Published in final edited form as:

Mol Phylogenet Evol. 2008 October ; 49(1): 136–152. doi:10.1016/j.ympev.2008.06.008.

Phylogenetic Relationships and the Evolution of Regulatory Gene Sequences in the Parrotfishes

Lydia L. Smith¹, Jennifer L. Fessler^{1,2}, Michael E. Alfaro^{1,3}, J. Todd Streelman⁴, and Mark W. Westneat^{*,1}

¹Department of Zoology, Field Museum of Natural History, Chicago, IL 60605-2496

²Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL

³Department of Ecology and Evolution, Washington State University, Pullman, WA

⁴School of Biology, Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology

Abstract

Regulatory genes control the expression of other genes and are key components of developmental processes such as segmentation and embryonic construction of the skull in vertebrates. Here we examine the variability and evolution of three vertebrate regulatory genes, addressing issues of their utility for phylogenetics and comparing the rates of genetic change seen in regulatory loci to the rates seen in other genes in the parrotfishes. The parrotfishes are a diverse group of colorful fishes from coral reefs and seagrasses worldwide and have been placed phylogenetically within the family Labridae. We tested phylogenetic hypotheses among the parrotfishes, with a focus on the genera *Chlorurus* and *Scarus*, by analyzing eight gene fragments for 42 parrotfishes and eight outgroup species. We sequenced mitochondrial *12s* rRNA (967 bp), *16s* rRNA (577 bp), and cytochrome b (477 bp). From the nuclear genome, we sequenced part of the protein-coding genes *rag2* (715 bp), *tmo4c4* (485 bp), and the developmental regulatory genes *otx1* (672 bp), *bmp4* (488 bp), and *dlx2* (522 bp). Bayesian, likelihood, and parsimony analyses on the resulting 4903 bp of DNA sequence produced similar topologies that confirm the monophyly of the scarines and provide a phylogeny at the species level for portions of the genera *Scarus* and *Chlorurus*. Four major clades of *Scarus* were recovered, with three distributed in the Indo-Pacific and one containing Caribbean/Atlantic taxa. Molecular rates suggest a Miocene origin of the parrotfishes (22 mya) and a recent divergence of species within *Scarus* and *Chlorurus*, within the past 5 million years. Developmentally important genes made a significant contribution to phylogenetic structure, and rates of genetic evolution were high in *bmp4*, similar to other coding nuclear genes, but low in *otx1* and the *dlx2* exons. Synonymous and nonsynonymous substitution patterns in developmental regulatory genes support the hypothesis of stabilizing selection during the history of these genes, with several phylogenetic regions of accelerated nonsynonymous change detected in the phylogeny.

© 2008 Elsevier Inc. All rights reserved.

*corresponding author: Department of Zoology, Field Museum of Natural History, Chicago, IL 60605-2496, mwestneat@fieldmuseum.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

phylogenetics; coral reef fishes; parrotfishes; regulatory genes; molecular clock; *bmp4*; *dlx2*; *otx1*

1. Introduction

Rapid progress in the molecular phylogenetics of diverse groups of organisms has created opportunities to combine phylogenetic studies with exploration of molecular evolution in a range of functionally or developmentally important genes. Regulatory genes are loci that control the expression of other genes, typically via DNA-binding transcriptional activity or cell-cell signalling. Despite strong conservation of functional domains across hundreds of millions of years of evolution (Manzanares et al., 2000), evidence exists of major changes in regulatory genes that may have produced important evolutionary changes in body plan and morphology (Lowe and Wray, 1997; Wagner et al., 2003). Many studies have used phylogenetic tools to generate evolutionary topologies of Hox genes and other regulatory loci (e.g., Kim et al., 2003; Purugganan, 1997; Santini and Bernardi, 2005; Wagner et al., 2005) yet there are few explorations of the phylogenetic utility of regulatory gene sequences in larger, multi-locus phylogenetic analyses. Several recent examples in arthropods (Cook et al., 2001) and plants (Barrier et al., 2001; Fan et al., 2004) have shown that regulatory loci can make strong contributions to phylogenetic resolution. This work, particularly in plants, has shown that regulatory gene sequences and their protein products do not necessarily evolve more slowly than structural genes, and in fact may evolve significantly faster (Barrier et al., 2001).

In this context, we are interested in addressing questions of sequence evolution in several of the regulatory genes in diverse groups of coral reef fishes. First, what is the phylogenetic utility of regulatory gene sequences in fishes, and at what level of divergence time do they contribute to phylogenetic resolution? Second, are rates of divergence in genes that regulate development of key anatomical features different relative to background levels of genetic change in diverse groups of reef fishes? Due to the conservatism of most regulatory loci, one might predict that rates of evolution in these genes would be low. However, rates of genetic change are rarely constant, and we predicted that we would detect periods in the phylogenetic history of regulatory genes in diverse clades during which evolution was accelerated. We also expected that some regulatory gene partitions would provide phylogenetic resolving power for species groups that is similar to other mitochondrial and nuclear genes. Alternatively, regulatory loci might have a slower evolutionary signal that is appropriate for resolving deeper phylogenetic relationships. By incorporating these important genes in phylogenetic analyses, we may contribute to broader questions of evolution in regulatory sequences and the generation of structural and functional diversity.

Labrid fishes display great morphological diversity in aspects of cranial morphology and function (Wainwright et al., 2004; Westneat et al., 2005). A particularly unusual group within the Labridae are the parrotfishes, with a modified skull, fused jaws for scraping algae off the substrate, and hypertrophied pharyngeal jaws for grinding their food. We sequenced regions of selected genes that play regulatory roles in the development of the head, teeth, and jaws of parrotfishes and assessed their levels of genetic divergence within parrotfishes in addition to those of more commonly used nuclear and mitochondrial loci. Our aims were to evaluate whether their rates of divergence provide appropriate phylogenetic signal to elucidate phylogenetic relationships within the parrotfishes and to explore how their patterns of molecular evolution compare with those of other, non-regulatory, coding genes. Eight gene fragments were sequenced: portions of *12s* and *16s* rRNA and cytochrome b (*cytb*) from the mitochondrial genome, and regions of the nuclear protein-coding genes *tmo4c4*,

Recombination activating gene-2 (*rag2*), Orthodenticle homeobox-1 (*otx1*), Bone morphogenetic protein-4 (*bmp4*), and Distal-less homeobox-2 (*dlx2*). The first five genes listed are frequently used in phylogenetic studies of fishes and other vertebrates; the final three genes are regulatory or developmentally important genes involved in development of the trophic system in fishes, but with as yet unproven phylogenetic utility among coral reef fishes.

The parrotfishes are a diverse clade of marine fishes that are circumtropical in distribution, but with the majority of taxa found in the Indo-Pacific region. Composed of about 90 species organized into 10 genera, more than half of the group are in the genus *Scarus*. Parrotfishes are known to be nested within the family Labridae (Streelman, et al., 2002, Westneat and Alfaro, 2005), with a long root branch distinguishing scarines from cheilines, their proposed sister group (Clements et al., 2004, Westneat and Alfaro, 2005). Using our diverse sample of genetic markers, a phylogenetic analysis of the Scaridae was conducted in order to confirm the monophyly of the family and its component genera; to test the division of Scaridae into seagrass-and coral-associated clades (Streelman et al., 2002) and the separation of *Chlorurus* from the remainder of *Scarus* (Bellwood, 1994); to investigate species relationships within *Scarus*; to examine patterns of evolution in color pattern and distribution within *Scarus*; and to assess phylogenetic utility of three regulatory genes compared to genes more commonly used in molecular systematics.

2. Materials and Methods

2.1 Species sampling and genes sequenced

DNA sequence data for 50 species of fishes were analyzed in this study, including 8 outgroup taxa from within the Labridae and 42 parrotfishes. We included at least one member of each scarine genus, with a focus on sampling within the two largest genera, *Chlorurus* (7 species sampled) and *Scarus* (24 species sampled). All specimens were collected by the authors or purchased at local fish markets, and most have a voucher specimen associated with the tissue (referenced in Genbank submission). Portions of eight genes were sequenced in the study, including three mitochondrial regions (*12s*, *16s*, and *cytb*) and 5 nuclear markers (*rag2*, *tmo4c4*, *bmp4*, *dlx2*, and *otx1*).

We sequenced fragments of 6 protein coding genes for this study, including three nuclear regulatory loci. *Tmo4c4* is a single-copy nuclear DNA locus first isolated by Streelman and Karl (1997) who found a high degree of similarity with a region of the human titin protein, a large protein that contributes to muscle assembly and resting tension. *Tmo4c4* has subsequently become popular for its utility in family-level phylogenetic studies of fishes (Farias et al., 2000, Lovejoy, 2000, Smith and Wheeler, 2004, Sparks and Smith 2004; Westneat and Alfaro, 2005). We generated 486 base pairs of *tmo4c4* sequence, encompassing all but the first 26 nucleotides of the fragment isolated by Streelman and Karl (1997).

Recombination activating gene-2 (*rag2*) consists of a single exon in vertebrates and works with RAG1 to synergistically activate V(D)J recombination, the process by which developing lymphocytes generate immune system diversity (Oettinger, et al., 1990). We sequenced 715 bp of 1600 bp of coding sequence, from roughly nucleotide 280–995 (from Hansen and Kaattari, 1996). All of the sequence used in this study is located in the proposed “functional core region” (Cuomo and Oettinger, 1994). Cytochrome b is a mitochondrial DNA gene widely used in phylogenetic analysis. It codes for an eight helix trans-membrane protein which serves as the main catalytic subunit of a respiratory chain enzyme (Esposti et al., 1993). We have a consensus region of 477 nucleotides from the 5' half of *cytb*. This corresponds with amino acid residues 18 through 176 of approximately 380 found in fishes.

Otx1 is a vertebrate homolog of the *Drosophila* gene orthodenticle (*otd*), a bicoid-class homeobox gene responsible for transcriptional regulation in anterior regions (Simeone et al., 2002). *Otx1* expression is mainly, but not exclusively, restricted to the head of developing embryos (Boncinelli et al., 1994) and is a vital determinant of brain development and early organization of the vertebrate head, eye and inner ear (Andreazzoli et al., 1997, Acampora et al., 1998, Morsli et al., 1999). We sequenced 672 nucleotides (excluding indels), including the final 7 amino acids of the homeodomain, the entire histidine-rich region, and one of the two C-terminal repeats (Simeone et al., 2002).

