



The complete mitochondrial genome of the black surfperch, *Embiotoca jacksoni*: Selection and substitution rates among surfperches (Embiotocidae)



Gary C. Longo^{a,*}, Brendan O'Connell^b, Richard E. Green^b, Giacomo Bernardi^a

^a Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, California 95060, USA

^b Department of Biomolecular Engineering, University of California, Santa Cruz, California 95064, USA

ARTICLE INFO

Article history:

Received 13 January 2016

Received in revised form 11 March 2016

Accepted 12 March 2016

Available online 16 March 2016

Keywords:

Embiotoca jacksoni

Mitochondrial genome

Embiotocidae

Mitochondrial divergence

Black surfperch

ABSTRACT

The complete 16,515 bp nucleotide sequence of the mitochondrial genome was determined for the black surfperch, *Embiotoca jacksoni* (Perciformes: Embiotocidae). The black surfperch mitochondrial genome contains 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, and the non-coding control region (D-loop), the gene order of which is identical to that observed in most vertebrates. The protein-coding gene sequences of *E. jacksoni* mitochondrial DNA were compared with two other embiotocid surfperches with available complete mitochondrial genomes, *Cymatogaster aggregata* and *Ditrema temminckii*. Across all mitochondrial protein-coding genes in surfperches the weighted average substitution rate was 2.079% per My and average dN/dS ratios for each protein-coding gene ranged from 0.016 in CytB to 0.608 in ND3. Substitution rates and dN/dS ratios were relatively high for ATP8 compared to other protein-coding genes. Although most protein-coding genes showed signals of purifying selection, we found evidence for positive selection in ND3 in *E. jacksoni*.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The black surfperch, *Embiotoca jacksoni*, is a reef dwelling marine fish in the family Embiotocidae (Perciformes) ranging from Northern California, USA to Central Baja California, Mexico. Embiotocid surfperches are endemic to the temperate North Pacific and over the last 13–18 My have radiated into diverse habitats including seagrass, kelp forests reefs, sandy bottoms, estuaries, and even freshwater (Bernardi and Bucciarelli, 1999; Longo and Bernardi, 2015; Tarp, 1952). Interestingly, surfperches have evolved a rare reproductive strategy in teleosts where they mate via internal fertilization and give birth to fully developed live young (viviparity). This natural history trait has made surfperches an attractive system for studying sperm competition, population genetics, and courtship displays (Bernardi, 2000; Cummings and Partridge, 2001; Reisser et al., 2009).

Mitochondrial DNA (mtDNA) has commonly been used for phylogenetic and population genetic analyses due to maternal inheritance, low effective population size, and high substitution rates compared to the

nuclear genome (Avice, 2004). Although the vertebrate mitochondrial genome is inherited as a single unit and rarely undergoes recombination, various regions exhibit markedly different substitution rates (Naylor and Brown, 1998). Consequently, regions with relatively high substitution rates (i.e., hypervariable regions) have been used to infer relatedness in taxa with shallow divergence times, while markers that exhibit slower substitution rates (e.g., 16S rRNA locus) are better suited for resolving deeper taxonomic questions (Bernardi and Crane, 1999; Pardo et al., 2005; Sturmbauer and Meyer, 1993). The analysis of a species' complete mitochondrial genome best illustrates the variability observed among loci. Here we present the complete mitochondrial genome of the black surfperch, *E. jacksoni*, and utilize two other available surfperch mitogenomes, *Cymatogaster aggregata* and *Ditrema temminckii*, to estimate substitution rates and selective pressure by inferring nonsynonymous (dN) and synonymous (dS) substitution rates across protein-coding genes.

2. Materials and methods

2.1. Sample collection and DNA extraction

The mitogenomes of *C. aggregata* (NC_009059) and *D. temminckii* (NC_009060) were downloaded from NCBI (Mabuchi et al., 2007). We

* Corresponding author at: Department of Ecology and Evolutionary Biology, University of California Santa Cruz, 100 Shaffer Road, Santa Cruz, California 95060, USA.

E-mail address: garycharleslongo@gmail.com (G.C. Longo).

