RESEARCH ARTICLES

Environmental Genomics: A Tale of Two Fishes[1]

Giuseppe Bucciarelli, Miriam Di Filippo,* Domenico Costagliola,* Fernando Alvarez-Valin,† Giacomo Bernardi,‡ and Giorgio Bernardi**

*Laboratory of Molecular Evolution, Stazione Zoologica Anton Dohrn, Villa Comunale, Napoli, Italy; †Sección Biomatemática, Facultad de Ciencias, Iguá, Montevideo, Uruguay; and ‡Department of Ecology and Evolutionary Biology, University of California, Santa Cruz

The influence of the environment on two congeneric fishes, *Gillichthys mirabilis* and *Gillichthys seta*, that live in the Gulf of California at temperatures of 10–25 °C, and up to 42–44 °C, respectively, was addressed by analyzing their genomes. Compared with *G. mirabilis*, *G. seta* showed some striking features. Substitution rates in the mitochondrial genes were found to be extremely fast, in fact faster than in noncoding control regions (D-loops), from which a divergence time of less than 0.66–0.75 Mya could be estimated. In the nuclear genome, 1) both AT → GC/GC → AT and transversion: transition ratios in coding sequences (CDSs) were relatively high; moreover, the ratios of nonsynonymous/synonymous changes (*Ka*/*Ks*) suggested that some genes were under positive selection; 2) DNA methylation showed a very significant decrease; and 3) a GC-rich minisatellite underwent a 4-fold amplification in the gene-rich regions. All these observations clearly indicate that the environment (temperature and the accompanying hypoxia) can rapidly mold the nuclear as well as the mitochondrial genome. The stabilization of gene-rich regions by the amplification of the GC-rich minisatellite and by the GC increase in nuclear CDSs is of special interest because it provides a model for the formation of the GC-rich and gene-rich isochores of the genomes of mammals and birds.

Introduction

Classically, sequence changes in the genome are visualized as resulting from point mutations and recombination. We found, however, that the vertebrate genomes underwent massive regional GC increases at the emergence of mammals and birds (Macaya et al. 1976; Thiery et al. 1976). We proposed (Bernardi and Bernardi 1986; see also Bernardi 2004, 2007) that these changes were due to the need of maintaining the thermodynamic stability of DNA, RNA, and proteins (GC-rich codons preferentially encode amino acids that stabilize proteins; Bernardi and Bernardi 1986; Costantini and Bernardi 2008). This "thermodynamic stability hypothesis" was supported by the finding that the GC increase affected only the gene-rich and not the gene-poor regions of the genome. Indeed, the gene-rich regions are characterized by an open chromatin structure (Saccone et al. 2002; Di Filippo and Bernardi 2008) and need an increased GC level to maintain stability at the increased body temperature, 37–41 °C, of mammals and birds, whereas the gene-poor regions are stabilized by their tightly closed chromatin structure. Because the thermodynamic stability hypothesis is based on a very general principle, one would expect to find it verified in other organisms as well. Indeed, a correlation was found between the optimal growth temperature and genome GC within prokaryotic families (Musto et al. 2004, 2006).

A critical test to determine whether an environmental factor, such as temperature, can affect the structure of the nuclear genome of vertebrates is provided here by comparing the compositional patterns, the DNA methylation, and the nucleotide substitutions in the nuclear genes of two congeneric gobies that live at very different temperatures *Gillichthys mirabilis* and *Gillichthys seta* (Huang and Bernardi 2001; Fields et al. 2002; see fig. 1). Nucleotide substitutions were also investigated in the mitochondrial genes and noncoding control regions (D-loops): Gobioid fishes were recently investigated in both phylogenesis and population genetics (Akihito et al. 2000, 2008).

The long-jawed mudsucker *G. mirabilis* inhabits saltwater creeks in coastal California, Baja California, and the northern Gulf of California. The short-jawed mudsucker *G. seta*, a paedomorphic variant of *G. mirabilis* (Barlow 1961), is restricted to the uppermost tide pools, in the northern Gulf of California, that are reached by sea water only rarely at the highest spring tides. Although *G. mirabilis* lives at 10–25 °C, *G. seta* experiences temperatures that may reach 42–44 °C (among the highest temperatures encountered by any fish; Nelson 2006) and an accompanying hypoxia. *Gillichthys mirabilis* was previously studied in its hypoxia-induced gene expression (Gracey et al. 2001) and its response to heat stress (Hochachka and Somero 2002; Cossins and Crawford 2005; Buckley et al. 2006).