Bone morphogenic protein-4 (*bmp4*) is a member of the transforming growth factor- β superfamily (Kingsley, 1994). *Bmp4* is a vertebrate homolog of *Drosophila* decapentaplegic (*dpp*) responsible for the establishment of dorsal-ventral polarity and segmentation in the fly (Irish and Gelbart, 1987). Similar patterning functions for *bmp4* have been recognized in zebrafish (Chen et al., 1997; Nikaido, 1997). *Bmp4* has also been implicated in the initiation and patterning of tooth formation (Peters and Balling 1999, Tucker et al., 1998), bone and joint development and repair (Kingsley, 1994; Cheng et al., 2003) and the evolution of skull diversity in cichlid fishes (Albertson et al., 2003; 2005) and Darwin's finches (Abzhanov, 2004). The *bmp4* coding region spans two exons to produce a protein of about 400 amino acids. We have 488 bps from the 5' region of exon 4 with all but 54 nucleotides from the pro-domain.

Dlx2 is one of multiple vertebrate homologs of the *Drosophila* homeodomain transcription factor *dll*, a key gene in neural crest and forebrain development (Holland et al., 1996). There are at least eight *dlx* genes in zebrafish; six of them, including *dlx2*, are found in inverted pairs linked to hox clusters (Stock et al., 1996). *Dlx2* is essential for the development of proximal regions of the first and second branchial arches, the oral jaws, and hyoid arch that form much of the feeding apparatus of fishes (Qiu et al., 1997). We have 159 bp of the 3' end of exon two, 117–159 bp of the subsequent variable-length intron, and 209 bp of the 5' region of exon three, including the entire homeobox domain (McGuinness, et al., 1996).

2.2 DNA extraction and amplification

Genomic DNA was extracted from 2–4 mm³ of tissue sample using a Puregene® DNA isolation kit (Gentra Systems, Minneapolis, MN), according to the Animal tissue DNA isolation Protocol. Briefly, tissue was homogenized in a low molarity salt buffer and digested overnight with proteinase K. After the addition of a protein precipitation solution, the sample was centrifuged to remove cellular material from the supernatant. DNA was then precipitated at room temperature in 100% isopropanol with 1.5 μ l of 20 mg ml⁻¹ glycogen. We pelleted the DNA by centrifugation on high speed followed by washing in 70% ethanol. DNA was rehydrated in Puregene® DNA hydration solution and then stored at 4°C.

Aliquots of genomic DNA isolates were used as templates for PCR to amplify double-stranded DNA products from the seven gene fragments. Each PCR had a reaction volume of 25 μ l and contained 1 μ l of DNA stock regardless of stock concentration (diluted 1:10 in some cases), 2.5 μ l 10X reaction buffer (100 mM Tris-HCl, 500 mM KCl, 15 mM MgCl₂, pH 8.3; Roche, Mannheim, Germany), 1.5 μ l of 2 mM each deoxynucleotide triphosphates, 0.1 μ l of 5 unit μ l⁻¹ *Taq* polymerase (Roche, Mannheim, Germany), and 1.0 μ l of each oligonucleotide primer, each at 10 μ M concentration. Primers used for amplification and sequencing are given in Table 1. Some samples amplified weakly using standard PCR, and were reamplified with Ready-to-go™ PCR beads (Amersham Pharmacia Biotech Inc., Piscataway, NJ).

Each *16s*, *12s* and *cytb* PCR included an initial denaturation step at 94° C for 2 min, followed by 30 cycles of PCR, each cycle including denaturation at 94° C for 30 s,

annealing at 49 or 60° C for 30 s, and extension at 72° C for 55 s. After a final extension step at 72° C for 2 min the PCR products were held at 4° C. Each nuclear gene PCR included an initial denaturation step at 95° C for 1 min, followed by 35 cycles of PCR, each cycle involving denaturation at 95° C for 30 s, annealing at 55° C for 1 min, and extension at 72° C for 90 s. The final extension step was at 72° C for 5 min. PCR reactions were performed using an MJ Research PTC-200 Peltier Thermal Cycler (MJ Research Inc., Watertown, Mass.).

PCR products were electrophoresed on 1% low melting agarose gels stained with ethidium bromide; bands were visualized under ultraviolet light then excised, melted at 70° C, and incubated with 1 µl GELase™ Agarose Gel-Digesting Preparation (Epicentre, Madison, WI) for at least 3 hours at 45° C. PCR products were cycle-sequenced using an ABI PRISM® Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit, version 3.0 or 3.1 (Applied Biosystems, Foster City, CA). Each reaction included 3.0 µl of buffer, 0.6 µl of primer, 1.0 µl of Big Dye™ reagent, 0.7–2.0 µl of PCR product template DNA, and enough water to bring the total reaction volume to 10 µl. The cycling protocol was 25 cycles of denaturation at 96° C for 10 s, annealing at 50° C for 5 s, and extension at 60° C for 4 min. Sequences of both strands were generated on an ABI PRISM® 3100 or 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA). All sequences have been deposited in GenBank with the following accession numbers: *12s* EU601178–EU601227, *16s* EU601228–EU601262, *tmo4c4* EU601263–EU601306, *Rag2* EU601307–EU601356, *cytb* EU601357–EU601405, *otx1* EU601406–EU601455, *bmp4* EU601456–EU601505, *dlx2* EU601506–EU601554.

2.3 Sequence alignment

Sequences from all genes were trimmed to the size of the smallest fragment to minimize the amount of missing data in the data matrix. Sequences for *tmo4c4*, *rag2*, *cytb*, *otx1*, *bmp4*, and the exonic regions of *dlx2* were aligned using Sequencher 4.2 (Gene Codes Corp). *12s* and *16s* sequences were aligned manually to a secondary structure for labrids (Westneat and Alfaro, 2005), using Se-Al v1.0a1 (Andrew Rambaut, University of Oxford). Gaps and ambiguously alignable regions were excluded from analysis. Se-Al was also used to align the intron sequences of *dlx2*; regions that could not be reliably aligned between taxa were also excluded. After excluding unalignable regions and trimming sequence ends, the character count for each gene was *12s* (743), *16s* (399), *tmo4c4* (485), *rag2* (715), *cytb* (477), *otx1* (666), *bmp4* (488) and *dlx2* (514), for a total of 4487 nucleotide characters.

2.4 Phylogenetic analysis

Phylogenetic analyses were conducted using PAUP* (4.0b10) or Mr. Bayes 3.1 (Ronquist and Huelsenbeck, 2003) on a quad-processor Apple Macintosh G5 computer or a distributed computer network (Illinois BioGrid). Heuristic parsimony searches to find the most parsimonious tree(s) were performed using TBR branch-swapping. One thousand random addition replicates were used to minimize the chance of finding only locally optimal trees (Maddison, 1991). All sites were equally weighted and gaps treated as missing characters. We used nonparametric bootstrapping (Felsenstein, 1985) to measure support of clades with 1000 pseudoreplicates and branch-swapping with two random addition replicates per pseudoreplicate.

For the maximum likelihood analysis the TRN+Γ+I substitution model (rmat = 1.0000 5.4800 1.0000 1.0000 7.5325) with invariable sites (pinvar = 0.5094) and among-site rate heterogeneity (shape=0.5298) was selected using hierarchical likelihood ratio tests implemented in Modeltest 3.7 (Posada and Crandall, 1998). A heuristic search with 100 random sequence additions was used to find the most likely tree.

We used MrBayes 3.1 to perform 4 independent runs of a Markov Chain Monte Carlo analysis with 6 chains running for 10 million generations on the Illinois BioGrid distributed computer network through DePaul University. The eight gene partitions were assigned separate parameters with their own model of sequence evolution, determined using ModelTest 3.7. We saved trees every 1000 generations for each run and used default (flat) priors for the transition/transversion rate ratio (Beta 1,1), branch length (Exp 10), alpha parameter of the gamma distribution for rate heterogeneity (Uniform 0.05–50), proportion of invariant sites (uniform 0–1), base frequencies (Dirichlet 1,1,1,1) and tree topology parameters (uniform over all possible topologies). Visual inspection of tree length, log-likelihood, and model parameters revealed that all parameters appeared to reach stationarity within 500,000 generations. To help ensure that stationarity was reached, we discarded the first 2 million generations (2000 sampled trees) from each run as burn-in and used the remaining 8 million generations (8000 sampled trees for each of the 4 runs) in all subsequent analysis. A majority rule consensus tree calculated from the post burn-in trees was constructed and used to determine the posterior probabilities of clades.

2.5 Regulatory genes, rates of molecular evolution, and variable clocks

To explore the phylogenetic utility of the regulatory genes and the topology supported by regulatory and non-regulatory genetic data, we partitioned the data into two subsets, one containing the regulatory genes and the other the non-regulatory molecular markers and each was analyzed with MrBayes 3.1 (3 independent runs each with 5 chains running for 10 million generations). We also examined rates of molecular evolution in the 8 gene sequences by generating phylograms of each partition on the Bayesian maximum probability topology (all data combined) to produce a phylogenetic map of the changes in gene sequence. These phylograms were used to assess branch lengths as a heuristic measure of the relative amounts of sequence evolution in each partition.

To examine the evolution of the six protein coding genes, we analyzed the amounts and rates of synonymous and non-synonymous change using both likelihood and Bayesian approaches. To explore nucleic acid substitution patterns under a likelihood framework, we used the Muse and Gaut (1994) codon model as implemented by the AnalyzeCodonData algorithm of the software package HyPhy (Kosokovsky Pond et al., 2005), in combination with Bayesian consensus tree. For each gene partition, we calculated the rates of synonymous (and nonsynonymous) substitution per site and visualized these rates along each branch of the phylogeny using the tree graphics utilities of HyPhy. Several branches of the phylogeny (root branch of the parrotfishes and root of *Scarus+Chlorurus*) appeared to have particularly high or variable rates of dN/dS in some genes relative to other branches, so the hypothesis was tested whether these branches had significantly different dN/dS ratio than the rest of the tree, for each gene, using the TestBranch dNdS module of HyPhy.

