.C. 2013. Cryptic population structure in the severely depleted cowcod, Sebastes levis. Canadian Journal of Fisheries and Aquatic Sciences. 71:(1): 81-92. doi: 10.1139/cjfas-2012-0510.

Huelsenbeck, J.P., and Ronquist F. 2001. MRBAYES; Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755. doi:10.1093/bioinformatics/17.8.754.

Hyde, J.R., and Vetter, R.D. 2007. The origin, evolution, and diversification of rockfishes of the genus Sebastes (Cuvier). Mol. Phylogenet. Evol. 44: 790–811. doi:10.1016/j.ympev.2006.12.026.

547 Hyde, J.R., and Vetter, R.D. 2009. Population genetic structure in the redefined vermilion
548 rockfish (Sebastes miniatus) indicates limited larval dispersal and reveals natural management
549 units. Canadian Journal of Fisheries and Aquatic Sciences. 66(9): 1569-1581.doi: 10.1139/F09-
550 104
551
552 Ingram, T 2011. Speciation along a depth gradient in a marine adaptive radiation. Proc. R. Soc.
553 B. 278: 613–618. doi:10.1098/rspb.2010.1127.
554
555 Ingram, T., and Kai Y. 2014. The geography of morphological convergence in the radiations of
556 Pacific Sebastes rockfishes. Am. Nat. 184: E115–E131. doi:10.1086/678053.
557
558 Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers.
559 Bioinformatics 24: 1403–1405. doi:10.1093/bioinformatics/btn129.
560
561 Jones, A.G., and Ardren, W.R. 2003. Methods of parentage analysis in natural populations. Mol.
562 Ecol. 12: 2511–2523. doi:10.1046/j.1365-294x.2003.01928.x.
563
564 Jukes, T.H., and Cantor, C.R. 1969. Evolution of protein molecules. In: Mammalian protein
565 metabolism, pp. 21–132. Academic Press, New York, N.Y.
566
567 Kai, Y., Nakayama, K., and Nakabo T. 2003. Molecular phylogenetic perspective on speciation
568 in the genus Sebastes (Scorpaenidae) from the Northwest Pacific and the position of Sebastes
569 within the subfamily Sebastinae. Ichthyological Res. 50: 239–244. doi:10.1007/s10228-003-
570 0163-9.
571
572 Kearse, M., Moir, R., Wilson, A. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M.,
573 Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes,
574 P., and Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software
575 platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649.
576 doi:10.1093/bioinformatics/bts199.
577
578 Kendall, A.W. 1991. Systematics and identification of larvae and juveniles of the genus
579 Sebastes. Environ. Biol. Fishes 30: 173–190. doi:10.1007/BF02296888.
580
581 Kendall, A.W. 2000. A historical review of Sebastes taxonomy and systematics. Mar. Fish. Rev.
582 62: 1–23.
583
584 Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions
585 through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111–120.
586 doi:10.1007/BF01731581.
587
588 Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics
589 Analysis Version 7.0 for bigger datasets. Mol. Biol. Evol. 33: 1870–1874.
590 doi:10.1093/molbev/msw054.
591
592 Li, Z., Gray, A.K., Love, M.S., Goto, A., and Gharrett, A.J. 2007. Are the subgenera of Sebastes
593 monophyletic? Biology, Assessment, and Management of North Pacific Rockfishes 23: 185–
594 206.
595

Li, Z., Gray, A.K., Love, M.S., Asahida, T., and Gharrett, A.J. 2006. Phylogeny of members of the rockfish (Sebastes) subgenus Pteropodus and their relatives. Can. J. Zool. 84: 527–536. doi:10.1139/Z06-022.

Li, B., and Kimmel., M. 2013. Factors influencing ascertainment bias of microsatellite allele sizes: impact on estimates of mutation rates. Genetics: 195: 563-572. doi: 10.1534/genetics.113.154161

Lischer, H.E.L., and Excoffier, L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28: 298–299. doi:10.1093/bioinformatics/btr642.

Moran, B.M., and Anderson, E.C. 2018. Bayesian inference from the conditional genetic stock identification model. Can. J. Fish. Aquat. Sci. 76: 551–560. doi:10.1139/cjfas-2018-0016.

Narum, S.R., Buonaccorsi, V.P., Kimbrell, C.A., and Vetter, R.D. 2004. Genetic divergence between gopher rockfish (Sebastes carnatus) and black and yellow rockfish (Sebastes chrysomelas). Copeia 4: 926–931. doi:10.1643/CG-02-061R2.

Nielsen, R., Hubisz, M.J., and Clark, A.G. 2004. Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. Genetics 168: 2373–2382. doi:10.1534/genetics.104.031039.

Ng T.C., and Anderson E.C. 2016. MICROHAPLOT. R software. https://cran.r-project.org/web/packages/microhaplot/index.html

Orr, J.W., and Blackburn, J.E. 2004. The dusky rockfishes (Teleostei: Scorpaeniformes) of the North Pacific Ocean: Resurrection of Sebastes variabilis (Pallas, 1814) and a redescription of Sebastes ciliatus (Tilesius, 1813). Fish. Bull. 102: 328–348.

Orr, J.W., and Hawkins, S. 2008. Species of the rougheye rockfish complex: resurrection of Sebastes melanostictus (Matsubara, 1934) and a redescription of Sebastes aleutianus (Jordan and Evermann, 1898) (Teleostei: Scorpaeniformes). Fish. Bull. 106: 111–134.