The sister relationship of these two species solves the problem with which we were confronted in our initial work (Bernardi and Bernardi 1986) on two other fishes living at high temperature, the Death Valley pupfish, *Cyprinodon salinus*, and the Lake Magadi tilapia, *Oreochromis alcalicus grahami*, which showed regional GC increases in their genomes, but could only be compared with evolutionarily distant species.

In this study, we used two experimental approaches, working at the genome level and at the level of orthologous coding sequences (CDSs), respectively.

Materials and Methods

Fish Samples and Nucleic Acid Preparation

Gillichthys mirabilis was collected from Estero Morhua, Sonora, Mexico, and *G. seta* from Estero la Pinta (a site ca. 5 km away). DNA and mRNA were prepared from either liver or muscle, using the method of Kay et al. (1952) and the Invitrogen Fast Track 2.0 Kit, respectively.

[1] This paper is respectfully dedicated to His Majesty the Emperor of Japan, a unique scholar of gobiid fishes.

Key words: body temperature, evolution, genomes, minisatellites, isochores, speciation.

E-mail: bernardi@szn.it.

Mol. Biol. Evol. 26(6):1235–1243. 2009
doi:10.1093/molbev/msp041
Advance Access publication March 6, 2009

FIG. 1.—The longjaw mudsucker, *Gillichthys mirabilis* (left), and the shortjaw mudsucker, *Gillichthys seta* (right), are the sole members of the genus. It has been postulated that *G. seta* is a paedomorphic of *G. mirabilis* (Barlow 1961; Bois and Jeffreys 1999). Both species are found intertidally, with *G. mirabilis* being predominantly in the lower intertidal in California and northern Gulf of California and *G. seta* being predominantly in the upper intertidal and restricted to the northern Gulf of California. Although *G. seta* inhabits waters that reach very high temperatures (up to 44 °C) and which are frequently low in oxygen, *G. mirabilis* behaviorally avoids temperatures higher that 23 °C (De Vlaming 1971). The images shown were taken from Smithsonian Tropical Research Institute and from Giacomo Bernardi (see also the link http://biogeodb.stri.si.edu/sftep/gallery.php?show=g&id=566).

Ultracentrifugation

Analytical ultracentrifugation was performed using a Beckman Optima XL-A ultracentrifuge. GC values were calculated from the buoyant densities according to Schildkraut et al. (1962). Shallow gradient ultracentrifugation (De Sario et al. 1995) was performed using a Beckman Optima XL-100K ultracentrifuge; 200 µg, 120 µg, and 50 µg of total DNA were run according to whether fractions had to be used for subsequent analytical ultracentrifugation analysis, cloning, or hybridization analysis, respectively.

Cloning of Shallow Gradient Fractions

Three *G. seta* DNA shallow gradient fractions averaging 44–45% GC were pooled, dialyzed against distilled water, digested for 3 h at 37 °C with restriction enzyme *Alu*I (one of the only two restriction enzymes tested that were able to degrade repeated DNA from *G. seta*, the other one being *Pst*I), and run in a 1.5% agarose gel. Gel slices corresponding to 400-2,500-bp fragments were cut off. DNA was extracted using the QIAquick Gel Extraction Kit by Qiagen, Milan, Italy and cloned using the EcoRV site of a pBluescript plasmid. Transformation was performed on electrocompetent *Escherichia coli* cells and positive clones were sequenced. Sequence similarity was calculated by the sequence identity matrix of the Bioedit program.

cDNA Preparation and Polymerase Chain Reaction Amplification

cDNAs from *G. mirabilis* and *G. seta* were synthesized on mRNAs templates using the Invitrogen Copy Kit; cDNA libraries were prepared by Invitrogen, Milan, Italy and cloned in pCMV-SPORT 6.1. Polymerase chain reaction (PCR) amplifications were performed in a 50-µl final volume of 1× PCR buffer (Roche, Milan, Italy) containing also $MgCl_2$, 0.5 µM primers, 0.2 mM dinucleotide triphosphate and 0.04 U/µl *Taq* polymerase.