To investigate the possible dates of origin of major parrotfish groups and estimate the timing of divergences among the parrotfishes, we used two methods that relaxed the assumptions of a strict molecular clock. Under a likelihood framework, the penalized likelihood (PL) module of the r8s 1.50 software of Sanderson (1997, 2002) was used to estimate the ages of scarine clades. The Bayesian MPP tree topology was used with the following parameters: algorithm=tn, rates = gamma, shape=0.53, and smoothing =4 (determined by cross validation). We also used CodonRates 1.0 (Seo et al., 2004) to analyze the rates of molecular evolution in protein coding genes in a Bayesian framework, reconstruct the rates of synonymous and nonsynonymous rates along branches of the parrotfish phylogeny, and estimate divergence times from codon substitution rates. We constrained two nodes in both analyses using fossil calibrations, fixing the root age of the labrid outgroup node at 55 million years old, based on the earliest fossil evidence of the Labridae (Bellwood, 1990) and constraining the minimum age of the parrotfish root node (split between seagrass and reef

clades) at 15 mya, based on the earliest fossil parrotfish, a *Calotomus* from the middle Miocene (Bellwood and Schultz, 1991).

3. Results

Phylogenetic analyses of eight gene sequences produced well resolved phylogenetic trees for most clades of parrotfishes (Fig. 1). Each of the gene fragments we used provided useful phylogenetic signal at multiple levels of phylogenetic resolution, with regulatory genes aiding in resolution of major clades within both outgroup and ingroup regions of the tree (Fig. 2). Bayesian, likelihood and parsimony methods produced similar topologies, with high support at many nodes. Results show a monophyletic scarine subfamily, most closely related to the cheiline labrids, and a well-resolved split between reef-associated taxa and a seagrass clade, confirming earlier results (Streelman et al., 2002; Westneat and Alfaro, 2005). The genera *Chlorurus* and *Scarus* were found to be monophyletic sister-groups, in agreement with morphological phylogenetics (Bellwood, 1990) and our previous molecular study (Streelman et al., 2002). The phylogeny illustrates several species pairs of *Chlorurus* and four main clades of *Scarus*, within which the topology has some areas of strong support and other relationships that are less well-resolved. Rates of synonymous and nonsynonymous change indicate stabilizing selection for all protein-coding loci in most parts of the phylogeny, and several branches had significantly higher or lower dN/dS ratios for several genes, indicating that dN/dS rates sped up and slowed down during parrotfish evolution. The results of rate analyses and estimation of clade ages using r8s and CodonRates were in agreement, showing that the main period of speciation in *Chlorurus* and *Scarus* is fairly recent, within the last 5 million years.

3.1 Molecular characterization and sequence divergence

The data set consisted of 4903 nucleotides of which 4487 were alignable, 1496 of the alignable bases were variable and 1121 were parsimony-informative. Base composition across the data set was fairly uniform (24.5% A, 28.9% C, 24.4% G, and 22.2% T). Within the *dlx2* intron, we found that *Calotomus carolinus* and *C. spinidens* share a 5 bp insertion, and *Scarus flavipectoralis* and *S. hypselopterus* share a 14 bp deletion.

Character optimizations of gene fragments onto the total evidence Bayesian topology inferred from the concatenated sequences (Fig. 3–Fig. 5) illustrate the total amounts of nucleotide substitution along various branches of the tree. Sequence divergence among the parrotfishes and labrid outgroups was generally high (Fig. 3–Fig. 5), with the exception of comparisons within *Chlorurus* and *Scarus*. Peak *12s* divergence (Fig. 3A) was 16.4%, occurring between *Nicholsina usta* and *Pseudodax mollucanus*, and within parrotfishes it was 13.1%, between *Nicholsina* and *Scarus frenatus*. Peak *16s* divergence was lower, 10.3% and 8.6% between scarines and outgroups and within scarines, respectively. Cytochrome b divergence (Fig. 3B) was highest (25.2%) between *Centrolabrus exoletus* and *Chlorurus capistratoides*, and was maximally 21.6% within parrotfishes. Resolution of the phylogeny among *Chlorurus* and *Scarus* at the species level was aided most by variability in *cytb* data set within the genera (Fig. 3B). Maximum *rag2* divergence (Fig. 4A) was between *Lachnolaimus maximus* and *Leptoscarus vaigiensis* (20.9%) and within scarines was 10.6% between *Cryptomus* and *Scarus coelestinus*. Peak sequence divergence of *tmo4c4* was 17.1% between *Epibulus* and *Hipposcarus*, and within parrotfishes was maximal (11.5%) between *Cryptomus* and *Hipposcarus*.

Regulatory genes showed similar, though slightly lower, levels of sequence divergence, with a maximal p-distance for *bmp4* of 15.4% (Fig. 4B), between *Lachnolaimus maximus* and *Leptoscarus vaigiensis*, and an intra-scarine maximal divergence of 7.5%. *Otx1* divergence was 10.4% between *Tautoga onitis* and *Nicholsina usta*, and within parrotfishes showed a

peak between *Nicholsina* and *Leptoscarus* of 8.6%. Maximal overall *dlx2* divergence (Fig. 5) was 13.7%, between *Centrolabrus* and *Chlorurus japonicus*, and was 7.1% between the scarine taxa *Cryptomus* and *Scarus coelestinus*. Most of this divergence is intronic however (Fig. 5), as the intron had a peak of 40% divergence (*Epibulus* to *Calotomus spinidens*) and a within parrotfish maximum of 23.4%, whereas the exonic regions were just 5.2% divergent overall (*Tautoga-Sparisoma*) and peaked at 2.5% within parrotfishes. Regulatory genes showed sequence divergences of zero or nearly zero between species pairs of *Chlorurus* and/or *Scarus*, and this was especially true of the *dlx2* exon (Fig. 5A).

3.2 Phylogeny of the parrotfishes

A diverse DNA sequence data set including both mitochondrial and nuclear loci, intron and exon regions, and protein-coding and ribosomal regions, provided excellent resolution for the topology of the scarine phylogeny (Fig. 1). Bayesian, parsimony and likelihood methods produced nearly identical topologies of relationship among the parrotfishes. Our sample of species showed that the genera *Scarus* and *Chlorurus* were monophyletic, and together they form a well-supported clade. The predominantly seagrass-associated genera (*Calotomus*, *Cryptotomus*, *Leptoscarus*, *Nicholsina*, and *Sparisoma*) are shown to be distinct from a reef-associated clade (*Bolbometapon*, *Cetoscarus*, *Chlorurus*, *Hipposcarus*, and *Scarus*). Three species pairs in the genus *Chlorurus* received strong support: *C. sordidus* + *C. japonensis*, *C. bleekeri* + *C. bowersi*, and *C. capistratoides* + *C. microrhinos*. The sister species to the rest of the *Chlorurus* included in the tree was *C. oedema*. Four multi-species clades within *Scarus* can be identified with high Bayesian posterior probability and greater than 80% maximum parsimony bootstrap support (Fig. 1, *Scarus* clades 1–4). *Scarus* clade 1 includes four Western Pacific species, including the sister pair *S. flavipectoralis* + *S. hypselopterus*. *Scarus* clade 2 is composed of 7 of the sampled species, with the pair *S. globiceps* + *S. oviceps* closely related to *S. rivulatus* and *S. quoyi*, the pair *S. chameleon* + *S. festivus* sister to them, and, *S. spinus* as the sister to the rest of clade 2. Clade 3 is a strongly supported subclade of the four Caribbean *Scarus* taxa included in the study (*S. coelestinus*, *S. guacamaia*, *S. iseri*, and *S. taeniopterus*), with *S. ghobban* positioned as tentative sister to the Caribbean clade. *Scarus* clade 4 is a group of 8 species in the tree, arranged in two subclades (Fig. 1), including mostly IndoPacific taxa and a single member that ranges to the eastern Pacific Ocean (*S. rubroviolaceus*).

3.3 Phylogenetic utility of regulatory genes

The partitioned analysis, in which the regulatory genes and non-regulatory genes were analyzed separately, show that the regulatory genes expressed a similar phylogenetic signal to the more commonly used mitochondrial and nuclear characters, with both regulatory and non-regulatory partitions capturing the topology of the combined data tree (compare Fig. 1 and Fig. 2). Both partitions display the same pattern of distinct coral and seagrass clades, of the monophyletic association of *Chlorurus* and *Scarus*, of the monophyly of *Sparisoma*, *Scarus*, and *Chlorurus*, and of the main *Scarus* sub-clades seen in the combined analyses. Strongly supported nodes in the two partitioned analyses (Fig. 2) are overwhelmingly in agreement with one another and with the combined analyses. A number of the primary nodes along the backbone of the tree are common between the combined and non-regulatory partitions, but are not strongly supported by the regulatory partition. Only one well-supported node in the non-regulatory partition and three nodes in the regulatory partition (Fig. 2) are in conflict with strongly supported nodes in the combined tree (Fig. 1).

3.4 Rates of molecular evolution: synonymous and non-synonymous change

Phylogenetic maps of gene substitutions (Fig. 3–Fig. 5) illustrate that substitutions have accumulated most along the root branch of the scarine and cheiline clades, least along *Scarus* and *Chlorurus* tip branches, and that the different genes vary in the degree to which

they contribute to branch lengths at particular locations in the phylogeny. For the protein coding genes, these substitution patterns were broken down into the rates of synonymous and non-synonymous changes within the phylogeny (Table 2). All six coding genes showed considerably higher rates of synonymous than non-synonymous change, resulting in dN/dS ratios well below 0.5, and usually below 0.1. A single exception is the rate ratio for *tmo4c4* at the reef clade of parrotfishes, which is greater than 1.0 (Table 2). The branches leading to the cheiline and scarine subfamily nodes had the highest rates of synonymous substitution, and rates tended to decrease at finer levels in the phylogeny. Many of the shorter branches among species had no non-synonymous changes, preventing the computation of rate ratios.

The dS and dN phylogenetic maps for *cytb* (Fig. 6), *rag2* (Fig. 7) and *bmp4* (Fig. 8), illustrate the various synonymous and non-synonymous nucleotide substitution patterns in different genetic markers. The phylogenetic information content of *cytb* was high for the species relationships among *Chlorurus* and *Scarus*, but that information is almost entirely due to silent substitutions at the codon level (Fig. 6A), with low branch lengths for changes that result in amino acid changes (Fig. 6B). For *rag2* and *bmp4* (Fig. 7, Fig. 8), dS is also much greater than dN, although the dN trees shows some nonsynonymous change among *Scarus* species, but low variability within *Chlorurus*. The dS and dN character map for *tmo4c4* was similar to that of *bmp4*, whereas those for *otx1* and *dlx2* show low dS and mostly zero dN branch lengths.