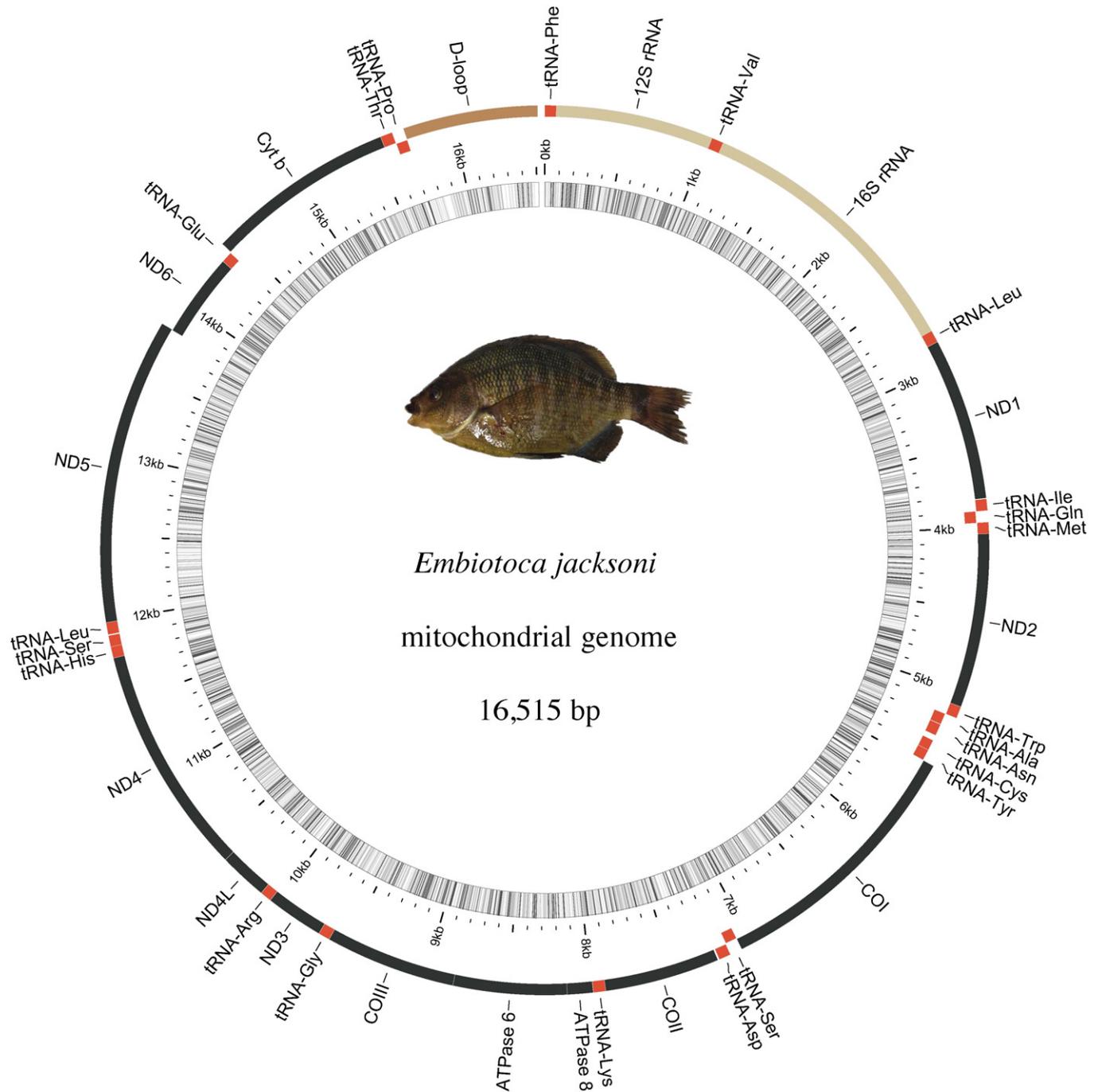


Fig. 1. Organizational map of the complete mitochondrial genome of the black surfperch, *Embiotoca jacksoni* (photograph of individual from which DNA was extracted). On the outer ring protein coding genes are shown in black, tRNAs in red, rRNAs in light brown, and the control region (D-loop) is in dark brown. Genes encoded on the light strand are inset relative to those on the heavy strand. The inner ring shows GC density with darker colors representing higher GC content.

generated a *de novo* mitochondrial genome of *E. jacksoni* using gill and liver tissue from a single individual freshly collected in Monterey, California (Fig. 1). Total DNA was extracted with Qiagen Blood and Cell Midi Kit following the manufacturer's protocol.