Hybridization

An aliquot of each shallow gradient fraction was diluted 70 times with 0.4 M NaOH to a final volume of 200 µl. In all, 100 µl of each diluted fraction was dot blotted on a positively charged nylon membrane (Hybond-N+, Amersham Pharmacia, Cologno Monzese, Milan, Italy). DNA probes (either cDNAs or PCRs) were radiolabeled using the Random Oligo Labeling method and [α-^{32}P]CTP and [α-^{32}P]ATP as radioactive nucleotide precursors.

Hybridization was performed overnight at 65 °C in a 1 M ethylenediaminetetraacetic acid (EDTA), 0.5 M Na_2HPO_4 (pH7.2), and 7% sodium dodecyl sulfate (SDS) solution. Filters were then washed once at room temperature in a 2× standard saline citrate (SSC) and 0.1% SDS solution, once at 65 °C for 30 min in a 2× SSC and 0.1% SDS solution and for an additional 30 min at the same temperature in a 0.5× SSC and 0.1% SDS solution. Hybridization intensity analysis was performed using a Typhoon Trio (Amersham Biosciences, Milan, Italy) and the associate intensity quantification software.

Gene Collection and Analysis

This operation was very laborious and time consuming because we needed a collection of CDS as completely sequenced as possible from both species. The strategy used can be described as follows: 1) 12 complete CDS of *G. mirabilis* available in GenBank were used as templates for designing primers for PCR amplification of the orthologous CDS from *G. seta* cDNA; 2) the other, partial, CDS were derived from our cDNA library of *G. mirabilis* (some of these sequences were also found as partial sequences in GenBank); 3) partial CDS were sequenced and used to find complete orthologous sequences in *Tetraodon nigroviridis*, *Takifugu rubripes*, *Danio rerio*, and *Oryzias latipes*; the heterologous fish sequences were then used to design degenerate primers for PCR amplification of the orthologous sequences present in *G. mirabilis* and *G. seta* cDNAs. In the case of *eno1* gene, two paralogous genes were found. It should be noted that the genes that were sequenced can be assumed to correspond to a representative set of *Gillichthys* genes because 12 *G. mirabilis* sequences were from GenBank and the other 56 sequences originated from our cDNA libraries of both fishes.

After alignment of the homologous genes of the two species, compositional (AT → GC; GC → AT) changes in

first, second, and third codon position along with nonsynonymous (NS) and synonymous (S) changes were calculated. The MEGA version 4.0 program (Tamura et al. 2007) was used to estimate the rates of synonymous (*Ks*) and nonsynonymous (*Ka*) distances of the coding region following Pamilo–Bianchi–Li method (Li 1993; Pamilo and Bianchi 1993).

Mitochondrial Genome Analysis

The complete mitochondrial genomes of *G. mirabilis* and *G. seta* were sequenced (see supplementary table 2, Supplementary Material online). In order to determine whether there was evidence of selection on mitochondrial genes, we used a comparative method, similar in concept to the classical neutrality test (Hudson et al. 1987). We took a region that is presumably under weak selection, the noncoding control region (the D-loop), and determined the ratio of divergences between a given gene and the control region between the two species. Under the hypothesis of purifying selection and neutrality, these ratios are expected to be smaller than 1 because the control region being a noncoding region is more free to vary than protein-coding genes.

Results

The Compositional Patterns of the Nuclear Genomes

The GC profiles of the genomes of *G. mirabilis* and *G. seta* were visualized by analytical cesium chloride centrifugation (fig. 2*A*). As usual with most fish genomes (Bernardi and Bernardi 1990; Bucciarelli et al. 2002), a fairly symmetrical DNA profile was observed for *G. mirabilis*. Although we do not have an outgroup for *Gillichthys* (the outgroup for this species pair is unknown), this result suggests that the profile of the ancestral genome must have looked like the genome of *G. mirabilis*. In contrast, *G. seta* exhibited a shoulder on the GC-rich side of the main peak. After the two DNAs had been subjected to preparative ultracentrifugation in shallow CsCl density gradients (fig. 2*B*), the analytical CsCl profiles of the first two fractions, largely corresponding to the main peak, were identical in both species (fig. 2*C* and *D*). In contrast, the two GC-rich fractions showed a gradual increase in buoyant density in *G. mirabilis*, whereas a shoulder at about 40% GC and a major peak at about 46% GC appeared in *G. seta* (fig. 2*E* and *F*).