Overall patterns of molecular evolution displayed by the developmental regulatory genes differed from one another, with *bmp4* usually having a rate similar to that of *cytb*, *rag2*, and *tmo4c4* in most regions of the tree, whereas *otx1* and exonic *dlx2* had much lower rates of both synonymous and non-synonymous changes. Non-synonymous changes in *otx1* and *dlx2* were rare, found only in the branch leading to the scarines. However, the dN/dS rate ratio of *dlx2* (0.33) is the highest of all six genes along the branch leading to the parrotfishes, significantly higher than the rate ratio of the rest of the phylogeny (Table 2). Four branches that were tested had accelerated dN/dS rate ratios (cheilines for *rag2*, scarines for *dlx2*, seagrass clade for *bmp4*, and reef clade for *tmo4c4*) and two had significantly lower rate ratios than the rest of the phylogeny (scarines for *bmp4* and *tmo4c4*).

3.5 Dates of origin and divergence times for the parrotfishes

Penalized likelihood (Sanderson, 2002) and the Bayesian approach in CodonRates (Seo et al., 2002) produced similar estimates of divergence times among the parrotfishes (Table 3: Fig. 9), with the exception of the split between cheilines and scarines which ranged from 36 mya (PL) to 51 mya (CR). Using an initial root age estimate of 55 mya and minimum age of the scarine clade of 15 mya yielded an age estimate for the Scarini at 17.3 (PL) to 18.0 (CR) mya. This is the date of the major split between the ancestors of the reef and seagrass clades (Fig. 9). The first divergence among the seagrass clade species sampled here is estimated to be 12–15 mya, and within the reef clade at 14–16 mya (Fig. 8, Table 3). The ancestral cladogenic event leading to the genera *Chlorurus* and *Scarus* was estimated at just 6–7 mya, and the origins of all *Scarus* and *Chlorurus* clades of the present-day were estimated to have occurred between 2 and 4 mya.

4. Discussion

The phylogenetic relationships among many species of parrotfishes are resolved with a combination of mitochondrial, nuclear, and regulatory gene fragments. Species-level resolution of 24 *Scarus* species and seven *Chlorurus* species suggests that this combination of genetic data provides the means to resolve phylogenetic questions at both deep and recent nodes in the parrotfishes. The developmental regulatory genes *otx1* and the *dlx2* exons were generally conserved, providing information primarily at deeper nodes of the phylogeny,

whereas *bmp4* and *dlx2* intron sequence were more variable and were phylogenetically useful at multiple hierarchical levels in the tree, similar to other nuclear genes (e.g., *rag2* and *tmo4c4*) that have been used more broadly in phylogenetic studies. We focus our interest below on the major patterns of parrotfish species phylogeny, the contribution of regulatory genes to phylogenetic resolution within parrotfishes, the evolutionary interpretation of patterns of nucleotide substitution observed, and the remarkably young ages of origin that we estimate for the parrotfishes.

4.1 Phylogeny of the parrotfishes

The higher-level pattern of parrotfish relationships resolved here agrees closely with that of Streebman *et al.* (2002) in regard to the split between seagrass and reef clades of scarines, and the ancestral splits of the genera *Bolbometopon*, *Cetoscarus*, and *Hipposcarus* at the base of the reef clade (Fig. 1). In addition, the present topology reinforces the morphological diagnosis of Bellwood (1994) regarding the genera *Chlorurus* and *Scarus* as being two distinct monophyletic lineages. The primary phylogenetic contribution of the present study is thus to advance our understanding of species relationships among the genera *Chlorurus* and *Scarus*. Most previous suggestions of relatedness among species have been based on color pattern and distribution information, so we summarize those characters that may be of interest in view of the current phylogenetic tree.

The earliest species to branch from the *Chlorurus* clade, *C. oedema*, has a relatively simple, dark color pattern and, like many *Chlorurus*, a somewhat restricted distribution, from Sri Lanka to the Philippines, and north to the Ryukyu Islands. The main *Chlorurus* clade sampled here, forming the sister-group to *C. oedema*, are more brightly and variably colored, and all have a cheekpatch of color in the terminal male coloration that *C. oedema* lacks. The species pair *C. sordidus* and *C. japanensis* share a disjunct color pattern on the caudal peduncle that abruptly changes from yellow (*C. japanensis*) or brownish (*C. sordidus*) to a light blue peduncle flecked with yellow. This species pair was suggested to be closely related by Choat and Randall (1986), and it also shares a pattern of the dorsal and anal fins in which the base and tips of the membranes are blue, with orange between, interrupted by a central blue band (also shared by *C. capistratoides*). *C. sordidus* is one of the most broadly distributed species of parrotfishes, ranging from the Red Sea south to Natal, South Africa and east to the Hawaiian islands, north to the Ryukyu Islands, and south to Perth, Australia, whereas *C. japanensis* is restricted to the region between the Ryukyu islands and Australia.

The sister pair *C. bleekeri* + *C. bowersi* in our analysis share particularly distinctive cheek patches of green/white (*C. bleekeri*) to solid green/blue (*C. bowersi*), a double chin strap of blue, and nearly identical fin colorations. These two species have disjunct Western Pacific distributions, with *C. bleekeri* occurring in the Moluccas north to the Marshall Islands, and south to the Great Barrier Reef and Vanuatu, whereas *C. bowersi* occupies the region from Indonesia, to the Philippines, Ryukyu Islands, and Palau. *Chlorurus troschelii* is also likely a member of this clade (Choat and Randall, 1986). Finally, the sister-group *C. capistratoides* + *C. microrhinos*, is well supported by molecular data but few color pattern characters confirm this relationship in an obvious way. *Cetoscarus microrhinos* is thought to be part of a species complex of *C. gibbus* in the Red Sea, *C. strongylocephalus* in the Indian Ocean and *C. microrhinos* in the west-central Pacific (Myers, 1999), so inclusion of all members of this group in future work may clarify the evolutionary history of this subclade.

The first clade to diverge from the *Scarus* radiation is *Scarus* clade 1 (Fig. 1), represented in this study by four relatively widespread parrotfish species. *Scarus psittacus* is the sister to the rest of this *Scarus* clade in our tree, and is one of the most broadly distributed species among all parrotfishes, occurring from the Red Sea to South Africa and east to the

Hawaiian, Marquesan, and Tuamotu islands. The other three species, *S. schlegeli*, *S. flavipectoralis* and *S. hypselopterus* are primarily Western Pacific with a distribution center around the Philippines, although *S. schlegeli* reaches the eastern Indian Ocean. The species pair *S. flavipectoralis* and *S. hypselopterus* share a TP color pattern in which the anterior half (slightly more in *S. hypselopterus*) of the body is orange and the caudal half is blue.

Scarus spinus is the sister species to the rest of *Scarus* clade 2 in our phylogeny, a relatively rare species distributed in the Western Pacific and eastern Indian Ocean. The species pair *S. chameleon* + *S. festivus* is next up the backbone of clade 2, two species with largely overlapping distributions from the Ryukyu to Lord Howe in the Pacific and west to the Indian Ocean, where *S. festivus* reaches to the African coast. *S. spinus*, *S. chameleon* and *S. festivus* share a characteristic deep head profile with a moderate or pronounced bulbous lump above the eye and a short snout that is particularly parrot-like. Choat and Randall (1986) identified the close relationship between *S. chameleon* and *S. festivus* based on meristics and color pattern. The four species remaining in *Scarus* clade 2 (*S. quoyi*, *S. rivulatus*, *S. globiceps* and *S. oviceps*) all have dark green-blue terminal males with complex facial lines and patterns, and the biogeography of this clade is intriguing because all four species have widespread distributions ranging from Indian Ocean to Western Pacific. The primary phases of *S. globiceps* and *S. rivulatus* are difficult to distinguish from one another. *S. globiceps* differs in that it is the only one with 3 scale rows and 5 or more scales in the ventral row. Previous workers (Choat and Randall, 1986) have remarked on the resemblance between *S. oviceps* and *S. dimidiatus* (here placed in *Scarus* clade 4) in both IP and TP color phases, but if our phylogeny is correct this would represent color pattern convergence.

The third major *Scarus* clade recovered in our analysis is a mostly Caribbean clade of parrotfishes, but is anchored by an extremely widespread species, *S. ghobban*, that does not itself range into the Caribbean. *S. ghobban* extends from the Red Sea and South Africa, north to southern Japan, south to Perth, New South Wales, and across to the Eastern Pacific, ranging from the Gulf of California to Ecuador. Due to its widespread distribution, its occupation of non-reef and deeper water habitats (Randall, 1986), and its apparent sister-group position to the Caribbean lineage, the population biogeography of *S. ghobban* would be of considerable interest. *Scarus taeniopterus* and *S. iseri* are closely related and form the next branches up the topology of clade 3. These two species share a striped IP color pattern and similar TP coloration, as well as almost identical distributions in the Western Atlantic and Caribbean. The species pair *S. coelestinus* + *S. guacamaia* are large, often blotchily colored parrotfishes that also share Western Atlantic distributions that extend far down the coast of South America.

Scarus clade 4 is composed of eight Indo-West Pacific taxa that often have distinctive color patterns in both the initial and terminal phases. The clade formed by *S. tricolor* + (*S. forsteni* + *S. rubroviolaceus*) are a case in point, in which the initial phases of all three species are highlighted by a rosy orange/red on tail and anal fin or across the body and fins, contrasted by dark scales in black, blue or green colors at midbody. The terminal phase of *S. forsteni* and *S. rubroviolaceus* share a characteristic dark blue saddle anteriorly, with a paler green body posteriorly. Although to our knowledge *S. rubroviolaceus* has not been proposed to be related to *S. forsteni* or *S. tricolor*, Choat and Randall (1986) reviewed the taxonomic confusion between *S. forsteni* and *S. tricolor*, indicative of their close relationship. *Scarus dimidiatus* and *S. scaber* (not sampled here) are thought to be closely related (Choat and Randall, 1986), and we recovered the species pair *S. frenatus* + *S. dimidiatus*. This is biogeographically suggestive as *S. frenatus* occurs from the Red Sea to Western Australia and throughout the Western Pacific, whereas its apparent close relatives *S. dimidiatus* and *S. scaber* are split geographically, with *S. scaber* in the Indian Ocean and *S. dimidiatus* restricted to Indonesia, Samoa, north to the Ryukyu Islands, and south to the Great Barrier

Reef. To round out our clade 4, *Scarus niger* is the sister to the species pair *S. altipinnis* + *S. prasiognathos*, three species with relatively deep bodies and similarly shaped caudal fins with a nearly lunate profile in large males. *S. niger* and *S. prasiognathos* are widely distributed Indo-Pacific taxa, whereas *S. altipinnis* is restricted to the Western Pacific. *S. altipinnis* and *S. prasiognathos* are both large species that have similar initial phase color patterns, and the TP males have a distinctive blue-green patch on the chin and throat. Choat and Randall (1986) suggested that these two species together with *S. falcipinnis* of the Western Indian Ocean form a species complex.

