Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 9271–9274, October 1993
Evolution

Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*

GIACOMO BERNARDI*, PAOLO SORDINO†, AND DENNIS A. POWERS

Department of Biological Sciences, Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950

Communicated by John C. Avise, July 19, 1993 (received for review June 9, 1993)

ABSTRACT **Molecular phylogenies using mitochondrial DNA and nuclear alleles of the lactate dehydrogenase B locus were found to be concordant for populations of *Fundulus heteroclitus* ranging from Canada to Florida. Both mitochondrial DNA and lactate dehydrogenase alleles show a clear separation between the northern individuals (from Nova Scotia and Maine) and the southern ones (from Georgia and Florida), with a mixed population found in the geographic intermediate (New Jersey). An historical isolation, possibly as ancient as 0.5–1 million years old, may have played a role in shaping the situation observed today.**

The relationship between phylogeny and geographic distribution constitutes the phylogeographic pattern of any species. Among the most striking of phylogeographic patterns are those that involve genealogical concordance. Conceptually, there are two types of genealogical concordances: gene–gene phylogenetic concordances and gene–geography phylogenetic concordances (1). Gene–gene phylogenetic concordances involve agreement among unlinked loci in the phylogenetic arrangements of alleles. Such concordances are likely to arise only when populations have been reproductively separated from one another for reasonably long periods of time. Gene–geography phylogenetic concordance involves geographic agreement between the gene phylogenies of codistributed species. In principle, the magnitude of phylogenetic divergence in a gene tree and the degree of spatial structure are independent variables among populations of a species. In the present study, we have used mitochondrial and nuclear DNA from the teleost fish *Fundulus heteroclitus* to determine gene phylogenies and evaluate their concordance.

Populations of *F. heteroclitus* are distributed nearly continuously from Canada to Florida. Northern and southern populations are known to freely interbreed in the laboratory and in nature. The contact zone between the northern and the southern populations is typified by morphological, behavioral, physiological, mtDNA, and nuclear gene clines (2–6).

Indirect measures based on mtDNA indicated that the gene flow among *F. heteroclitus* populations from an 8-km stretch of Chesapeake Bay is strong enough to prevent differentiation due to genetic drift (7). A recent analysis of mtDNA restriction site polymorphisms showed, however, a clear separation between northern and southern population assemblages (6).

The *LDH-B* locus, which encodes the heart-type lactate dehydrogenase, varies clinically along a latitudinal gradient. Populations at the latitudinal extremes are virtually fixed for two different codominant allozymes (4) ($LDH\text{-}B^a$ in southern populations, $LDH\text{-}B^b$ in northern populations). These *LDH-B* allozymes are kinetically different for several characteristics such as substrate affinities, reaction rates, heat stabilities, and inhibition constants (8–10). Fish with the northern allozyme swim faster than their southern counterparts at 10°C (11), and the metabolic rates of developing embryos are also different for the two forms (12). Taken together, these data strongly suggest that this locus is affected by environmental temperature in an adaptationally important manner.

To evaluate the phylogenetic concordance between mitochondrial and nuclear DNA genealogies, we have (*i*) analyzed restriction fragment length polymorphism (RFLP) data from 10 mtDNA haplotypes taken from the literature (6); (*ii*) PCR amplified and directly sequenced 8 mitochondrial cytochrome *b* genes; and (*iii*) PCR amplified, cloned, and sequenced 19 alleles of the *LDH-B* nuclear locus, from individual fish sampled in different localities from the northern to the southern extremes of the distribution of the species.‡