The Distribution of Genes in the Nuclear Genome

When cDNAs from both fishes were hybridized on shallow gradient fractions from both DNAs (fig. 3), the hybridization profiles in *G. mirabilis* largely followed the main DNA band, only a small shoulder appearing in the GC-rich range (44–47% GC). In contrast, the *G. seta* profiles showed a major peak in the GC-rich fractions (44–45% GC) and a minor peak corresponding to the main DNA band. Only small compositional differences were found in the GC levels of orthologous CDSs and such differences could not justify the shift of *G. seta* genes to higher densities. We explored, therefore, the possibility that the shift was due to the expansion of interspersed GC-rich sequences.

The Expansion of a Minisatellite in the *G. seta* Genome

When the GC-rich DNA fractions of *G. seta* were cloned after *Alu*I digestion, 10% of the 1,000 clones analyzed, consisted of arrays of a 38-bp tandem repeat sequence (5′-CTGGTTTGGGTTGGACCTGTTTCAGTCCCGTGT-GAGTC-3′) that exhibited a similarity of 80% among each other. The repeats showed a comparatively very high GC level (52% vs. 38% for the main peak), a very strong strand asymmetry (A: 10.8%, G: 33.1%, C: 18.7%, and T: 37.4%), sizes in the 180- to 660-bp range, and short internal repeats (see supplementary fig. 1, Supplementary Material online). Specific primers designed on the basis of *G. seta Alu*I repeats allowed us to sequence the corresponding family from *G. mirabilis*, which showed an interspecific similarity of 81%. The tandem repetition of a 38-bp sequence, itself made of shorter repeats, showed that the *Alu*I sequence was a typical minisatellite. Even if this GC-rich minisatellite will also be called, for the sake of convenience, "*Alu*I sequence" or "*Alu*I repeat," after the restriction enzyme that cut it, it has absolutely nothing to do with the Alu sequences of the primate genome.

When *Alu*I repeat from *G. seta* and *G. mirabilis* were hybridized on the shallow gradient fractions of the corresponding species, they produced a hybridization profile characterized by a minor peak at 40–45% GC and a major peak around 47% GC (fig. 4). The amount of the *Alu*I sequences in the two genomes were then quantified by hybridization with total DNAs from both species and shown to be 2.5 times more abundant in *G. seta* compared with *G. mirabilis*. If one considers, however, that changes essentially concerned the gene-rich regions (see fig. 4), the increase could be estimated to be at least 4-fold in these regions. In contrast, the cloned DNA fragments obtained from PstI digestion (another family of repeats; see supplementary fig. 2 [Supplementary Material online] for the sequence) did not show any difference in amount in the two fishes and its hybridization followed the profile of the main peak (fig. 4). The presence of similar or identical *Alu*I repeats in intergenic sequences of Zebrafish, Stickleback, Medaka, and Tetraodon and the existence of "single-copy" sequences flanking the *Alu*I repeats in *G. seta* indicated that they were interspersed in the genome.

Nucleotide Substitutions in Nuclear Genes

In all, 34 pairs of orthologous nuclear CDS from the two species were investigated (see supplementary tables 1 and 2, Supplementary Material online). Their list is shown in table 1 along with their sequenced sizes (49,331 bp out of a total size of 57,262 bp), the total number of nonsynonymous (NS) and synonymous (S) changes, the corrected proportion of nonsynonymous (*Ka*) and synonymous (*Ks*) changes, the number of transitions (tr) and transversions (tv), and the number of the compositional changes (AT → GC, GC → AT). A presentation of all nucleotide changes is given in supplementary table 3 (Supplementary Material online).