4.2 Evolutionary trends in mitochondrial, nuclear, and nuclear regulatory genes

The higher level splits among labrid subgroups and the distinctions between parrotfish genera were supported by each of the 8 genes used here, including regulatory genes. However, the strength of phylogenetic signal at the species level was less balanced, with cytochrome b, a gene often used in population genetic studies in fishes (Carmona et al., 2000; Feulner et al., 2007), providing half of the sites informative for parsimony within the 24 *Scarus* taxa. At the species level, the “standard” phylogenetic markers (*12s*, *16s*, *tmo4c4*, and *rag2*) and regulatory gene fragments (*bmp4*, *dlx2*, and *otx1*), showed similar resolving power and similar phylogenetic structure (Fig. 2). However, regulatory partitions showed a wide range of variability in substitution rates, with *bmp4* contributing much more information at the species level than the *otx1* or *dlx2* exon partitions (Fig 3–Fig 4). This result is similar to comparisons of *bmp4*, *otx1*, and other regulatory genes in African cichlid fishes (Terai et al., 2002) in which *bmp4* showed the highest rates of both synonymous and non-synonymous change.

The utility of the regulatory genes for phylogenetic questions differs among hierarchical levels, precisely the pattern of information content that phylogeneticists may be looking for when addressing large-scale and finer scale questions in their analyses. *Bmp4* showed dS and dN levels at many nodes of the tree similar to that of *tmo4c4* (Table 2), a gene fragment which has been popular for family-level phylogenetic studies of fishes (e.g., Near et al., 2004; Sparks and Smith, 2004; Westneat and Alfaro, 2005) since its characterization by Streelman and Karl (1997). We conclude that *bmp4* sequence represents a useful additional nuclear gene that will be informative in studies of family-level systematics in fishes. In contrast, *otx1* accumulates mutations much more slowly than *bmp4* and so may be more valuable in studies with a greater time depth, for example among families within the Perciformes, a problem gaining increased attention with a wide array of genetic and genomic data (Miya et al., 2003; Smith and Craig, 2007). Similarly, the two *dlx2* exon fragments that we sampled evolve slowly, no doubt in part because almost half of our nucleotides are from sequence of the homeodomain. The intervening intron sequence, however, displays considerable variation among parrotfishes (Fig. 5B), but alignment with the other labrid outgroups was problematic. This segment is perhaps best suited to phylogenetic investigation among closely related species or perhaps even populations, but it remains to be seen if other taxa will display such significant divergence in this region.

Phylogenetic analysis of substitution patterns in protein coding genes can reveal regions of both the genes and the phylogenies in which an increase in nonsynonymous change serves as a signal for selection (Guindon et al., 2004; Nei, 2005; Seo et al., 2004). All of the genes examined here showed higher rates of synonymous than non-synonymous change, resulting in dN/dS ratios well below 0.5, and usually below 0.1 (Table 2), suggesting that a process of purifying or stabilizing selection has acted on these molecules. A single exception is the rate ratio for *tmo4c4* at the reef clade of parrotfishes, which is greater than 1.0 (Table 2), indicating that there was a burst of nonsynonymous change along this branch that may be indicative of positive selection for protein evolution in this gene. Despite an overall low substitution rate for regulatory genes, it is suggestive that dS and dN were highly variable on

the phylogeny, and that the branches leading to the cheiline and scarine subfamily nodes had the highest rates of nucleotide substitution, including both dS and dN. For example, the *bmp4* dS and dN trees (Fig. 8) illustrate this variability, with long branch lengths to the cheiline and scarine clades. Also, it is noteworthy that the dN/dS rate ratio of *dlx2* leading to the parrotfishes (0.33) is the highest of all six genes along this part of the phylogeny. An elevated level of variation in regulatory genes has often been shown to be associated with morphological diversity in plants (Aagard et al., 2006; Barrier et al., 2001). Recent exploration of skull function of labrid fishes in the context of a phylogeny (Westneat et al., 2005) has shown that the cheiline and scarine labrids are a hot-spot of structural and functional evolution in these fishes, and *dlx2* is known to regulate formation of pharyngeal elements of the feeding apparatus during development (Qiu et al., 1997; Stock et al., 1996). The discovery of elevated dN/dS ratios at major nodes of the parrotfishes for *dlx2* and *bmp4* (Table 2) suggests that these phylogenetic branches may be an intriguing place to search for changes in developmental patterns in the skull and pharyngeal arches. If the present study can be broadened to include multiple clades of fishes with differential levels of morphological and mechanical divergence, we may discover whether evolutionary changes in functional morphology are correlated with higher rates of change in the developmentally important signaling molecules produced by regulatory genes.

4.3 Divergence times in parrotfish evolution

The evolutionary age of the parrotfishes appears to be younger than ages previously proposed for parrotfishes and for some other reef fish clades with late Oligocene and early Miocene crown group ages (Alfaro et al., 2007; Bernardi et al., 2000; McCafferty et al., 2002; Streelman et al., 2002; Klanten et al., 2004). The two approaches to age estimation (ML and Bayesian) converged on ages of 17–18 mya for the root scarine node, much more recently than the 42 million year age arrived at using a transversion rate of 0.14% per million years for *12s* and *16s* genes (Streelman et al., 2002). The age estimates of major nodes in the parrotfish phylogeny depend upon the assigned age of the root node, the frequency and pattern of nucleotide substitutions, and the assumptions of the algorithms used to reconstruct nodal ages throughout the tree. The estimate of root node age for the Labridae at 55 mya and a minimum age for parrotfishes at 15 mya are the only independent estimates of age used, based on the fossil labrid genus *Phyllopharyngodon* (Bellwood, 1990) from Monte Bolca and a fossil *Calotomus* from the middle Miocene (Bellwood and Schultz, 1991). The origin of the root of the Labridae may in fact be anywhere from 50 to as old as 90 ma, according to estimates based on plate tectonics, dating of reef fish lineages with molecular clocks and patterns of fish otolith preservation (Bellwood and Wainwright, 2002). Sensitivity analyses reveal that, for example, constraining the labrid root to be 75 mya effectively multiplies the age estimates of Table 3 by a factor of about 1.5, and such scenarios should be considered until the dates of labrid origin are more certain.

We estimated the seagrass clade to be 12–15 my old, with the base of the Caribbean clade to be 9.75–11.6 my old. Bernardi *et al.* (2000) used *12s* and *16s* mitochondrial rDNA and a rate of 0.14% transition divergence per million years (Ritchie et al., 1996) to place the origin of *Sparisoma* at 14.3–35 million years ago. A more recent analysis of Caribbean *Nicholsina* and *Sparisoma* (Robertson et al., 2006), placed the age of the most common ancestor of these two genera in a range of 10–25 my, based on divergence rates of cytochrome b from 1–2.5% per million years. Our reconstruction would suggest that their scenarios for *Sparisoma* biogeography using the faster end of the substitution rate scale and the more recent origin and diversification dates for *Sparisoma* may be correct. This is intriguing because it places the early splits among the Caribbean seagrass clade taxa fairly shortly after the closure of the Tethys Sea 12–18 mya and nearly synchronously with the formation of the Amazon barrier, about 11 mya (Hoorn, 1996). As Robertson *et al.* (2006) illustrate using

Sparisoma diversification, the Amazon outflow is associated with species boundaries and genetic differentiation of trans-Amazon species pairs.

The clade of reef dwelling parrotfishes is estimated to have begun diversifying 14–16 mya, also in the temporal vicinity of the Tethyan closure events. However, the split between *Chlorurus* and *Scarus* appears to be considerably more recent, at 6.0–7.4 mya, and the genera *Chlorurus* and *Scarus* underwent most of their diversification within the past 3–5 million years (Fig. 9). It is thought provoking to consider that the global distribution of all currently living *Scarus* have diverged from one another in just the past 5 million years, and even more recently for some clades. *Scarus* clades 2, 3, and 4 (Fig. 1, Fig. 8) have root nodes at between 2 and 3 mya, thus implying that most *Scarus* species are recent products of speciation. A particularly informative age estimate is that of *Scarus* clade 3, including *S. ghobban* plus the Caribbean clade, at 2.4–3.5 mya. This range includes the likely date of the complete closure of the isthmus of Panama at approximately 3.1 – 3.5 mya (Coates and Obando, 1996), a possible scenario for the isolation of the Caribbean lineage from the ancestors of the widespread species *S. ghobban* that ranges to the Eastern Pacific today.

The pantropical distribution and the relatively recent ages of divergence of the 4 main clades of *Scarus* imply that fluctuations in sea level (Haq et al., 1987) and patterns of differential cooling of the oceans during the Pliocene and Pleistocene may be the driving forces behind the rapid radiation in this genus, that today is largely restricted to the complex reefs built by hard corals. This is similar to the pattern of rapid speciation in Caribbean *Sparisoma* that took place 2.8–5.6 million years ago (Bernardi et al., 2000; Robertson et al., 2006), corresponding with periods of sealevel fluctuation and extinction of corals in the Caribbean 1–4 million years ago (Johnson et al., 1996). Alternatively, processes of ecological speciation and divergence due to sexual selection remain a possible explanation for the rapid radiation of parrotfishes, which all have pelagic larval phases and highly similar morphology, differing primarily in color pattern and behavior. Isolation of populations may be driven by local environmental or behavioral factors rather than large-scale geographic boundaries, and Rocha *et al.* (2005) outline several clear cases of potential ecological speciation in labrid fishes. In addition, the protogynous mating system of parrotfishes, where species aggregate and have male-dominated harem systems organized by color recognition, has been proposed as a possible driving force for speciation via sexual selection mechanisms in this group (Streelman et al., 2002). Intriguingly, recent work (Boughman et al., 2005) has found that ecological and sexual selection factors may operate in tandem to drive species divergence. To test the various alternative and possibly interactive mechanisms of allopatric vs. local speciation will require population genetic sampling of recently evolved species pairs such as *S. altipinnis* + *S. prasiognathos* and *S. festivus* + *S. chameleon* (Table 3), each with estimated ages at less than 1 million years. Detailed species sampling with inclusion of population samples, particularly for widespread species, will help resolve long-standing questions of both phylogenetic pattern and evolutionary process in this charismatic group of reef fishes.