Paetkau, D., Calvert, W., Stirling, I., and Strobeck, C. 1995) Microsatellite analysis of population structure in Canadian polar bears. Mol. Ecol. 4: 347–354. doi:10.1111/j.1365-294x.1995.tb00227.x.

Paradis, E. 2010. Pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26: 419–420. doi:10.1093/bioinformatics/btp696.

Pearse, D.E., Wooninck, L., Dean, C.A., and Garza, J.C. 2007. Identification of northeastern Pacific rockfish using multilocus nuclear DNA genotypes. Trans. Am. Fish. Soc. 136: 272–280. doi:10.1577/T06-051.1.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. 2012. Double Digest RADseq: an inexpensive method for *De Novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7(5): e37135. doi:10.1371/journal.pone.0037135

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambaut, A. 2016. FigTree, version 1.4.3. Computer program distributed by the author, URL http://tree.bio.ed.ac.uk/software/figtree.

Rannala, B., and Mountain, J.L. 1997. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. U.S.A. 94: 9197–9201. doi:10.1073/pnas.94.17.9197.

Shaw, K.L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc. Natl. Acad. Sci. U.S.A. 99: 16122–16127. doi:10.1073/pnas.242585899.

Stephens, M., and Donnelly, P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73: 1162–1169. doi:10.1086/379378.

Tajima, F., and Nei, M. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. 1: 269–285. doi:10.1093/oxfordjournals.molbev.a040317.

Tamura, K., and Kumar, S. 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol. Biol. Evol. 19: 1727–1736. doi:10.1093/oxfordjournals.molbev.a003995.

Tamura, K., and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512–526. doi:10.1093/oxfordjournals.molbev.a040023.

Tamura, K., Nei, M., and Kumar, S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U.S.A. 101: 11030–5. doi:10.1073pnas.0404206101.

Tribuzio, C.A. and Echave, K.B. 2019. Assessment of the Other Rockfish stock complex in the Gulf of Alaska. North Pacific Fishery Management Council Gulf of Alaska Stock Assessment and Fishery Evaluation Report: 1177-1222. URL https://apps-afsc.fisheries.noaa.gov/refm/docs/2019/GOAorock.pdf

Vowles, E.J., and Amos, W. 2006. Quantifying ascertainment bias and species-specific length differences in human and chimpanzee microsatellites using genome sequences. Mol. Biol. Evol. 23: 598–607. doi:10.1093/molbev/msj065.

Wakeley, J., Nielsen, R., Liu-Cordero, S.N., and Ardlie, K. 2001. The discovery of single-nucleotide polymorphisms—and inferences about human demographic history. Am. J. Hum. Genet. 69: 1332–1347. doi:10.1086/324521.

Wallace et al., 2022. Resolving the phylogenetic relationship among recently diverged members of the rockfish subgenus Sebastosomus. Molecular Phylogenetics and Evolution. 173: doi:10.1016/j.ympev.2022.107515

Wickham, H., François, R., Henry, L., and Müller, K. 2018. dplyr: A Grammar of Data Manipulation. R package version. URL https://cran.r-project.org/web/packages/dplyr/index.html.

