Molecular Phylogenetics and Evolution 50 (2009) 401–404

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Stickleback phylogenies resolved: Evidence from mitochondrial genomes and 11 nuclear genes

Ryouka Kawahara [a,*], Masaki Miya [b], Kohji Mabuchi [a], Thomas J. Near [c], Mutsumi Nishida [a]

[a] *Department of Marine Bioscience, Ocean Research Institute, University of Tokyo, 1-15-1 Minamidai, Nakano-ku, Tokyo 164-8639, Japan*
[b] *Natural History Museum and Institute, Chiba, 955-2 Aoba-cho, Chuo-ku, Chiba 260-8682, Japan*
[c] *Department of Ecology and Evolutionary Biology and Peabody Museum of Natural History, Yale University, New Haven, CT 06520, USA*

ARTICLE INFO

Article history:
Received 28 August 2008
Accepted 20 October 2008
Available online 5 November 2008

1. Introduction

The threespine Stickleback, *Gasterosteus aculeatus*, is emerging as a model system for studying the molecular basis of morphological variation in vertebrates (Kingsley and Peichel, 2007; Kitano et al., 2008). Many populations of *G. aculeatus* have colonized freshwater habitats from marine ancestors and rapidly adapted to distinct habitats and resources. This pattern of parallel speciation has attracted the attention of evolutionary biologists as an outstanding system for investigating patterns and mechanisms of adaptive radiation and ecological speciation (Schluter, 2000; Schluter and McPhail, 1992).

A resolved phylogenetic hypothesis for the seven recognized stickleback (Gasterosteidae) species is critical for identifying appropriate contrasts to investigate patterns of adaptive divergence observed in *G. aculeatus*, and to place diversification of the entire clade in a comparative context. However, phylogenetic relationships appear unresolved in spite of a number of attempts based on morphology (Bowne, 1994; Keivany and Nelson, 2004), behavior (McLennan, 1991, 1993), mtDNA gene sequences (Mattern, 2004), and various combinations of these data types (Mattern and McLennan, 2004; McLennan and Mattern, 2001). To examine the phylogenetic relationships of Gasterosteidae, we performed phylogenetic analyses using DNA sequence datasets that include whole mitochondrial genomes (14,807 bp) and 11 single-copy nuclear genes (8703 bp).

2. Materials and methods

2.1. Taxonomic sampling

Phylogenetic analyses included at least one species from each of the eight genera of suborder Gasterosteoidei as defined in a recent phylogenetic analysis of teleost mitogenome sequence data (Kawahara et al., 2008). A list of the nine species examined in this study is provided in Table 1, along with DDBJ/EMBL/GenBank accession numbers and references of mitogenome sequences. In addition to the whole mitogenome sequences previously determined for three gasterosteoid species (*Gasterosteus aculeatus*, *Hypoptychus dybowskii*, and *Aulorhynchus flavidus*), we sequenced the whole mitogenome for an additional six species. Eleven nuclear genes (rag1, MLL, zic1, myh6, Ptr, tbr1, ENC1, gylt, SH3PX3, plagl2, and serb2; Supplementary Table 1) were sequenced for the nine species listed in Table 1.

Locality information is as follows: *Hypoptychus dybowskii* and *Aulichthys japonicus*, Otsuchi Bay, Iwate Prefecture, Japan. *Aulorhynchus flavidus*: Vancouver Aquarium, Vancouver, British Columbia, Canada. *Culaea inconstans*: the Arboretum at the University of Guelph, Guelph, Ontario, Canada. *Apeltes quadracus* and *Gasterosteus wheatlandi*: Shediac River, New Brunswick, Canada. *Gasterosteus aculeatus* and *Pungitius pungitius*: Akkeshi, Hokkaido, Japan. *Spinachia spinachia*: Amager Fælled, Denmark.

2.2. DNA extraction, PCR and sequencing

Total genomic DNA was extracted from frozen or ethanol-preserved specimens using the Aquapure genomic DNA isolation kit (Bio-Rad Laboratories, Inc., Hercules, CA, USA), following each manufacturer's protocol.

The mitogenomes of the six gasterosteoid species were amplified in their entirety using a long PCR technique (Cheng et al., 1994). Long PCR primers and reaction conditions followed previous studies (Inoue et al., 2001; 2004; Kawaguchi et al., 2001). Long PCR products diluted with TE buffer (1:10–100) subsequently served as templates for short PCR reactions. We used fish-versatile PCR primers in various combinations to amplify contiguous, overlapping segments of the entire mitogenome for each of the six species (locations and sequences of primers available upon request to

* Corresponding author. Fax: +81 3 5351 6488.
E-mail address: kawaharar@ori.u-tokyo.ac.jp (R. Kawahara).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.10.014

Table 1
List of species and accession numbers of mitogenomes examined in this study.