MATERIALS AND METHODS

DNA Amplifications and Sequences. *Lactate dehydrogenase.* The fish in this study were collected from Nova Scotia, Maine, New Jersey, Georgia, and Florida. All allelic *LDH-B* sequences were amplified by PCR after cDNA synthesis from total RNAs extracted from liver or heart tissues. The PCR products were cloned in the phagemid Bluescript (Stratagene). One to eight clones per individual were sequenced using Sequenase 2.0 (United States Biochemical) on both strands. A congeneric species living on the West Coast of the United States, *Fundulus parvipinnis* (Santa Barbara, CA), was used as an outgroup. The PCR amplifications used the following primers: external primers, −10H, 5′-GCA AAGCTT TCCCAGACTTTTCTCCT-3′; +10B, 5′-GCA GGATCC CTGTGAGACGACGGG-3′ leading to a 1023-bp product; internal primers, 18H, 5′-GCA AAGCTT ATG TCC TCA GTC CTG CAG-3′; 986B, 5′-GCA GGATCC CAG GTC CTT CAG GTC CTT C-3′ leading to a 968-bp product, where underlined sequences represent restriction sites added to the primer, and numbers indicate the position of the nucleotide at the 3′ end in the *F. heteroclitus* sequence. The following sequencing primers were used: 142F, 5′-GAG CTC GCT CTG GTG GAC-3′; 343F, 5′-GTC AAC GTC TTC AAG TGC-3′; 553F, 5′-TCC TTC AAC GGC TGG GTG-3′; 808F, 5′-CGC GTC CAC CCC GTC TCC-3′; 181R, 5′-CAG GTC GAT GAT CTC CCC-3′; 376R, 5′-GTT GGG GCT GTA CTT TAT-3′; 583R, 5′-GGG AAC GCT GGT GTC TCC-3′; 850R, 5′-GAA GAC CTC CTC GCC GAT-3′, where numbers indicate the position of the nucleotide at the 3′ end, and F and R refer to primers that bind to the noncoding and coding

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RFLP, restriction fragment length polymorphism; MP, maximum parsimony; NJ, neighbor joining; UPGMA, unweighted pair group method with arithmetic mean; Mya, million years ago.

*To whom reprint requests should be addressed.
†Present address: Stazione Zoologica Anton Dohrn, Naples, 80121, Italy.

‡The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L23771–L23797).

strands, respectively. Two additional primers, binding the vector, T3 (5′-ATTAA CCCTC ACTAA AG-3′) and T7 (5′-AATAC GACTC ACTAT AG-3′) were also used for sequencing.

Cytochrome b. A 307-bp (excluding the priming sites) fragment of the cytochrome *b* gene was PCR amplified. The primers and protocols used for the amplification were from ref. 13. Direct sequencing was performed using the fmol sequencing system (Promega) or after isolating a strand linked to magnetic Dynabeads (Dynal, Great Neck, NY) and then using Sequenase 2.0 (United States Biochemical). Whenever possible, the same individuals were sequenced for *LDH-B* alleles and cytochrome *b*. These individuals were Nova Scotia 1, Nova Scotia 2, Georgia 2, and Florida 1 (see also Figs. 1 and 2).

Phylogenetic Analyses. Tree topologies were obtained by three different methods: maximum parsimony (MP), neighbor joining (NJ), and unweighted pair group method with arithmetic mean (UPGMA). The outgroup *F. parvipinnis* was used to root the phylogenetic trees. The tree topology and branch length from a maximum parsimony analysis were estimated by the computer program PAUP using a heuristic search (16). NJ (not shown) and UPGMA (not shown) phylogenetic trees were generated by using the PHYLIP package (17), version 3.4. The degree of support for internal branches of the trees was further assessed by bootstrapping with 2000 replicates (18) performed with PAUP for MP, or SEQBOOT, DNADIST, and CONSENSE (from PHYLIP) for NJ and UPGMA.

Time of Divergence. The divergence between *LDH-B* alleles from northern and southern populations was calculated by using only substitutions at 4-fold degenerate sites to minimize the influence of selection in measuring molecular differences. The divergence between *F. heteroclitus* and *F. parvipinnis*, estimated to have occurred at the beginning of the Pliocene, 5.3 million years ago (Mya) (19), before the closing of the Isthmus of Panama, some 3 Mya (20), was used for external calibration. The divergence for the cytochrome *b* sequences was estimated by using the standard rate of mutation for fish cytochrome *b* of 2.5% per million year (21). The time of divergence between northern and southern mtDNA haplotypes was from the literature (6).

RESULTS

***LDH-B* Alleles.** Two codominant *LDH-B* alleles have been observed by starch gel electrophoresis—*LDH-B*^{a} in southern populations and *LDH-B*^{b} in northern populations (4). Two amino acid differences were found between the two alleles (ref. 22; this study). Position 185 is an alanine in the southern allele and a serine in the northern one and is responsible for a higher stability of the protein at elevated temperature (22). Position 311 is an aspartic acid in the southern allele and an alanine in the northern one and accounts for the charge difference detected in allozymes studied. In the case of geographically intermediate populations (New Jersey), the four possible combinations of amino acids in positions 185 and 311 were observed (i.e., Ser-Ala, Ala-Asp, Ser-Asp, and Ala-Ala). This observation indicates that recombination events may have taken place in the contact zone between the northern and southern alleles. The presence of 31.9% "recombinant" alleles (Ser-Asp and Ala-Ala) in the New Jersey population was recently determined by allele-specific PCR analysis (23), possibly showing a high level of recombination.