2.2. Sequencing and assembly

Genomic DNA was sheared and used to make an Illumina sequencing library (Meyer and Kircher, 2010). Briefly, the DNA was sonicated

to approximately 300 bp and end repaired. Next, sequencing adapters were ligated to both ends of the DNA. The adapters were filled-in, and the DNA was amplified in an indexing PCR. After library preparation, the library size distribution was confirmed by agarose gel electrophoresis, and the library was size-selected with a Sage Science BluePippin with a 2% agarose gel cassette. The library was sequenced on an Illumina HiSeq 2500 with 2×150 PE rapid run chemistry. From these raw data we assembled the mitochondrial genome using a reference-guided, iterative assembly approach (<https://github.com/mpieva/mapping->

iterative-assembler) designed to accurately assemble short, circular genomes from sequence data which may have many irrelevant reads in it (Briggs et al., 2009). The final assembly is supported by an average of 42.3 fold sequence coverage. No position was supported by fewer than 15 reads. The complete mitochondrial genome sequence of the black surfperch has been deposited in GenBank (accession number KU530212).

2.3. Sequence analyses

We compared substitution rates across protein-coding genes among the three available surfperch mitogenomes, *C. aggregata* (CAG), *D. temminckii* (DTE), and *E. jacksoni* (EJA). Individual genes were aligned and for consistent comparisons, one substitution model (Kimura 2-parameter) was used to estimate pairwise distances for all protein-coding loci using PAUP v4.0a (Swofford, 2003). Substitution rates (distance per My) were calculated for each species pair using the following estimated times to most recent common ancestor: 10 My for CAG/EJA and CAG/DTE, 7 My for DTE/EJA (Longo and Bernardi, 2015).

We estimated branch specific dN/dS ratios in a phylogenetic framework for the entire sequence of each protein-coding gene in the mitochondrial genome. A pruned surfperch phylogeny was constructed using the tree editing software Archaeopteryx (Han and Zmasek, 2009) and appended to each sequence alignment with branch lengths from a previously published embiotocid phylogeny based on 523 RADseq markers (Longo and Bernardi, 2015). The Branch-site random effects likelihood (Branch-site REL) method was used to compute average dN/dS for each branch in the pruned phylogeny as implemented in HyPhy (Kosakovsky Pond et al., 2011). Branch-site REL method detects lineages where a proportion of sites evolve with dN/dS > 1 but makes no assumptions about which lineages those are or about what happens to the rest of the lineages.

3. Results and discussion

3.1. Genome organization

The complete mitochondrial genome of *E. jacksoni* was determined to be 16,515 bp and contained 22 tRNA genes, 13 protein-coding genes, two rRNA genes, and a control region (D-Loop) in the same organization as for most fish genomes (Fig. 1, Table 1) including the two other published surfperch genomes (Iwasaki et al., 2013; Mabuchi et al., 2007). As commonly found in other vertebrates, most genes were encoded on the heavy strand except for eight tRNAs (Gln, Ala, Asn, Cys, Tyr, Ser, Glu, and Pro) and NADH dehydrogenase subunit 6 (ND6) (Fig. 1, Table 1). Black surfperch mitochondrial open reading frames overlapped in five locations for a total 23 bp (Table 1). Nucleotide frequencies in the *E. jacksoni* mitochondrial genome were A = 27.67%, C = 27.76%, G = 16.48%, and T = 28.09% (GC content = 44.24%).

