LASER WORDS | JFB | jfb˙2378 | B | Dispatch: September 10, 2009 | Journal: JFB | CE: Narayanan
Journal Name | Manuscript No. | Author Received: | No of pages: 7 | TS: Backiaraj

Journal of Fish Biology (2009) **75**, 0–0
doi:10.1111/j.1095-8649.2009.02378.x, available online at www.interscience.wiley.com

BRIEF COMMUNICATION

A cryptic lineage within the pupfish *Cyprinodon dearborni* suggests multiple colonizations of South America

R. A. Haney*†, B. J. Turner‡ and D. M. Rand§

**Department of Organismal Biology and Anatomy, University of Chicago, 1057 E 57th St., Chicago, IL 60637, U.S.A., ‡Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, U.S.A. and §Department of Ecology and Evolutionary Biology, Box G-W, Brown University, Providence, RI 02912, U.S.A.*

(Received 27 October 2008, Accepted 8 June 2009)

The coastal South American species *Cyprinodon dearborni* contains two lineages distinct at both mitochondrial and nuclear loci. One appears to be a long-term South American endemic, whereas the other is a more recent colonizer related to the widespread *Cyprinodon variegatus*.

Key words: *Cyprinodon*; mitochondrial DNA; nuclear DNA; phylogenetics; phylogeography; pupfish.

Exceptions to well-marked biogeographic patterns can provide insights into factors that may have played an important role in the diversification of a group of organisms. The pupfish *Cyprinodon dearborni* (Meek) is a conspicuous exception to a very strong north/south complementary pattern: the oviparous killifishes of North and South America are sharply distinctive, and there is almost no overlap between them. Originally described from the island of Curaçao, off the western Venezuelan coast (Meek, 1909), *C. dearborni* is not a well-studied species. Its reported distribution extends coastwise from the Goajira peninsula in Colombia to Isla de Margarita, Venezuela, and includes Lago Maracaibo and environs, the islands of Curacao, Aruba, Bonaire and Los Roques, and several interior Venezuelan lakes. Unconfirmed records from northwest Colombia, Trinidad and Guyana (Wildekamp, 1995) suggest the possibility of an even more extensive distribution along the South American coast. Besides *C. dearborni*, and the endemic genus *Orestias* of the Andean altiplano, all other pupfishes (including all other species of *Cyprinodon*) and related genera (*e.g. Jordanella*) are nearctic or Caribbean in distribution (Nirchio *et al.*, 2003). In fact, most South American killifishes are rivulids (*Rivulus*, *Cynolebias* and

AQ1 †Author to whom correspondence should be addressed. Tel.: 401 632 9316; email: robert.a.haney@gmail.com

allied genera) and are more closely related to the aplocheilid killifishes of Africa and Asia than to any nearctic group (Parenti, 1981).

Intuitively, its novel distribution relative to its congeners implicitly suggests that *C. dearborni* probably stems from a single colonization of the South American coast by some ancestral (Caribbean) pupfish. Previous molecular data (Echelle *et al*., 2005, 2006), derived from a specimen from Bonaire, were consistent with a single, recent colonization event, for they placed the species within a well-resolved 'maritime clade' of North American coastal and Caribbean forms related to *Cyprinodon variegatus* (*sensu lato*). This colonization event would seemingly have occurred across a gap in the current distribution of the maritime clade in the Southern Caribbean, as *Cyprinodon artifrons* (Hubbs), a coastal species with an intervening distribution, is far more divergent from members of the maritime clade (Echelle *et al*., 2005). As sampling to date, however, has been limited, diversity in *Cyprinodon* along the South American coastline may have been underestimated.

In this study, both mitochondrial and nuclear sequence data were used to extend knowledge of South American diversity in *Cyprinodon*. Specifically, these were used to test for the presence of additional lineages of *Cyprinodon* on the South American coastline and to explore their position within the phylogeny of the genus. Furthermore, the monophyly of *C. dearborni* was tested and biogeographic scenarios revised for the origins of South American populations of *Cyprinodon*.