Figures

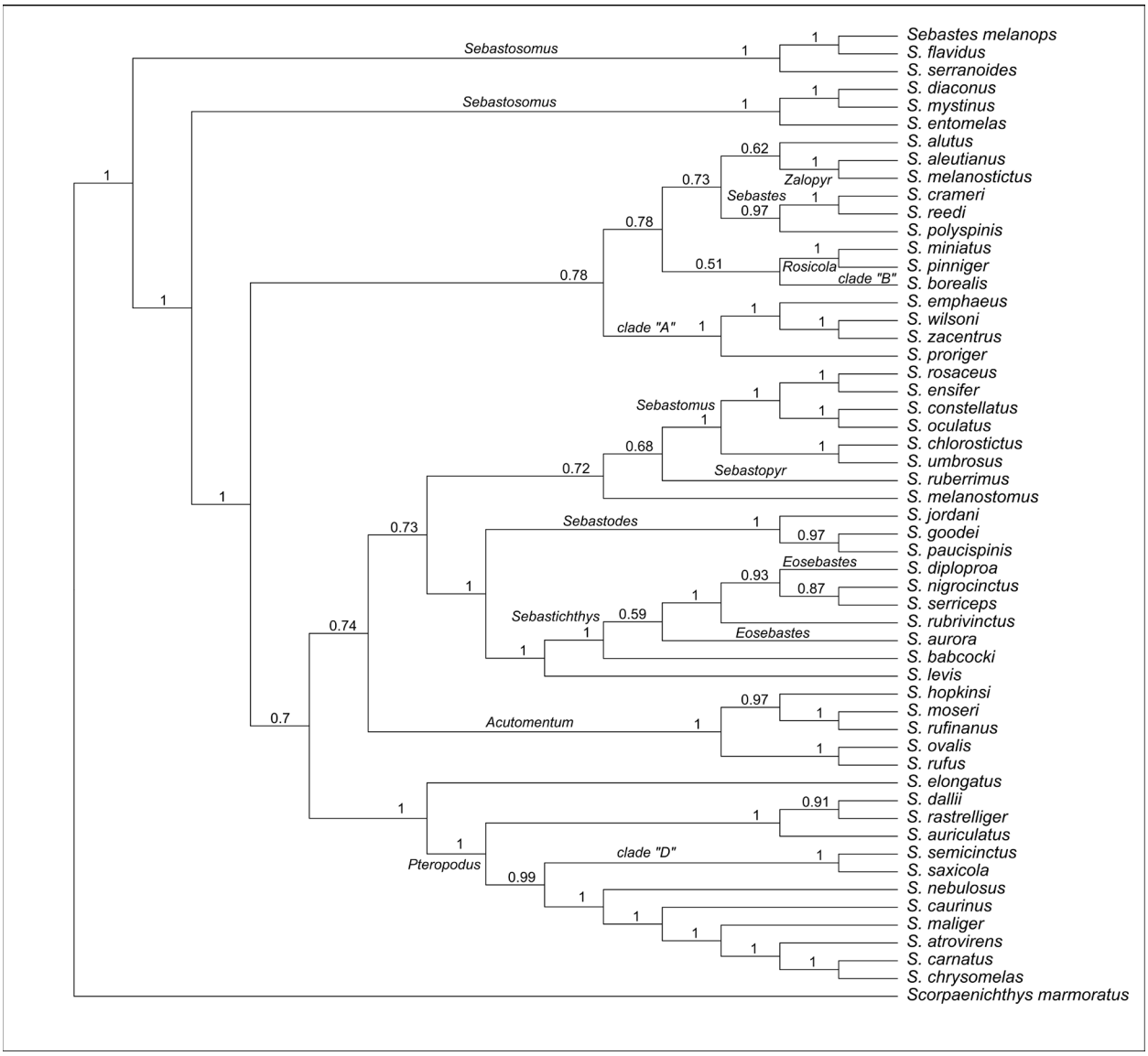
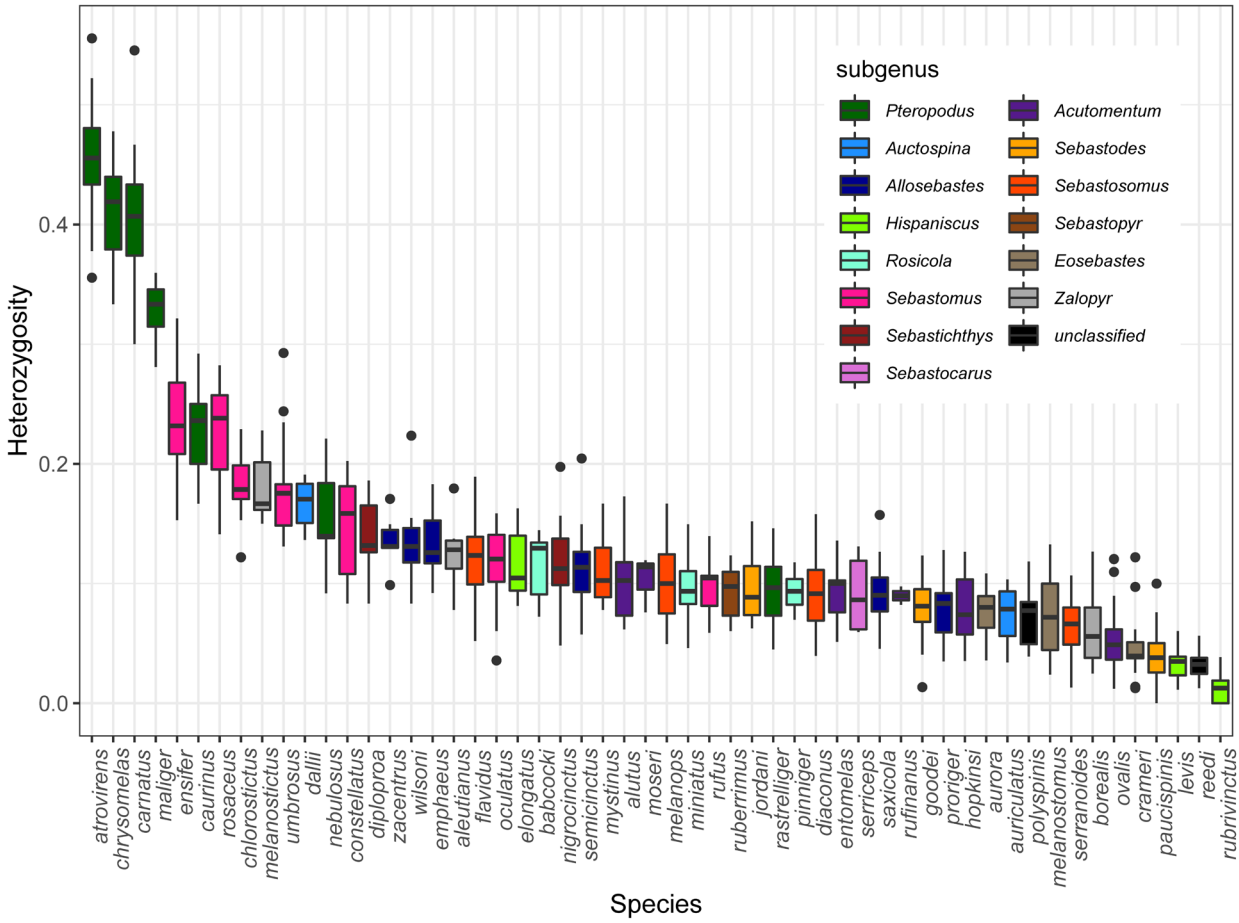