3.2. Substitution rates and dN/dS

Pairwise substitution rates across embiotocid mitochondrial genes varied from 1.38% per My in CO1 (CAG × EJA) to 5.28% per My in ATP8 (DTE × EJA) (Fig. 2). The average weighted substitution rate across all mitochondrial protein-coding genes in surfperches was 2.079% per My (Fig. 2). In animal mitochondrial genomes, ATP8 generally exhibits a high substitution rate and dN/dS compared to other protein-coding genes (Castellana et al., 2011; Oliveira et al., 2008; Zardoya and Meyer, 2007). This elevated substitution rate can either be explained by positive selection or relaxed selection on ATP8 (Castellana et al., 2011). For example, ATP8 is missing in some metazoan mitochondrial genomes (Breton et al., 2010), which would support elevated substitution rates due to a relaxation in selective pressure (potentially followed

Table 1 Please check Table 1 header if presented correctly and amend as necessary
Organization of the mitochondrial genome of *Embiotoca jacksoni*.

Gene	Position		Intergenic nucleotides (bp)	Strand
	Start	End		
tRNA-Phe	1	69	0	H
12S rRNA	70	1017	0	H
tRNA-Val	1018	1089	0	H
16S rRNA	1090	2780	0	H
tRNA-Leu	2781	2854	0	H
ND1	2855	3829	0	H
tRNA-Ile	3834	3903	4	H
tRNA-Gln	3903	3973	-1	L
tRNA-Met	3973	4041	-1	L
ND2	4042	5087	0	H
tRNA-Trp	5088	5159	0	H
tRNA-Ala	5161	5229	1	L
tRNA-Asn	5232	5304	2	L
tRNA-Cys	5339	5404	34	L
tRNA-Tyr	5405	5472	0	L
COI	5474	7069	1	H
tRNA-Ser	7070	7140	0	L
tRNA-Asp	7144	7215	3	H
COII	7225	7915	9	H
tRNA-Lys	7916	7989	0	H
ATP8	7991	8155	1	H
ATP6	8146	8828	-10	H
COIII	8829	9613	0	H
tRNA-Gly	9614	9683	0	H
ND3	9684	10,032	0	H
tRNA-Arg10033	10,101	0	H	
ND4L	10,102	10,398	0	H
ND4	10,392	11,772	-7	H
tRNA-His11773	11,841	0	H	
tRNA-Ser(2)	11,842	11,909	0	H
tRNA-Leu(2)	11,916	11,988	6	H
ND5	11,989	13,827	0	H
ND6	13,824	14,345	-4	L
tRNA-Glu	14,346	14,414	0	L
CYTB	14,420	15,560	5	H
tRNA-Thr	15,561	15,633	0	H
tRNA-Pro	15,635	15,704	1	L
D-Loop	15,705	16,515	0	H

by complete loss) in some lineages. In surfperches ATP8 yields the highest average substitution rate (3.49% per My) and second highest dN/dS (0.198) among mitochondrial protein-coding genes (Figs. 2, 3). These relatively high values for ATP8 in surfperches are likely due to a relaxation in selective pressure as opposed to positive selection as dN/dS < 1 in all lineages.

Branch-site REL analyses for positive selection in all mitochondrial protein-coding genes resulted in dN/dS ranging from 0 in ND4 of *C. aggregata* to 1.76 in ND3 of *E. jacksoni* (Fig. 3). The relatively high dN/dS value observed in ND3 of the black surfperch suggests this protein may be experiencing positive selection with dN/dS > 1 (Fig. 3). ND3, as well as ND1, ND2, ND4, ND4L, ND5, and ND6, each code for a single protein component in NADH dehydrogenase complex 1, which consists of around 45 total subunits (~39 proteins are derived from nuclear genes) in Metazoa (Carroll et al., 2006; Castellana et al., 2011). Complex 1 is the first of several components in the oxidative phosphorylation pathway that generates ATP and is crucial for cellular energy production. Therefore variations in mitochondrial protein coding genes involved in this pathway potentially affect metabolic performance. ND3 has been shown to be under selection in several systems (Kennedy and Nachman, 1998). For example, a recent study analyzed mtDNA sequences from hundreds of Atlantic salmon individuals from many populations and found a pattern of positive selection in ND1, ND3, and ND4 that correlated with latitudinal gradients (Consuegra et al., 2015). Interestingly, similar results were