Time of Divergence. The sequence divergence between *F. heteroclitus* and *F. parvipinnis* was 4.5% at the 4-fold degenerate sites for the *LDH-B* alleles. Within *F. heteroclitus*, the divergence between northern and southern populations was 0.43%. The inferred time of divergence between these populations was therefore estimated to be ≈1/10th of the time of divergence between the two species, 0.5 Mya.

The sequence divergence for the cytochrome *b* between the northern and southern populations was estimated to have occurred 0.6 Mya, while this divergence has been estimated to have occurred ≈1 Mya, when RFLP data were used (6).

Phylogenetic Relationships. Phylogenetic relationships obtained for the *LDH-B* locus were compared with phylogenetic relationships obtained with mtDNA RFLP data (6) and cytochrome *b* sequences. However, since there is a possibility of recombination events in the New Jersey population, these data were not included in the present analysis. In this

FIG. 1. Variable nucleotides from the coding region of the *LDH-B* locus in *F. heteroclitus* and *F. parvipinnis*. First row contains nucleotide positions numbered according to ref. 14. Second row contains the consensus nucleotide for each site. Nucleotides identical to the consensus are shown as a dot. Each substitution relative to the consensus nucleotide at a site is classified as either synonymous (s) or replacement (r), following ref. 15, and is indicated above the sequences for the polymorphic sites or under the sequences for the fixed sites. Localities of collection are in the first column; sample numbers are in the second column. Letters a and b designate different alleles found in the same individual.

```
                                          11 1 11111 1 111111111122222222222222222222223
                 1111222233333444444677788900 0 11223 3 445556789911223333455566677888990
Positions        70349268912689012369703625446 9 28143 7 561473246917032567703958912039251
-------------------------------------------------------------------------------------------
Consensus        CCCTCGTAATGAATCCTAACTTTCTTCGC T CATAC C TAACCGCTCTACTATAGACATCTCTAGATTTTT

NovaScotia 1     ............................. . ..C.. . .................................
           2     ............................. . ..C.. . .................................
Maine      3     ............................. . ..... . ..........G......................
           4     ............................. . ..... . .................................

Georgia    2     ............................. C ..... T ..G..............................
           3     ................G............ C ..... T ..G........................G.....
           4     ............................T C ..... T ..G..............................
Florida    1     ............................. C ..... T .....................T...........

F.parvipinnis    AAACAACTGGTGGATT.GCACCCTCCTA. . TTCGT T AT.TTAACTC.TATCGACTGA.ATCTC.CCAGC
-------------------------------------------------------------------------------------------
```

FIG. 2. Variable nucleotides from the sequenced region of the cytochrome *b* gene in *F. heteroclitus* and *F. parvipinnis*. First row contains nucleotide positions numbered 1–307 (i.e., the length of the sequenced fragment). Second row contains the consensus nucleotide for each site. Nucleotides identical to the consensus are shown as a dot. Localities of collection are in the first column, sample numbers are in the second column.

study, Nova Scotia and Maine represent the northern population of *F. heteroclitus*, while Georgia and Florida represent the southern population. The three methods used to reconstruct phylogenies, MP, NJ, and UPGMA, gave similar tree topologies. Therefore, only one method (MP) is represented here (Fig. 3). Three, five, and seven most parsimonious trees were obtained for the *LDH-B* alleles, cytochrome *b*, and RFLP data, respectively. Their lengths were 70 steps, 70 steps, and 16 steps, and their consistency indexes were 0.90, 0.98, and 1.0, respectively. A phylogram corresponding to the consensus of the most parsimonious trees for each set of data is represented in Fig. 3. The three methods used for reconstructing phylogenies consistently separated the northern *LDH-B* alleles from the southern ones with bootstrap values close to 70%. Bootstrap values were >90% when the same analysis was performed with only the extreme localities (Florida and Nova Scotia). Phylogenies based on mtDNA RFLP and cytochrome *b* sequence data also distinctly separated the northern population of *F. heteroclitus* from the southern one. Indeed, statistical analysis supported the separation with bootstrap values of 85% and 100% for cytochrome *b* and RFLP data, respectively. *LDH-B* and mtDNA phylogenies therefore showed striking similarities, indicating a strong gene–gene concordance (Fig. 3).