Acknowledgments

Field work for specimen collection was aided by the Lizard Island Research Station, Bureau of Fisheries and Aquatic Research of the Philippines, Mark McGrouther, Jeff Leis, Randy Mooi, Tom Trinski, Jeff Williams, Kent Carpenter, Luz Regis, Jeff Janovetz, Brad Wright, Aaron Rice, Jim Cooper, and others on expeditions to the northern Great Barrier Reef, Vanuatu, Solomon Islands, and the Philippines. DNA sequencing was performed in the Pritzker Laboratory of Molecular Evolution at the Field Museum of Natural History. Thanks to Rob Vogelbacher and David Angulo for access to the Illinois BioGrid through DePaul University. This research was primarily funded by NSF 0235307 to M. Westneat and an NSF/Alfred P. Sloan Foundation Postdoctoral Fellowship (NSF DBI 9803946) grant to J. T. Streelman, with support during analysis and manuscript preparation by NSF DEB 0445453 to M. Alfaro and by John D. and Catherine T. MacArthur Foundation support of the Encyclopedia of Life.

Literature Cited

- Aagaard JE, Willis JH, Phillips PC. Relaxed selection among duplicate floral regulatory genes in Lamiales. *J. Mol. Evol.* 2006; 63:493–503. [PubMed: 17021928]
- Abzhanov A, Protas M, Grant R, Grant PR, Tabin CJ. *Bmp4* and morphological variation of beaks in Darwin's finches. *Science*. 2004; 305:1462–1465. [PubMed: 15353802]
- Acampora D, Avantsaggiato V, Tuorto F, Briata P, Corte G, Simeone A. Visceral endoderm-restricted translation of *otx1* mediates recovery of *otx2* requirements for specification of anterior neural plate and normal gastrulation. *Development*. 1998; 125:5091–5104. [PubMed: 9811592]
- Albertson RC, Streelman JT, Kocher TD. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *PNAS*. 2003; 100:5252–5257. [PubMed: 12704237]
- Albertson RC, Streelman JT, Kocher TD, Yelick PC. Integration and evolution of the cichlid mandible: the molecular basis of alternative feeding strategies. *PNAS*. 2005; 102:16287–16292. [PubMed: 16251275]
- Alfaro ME, Santini F, Brock CD. Do reefs drive diversification in marine teleosts? Evidence from the pufferfishes and their allies (Order Tetraodontiformes). *Evolution*. 2007; 61–69:2104–2126.
- Andreazzoli M, Pannese M, Boncinelli E. Activating and repressing signals in head development: the role of *Xotx1* and *Xotx2*. *Development*. 1997; 124:1733–1743. [PubMed: 9165121]
- Barrier M, Robichaux RH, Purugganan MD. Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Natl. Acad. Sci. USA*. 2001; 98:10208–10213. [PubMed: 11517318]
- Bellwood DR. A new fossil fish *Phyllopharyngodon longipinnis* gen. et sp. nov. (Family Labridae) from the Eocene, Monte Bolca, Italy. *Studi e Ricerche sui Giacimenti Terziari di Bolca, Museo Civico di Storia Naturale di Verona*. 1990; 6:149–160.
- Bellwood DR. A phylogenetic study of the parrotfishes family Scaridae (Pisces: Labroidei), with a revision of genera. *Rec. Austr. Mus. Suppl.* 1994; 20:1–86.
- Bellwood DR, Schultz O. A review of the fossil record of the parrotfishes (Labroidei: Scaridae) with a description of a new *Calotomus* species from the middle Miocene (Badenian) of Austria. *Ann. Naturhist. Mus. Wien*. 1991; 92:55–71.
- Bellwood, DR.; Wainwright, PC. The history and biogeography of fishes on coral reefs. In: Sale, PF., editor. *Coral reef fishes: dynamics and diversity in a complex ecosystem*. San Diego: Academic Press; 2002. p. 5-32.
- Bernardi G, Robertson DR, Clifton KE, Azzurro E. Molecular systematics, zoogeography, and evolutionary ecology of the Atlantic parrotfish genus *Sparisoma*. *Mol. Phylogenet. Evol.* 2000; 15:292–300. [PubMed: 10837158]
- Boncinelli E, Gulisano M, Broccoli V. *Emx* and *Otx* homeobox genes in the developing mouse brain. *J. Neurobiol.* 1993; 24:1356–1366. [PubMed: 7901323]
- Boughman JW, Rundle HD, Schluter D. Parallel evolution of sexual isolation in sticklebacks. *Evolution*. 2005; 59:361–373. [PubMed: 15807421]
- Carmona JE, Domínguez J, Doadrio I. Congruence between allozyme and cytochrome b gene sequence data in assessing genetic differentiation within the Iberian endemic *Chondrostoma lemmingii* (Pisces: Cyprinidae). *Nature*. 2000; 84:721–732.
- Cheng H, Jiang W, Phillips FM, Haydon RC, Peng Y, Zhou L, Luu HH, An N, Breyer B, Vanichakarn P, Szatkowski JP, Park JY, He T-C. Osteogenic activity of the fourteen types of human bone morphogenetic proteins (BMPs). *J. Bone Joint Surg. A*. 2003; 85:1544–1552.
- Chen JN, van Eeden FJ, Warren KS, Chin A, Nusslein-Volhard C, Haffter P, Fishman MC. Left-right pattern of cardiac *bmp4* may drive asymmetry of the heart in zebrafish. *Development*. 1997; 124:4373–4382. [PubMed: 9334285]
- Choat JH, Randall JE. A review of the parrotfishes (family Scaridae) of the Great Barrier Reef of Australia with description of a new species. *Rec. Australian Museum*. 1986; 38(4):175–228.
- Clements KD, Alfaro ME, Fessler JL, Westneat MW. Relationships of the temperate Australasian labrid tribe Odacini. *Mol. Phylogenet. Evol.* 2004; 32:575–587. [PubMed: 15223039]
- Coates, AG.; Obando, JA. The geologic evolution of the central American Isthmus. In: Jackson, JBC.; Budd, AF.; Coates, AG., editors. *Evolution and Environment in Tropical America*. Univ. Chicago: Chicago Press; 1996. p. 21-56.

- Cook CE, Smith ML, Telford MJ, Bastianello A, Akam M. Hox genes and the phylogeny of the arthropods. *Curr. Biol.* 2001; 11:759–763. [PubMed: 11378385]
- Cuomo CA, Oettinger MA. Analysis of regions of *rag-2* important for V(D)J recombination. *Nucleic Acids Res.* 1994; 22:1810–1814. [PubMed: 8208604]
- Esposti MD, De Vries S, Crimi M, Ghelli A, Patarnello T, Meyer A. Mitochondrial cytochrome b: evolution and structure of the protein. *Biochim. Biophys. Acta.* 1993; 1143:243–271. [PubMed: 8329437]
- Fan C, Purugganan MD, Thomas DT, Wiegmann BM, Xiang QY. Heterogeneous evolution of the Myc-like Anthocyanin regulatory gene and its phylogenetic utility in *Cornus* L. (Cornaceae). *Mol. Phylogenet. Evol.* 2004; 33:580–594. [PubMed: 15522789]
- Farias IP, Ortí G, Meyer A. Total evidence: molecules, morphology, and the phylogenetics of cichlid fishes. *J. Exp. Zool.* 2000; 288:76–92. [PubMed: 10750055]
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 1985; 39:783–791.
- Feulner PGD, Kirschbaum F, Mamonekene V, Ketmaier V, Tiedemann R. Adaptive radiation in African weakly electric fish (Teleostei: Mormyridae: *Campylomormyrus*): a combined molecular and morphological approach. *J. Evol. Biol.* 2007; 20:403–414. [PubMed: 17210033]
- Froese, R.; Pauly, D. A Global Information System on Fishes: concepts, design and data sources. Los Baños, Laguna, Philippines: ICLARM; 2000. FishBase 2000; p. 344(www.fishbase.org)
- Goldman N, Yang ZH. Codon-based model of nucleotide substitution for protein-coding DNA-sequences. *Mol. Biol. Evol.* 1994; 11:725–736. [PubMed: 7968486]
- Guindon S, Rodrigo AG, Dyer KA, Huelsenbeck JP. Modeling the site-specific variation of selection patterns along lineages. *PNAS.* 2004; 101:12957–12962. [PubMed: 15326304]
- Hansen JD, Kaattari SL. The recombination activating gene 2 (*rag2*) of the rainbow trout *Oncorhynchus mykiss*. *Immunogenetics.* 1996; 44:203–211. [PubMed: 8662087]
- Haq BU, Hardenbol J, Vail PR. Chronology of fluctuating sealevels since the Triassic. *Science.* 1987; 235:1156–1167. [PubMed: 17818978]
- Holland N, Panganiban G, Henyey E, Holland LZ. Sequence and developmental expression of *AmphiDll*, an amphioxus *Distalless* gene transcribed in the ectoderm, epidermis and nervous system: insights into the evolution and forebrain of craniates. *Development.* 1996; 122:2911–2920. [PubMed: 8787764]
- Hoorn C. Miocene deposits in the Amazon foreland basin. *Science.* 1996; 273:122–125.
- Irish VF, Gelbart WM. The decapentaplegic gene is required for dorsoventral patterning of the *Drosophila* embryo. *Genes Dev.* 1987; 1:868–879. [PubMed: 3123323]
- Johnson KG, Budd AF, Stemmann TA. Extinction selectivity and ecology of Neogene Caribbean corals. *Paleobiology.* 1995; 21:52–73.
- Kim C-B, Shashikant CS, Sumiyama K, Wang WCH, Amemiya CT, Ruddle FH. Phylogenetic analysis of the mammalian *Hoxc8* non-coding region. *J. Struc. Func. Genomics.* 2003; 3:195–199.
- Kingsley DM. The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. *Genes and Development.* 1994; 8:133–146. [PubMed: 8299934]
- Klanten SO, Van Herwerden L, Choat JH, Blair D. Patterns of lineage diversification in the genus *Naso* (Acanthuridae). *Mol Phylogenet Evol.* 2004; 32:221–235. [PubMed: 15186809]
- Kosakovsky-Pond SL, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* 2005; 21:676–679. [PubMed: 15509596]
- Lovejoy NR. Reinterpreting recapitulation: Systematics of needlefishes and their allies (Teleostei:Beloniformes). *Evolution.* 2000; 54:1349–1362. [PubMed: 11005301]
- Lowe CJ, Wray GA. Radical alterations in the roles of homeobox genes during echinoderm evolution. *Nature.* 1997; 389:718–721. [PubMed: 9338781]
- Maddison DR. The discovery and importance of multiple islands of most-parsimonious trees. *Syst. Zool.* 1991; 40:315–323.
- Manzanares M, Wada H, Itasaki N, Trainor PA, Krumlauf R, Holland PW. Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head. *Nature.* 2000; 408:854–857. [PubMed: 11130723]