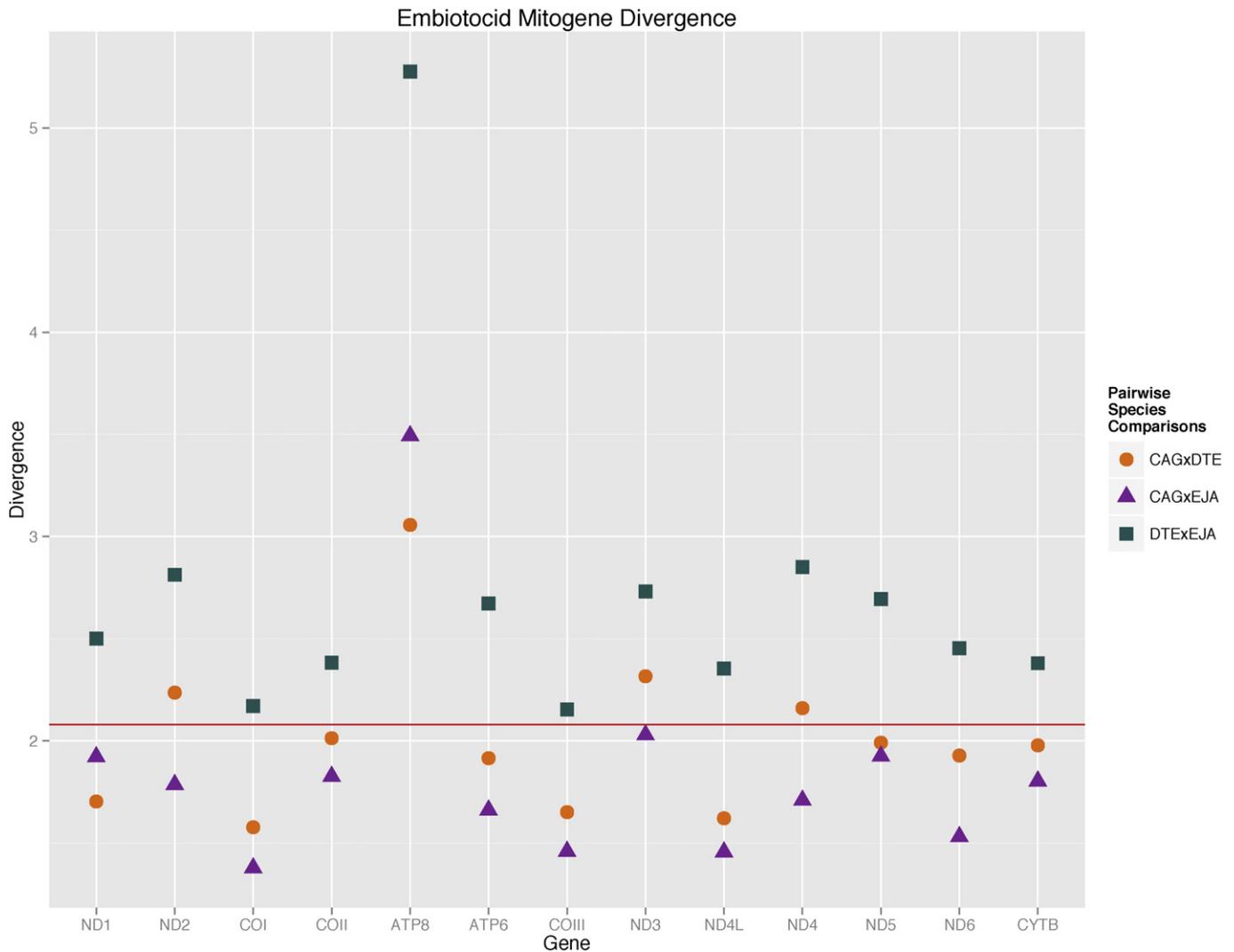


Fig. 2. Pairwise divergence rates across mitochondrial protein-coding genes in three embiotocid surfperches: *Cymatogaster aggregata* (CAG), *Ditrema temminckii* (DTE), and *Embiotoca jacksoni* (EJA). The red line represents the weighted average substitution rate across all mitochondrial protein-coding genes. Gene order reflects position in mitogenome.