Cyprinodon artifrons was sampled from Carrie Bowe Caye, Belize. *Cyprinodon dearborni* was sampled from Laguna de Restinga on Isla de Margarita in eastern Venezuela and from Bonaire. Samples of *C. variegatus* from two nominal subspecies (*C. v. ovinus* and *C. v. variegatus*) were collected at four locations (Tuckerton, NJ; Georgetown, SC; Sapelo Island, GA; Port Fourchon, LA) along the Atlantic and Gulf coasts of the United States (Haney *et al*., 2007).

A fragment of the mitochondrial NADH dehydrogenase subunit 2 (ND2) locus was amplified from 16 individuals of *C. dearborni* from Isla de Margarita, one individual of *C. dearborni* from Bonaire and 18 *C. artifrons*, using primers L4633 and H5334 from Miya & Nishida (1999). A fragment of the mitochondrial control region was amplified from these individuals using primers A and B from Lee *et al*. (1995). Mitochondrial amplicons were sequenced in the forward direction. Composite mitochondrial haplotypes were constructed for each individual. Composite haplotypes from *C. variegatus* and other species of *Cyprinodon*, including a single mitochondrial haplotype from the *C. dearborni* Bonaire sample, were derived from previous studies (Echelle *et al*., 2005, 2006; Haney *et al*., 2007). A fragment of the recombination activating gene 1 (RAG1) locus was amplified using the primers RAGB and RAGCpup (Carson & Dowling, 2006) from 10 individuals of *C. dearborni* from Isla de Margarita, two individuals of *C. dearborni* from Bonaire, 11 of *C. artifrons* and 38 individuals from four populations of *C. variegatus* (two from each subspecies: *C. v. ovinus* and *C. v. variegatus*). RAG1 was sequenced in both directions. Only high quality reads (phred score of 30 or above) as defined by default settings in Sequencher 4·5 (Gene Codes Corporation, Ann Arbor, MI, U.S.A.) were used, and heterozygous sites were used when multiple peaks were apparent in both directions, with secondary peaks at least 60% of primary. Most heterozygous individuals were only so at a single site, making determination of the phase of the two alleles present unambiguous. For two multiple heterozygotes, PHASE 2·1·1 (Stephens *et al*., 2001) was used to resolve the phase of each allele. This program implements a Markov

chain Monte-Carlo method to estimate the phase of haplotypes in the sample from unphased genotypic data. Homozygous individuals were identified as 'known phase' to improve resolution.

Recombination in the nuclear sequence data was tested for by estimating the minimum number of recombination events using the four-gamete test of Hudson & Kaplan (1985) in DnaSP 4·0 (Rozas & Rozas, 1999).

Best-fit models of nucleotide substitution for sequence data were determined using Modeltest 3·5 (Posada & Crandall, 1998) and the Akaike information criterion (AIC). Neighbor-joining (NJ) trees were assembled in MEGA4 (Tamura *et al*., 2007), maximum-likelihood (ML) phylogenies were constructed using PhyML 2·4·4 (Guindon & Gascuel, 2003) and Bayesian analysis was performed using MrBayes 3·1 (Ronquist & Huelsenbeck, 2003). Nodal support was estimated with 5000 bootstrap pseudoreplicates for the neighbour-joining phylogeny, with 1000 bootstrap pseudoreplicates for the ML tree and via posterior probability values in the Bayesian analysis.

Phylogenetic reconstruction was performed for mitochondrial data obtained for this study together with a subset of 24 two-loci haplotypes from previous studies representing all known major clades of *Cyprinodon* (Echelle *et al*., 2005). Analysis with nuclear sequences used all distinct alleles identified in this study, in addition to RAG1 sequences from the species *C. bifasciatus* and *C. atrorus* from a previous study (Carson & Dowling, 2006).

The significance of the difference in likelihood between the ML and *C. dearborni* monophyly constraint trees was assessed with the Shimodaira-Hasegawa (SH) test (Shimodaira & Hasegawa, 1999) implemented in PAUP* 4·04b10 (Swofford, 1998).