- McCafferty S, Bermingham E, Quenouille B, Planes S, Hoelzer G, Asoh K. Historical biogeography and molecular systematics of the Indo-Pacific genus *Dascyllus* (Teleostei: Pomacentridae). *Mol. Ecol.* 2004; 11:1377–1392. [PubMed: 12144659]
- McGuinness T, Porteus MH, Smiga S, Bulfone A, Kingsley C, Qiu M, Liu JK, Long JE, Xu D, Rubenstein JL. Sequence, organization, and transcription of the *Dlx-1* and *Dlx-2* locus. *Genomics.* 1996; 35:473–485. [PubMed: 8812481]
- Myers, RF. Micronesian reef fishes: a comprehensive guide to the coral reef fishes of Micronesia. 3rd revised and expanded edition. Barrigada, Guam: Coral Graphics; 1999. p. 330
- Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM, Nishida M. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 2003; 26:121–138. [PubMed: 12470944]
- Morsli H, Tuorto F, Choo D, Postiglione MP, Simeone A, Wu DK. *otx1* and *Otx2* activities are required for the normal development of the mouse inner ear. *Development.* 1999; 126:2335–2343. [PubMed: 10225993]
- Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with applications to the chloroplast genome. *Mol. Biol. Evol.* 1994; 11:715–725. [PubMed: 7968485]
- Near TJ, Bolnick DI, Wainwright PC. Investigating phylogenetic relationships of sunfishes and black basses (Actinopterygii: Centrarchidae) using DNA sequences from mitochondrial and nuclear genes. *Molec. Phylogenet. Evol.* 2004; 32:344–357. [PubMed: 15186819]
- Nei M. Selectionism and neutralism in molecular evolution. *Mol. Biol. Evol.* 2005; 22:2318–2342. [PubMed: 16120807]
- Nikaido M, Tada M, Saji T, Ueno N. Conservation of BMP signaling in zebrafish mesoderm patterning. *Mech. Dev.* 1997; 61:75–88. [PubMed: 9076679]
- Oettinger MA, Schatz DG, Gorka C, Baltimore D. Rag-1 and Rag-2, adjacent genes that synergistically activate V(D)J recombination. *Science.* 1990; 248:1517–1523. [PubMed: 2360047]
- Parenti P, Randall JE. An annotated checklist of the species of the labroid fish families Labridae and Scaridae. *Ichthyol Bull. (J.L.B. Smith Inst.).* 2000; 68:1–97.
- Peters H, Balling R. Teeth: where and how to make them. *Trends Genet.* 1999; 15:59–65. [PubMed: 10098408]
- Posada D, Crandall KA. Modeltest: Testing the model of DNA substitution. *Bioinformatics.* 1998; 14:817–818. [PubMed: 9918953]
- Purugganan MD. The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. *J. Molec. Evol.* 1997; 45:392–396. [PubMed: 9321418]
- Qiu M, Bulfone A, Ghattas I, Meneses JJ, Christensen L, Sharpe PT, Presley R, Pedersen RA, Rubenstein JL. Role of the *Dlx* homeobox genes in proximodistal patterning of the branchial arches: mutations of *Dlx-1*, *Dlx-2*, and *Dlx-1* and *-2* alter morphogenesis of proximal skeletal and soft tissue structures derived from the first and second arches. *Dev. Biol.* 1997; 185:165–184. [PubMed: 9187081]
- Randall, JE. Scaridae. In: Smith, MM.; Heemstra, PC., editors. *Smiths' sea fishes*. Berlin: Springer-Verlag; 1986. p. 706-714.
- Ritchie PA, Bargelloni L, Meyer A, Taylor JA, Macdonald JA, Lambert DM. Mitochondrial phylogeny of trematomid fishes (Nototheniidae, Perciformes) and the evolution of Antarctic fish. *Mol. Phylogenet. Evol.* 1996; 5:383–390. [PubMed: 8728396]
- Robertson DR, Karg F, de Moura RL, Victor BC, Bernardi G. Mechanisms of faunal enrichment in Atlantic parrotfishes. *Mol. Phylogenet. Evol.* 2006; 40:795–807.
- Rocha LA, Robertson DR, Roman J, Bowen BW. Ecological speciation in tropical reef fishes. *Proc. R. Soc. B.* 2005; 272:573–579.
- Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003 Aug 12; 19(12):1572–1574. [PubMed: 12912839]
- Sanderson MJ. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 1997; 14:1218–1231.

- Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 2002; 19:101–109. [PubMed: 11752195]
- Santini S, Bernardi G. Organization and base composition of *Tilapia* Hox genes: implications for the evolution of Hox clusters in fish. *Gene*. 2005; 346:51–61. [PubMed: 15716008]
- Seo T-K, Thorne JL, Kishino H. Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences. *Mol. Biol. Evol.* 2004; 21:1201–1213. [PubMed: 15014159]
- Simeone A, Puelles E, Acampora D. The Otx family. *Curr Opin. Genet. Dev.* 2002; 12:409–415. [PubMed: 12100885]
- Smith WL, Craig MT. Casting the percomorph net widely: The importance of broad taxonomic sampling in the search for the placement of serranid and percoid fishes. *Copeia*. 2007:35–55.
- Smith WL, Wheeler WC. Polyphyly of the mail-cheeked fishes (Teleostei: Scorpaeniformes): evidence from mitochondrial and nuclear sequence data. *Mol. Phylogenet. Evol.* 2004; 32:627–646.
- Sparks JS, Smith WL. Phylogeny and biogeography of the Malagasy and Australasian rainbowfishes (Teleostei: Melanotaenioidei): Gondwanan vicariance and evolution in freshwater. *Mol. Phylogenet. Evol.* 2004; 33:719–734. [PubMed: 15522799]
- Stock DW, Ellies DL, Zhao Z, Ekker M, Ruddle FR, Weiss KM. The evolution of the vertebrate Dlx gene family. *Proc. Natl. Acad. Sci. USA*. 1996; 93:10858–10863. [PubMed: 8855272]
- Streelman JT, Alfaro M, Westneat MW, Bellwood DR, Karl SA. Evolutionary history of the parrotfishes: biogeography, ecomorphology, and comparative diversity. *Evolution*. 2002; 56:961–971. [PubMed: 12093031]
- Streelman TJ, Karl SA. Reconstructing labroid evolution with single-copy nuclear DNA. *Proc. R. Soc. Lond. B*. 1997; 264:1011–1020.
- Swofford, DL. PAUP* 4.00: Phylogenetic Analysis Using Parsimony (*and other methods) Version 4.0. Mass: Sinauer Assoc.; 2000.
- Terai Y, Morikawa N, Okada N. The evolution of the pro-domain of bone morphogenetic protein 4 (*bmp4*) in an explosively speciated lineage of East African cichlid fishes. *Mol. Biol. Evol.* 2002; 19:1628–1632. [PubMed: 12200490]
- Tucker AS, Matthews KL, Sharpe PT. Transformation of tooth type induced by inhibition of BMP signaling. *Science*. 1998; 282:1136–1138. [PubMed: 9804553]
- Wagner GP, Amemiya C, Ruddle F. Hox cluster duplications and the opportunity for evolutionary novelties. *PNAS*. 2003; 100:14603–14606. [PubMed: 14638945]
- Wagner GP, Takahashi K, Lynch V, Prohaska SJ, Fried C, Stadler PF, Amemiya C. Molecular evolution of duplicated ray finned fish HoxA clusters: Increased synonymous substitution rate and asymmetrical co-divergence of coding and non-coding sequences. *J. Mol. Evol.* 2005; 60:665–676. [PubMed: 15983874]
- Wainwright PC, Bellwood DR, Westneat MW, Grubich JR, Hoey AS. A functional morphospace for labrid fishes: patterns of diversity in a complex biomechanical system. *Biol. J. Linn. Soc.* 2004; 82:1–25.
- Westneat MW. Feeding, function, and phylogeny: Analysis of historical biomechanics and ecology in labrid fishes using comparative methods. *Syst. Biol.* 1995; 44:361–383.
- Westneat MW, Alfaro ME, Wainwright PC, Bellwood DR, Grubich JR, Fessler J, Clements KD, Smith L. Local phylogenetic divergence and global evolutionary convergence of skull biomechanics in reef fishes of the family Labridae. *Proc. Roy. Soc. B*. 2005; 272:993–1000.
- Westneat MW, Alfaro ME. Phylogenetic relationships and evolutionary history of the reef fish family Labridae. *Mol. Phylogenet. Evol.* 2005; 36:370–390. [PubMed: 15955516]

Figure 1.

Phylogenetic hypothesis for the parrotfishes calculated using Bayesian inference. Maximum posterior probability topology resulting from 4 runs of a MCMC analysis each with 6 chains running for 10 million generations. Posterior probabilities are shown for each node (computed from majority rule consensus of post burn-in trees), and the black dots indicate support at the 90% level or greater from a 1000 repetition parsimony bootstrap analysis. The root of the seagrass and reef clades and four clades of the genus *Scarus* are identified. Photos are by John E. Randall (via Fishbase.org) with permission, identified by initials of genus and species.

Figure 2.