Classification	Species	Accession Nos.	References
Hypoptychidae	*Hypoptychus dybowskii*	AP004437	Miya et al. (2003)
Aulorhynchidae	*Aulichthys japonicus*	AB445127	This study
Aulorhynchidae	*Aulorhynchus flavidus*	AP009196	Kawahara et al. (2008)
Gasterosteidae	*Apeltes quadracus*	AB445126	This study
Gasterosteidae	*Culaea inconstans*	AB445125	This study
Gasterosteidae	*Gasterosteus wheatlandi*	AB445129	This study
Gasterosteidae	*Gasterosteus aculeatus*	AP002944	Miya et al. (2001)
Gasterosteidae	*Pungitius pungitius*	AB445130	This study
Gasterosteidae	*Spinachia spinachia*	AB445128	This study

R.K.). Short PCR reaction conditions followed Miya and Nishida (1999).

The 11 nuclear genes for each nine species of gasterosteoids were amplified using PCR primers listed in Supplementary Table 2. Based on the information of threespine stickleback genome database (Ensembl [http://www.ensembl.org/index.html]), these genes were confirmed as single-copy. Conditions for PCR followed protocols presented in previous studies (Li et al., 2007; Venkatesh et al., 1999).

Double-stranded PCR products of mitogenome and nuclear genes were purified using an Exosap-IT enzyme reaction (GE Healthcare Bio-Sciences Corp., Piscataway, NT, USA). These were subsequently used for direct cycle sequencing with dye-labeled terminators (Big Dye terminator ver. 3.1, Applied Biosystems, Foster City, CA, USA) and the same primers as those used in the PCRs. All sequencing reactions were performed according to the manufacturer's instructions. Labeled fragments were analyzed on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems).

2.3. Sequences editing and alignment

The sequence electrophenograms were edited and the consensus sequences concatenated with the computer program ATGC Ver. 4.0 (Genetyx Corporation, Tokyo, Japan). All sequences were deposited in DDBJ/EMBL/GenBank (Accession Nos. AB445125–AB445229; Table 1 and Supplemental Table 1). We combined the six newly-determined mitogenome sequences with the three previously published sequences to construct multiple sequence alignments. For each individual protein coding gene, we manually aligned the sequences for the nine species, with reference to the translated amino acid sequence using MacClade (Maddison and Maddison, 2000). All positions that included stop codons, inferred gaps, and ambiguously aligned regions were excluded from the subsequent phylogenetic analyses. The ND6 gene was excluded because of its heterogeneous base composition and consistently poor phylogenetic performance (Miya and Nishida, 2000). The 22 tRNA genes were aligned manually, and ambiguously aligned positions and inferred gaps were excluded. The 12S and 16S rRNA sequences were aligned using the software Proalign ver. 0.5 (Löytynoja and Millinkovitch, 2003) with default settings. Regions with posterior probabilities of $\leqslant$70% were excluded from the subsequent phylogenetic analyses. The aligned mitogenome dataset included 14,807 nucleotide positions, comprising 12 protein coding genes (10,776 nucleotide positions), two rRNA genes (2518 nucleotide positions), and 22 tRNA genes (1513 nucleotide positions).

We also aligned each nuclear gene sequence with reference to the translated amino acid sequence. Positions that included stop codons, gaps, and ambiguously aligned regions were excluded from the subsequent phylogenetic analyses. Aligned sequences of the 11 nuclear genes for nine species were concatenated and this dataset contained 8703 nucleotide positions.

2.4. Phylogenetic analysis

Phylogenetic trees were estimated using partitioned maximum likelihood and partitioned Bayesian methods. These analyses were performed independently for the mitogenome and nuclear gene datasets. For the mitogenome dataset, we set five partitions that corresponded to each of the three codon positions in the protein coding genes and a single partition for each of the tRNA and rRNA genes. For the nuclear gene dataset, we designated 33 partitions that correspond to three codon positions in each of the 11 nuclear genes. We also investigated using 11 partitions to reflect that each nuclear gene is independent locus and three partitions based on codon positions. However, the phylogenies estimated from all of the partitioned datasets were identical.