also found in Pacific salmon (Garvin et al., 2011) and killer whales (Foote et al., 2011) suggesting these patterns of positive selection in mitochondrial genes could be driven by different metabolic demands in the lower temperatures of higher latitudes. Surfperches are certainly a good system to examine this pattern as they range from warmer waters off Southern California and Baja, Mexico to cooler waters off Alaska, but here we only have data from three individuals each representing separate species. Dissimilar metabolic demands across surfperch species may also be driving the observed $dN/dS > 1$ in *E. jacksoni*, although additional data will be needed to confirm this pattern. On the other hand, purifying selection seems to be particularly strong in CytB and COIII, where dN/dS is low and narrow among comparisons (Fig. 3). Indeed, due to their crucial role in cellular respiration, mitochondrial protein-coding genes are expected to be mostly under purifying selection ($dN/dS < 1$), which we observe here.

4. Conclusion

While the mitochondrial genome is often seen as a single non-recombining unit, striking differences in substitution rates within

the genome are observed, which in the case of surfperches translates into a more than twofold difference between the slowest and fastest genes (1.71%–3.94% per My). A number of explanations have been proposed to explain these differences, including differential selective pressure (Castellana et al., 2011; Consuegra et al., 2015; Popadin et al., 2012). Here, we examined dN/dS ratios among embiotocids and found most protein-coding genes exhibiting signals of purifying selection with some evidence for positive selection in the protein-coding gene ND3 in *E. jacksoni* when compared to *C. aggregata* and *D. temminckii*. Our results, however, should be interpreted with caution because the availability of only three surfperch mitochondrial genomes may have resulted in low power to detect purifying selection.

Acknowledgments

This project was funded by the Lerner-Gray Fund for Marine Research, Myers Trust, and the UCSC Ecology and Evolutionary biology research and travel fund. We would like to thank Andrew Kim for his dive support in Monterey.