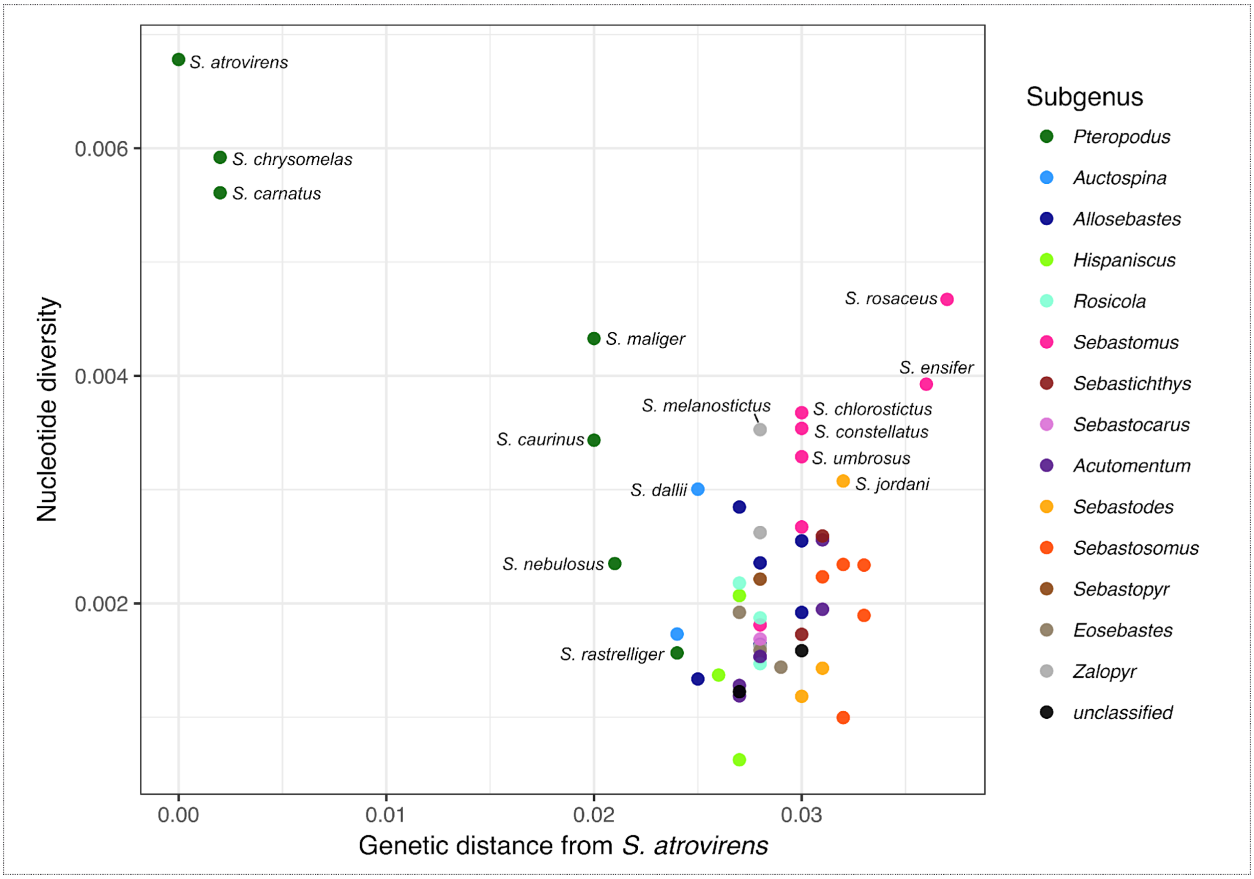
Fig. 1: Consensus tree estimated using the General Time Reversible (GTR) model and Bayesian posterior analysis for 54 *Sebastes* species and one member of the sister genus, *Scorpaenichthys*. Genetic data includes 96 nuclear markers. Bayesian posterior probabilities above 50 are indicated at nodes.



706

707 Fig. 2: Heterozygosity for 54 *Sebastes* species genotyped with 96 nuclear genetic
708 markers. Mean internal heterozygosity per species is indicated as the dark bar inside
709 the box. Boxes represent the 25th and 75th percentiles (first and third quartiles) and
710 whiskers extend to the smallest and largest values 1.5 times the distance between the
711 first and third quartiles. Points beyond that distance are plotted individually. Fill colors
712 indicate subgenera classification consistent with designations in Table 1.

713



714

715 Fig. 3: Genetic distance from *Sebastes atrovirens* and nucleotide diversity for 54
716 *Sebastes* species classified to subgenus. Genetic data includes 96 nuclear markers.
717 Genetic distance is measured as pairwise differences (base differences per site,
718 excluding gaps and missing data). Names for species with nucleotide diversity >0.003
719 and all members of the *Pteropodus* subgenus are shown.
720

721 **Tables**

722 Table 1. Number of samples per *Sebastes* species included in the self-assignment and
723 phylogenetic analyses. Mean nucleotide diversity, mean internal heterozygosity and
724 nominal subgenera classification is included.