Maximum likelihood analyses were conducted with RAxML ver. 7.0.3 (Stamatakis, 2006). We selected GTRMIX as a nucleotide substitution model. In this model, a tree inference (search for a good topology) was performed under the GTRCAT model and the final tree topology was evaluated under the GTRGAMMA model such that it yields stable likelihood values. We performed 100 RAxML runs and found the best ML tree by comparing the likelihood scores. To evaluate the robustness of the ML tree, 1000 bootstrap pseudoreplicates were calculated for each data set. The GTRCAT model was selected for the bootstrap analyses.

Partitioned Bayesian phylogenetic analyses were conducted with MrBayes ver. 3.1.2 (Ronquist and Huelsenbeck, 2003) on the mitogenome and nuclear gene datasets. The GTR + I + G model (Yang, 1994) was assigned to each of the designated data partitions. Other parameter settings in MrBayes are as described in the previous study (Kawahara et al., 2008). The Markov chain Monte Carlo process was set so that four chains (three heated and one cold) ran simultaneously. We conducted two independent runs for each data set and continued for 5×10^6 cycles. Convergence (lack of improvement in the likelihood score) was evaluated graphically and all trees and parameters discarded before convergence was reached as a "burn-in". After confirming agreement of the estimated parameters between the two independent runs, we pooled all trees after the "burn-in" period from the two independent runs. Posterior probabilities of the phylogenies and their branches were estimated on the basis of the pooled trees.

3. Results and discussion

Four phylogenetic analyses using maximum likelihoood and Bayesian phylogenetic methods for the mitogenome and nuclear gene datasets all resulted in an identical tree topology with high maximum likelihood bootstrap and Bayesian posterior probability support for most nodes (bootstrap $\geqslant$90%, posterior probability $\geqslant$99%). Using *Hypoptychus dybowskii* as an outgroup (Kawahara et al., 2008), the six sampled gasterosteid species were monophyletic in both the mitogenome and nuclear gene phylogenies, and most closely related to one of the two aulorhynchids, *Aulorhynchus flavidus*, rendering Aulorhynchidae paraphyletic with respect to Gasterosteidae (Fig. 1). However, BA posterior probability for this clade was not significant in the mitogenome phylogenetic analysis (Fig. 1). A monophyletic group comprising the two sampled *Gasterosteus* species (*G. aculeatus* and *G. wheatlandi*) was the sister lineage to all of the other gasterosteid species. All of the previously published phylogenetic hypotheses for Gasterosteidae represented *Spinachia spinachia* as the sister species of all other extant gasterosteid species (Bowne, 1994; Keivany and Nelson, 2004; Mattern, 2004; Mattern and McLennan, 2004; McLennan, 1991, 1993; McLennan and Mattern, 2001). Our phylogenetic analyses result in a derived phylogenetic placement of *S. spinachia* and this species is most closely related to *Apeltes quadracus*, and this clade is the