A comparison of phylogenetic signal in the data partitioned by regulatory (*otx1*, *dlx2*, *bmp4*) and non-regulatory (*12s*, *16s*, *cytb*, *tmo4c4*, *rag2*) genes. Topologies illustrated are the majority-rule consensus tree of the 95% credible set of trees from 3 runs of a MCMC analysis each with 5 chains running for 10 million generations. Black dots (•) indicate strongly supported nodes at which both partitioned trees agree with each other and with the combined data tree. Strongly supported nodes that agree with the combined data tree but are not strongly supported in the other partition tree (+) are less frequent, and strongly supported nodes that disagree with the combined data tree (*) are rare.

Figure 3. Phylogram of rates of molecular evolution in the (A) 12s and (B) cytochrome b data partitions illustrated on the Bayesian consensus tree of all concatenated sequences. Numbers on selected branches indicate that the branchlength is N times the length shown.

Figure 4. Rates of molecular evolution in the (A) *rag2* and (B) *bmp4* data partitions illustrated on the Bayesian consensus tree of all concatenated sequences. Numbers on selected branches indicate that the branchlength is N times the length shown.

Figure 5. Rates of molecular evolution in the (A) *dlx2* exon and (B) *dlx2* intron data partitions illustrated on the Bayesian consensus tree of all concatenated sequences. Numbers on selected branches indicate that the branchlength is N times the length shown.

Figure 6. Synonymy (A) and non-synonymous (B) substitution rate trees for cytochrome b, calculated using the Bayesian consensus topology and the Muse-Gaut (1994) codon substitution model implemented in HyPhy (Kosokovsky Pond et al., 2005).

Figure 7. Synonymous (A) and non-synonymous (B) substitution rate trees for *rag2*, calculated using the Bayesian consensus topology and the Muse-Gaut (1994) codon substitution model.

Figure 8. Synonymous (A) and non-synonymous (B) substitution rate trees for *bmp4*, calculated using the Bayesian consensus topology and the Muse-Gaut (1994) codon substitution model.

Figure 9.

Chronogram illustrating the ages of origin and diversification of parrotfish phylogenetic groups, calculated using the Bayesian inference topology and the Bayesian posterior probability algorithms for mean and variance of nodal ages implemented in CodonRates 1.0 (Seo et al., 2004) assuming a root age of the labrid outgroups of 55 million years ago.

Table 1

Primers and primer sequences used for amplification and sequencing in this study.

Gene	Primer name	Primer sequence
12s rDNA	12s-phe22	AGC ATA ACA CTG AAG ATG TTA
	12s53F	CAC AAA GGC TTG GTC CTG ACT TT
	12s489F	CTG GGA TTA GAT ACC CCA CTA TGC
	12s613R	TCG GTT CTA GAA CAG GCT CCT CTA G
	12s991R	GGT ACA CTT ACC ATG TTA CGA CT
16s rDNA	16SAR	CGC CTG TTT ATC AAA AAC AT (Palumbi, 1996)
	16SBR	CCG GTC TGA ACT CAG ATC ACG T (Palumbi, 1996)
tmo4C4	Tmo-f1-5	CCT CCG GCC TTC CTA AAA CCT CTC (Streelman and Karl, 1997)
	Tmo-f1-6	GAA AAG AGT GTT TGA AAA TGA
	Tmo-r1-3	CAT CGT GCT CCT GGG TGA CAA AGT (Streelman and Karl, 1997)
rag2	RAG2-pfish38F	GGC TCC CAG AGA GTT ACC TCA THC A
	RAG2-F1	GAT GGC CTT CCC TCT GTG GGT AC
	RAG2-pfish407F	GGT CAC TCC ATC CTG TCC GAC T
	RAG2-pfish535R	CTC ACA GGA RAG CAA AGG GCT ACC C
	RAG2-R3	GAT GGC CTT CCC TCT GTG GGT AC
cytb	Lcytb-ch	TTC TAA AAT CGC AAA CGA CGC
	LcytbScar1	GAA ACC ACC GTT GTT ATT CAA CTA
	LcytbScar2	TAA TGG CAA GTC TCC GAA AAA CTC
	Cytb R580	AGR AAG TGG AAK GCR AAG AAT CG
	HcytbScar615	GGA CTA CGG CAG CTG CRA TGA C
	HcytbScar615b	AGA ACT ACG GCA GCT GCR ATA AC
	HcytbScar750	GAG AGC AAT AAG GAG GGC TGC
otx1	Otx1-F1	TCC AAG CAG TCT GTG TGG TGT TAA A
	Otx1-R1	TTT GGT TCA AGA ACC GGA GG
bmp4	Bmp4-2Fa	TCT YAT YTC AGA GCA CAT GGA GAG G
	Bmp4-2Fb	AAC CTC ACC AGC ATT CCA GA
	Bmp4-2R	ATC GCT GAA GTC CAC GTA C
dlx2	Dlx2-F760	GAA GAG AGY GAG CCA GAA ATC
	Dlx2-R2	AGT TTG CCA AAA ACG ACG AA

Table 2

Results of analysis of synonymous and non-synonymous change in 6 protein coding genes at the major nodes in the phylogeny of the parrotfishes calculated using the codon model of Muse and Gaut (1994). Genes are not listed if all cell entries were zero and ratios are left blank if undefined.

Node	Gene	dS	dN	dN/dS
Cheilines	<i>cytb</i>	0.1742	0.0124	0.071
	<i>rag2</i>	0.2154	0.0328	$\wedge$ 0.152
	<i>tmo4c4</i>	0.2590	0.0085	0.033
	<i>bmp4</i>	0.2269	0.0139	0.061
	<i>dlx2</i>	0.0228	0	
	<i>otx1</i>	0.1055	0	
Scarines	<i>cytb</i>	0.2805	0.0053	0.019
	<i>rag2</i>	0.3676	0.0325	0.088
	<i>tmo4c4</i>	0.4053	0.0056	*0.014
	<i>bmp4</i>	0.2925	0.0085	*0.029
	<i>dlx2</i>	0.0353	0.0117	$\wedge$ 0.331
	<i>otx1</i>	0.2102	0.0020	0.009
Seagrass	<i>rag2</i>	0.0505	0.0046	0.090
	<i>tmo4c4</i>	0.0405	0.0085	0.210
	<i>bmp4</i>	0.0185	0.0057	$\wedge$ 0.309
	<i>otx1</i>	0.0609	0	
Reef	<i>cytb</i>	0.1359	0	
	<i>rag2</i>	0.0000	0.0039	
	<i>tmo4c4</i>	0.0023	0.0028	$\wedge$ 1.215
Chlorurus +Scarus	<i>cytb</i>	0.0647	0.0032	0.049
	<i>rag2</i>	0.0462	0.0023	0.049
	<i>tmo4c4</i>	0.0535	0.0028	0.053
	<i>bmp4</i>	0.0411	0.0057	0.139
	<i>dlx2</i>	0.0114	0	
	<i>otx1</i>	0.0483	0	
Chlorurus	<i>cytb</i>	0.3327	0	
	<i>rag2</i>	0.0294	0.0039	0.134
	<i>tmo4c4</i>	0.0783	0	
	<i>bmp4</i>	0.0550	0	
	<i>dlx2</i>	0.0115	0	
	<i>otx1</i>	0.0037	0	
Scarus	<i>cytb</i>	0.0412	0	
	<i>rag2</i>	0.0142	0.0020	0.138
	<i>tmo4c4</i>	0.0582	0	

Node	Gene	dS	dN	dN/dS
	<i>otx1</i>	0.0250	0	

Branches with a significantly higher (^) or lower(*) dN/dS ratio than the rest of the phylogeny are indicated.

Table 3

Lineage rate/time analysis for the scarine phylogeny, showing the estimated ages for major clades using penalized likelihood, PL (r8s software, Sanderson, 2001) and Bayesian analysis incorporating a codon model, CR (CodonRates, Seo *et al.*, 2004).

Node	PL Age (mya)	Local Rate	CR Age (mya)	St. Dev.
Labridae (fixed)	55.00		56.43	5.19
Scarines + Cheilines	35.78	0.0074	51.09	5.07
Scarini (min 15 mya)	17.29	0.0076	18.01	2.99
Seagrass Clade	14.52	0.0092	12.63	2.34
<i>Cryptotomus</i> + <i>Nicholsina</i> + <i>Sparisoma</i>	11.66	0.0083	9.75	1.98
<i>Cryptotomus</i> + <i>Nicholsina</i>	8.01	0.0083	6.63	1.56
<i>Calotomus carolinus</i> + <i>spinidens</i>	0.57	0.0062	0.05	0.06
<i>Sparisoma</i>	3.99	0.0056	1.77	0.52
<i>Sparisoma chrysotermum</i> + <i>viride</i>	3.54	0.0028	1.36	0.46
Reef Clade	14.43	0.0034	16.39	2.89
<i>Bolbometopon</i> + <i>Cetoscarus</i>	9.19	0.0031	10.34	2.25
<i>Chlorurus</i> + <i>Scarus</i>	7.36	0.0052	5.95	1.36
<i>Chlorurus</i>	3.05	0.0040	2.09	0.64
<i>Chlorurus bleekeri</i> + <i>bowersi</i>	1.30	0.0062	0.99	0.38
<i>Chlorurus capistratoides</i> + <i>microrhinos</i>	0.70	0.0035	0.08	0.06
<i>Scarus</i>	5.68	0.0061	4.54	1.08
<i>Scarus</i> Clade 1	4.47	0.0043	3.56	0.94
<i>Scarus flavipectoralis</i> + <i>hypslepterus</i>	1.65	0.0066	1.31	0.44
<i>Scarus</i> Clade 2	3.70	0.0046	3.36	0.91
<i>Scarus festivus</i> + <i>chameleon</i>	0.46	0.0038	0.21	0.05
<i>Scarus globiceps</i> + <i>oviceps</i>	0.24	0.0025	0.10	0.09
<i>Scarus</i> Clade 3 (Caribbean + <i>ghobban</i>)	3.48	0.0050	2.43	0.68
<i>Scarus coelstinus</i> + <i>guacamaia</i>	1.07	0.0036	0.54	0.24
<i>Scarus</i> Clade 4	3.29	0.0064	2.26	0.64
<i>Scarus dimidiatus</i> + <i>frenatus</i>	1.97	0.0033	1.55	0.50
<i>Scarus altipinnis</i> + <i>prasiognathos</i>	0.73	0.0028	0.20	0.15