Species	samples	nucleotide diversity	heterozygosity	subgenus
Sebastes atrovirens	32	0.00678	0.4578	*Pteropodus*
S. chrysomelas	32	0.00592	0.4086	*Pteropodus*
S. carnatus	32	0.00561	0.4032	*Pteropodus*
S. rosaceus	19	0.00467	0.2275	*Sebastomus*
S. maliger	16	0.00433	0.3285	*Pteropodus*

S. ensifer	19	0.00393	0.2362	*Sebastomus*
S. chlorostictus	15	0.00368	0.1829	*Sebastomus*
S. constellatus	16	0.00354	0.1473	*Sebastomus*
S. melanostictus	7	0.00353	0.1814	*Zalopyr*
S. caurinus	9	0.00343	0.2336	*Pteropodus*
S. umbrosus	18	0.00329	0.1788	*Sebastomus*
S. jordani	15	0.00308	0.0956	*Sebastodes*
S. dallii	6	0.00300	0.1667	*Auctospina*
S. zacentrus	6	0.00285	0.1350	*Allosebastes*
S. oculatus	18	0.00267	0.1163	*Sebastomus*
S. aleutianus	9	0.00262	0.1250	*Zalopyr*
S. diploproa	16	0.00259	0.1388	*Sebastichthys*
S. alutus	24	0.00256	0.1038	*Acutomentum*
S. wilsoni	16	0.00255	0.1345	*Allosebastes*
S. semicinctus	17	0.00236	0.1136	*Allosebastes*
S. nebulosus	21	0.00235	0.1548	*Pteropodus*
S. mystinus	25	0.00234	0.1068	*Sebastosomus*
S. diaconus	26	0.00234	0.0932	*Sebastosomus*
S. melanops	31	0.00223	0.1021	*Sebastosomus*
S. ruberrimus	13	0.00221	0.0958	*Sebastopyr*
S. miniatus	20	0.00218	0.0976	*Rosicola*
S. elongatus	16	0.00207	0.1149	*Hispaniscus*
S. hopkinsi	24	0.00195	0.0790	*Acutomentum*
S. aurora	14	0.00192	0.0770	*Eosebastes*
S. emphaeus	20	0.00192	0.1341	*Allosebastes*
S. flavidus	31	0.00189	0.1189	*Sebastosomus*
S. babcocki	11	0.00187	0.1146	*Rosicola*
S. rufus	11	0.00181	0.0970	*Sebastomus*
S. auriculatus	16	0.00173	0.0734	*Auctospina*
S. nigrocinctus	29	0.00173	0.1145	*Sebastichthys*
S. serriceps	5	0.00169	0.0915	*Sebastocarus*
S. proriger	31	0.00164	0.0790	*Allosebastes*
S. borealis	8	0.00163	0.0637	*Zalopyr*
S. melanostomus	20	0.00159	0.0726	*Eosebastes*
S. reedi	14	0.00158	0.0313	*unclassified*
S. rastrelliger	12	0.00156	0.0945	*Pteropodus*
S. ovalis	31	0.00153	0.0521	*Acutomentum*
S. entomelas	15	0.00153	0.0926	*Acutomentum*
S. pinniger	16	0.00147	0.0944	*Rosicola*
S. crameri	19	0.00144	0.0461	*Eosebastes*

S. paucispinis	32	0.00143	0.0410	*Sebastodes*
S. levis	32	0.00137	0.0325	*Hispaniscus*
S. saxicola	16	0.00134	0.0904	*Allosebastes*
S. moseri	3	0.00128	0.1031	*Acutomentum*
S. polyspinis	8	0.00122	0.0731	*unclassified*
S. rufinanus	2	0.00119	0.0899	*Acutomentum*
S. goodei	32	0.00118	0.0803	*Sebastodes*
S. serranoides	32	0.00100	0.0643	*Sebastosomus*
S. rubrivinctus	19	0.00063	0.0120	*Hispaniscus*
Total samples	**997**			