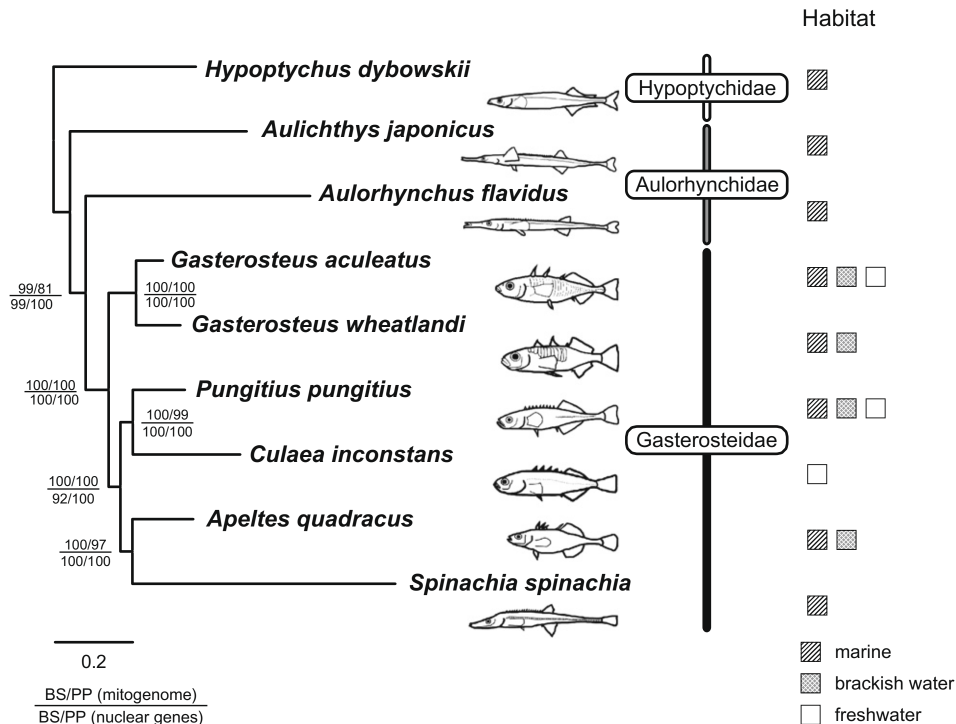

Fig. 1. Stickleback phylogeny resulting from maximum likelihood (ML) analysis of the mitogenome dataset (with branch lengths proportional to number of substitutions per site). All four phylogenetic analyses using partitioned ML and Bayesian (BA) methods on the mitogenome and nuclear gene datasets resulted in this tree topology. Numbers beside internal nodes provide support values from each of the four analyses: numbers above horizontal line are ML bootstrap values from 1000 pseudoreplicates and BA posterior probabilities resulting from analyses of the mitogenome dataset and numbers below horizontal line are those resulting from analyses of the 11 nuclear gene dataset. Boxes on the right side of the tree indicate the habitat of the gasterosteoid species. *Gasterosteus aculeatus* and *Pungitius pungitius* are widely distributed in marine, freshwater, and brackish habitats, *Gasterosteus wheatlandi* and *Apeltes quadracus* occur in marine and brackish habitats, whereas the other species are restricted to freshwater (*Culaea inconstans*) or marine habitats (*Hypoptychus dybowskii*, *Aulichthys japonicus*, *Aulorhynchus flavidus*, and *Spinachia spinachia*).

sister lineage of a clade comprising *Pungitius pungitius* and *Culaea inconstans* (Fig. 1).

There have been several studies investigating gasterosteid phylogeny and collectively have used a diverse array of character types, including morphology and behavior; however, the only study using DNA gene sequence data has been limited to mitochondrial genes (Mattern, 2004; Mattern and McLennan, 2004). The gasterosteid phylogeny presented in Mattern (2004) was based on partial sequences of five mitochondrial genes and two of the internal nodes were supported with low parsimony bootstrap support values, and is very different from the phylogeny resulting from our analyses (Fig. 1). The differences in the phylogeny presented in Mattern (2004) and one resulting from analyses of our mitogenome dataset is most likely the result from the difference in the size of the DNA datasets, with 2879 base pairs used by Mattern (2004) and 14,807 base pairs in our whole mitogenome dataset. Interestingly, we found that phylogenies inferred from each of the individual mitochondrial genes were all different from the phylogeny resulting from analysis of the whole mitogenome dataset (see Supplementary Fig. 1); however, we ruled out the possibility of nuclear encoded mtDNA pseudogenes due to the use of long PCR and no interruption of reading frames among the 13 protein coding genes. The exact congruence in the phylogenies inferred from the mitogenome dataset and the partitioned ML and BA analyses of the 11 sampled nuclear genes provided substantial confidence to the phylogenetic hypothesis presented in Fig. 1.

4. Conclusion

Our analyses of mitogenome and nuclear gene datasets allow the presentation of a new phylogenetic framework for gasterosteid fishes, and a novel historical perspective for investigating the diversification of this clade. One of the most striking results of our new phylogenetic analyses is the phylogenetic placement of *Spinachia spinachia*, where most previous phylogenetic hypotheses have identified *S. spinachia* as the earliest extant gasterosteid lineage, and similarities between *S. spinachia* and the two aulorhynchid species (*Aulorhynchus flavidus* and *Aulichthys japonicus*) were hypothesized to reflect the ancestral condition in gasterosteids (Mattern and McLennan, 2004; McLennan, 1991, 1993; McLennan and Mattern, 2001). However, our new phylogeny indicates that the similarities shown in morphological and behavioral characters between *S. spinachia* and aulorhynchids probably resulted from the independent adaptation to similar marine habitats (Fig. 1).

Acknowledgments

We thank David L.G. Noakes, Beren W. Robinson, Kathryn Peiman, R.A. Curry, Marc Skinner, Simon C. Courtenay, Peter Rask Møller, Takuji Oyama, Izumi Akagawa, and Hiroshi Takahashi for helping in sample collection or providing samples, and Jun G. Inoue, Sébastian Lavoué, and Yusuke Yamanoue for valuable discussion and suggestions. This study was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology (15380131, 17207007, and 19207007), and the United States National Science Foundation (DEB-0716155).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.10.014.

References

Bowne, P.S., 1994. Systematics and morphology of the Gasterosteiformes. In: Bell, A.M., Foster, S.A. (Eds.), The Evolutionary Biology of the Threespine Stickleback. Oxford University Press, Oxford, pp. 28–60.