The mitochondrial alignment was 1013 bp and consisted of sequence from two loci (control region: 353 or 354 bp; ND2: 651 bp) that defined 51 distinct composite haplotypes from 51 individuals of *Cyprinodon*. Thirteen distinct control region haplotypes and 10 distinct ND2 haplotypes from *C. dearborni* and *C. artifrons* were newly obtained and were deposited in Genbank under accession numbers GQ180997-GQ181008 for the control region and GQ181009-GQ181018 for ND2.

This dataset had 329 variable and 252 parsimony-informative sites. The mtDNA tree indicates two distinct lineages occurring within *C. dearborni*, hereafter referred to as Lineage 1 and Lineage 2. Lineage 1 contains sequences from both Bonaire individuals and receives strong support from both NJ and ML bootstrap values and from Bayesian posterior probability (Fig. 1). This distinct lineage clearly falls within the strongly supported maritime clade, together with Atlantic and Gulf coast populations of the wide-ranging *C. variegatus* and several endemic species of the West Indies (Echelle *et al*., 2005). The net Tamura-Nei gamma-corrected distance between Lineage 1 and *C. variegatus* (not including sequences of *C. v. ovinus*) is 2·4%. A second strongly supported clade (100% NJ bootstrap, 100% ML bootstrap and 1·00 posterior probability), however, is found within *C. dearborni* and includes all haplotypes sampled from Isla de Margarita. This lineage (Lineage 2) clearly falls outside the maritime clade. The net mean distance between the two clades of *C. dearborni* under a Tamura-Nei nucleotide substitution model with rate variation among sites was 10·21%. A deeper clade containing Lineage 2, along with the maritime and Great Plains clades (2A and 2B) of Echelle *et al*. (2005), is strongly supported by all three phylogenetic methods.