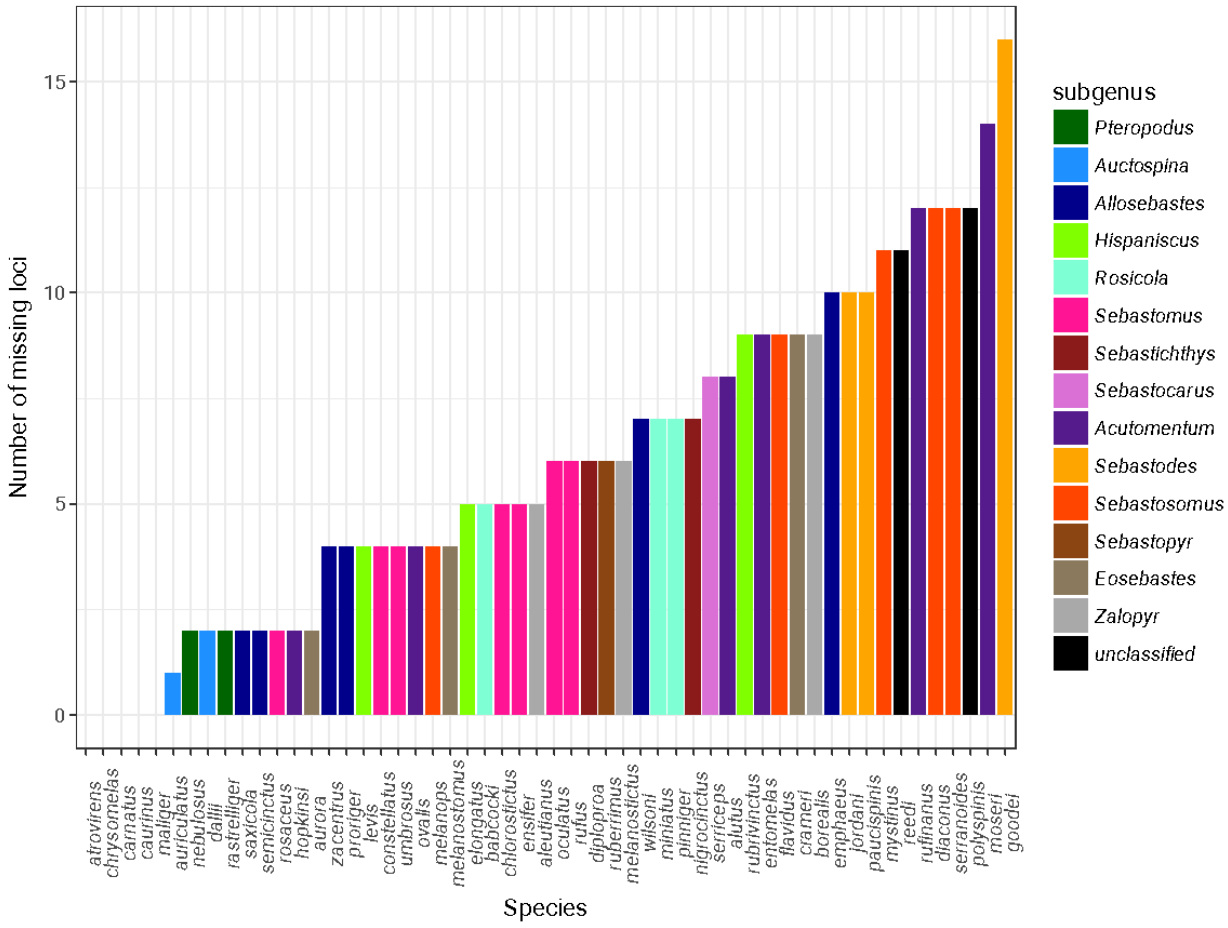
725

Supplemental Material

Table S1. Number of samples per *Sebastes* species included in the variant call format (VCF) file used to call genotypes.

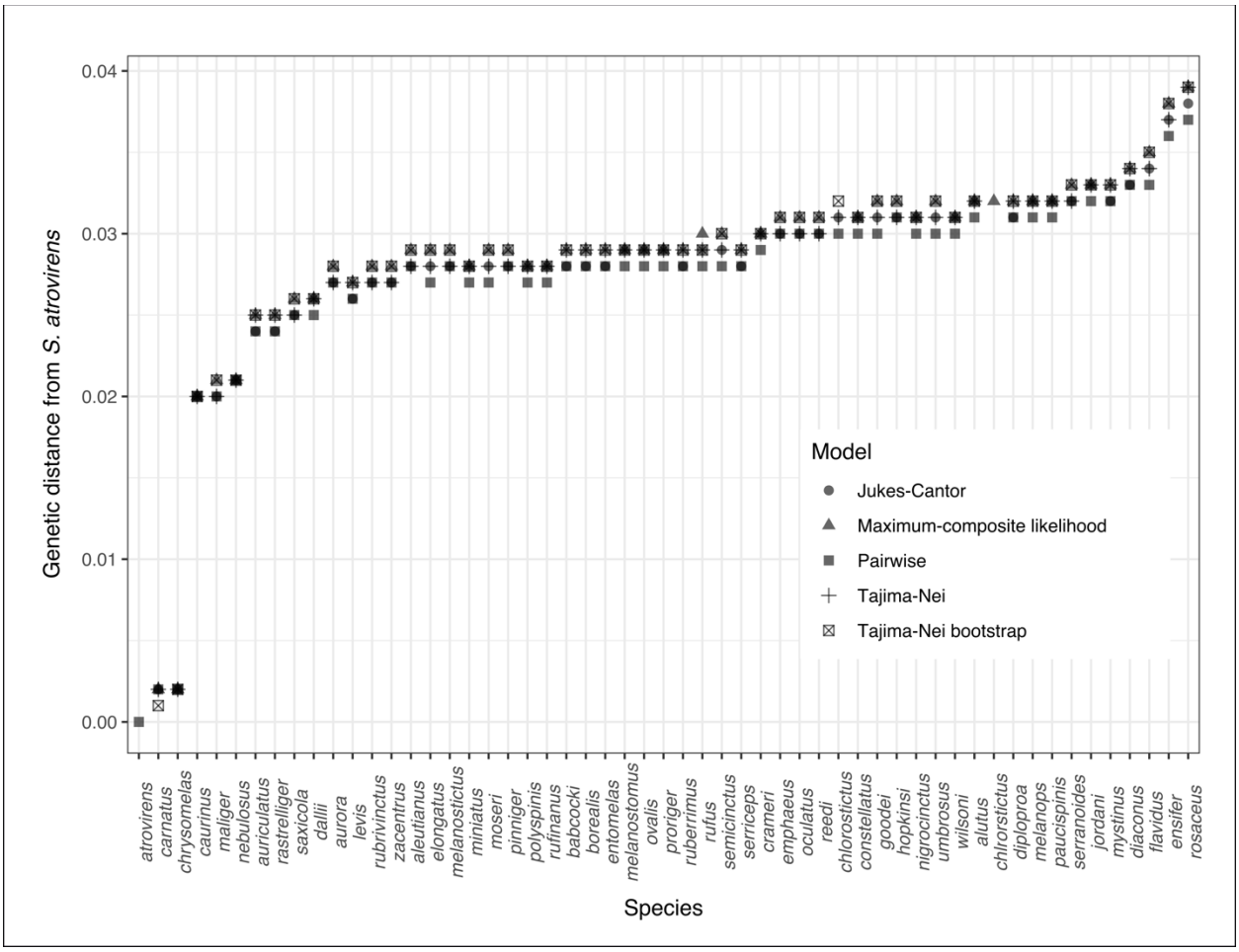
Species	samples
S. atrovirens	242
S. chrysomelas	113
S. goodei	65
S. levis	48
S. ovalis	36
S. serranoides	33
S. flavidus	32
S. hopkinsi	32
S. melanops	32
S. paucispinis	32
S. proriger	31
S. caurinus	30
S. carnatus	29
S. diaconus	26
S. emphaeus	26
S. ensifer	26
S. mystinus	25
S. alutus	24
S. miniatus	23
S. crameri	22
S. nebulosus	22
S. melanostomus	20
S. rosaceus	20
S. umbrosus	20
S. rubrivinctus	19
S. semicinctus	19
S. auriculatus	16
S. aurora	16
S. babcocki	16
S. constellatus	16
S. diploproa	16
S. elongatus	16
S. entomelas	16
S. jordani	16