DISCUSSION

The separation observed at the mtDNA haplotype level has been proposed to be the result of secondary intergradation

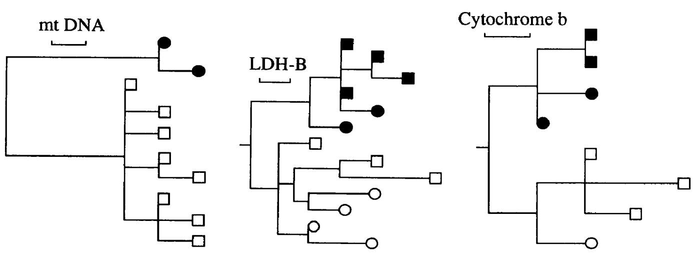

FIG. 3. Phylogenetic relationships of populations of *F. heteroclitus* using *LDH-B* allele sequences, cytochrome *b* sequences, and mtDNA data. *LDH-B* alleles from four localities of *F. heteroclitus* are shown. One allele from Georgia and one from Maine were from the literature (refs. 22 and 14, respectively). mtDNA RFLP data are from the literature (6). Complete *LDH-B* sequences (excluding the primer regions, positions 1–18 and 984–1005) and a 307-bp fragment of the cytochrome *b* gene were used in the phylogenetic analysis. Bars in each tree represent one step. ■, Nova Scotia; ●, Maine; □, Georgia; ○, Florida.

(6). The results obtained with the cytochrome *b* sequences are in agreement with these conclusions and are not unique. Indeed, eight other marine fish species have been shown to have similar genetic breaks along the eastern shore of the United States without recourse to selective arguments (24). However, the gene–gene concordance in *F. heteroclitus* does not imply that the same processes are responsible for the genetic divergence observed for both nuclear and mtDNA. The genetic divergence found in *F. heteroclitus LDH-B* alleles could have originated by two different mechanisms. The first would require a historical subdivision of the species into selectively different environments (e.g., cold north vs. warm south) with directional selection of the isolates. The clines would then arise via migration and gene flow after the removal of the isolating barriers. The second mechanism would require that clines originated in response to selective pressures imposed by an ecological gradient. In other words, the relative fitness for each genotype changed in concert with the environmental gradient.

To determine whether selective pressures play a role in shaping the observed divergences, we have analyzed our *LDH-B* allele data with two different approaches. We first examined the ratios of nonsynonymous to synonymous substitutions (Fig. 1). If protein evolution occurs by neutral processes, the ratio found between *F. heteroclitus* and *F. parvipinnis* is expected to be the same as the ratio obtained within *F. heteroclitus* (15); 15.5% of the fixed differences between species were found to be nonsynonymous, and 29.5% of the polymorphisms within a species were nonsynonymous substitutions. A *G* test of independence (with the Williams correction for continuity) (15, 25) showed that the difference between the two ratios was not statistically significant ($G = 1.3$; $P > 0.1$). The second approach was to perform a phylogenetic analysis of our data after removing the positions known to be under selective constraints (e.g., the nonsynonymous substitutions corresponding to amino acids 185 and 311). If the same phylogeny is obtained with such an analysis, it would suggest that the positions 185 and 311 are not determinant in the observed divergence between the two populations. Conversely, if the phylogeny is changed, it would mean that selection might be responsible for the divergence. When the nonsynonymous substitutions were removed from the analysis, the same phylogenetic tree was found. This phylogeny, however, was weakly supported after a bootstrap analysis. While *LDH-B* alleles have been shown to be under selective constraints (26), our results suggest that natural selection, by itself, does not entirely account for the original genetic separation between northern and southern populations of *F. heteroclitus*.