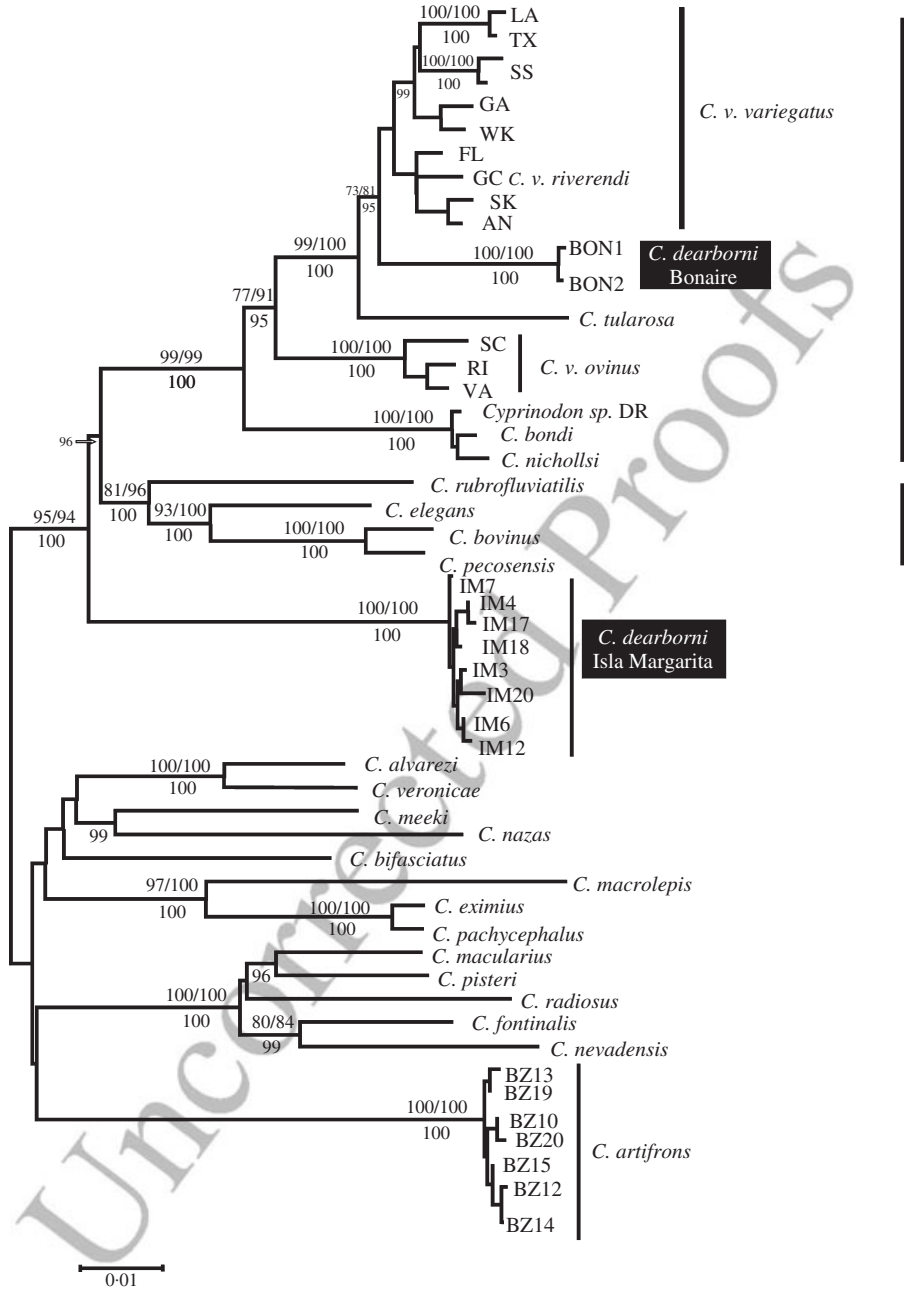

FIG. 1. Midpoint-rooted neighbour-joining tree based on 1013 bp of mitochondrial sequence. Numbers above branches represent neighbour-joining and maximum-likelihood bootstrap values above 70% separated by a slash. Below branches are Bayesian posterior probabilities greater than 0·95, multiplied by 100. The following three sets of values were omitted from the figure due to space considerations: GA + WK = 88/84, 100, FL + GC + SK + AN = 92/90, SK + AN = 93/95, 100.

The fit of the recovered ML topology was compared to a tree with *C. dearborni* constrained to be monophyletic. The result was a significantly lower likelihood for the constraint tree and a rejection of *C. dearborni* monophyly (SH-test, $P < 0{\cdot}001$).

Considering the RAG1 nuclear locus, 294 bp of sequence from 61 individuals yielded seven distinct alleles of the 122 total sampled. All distinct nuclear alleles obtained for this study were deposited in Genbank under accession numbers GQ180990-GQ180996. No recombination is inferred in the nuclear sequence data, as the estimate of the minimum number of recombination events is zero. There were four alleles identified in *C. variegatus*, two in *C. artifrons* and three in *C. dearborni*. One of the three alleles found in *C. dearborni* was identical to a common allele in *C. variegatus*. This allele, however, was only sampled in two individuals at the Bonaire site, in individuals with maritime clade mitochondrial DNA. Similarly, two different alleles were sampled from *C. artifrons*, but one was identical to a second common allele in *C. variegatus*. Although Lineage 1 of *C. dearborni* appeared fixed for one allele that was common in all sampled populations of *C. variegatus*, both alleles found in 10 individuals of Lineage 2 had a fixed nucleotide difference from all *C. variegatus* alleles.

In the alignment of the seven distinct alleles, there were six variable and two parsimony-informative sites. When three alleles from *C. bifasciatus* and *C. atrorus* from a previous study (Carson & Dowling, 2006) were aligned, there were 11 variable and four parsimony-informative sites in the alignment.

Three nodes receive marginal to solid support in phylogenetic analysis of the nuclear RAG1 locus (Fig. 2). The node joining the two alleles from *C. dearborni* Lineage 2 is supported by a bootstrap value of 65% from NJ bootstrap analysis, 67% from ML bootstrap and a posterior probability of 0·99.