S. maliger	16
S. oculatus	16
S. pinniger	16
S. rastrelliger	16
S. reedi	16
S. saxicola	16
S. wilsoni	16
S. chlorostictus	15
S. dalli	15
S. zacentrus	15
S. ruberrimus	13
S. rufus	12
S. moseri	10
S. aleutianus	9
S. helvomaculatus	9
S. rufinanus	9
S. borealis	8
S. polyspinis	8
S. serriceps	8
S. melanostictus	7
S. simulator	7
S. nigrocinctus	6
S. brevispinis	1
S. rosenblatti	1
Total	**1440**



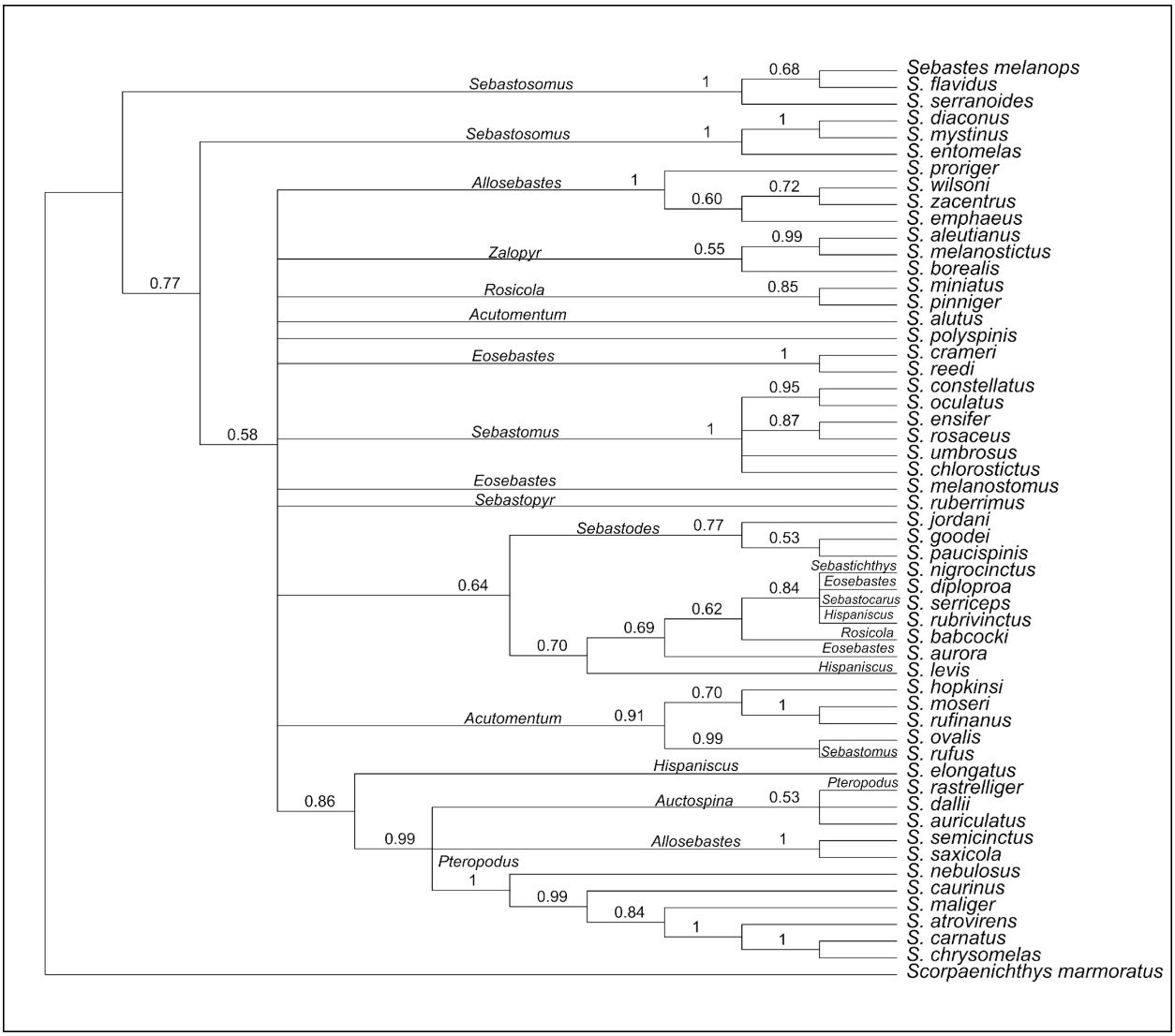
729

730 Fig. S1: Number of genetic markers (loci) that do not amplify in all samples for a given
731 *Sebastes* species out of 96 total markers. Species are ranked from fewest-missing to
732 most-missing loci. Color corresponds to subgenus. No loci are missing entirely from *S.*
733 *atrovirens, S. chrysomelas, S. carnatus, S. caurinus*, or *S. maliger*.