Some arguments favor the hypothesis of a historical isolation. The gene–gene concordance strongly supports a long period of separation between populations, accompanied by reproductive isolation (1, 27). The estimated dates of divergence between populations and within populations were found to be ancient and roughly corresponding between the two phylogenies. The divergence between northern and southern populations of *F. heteroclitus* was estimated to have occurred ≈0.5 Mya when using *LDH-B* allele sequences, 0.6 Mya when using cytochrome *b* sequences, and 1 Mya when using RFLP data. A geological event might be the original cause for isolation of the northern and the southern populations. Glacial events and/or sea level changes have been proposed as possible causes for the creation of natural "barriers" along the East Coast of the United States and are known to have been subject to important variations in the last million years (28).

Concordant molecular phylogenies were found when using *LDH-B* allele sequences, cytochrome *b* sequences, and mtDNA types from northern and southern populations of *F. heteroclitus*. Our data suggest that a historical isolation of the populations in selectively different environments must have played an important role in shaping the geographical distribution of both mtDNA types and *LDH-B* alleles observed today.

This manuscript is dedicated to the memory of Dr. Galeazzo Riva. We thank J. C. Avise, M. Hare, S. R. Palumbi, T. Lauerman, J. M. Quattro, D. W. Stock, W. S. Nelson, L. DiMichele, and N. L. Crane for discussion; A. Meyer for making a cytochrome *b* sequence from a New York individual available; and L. DiMichele for providing the fish samples. This work was supported by National Science Foundation Grant BSR90-22648 to D.A.P.

1. Avise, J. C. & Ball, R. M., Jr. (1990) *Oxford Surv. Evol. Biol.* **7,** 45–67.
2. Able, K. W. & Castagna, M. (1975) *Am. Zool.* **26,** 145–157.
3. Morin, R. P. & Able, K. W. (1983) *Copeia* 726–740.
4. Powers, D. A. & Place, A. R. (1978) *Biochem. Genet.* **16,** 593–607.
5. Place, A. R. & Powers, D. A. (1978) *Biochem. Genet.* **16,** 577–591.
6. Gonzales-Villasenor, L. I. & Powers, D. A. (1990) *Evolution* **44,** 27–37.
7. Brown, B. L. & Chapman, R. W. (1991) *Evolution* **45,** 1147–1161.
8. Place, A. R. & Powers, D. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 2354–2358.
9. Place, A. R. & Powers, D. A. (1984) *J. Biol. Chem.* **259,** 1299–1308.
10. Place, A. R. & Powers, D. A. (1984) *J. Biol. Chem.* **259,** 1309–1318.
11. DiMichele, L. & Powers, D. A. (1982) *Nature (London)* **296,** 563–564.
12. DiMichele, L. & Powers, D. A. (1982) *Science* **216,** 1014–1016.
13. Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X. & Wilson, A. C. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6196–6200.
14. Crawford, D. L., Constantino, H. R. & Powers, D. A. (1989) *Mol. Biol. Evol.* **6,** 369–383.
15. McDonald, J. H. & Kreitman, M. (1991) *Nature (London)* **351,** 652–654.
16. Swofford, D. L. (1991) *Ill. Nat. Hist. Surv.,* Manual Version 3.0.
17. Felsenstein, J. (1989) PHYLIP (University Herbarium, Univ. of California, Berkeley), Manual Version 3.4.
18. Felsenstein, J. (1985) *Evolution* **39,** 783–791.
19. Griffith, R. W. (1972) Ph.D. thesis (Yale Univ., New Haven, CT).
20. Keigwin, L. D., Jr. (1978) *Geology* **6,** 630–634.
21. Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. (1990) *Nature (London)* **347,** 550–553.
22. Lauerman, T. (1990) Ph.D. thesis (Johns Hopkins Univ., Baltimore).
23. Powell, M. A., Crawford, D. L., Lauerman, T. & Powers, D. A. (1992) *Mar. Mol. Biol. Biotechnol.* **1,** 391–396.
24. Avise, J. C. (1992) *Oikos* **63,** 62–76.
25. Sokal, R. R. & Rohlf, F. J. (1981) *Biometry* (Freeman, San Francisco).
26. Powers, D. A., Smith, M., Gonzalez-Villasenor, I., DiMichele, L., Crawford, D. L., Bernardi, G. & Lauerman, T. (1993) *Oxford Surv. Evol. Biol.* **9,** 81–157.
27. Avise, J. C. (1989) *Evolution* **43,** 1192–1208.
28. Cronin, T. M., Szabo, B. J., Ager, T. A., Hazel, J. E. & Owens, J. P. (1981) *Science* **211,** 233–240.