The monophyly of *C. dearborni*, as currently described, is clearly rejected by molecular data. First, two divergent mitochondrial lineages occur within this species.

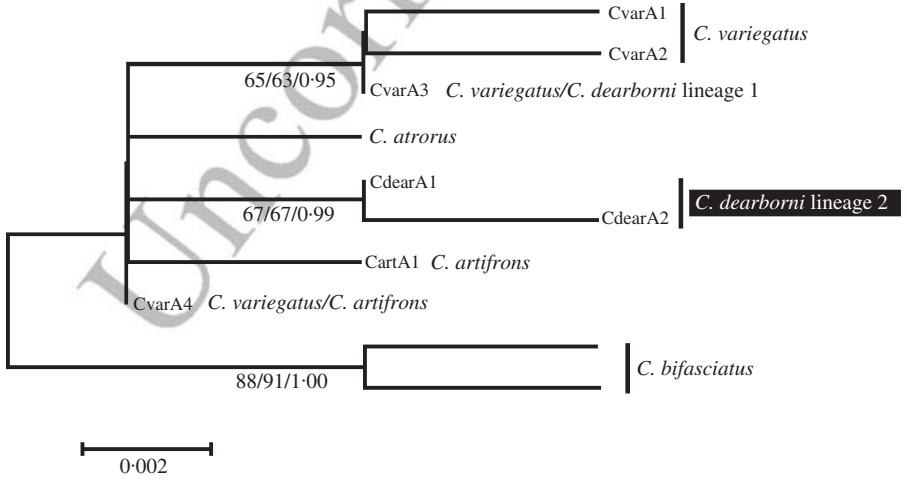

FIG. 2. Midpoint-rooted neighbour-joining tree based on 294 bp of nuclear RAG1 sequence. Numbers separated by slashes below branches represent NJ and ML bootstrap values, followed by Bayesian posterior probabilities.

One of these lineages falls within the maritime clade of Echelle *et al*. (2005) and is closely related to *C. variegatus*. A second, previously unrecognized lineage, however, also occurs within *C. dearborni* and falls outside the maritime clade, occurring as a basal branch within clade 2 of Echelle *et al*. (2005). Second, sequence data from the nuclear RAG1 locus also support a distinction between Lineage 1 and Lineage 2 of *C. dearborni*. The molecular evidence, from independent loci, suggests long-term isolation of the two lineages within *C. dearborni* and indicates that the South American lineage identified in this study may represent a previously unrecognized species distinct from the previously recognized lineage of *C. dearborni* and all other taxa that comprise the maritime clade of *Cyprinodon*. Further data from other sources, however, are necessary to confirm the status of this newly recognized lineage. Interestingly, an earlier allozyme study (Alfonsi *et al*., 2003) identified two genetically distinct populations from two sites in eastern Venezuela, which also showed morphological differentiation. It is tempting to speculate that these populations might represent the same lineages recognized from the sequence data presented here, but a lack of overlap of either molecular markers or sampling localities precludes firm conclusions. This newly identified lineage is also clearly divergent from *C. artifrons*, whose distribution is closest geographically to that of *C. dearborni*. Although the phylogenetic relationships within this genus have been investigated by molecular methods in previous studies (Echelle *et al*., 2005; 2006; Haney *et al*., 2007), this result suggests that further diversity may remain to be discovered, particularly in tropical regions. It also highlights the need for taxonomic revision, as the current taxonomy within the genus *Cyprinodon* in many cases, particularly within the maritime clade, does not appear to reflect evolutionary history, as pointed out by Echelle *et al*. (2006).