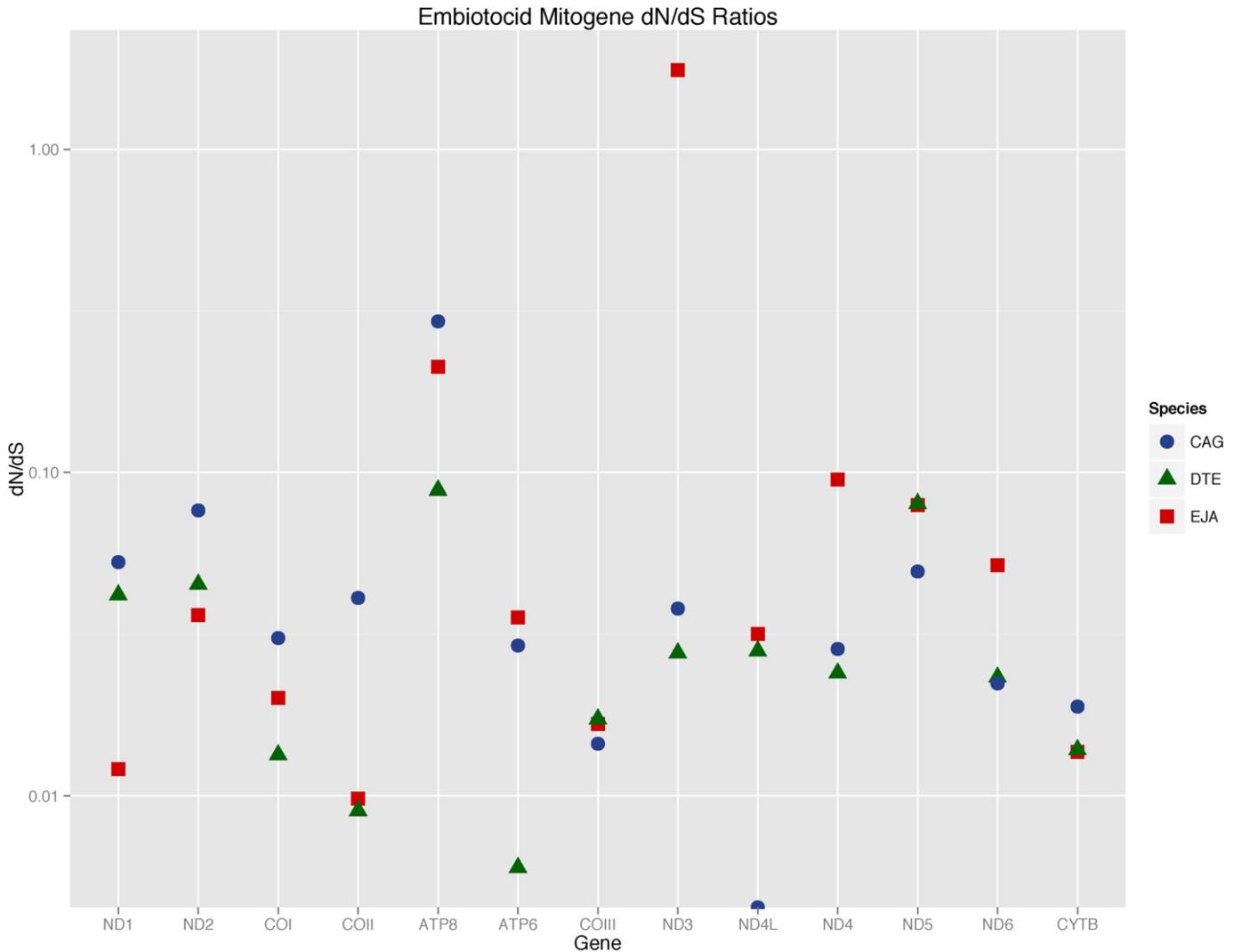


Fig. 3. Branch specific dN/dS ratios (\log_{10}) across mitochondrial protein-coding genes in three embiotocid surfperches: *Cymatogaster aggregata* (CAG), *Ditrema temminckii* (DTE), and *Embiotoca jacksoni* (EJA). Gene order reflects position in mitogenome.

References

- Avise, J.C., 2004. *Molecular Markers, Natural History, and Evolution*. second ed. Chapman and Hall, New York.
- Bernardi, G., 2000. Barriers to gene flow in *Embiotoca jacksoni*, a marine fish lacking a pelagic larval stage. *Evolution* (N. Y.) 54, 226–237.
- Bernardi, G., Bucciarelli, G., 1999. Molecular phylogeny and speciation of the surfperches (Embiotocidae, Perciformes). *Mol. Phylogenet. Evol.* 13, 77–81.
- Bernardi, G., Crane, N.L., 1999. Molecular phylogeny of the humbug damselfishes inferred from mtDNA sequences. *J. Fish Biol.* 54, 1210–1217.
- Breton, S., Stewart, D.T., Hoeh, W.R., 2010. Characterization of a mitochondrial ORF from the gender-associated mtDNAs of *Mytilus* spp. (Bivalvia: Mytilidae): identification of the “missing” ATPase 8 gene. *Mar. Genomics* 3, 11–18. <http://dx.doi.org/10.1016/j.margen.2010.01.001>.
- Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V.B., Golovanova, L.V., de la Rasilla, M., Forcia, J., Rosas, A., Pääbo, S., 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325, 318–321. <http://dx.doi.org/10.1126/science.1174462> (80-).
- Carroll, J., Fearnley, I.M., Skehel, J.M., Shannon, R.J., Hirst, J., Walker, J.E., 2006. Bovine complex I is a complex of 45 different subunits. *J. Biol. Chem.* 281, 32724–32727. <http://dx.doi.org/10.1074/jbc.M607135200>.
- Castellana, S., Vicario, S., Saccone, C., 2011. Evolutionary patterns of the mitochondrial genome in metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes. *Genome Biol. Evol.* 3, 1067–1079. <http://dx.doi.org/10.1093/gbe/evr040>.
- Consuegra, S., John, E., Verspoor, E., de Leaniz, C.G., 2015. Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genet. Sel. Evol.* 47, 58. <http://dx.doi.org/10.1186/s12711-015-0138-0>.
- Cummings, M.E., Partridge, J.C., 2001. Visual pigments and optical habitats of surfperch (Embiotocidae) in the California kelp forest. *J. Comp. Physiol. A Sensory, Neural, Behav. Physiol.* 187, 875–889. <http://dx.doi.org/10.1007/s00359-001-0258-6>.
- Footo, A.D., Morin, P.A., Durban, J.W., Pitman, R.L., Wade, P., Willerslev, E., Gilbert, M.T.P., da Fonseca, R.R., 2011. Positive selection on the killer whale mitogenome. *Biol. Lett.* 7, 116–118. <http://dx.doi.org/10.1098/rsbl.2010.0638>.
- Garvin, M.R., Bielawski, J.P., Gharrett, A.J., 2011. Positive darwinian selection in the piston that powers proton pumps in Complex I of the mitochondria of Pacific salmon. *PLoS One* 6. <http://dx.doi.org/10.1371/journal.pone.0024127>.
- Han, M.V., Zmasek, C.M., 2009. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinf.* 10, 356. <http://dx.doi.org/10.1186/1471-2105-10-356>.
- Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T.P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M., Nishida, M., 2013. MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol. Evol.* 30, 2531–2540. <http://dx.doi.org/10.1093/molbev/mst141>.
- Kennedy, P., Nachman, M.W., 1998. Deleterious mutations at the mitochondrial ND3 gene in South American marsh rats (*Holochilus*). *Genetics* 150, 359–368.
- Kosakovskiy, S.L., Murrell, B., Fournier, M., Frost, S.D.W., Delpont, W., Scheffler, K., 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* 28, 3033–3043. <http://dx.doi.org/10.1093/molbev/msr125>.
- Longo, G., Bernardi, G., 2015. The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. *Mol. Phylogenet. Evol.* 88, 55–63. <http://dx.doi.org/10.1016/j.ympev.2015.03.027>.
- Mabuchi, K., Miya, M., Azuma, Y., Nishida, M., 2007. Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. *BMC Evol. Biol.* 7, 10. <http://dx.doi.org/10.1186/1471-2148-7-10>.
- Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010. <http://dx.doi.org/10.1101/pdb.prot5448> (pdb.prot5448–pdb.prot5448).