Cheng, S., Higuchi, R., Stoneking, M., 1994. Complete mitochondrial genome amplification. Nat. Genet. 7, 350–351.

Inoue, J.G., Miya, M., Aoyama, J., Ishikawa, S., Tsukamoto, K., Nishida, M., 2001. Complete mitochondrial DNA sequence of the Japanese eel *Anguilla japonica*. Fish Sci. 67, 118–125.

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2004. Mitogenomic evidence for the monophyly of elopomorph fishes (Teleostei) and the evolutionary origin of the leptocephalus larva. Mol. Phylogenet. Evol. 30, 274–286.

Kawaguchi, A., Miya, M., Nishida, M., 2001. Complete mitochondrial DNA sequence of *Aulopus japonicus* (Teleostei: Aulopiformes), a basal Eurypterygii: longer DNA sequences and higher-level relationships. Ichthyol. Res. 48, 213–223.

Kawahara, R., Miya, M., Mabuchi, K., Lavoue, S., Inoue, J.G., Satoh, T.P., Kawaguchi, A., Nishida, M., 2008. Interrelationships of the 11 gasterosteiform families (sticklebacks, pipefishes, and their relatives): a new perspective based on whole mitogenome sequences from 75 higher teleosts. Mol. Phylogenet. Evol. 46, 224–236.

Keivany, Y., Nelson, J.S., 2004. Phylogenetic relationships of sticklebacks (Gasterosteidae), with emphasis on ninespine sticklebacks (*Pungitius* spp.). Behaviour 141, 1485–1497.

Kingsley, D.M., Peichel, C.L., 2007. The molecular genetics of evolutionary change in sticklebacks. In: Östlund-Nilsson, S., Mayer, I., Huntingford, F.A. (Eds.), Biology of the Three-spined Stickleback. CRC Press, Boca Raton, FL, pp. 41–81.

Kitano, J., Bolnick, D.I., Beauchamp, D.A., Mazur, M.M., Mori, S., Nakano, T., Peichel, C.L., 2008. Reverse evolution of armor plates in the threespine stickleback. Curr. Biol. 18, 769–774.

Löytynoja, A., Millinkovitch, M.C., 2003. A hidden Markov model for progressive multiple alignment. Bioinformatics 19, 1505–1513.

Li, C.H., Orti, G., Zhang, G., Lu, G.Q., 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol. Biol. 7, 44.

Maddison, W.P., Maddison, D.R., 2000. MacClade, Version 4. Sinauer, Sunderland, MA.

Mattern, M.Y., 2004. Molecular phylogeny of the Gasterosteidae: the importance of using multiple genes. Mol. Phylogenet. Evol. 30, 366–377.

Mattern, M.Y., McLennan, D.A., 2004. Total evidence phylogeny of Gasterosteidae: combining molecular, morphological and behavioral data. Cladistics 20, 14–22.

McLennan, D.A., 1991. Integrating phylogeny and experimental ethology: from pattern to process. Evolution 45, 1773–1789.

McLennan, D.A., 1993. Phylogenetic relationships in the Gasterosteidae: an updated tree based on behavioral characters with a discussion of homoplasy. Copeia 1993, 318–326.

McLennan, D.A., Mattern, M.Y., 2001. The phylogeny of the Gasterosteidae: combining behavioral and morphological data sets. Cladistics 17, 11–27.

Miya, M., Nishida, M., 1999. Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): first example of transfer RNA gene rearrangements in bony fishes. Mar. Biotechnol. 1, 416–426.

Miya, M., Nishida, M., 2000. Molecular systematics of the deep-sea fish genus *Gonostoma* (Stomiiformes: Gonostomatidae): two paraphyletic clades and resurrection of Sigmops. Copeia, 378–389.

Miya, M., Kawaguchi, A., Nishida, M., 2001. Mitogenetic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Phylogenet. Evol. 18, 1993–2009.

Miya, M., Takeshima, H., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M., Nishida, M., 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phylogenet. Evol. 26, 121–138.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Schluter, D., 2000. Ecological character displacement in adaptive radiation. Am. Nat. 156, S4–S16.

Schluter, D., McPhail, D.J., 1992. Ecological character displacement and speciation in sticklebacks. Am. Nat. 140, 5–108.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Venkatesh, B., Ning, Y., Brenner, S., 1999. Late changes in spliceosomal introns define clades in vertebrate evolution. Proc. Natl. Acad. Sci. USA 96, 10267–10271.

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39, 306–314.