Resolution of the zoogeographic and population-level relationships of the two lineages within nominal '*Cyprinodon dearborni*' must await more extensive sampling, but it seems clear that there were at least two successful colonizations of coastal South America by ancestral pupfishes, and that one of these occurred much earlier than the other. Divergence of the maritime and Great Plains clades (2A and 2B; see Fig. 1) within clade 2 occurred between three and six million years ago (Echelle *et al*., 2005). The position of Lineage 2 of *C. dearborni* within the phylogeny suggests that it diverged from the other lineages within clade 2 on a similar time frame. The more recent colonization of South America by a maritime clade lineage apparently occurred during Pleistocene diversification within this clade (Echelle *et al*., 2006). These colonization events have contributed to the extensive diversity occurring within *Cyprinodon*.

We thank J. Scanlan and M. Nirchio for specimens of *C. dearborni* from Bonaire and the Isla de Margarita, respectively, and B. Silliman for specimens of *C. variegatus*. This work was supported by NOAA-NERRS and Rhode Island EPSCoR fellowships to R.A.H. and NSF (DEB 0108500 and 0343464) to D.M.R.

References

Alfonsi, C., Lopez, H. & Perez, J. E. (2003). Genetic and morphological characteristics of the populations of *Cyprinodon dearborni* (Atherinomorpha: Cyprinodontidae) in Chacopata and Laguna de los Patos, Venezuela. *Revista de Biologia Tropical* **51**, 7–15.

Carson, E. W. & Dowling, T. E. (2006). Influence of hydrogeographic history and hybridization on the distribution of genetic variation in the pupfishes *Cyprinodon atrorus* and *C. bifasciatus*. *Molecular Ecology* **15**, 667–679.

Echelle, A. A., Carson, E. W., Echelle, A. F., Van Den Bussche, R. A., Dowling, T. E. & Meyer, A. (2005). Historical biogeography of the New-World pupfish genus *Cyprinodon* (Teleostei: Cyprinodontidae). *Copeia* **2005**, 320–339.

Echelle, A. A., Fuselier, L., Van Den Bussche, R. A., Rodriguez, C. M. L. & Smith, M. L. (2006). Molecular systematics of Hispaniolan pupfishes (Cyprinodontidae: *Cyprinodon*): implications for the biogeography of insular Caribbean fishes. *Molecular Phylogenetics and Evolution* **39**, 855–864.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**, 696–704.

Haney, R. A., Silliman, B. R., Fry, A. J., Layman, C. A. & Rand, D. M. (2007). The Pleistocene history of the sheepshead minnow (*Cyprinodon variegatus*): non-equilibrium evolutionary dynamics within a diversifying species complex. *Molecular Phylogenetics and Evolution* **43**, 743–754.

Hudson, R. R. & Kaplan, N. L. (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.

Lee, W., Conroy, J., Howell, W. H. & Kocher, T. D. (1995). Structure and evolution of teleost mitochondrial control region. *Journal of Molecular Evolution* **41**, 54–66.

Meek, S. E. (1909). New species of fishes from tropical America. *Field Columbian Museum, Zoological Series* **7**, 207–211.

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

Nirchio, M., Cequea, H. & Turner, B. J. (2003). Karyotypic characterization and nucleolus organizer regions in *Cyprinodon dearborni* (Meek, 1909) from Venezuela. *Interciencia* **28**, 352–354.

Parenti, L. R. (1981). A phylogenetic and biogeographic analysis of Cyprinodontiform fishes (Teleostei, Atherinomorpha). *Bulletin of the American Museum of Natural History* **168**, 335–557.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Rozas, J. & Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.

Shimodaira, H. & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**, 1114–1116.

Stephens M., Smith N. J. & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978–989.

Swofford, D. L. (1998). *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, version 4.04b10. Sunderland, MA: Sinauer Associates.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599.

Wildekamp, R. H. (1995). *A World of Killies, Atlas of the Oviparous Cyprinodontiform Fishes of the World*, Vol II. Mishawaka, IN: American Killifish Association.

QUERIES TO BE ANSWERED BY AUTHOR

Queries from the Copyeditor:

AQ1. Please provide fax number in details.

AQ2. We have inserted ‘with’ before ‘1000 bootstrap’. Please confirm if correct.

AQ3. Please check if ‘obtained for this study’ should be ‘obtained from this study’ ?