- Naylor, G.J.P., Brown, W.M., 1998. Amphioxus mitochondrial DNA, Chordate Phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* 47, 61–76.
- Oliveira, D.C.S.G., Raychoudhury, R., Lavrov, D.V., Werren, J.H., 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *nasonia* (hymenoptera: pteromalidae). *Mol. Biol. Evol.* 25, 2167–2180. <http://dx.doi.org/10.1093/molbev/msn159>.
- Pardo, B.G., Machordom, A., Foresti, F., Porto-Foresti, F., Azevedo, M.F.C., Bañón, R., Sánchez, L., Martínez, P., 2005. Phylogenetic analysis of flatfish (Order Pleuronectiformes) based on mitochondrial 16 s rDNA sequences. *Sci. Mar.* 69, 531–543. <http://dx.doi.org/10.3989/scimar.2005.69n4531>.
- Popadin, K.Y., Nikolaev, S.I., Junier, T., Baranova, M., Antonarakis, S.E., 2012. Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes. *Mol. Biol. Evol.* 30, 347–355. <http://dx.doi.org/10.1093/molbev/mss219>.
- Reisser, C.M.O., Beldade, R., Bernardi, G., 2009. Multiple paternity and competition in sympatric congeneric reef fishes, *Embiotoca jacksoni* and *E. lateralis*. *Mol. Ecol.* 18, 1504–1510. <http://dx.doi.org/10.1111/j.1365-294X.2009.04123.x>.
- Sturmbauer, C., Meyer, A., 1993. Mitochondrial phylogeny of the endemic mouthbrooding lineages of cichlid fishes from Lake Tanganyika in Eastern Africa. *Mol. Biol. Evol.* 10, 751–768.
- Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).
- Tarp, F.H., 1952. Fish Bulletin No. 88. A Revision of the Family Embiotocidae (The Surfperches). State of California Department of Fish and Game.
- Zardoya, R., Meyer, A., 2007. Phylogenetic Performance of Mitochondrial Protein-Coding Genes in Resolving Relationships among Vertebrates.