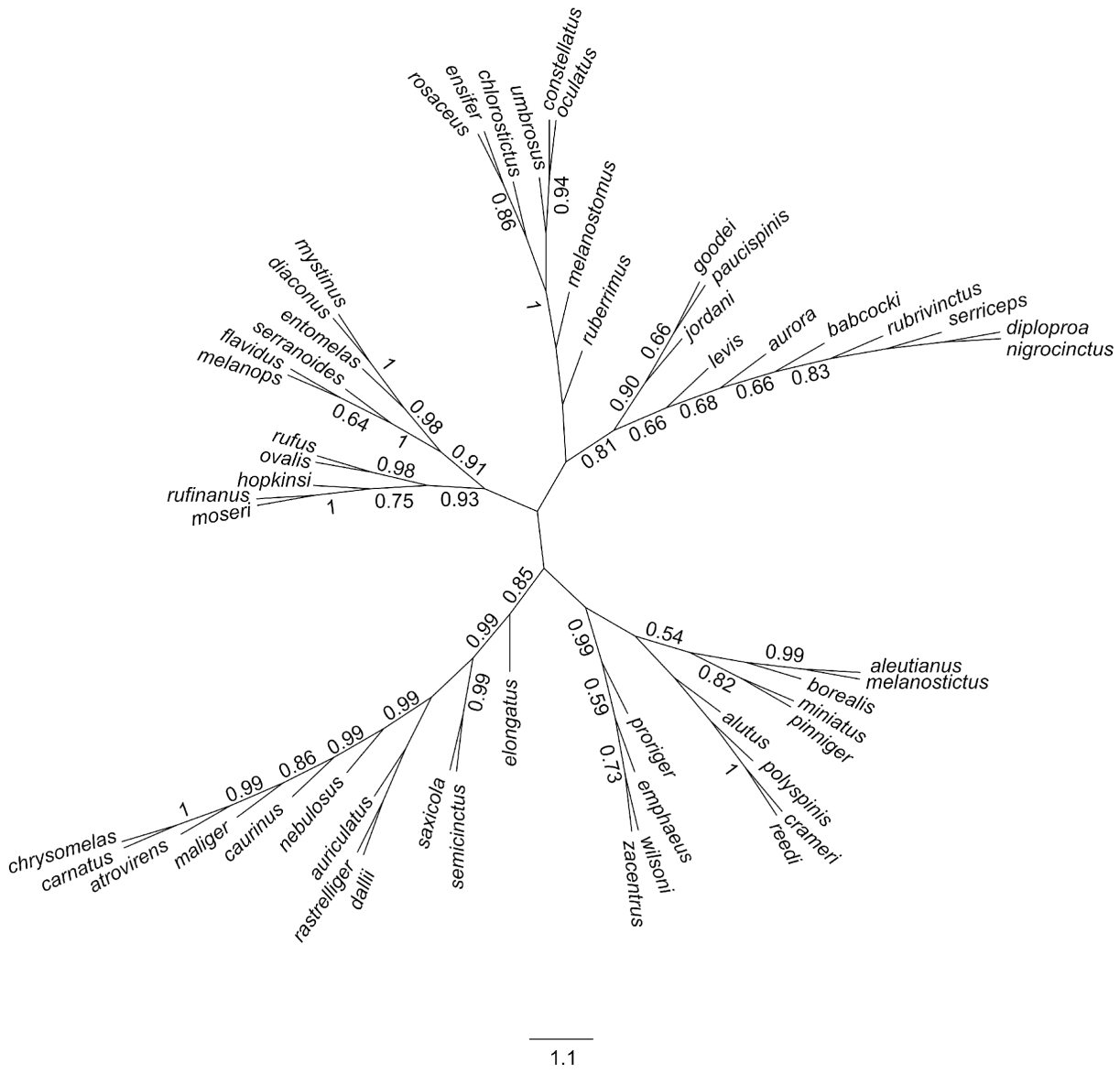
734

735 Fig. S2: Genetic distances are base substitutions per site between each *Sebastes*
736 species and *S. atrovirens*. Nucleotide substitution models are indicated by symbols. The
737 Jukes-Cantor (Jukes and Cantor 1969), Tajima-Nei (Tajima and Nei 1984), and Kimura
738 2-parameter (Kimura 1980) models included uniform rates among sites. The maximum
739 composite like- lihood model (Tamura, Nei, and Kumar 2004) and Tajima-Nei model
740 with 1000 bootstraps included rate variation among sites modeled with a gamma
741 distribution and differences in sequence composition (Tamura and Kumar 2002). The
742 dataset included 10,695 sites, excluding gaps and missing data and analyses were
743 conducted in MEGA v. 7.0.26.



744

745 Fig. S3: Rooted maximum-likelihood tree estimated with the General Time Reversible
746 model and 1000 bootstrap replicates for 54 *Sebastes* species and one member of the
747 sister genus, *Scorpaenichthys*. Genetic data includes 96 nuclear markers. Bootstrap
748 values above 50 are indicated above nodes on relevant branches. Analyses included
749 11,622 sites and retained gaps and missing data.



750

751 Fig. S4: Unrooted maximum-likelihood tree estimated using the General Time
752 Reversible model with 1000 bootstrap replicates for 54 *Sebastes* species. Bootstrap
753 support above 0.50 is shown on branches. The analysis included 11,622 sites and
754 retained gaps and missing data.