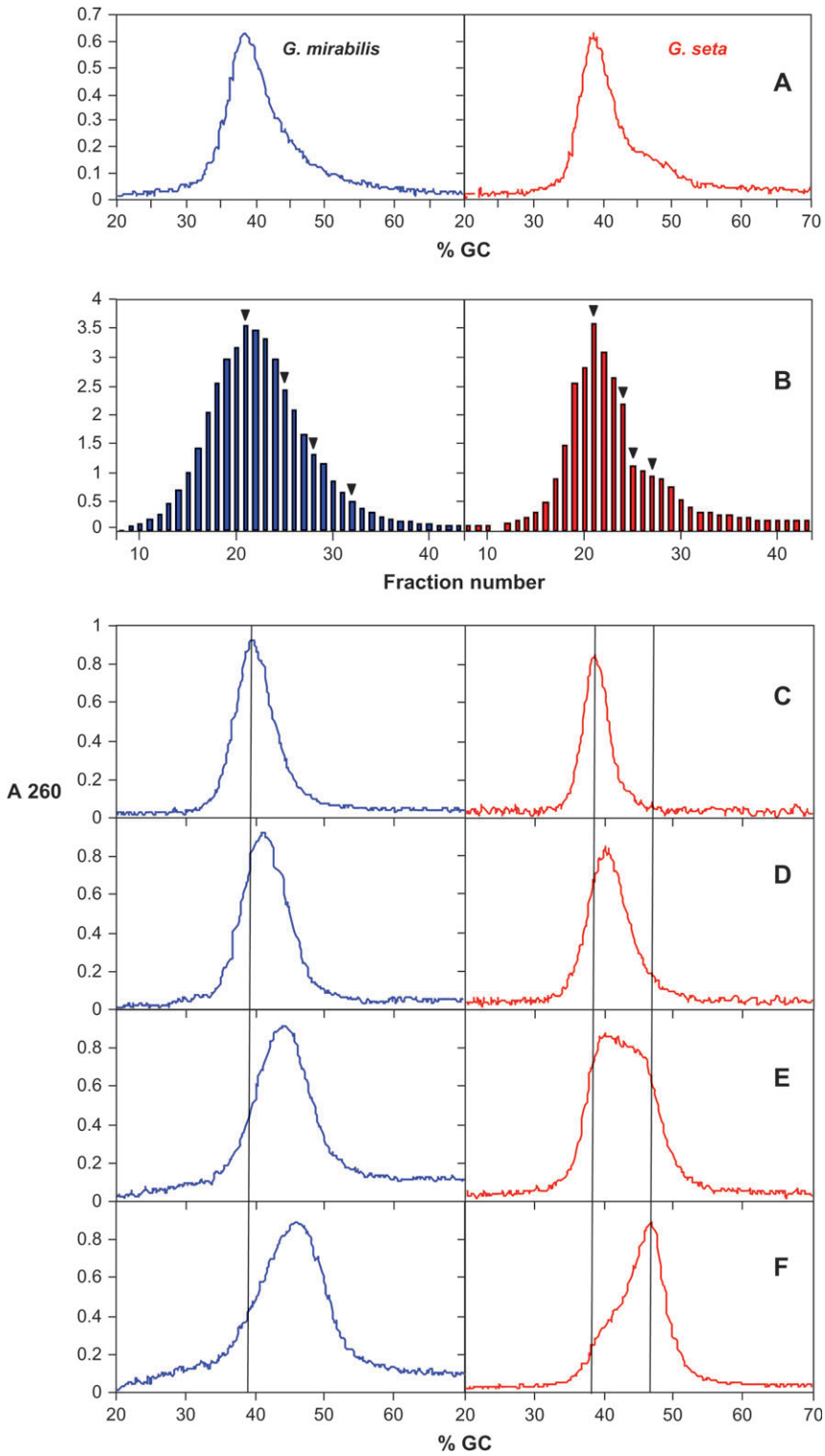

FIG. 2.—(*A*) Analytical ultracentrifugation profiles of DNAs from *Gillichthys mirabilis* and *Gillichthys seta*; buoyant density values were converted to GC values (see Materials and Methods). Optical density values at 260 nm (A260) are given on the ordinate. (*B*) Preparative shallow gradient profiles of DNAs; in this case, abscissa values are fraction numbers. (*C–F*) Analytical ultracentrifugation profiles of shallow gradient fractions indicated by the arrows in (*B*). The two vertical lines correspond to the modal buoyant density of the main peak and of the heavy fraction of *G. seta*, respectively. Blue color corresponds to *G. mirabilis* and red color to *G. seta*.

The data of table 1 indicate that at least two *Ka*/*Ks* values, 0.43 (dio1) and 0.69 (E3), may be indicative of an acceleration in amino acid substitution rates, very likely due to positive selection (see Tang and Wu 2006). Moreover, the transition:transversion ratio was well below the usual 2–10 range in 18 genes (~53% of the sample). Another remarkable feature exhibited by *G. seta* genes was the overall excess of AT → GC over GC → AT changes

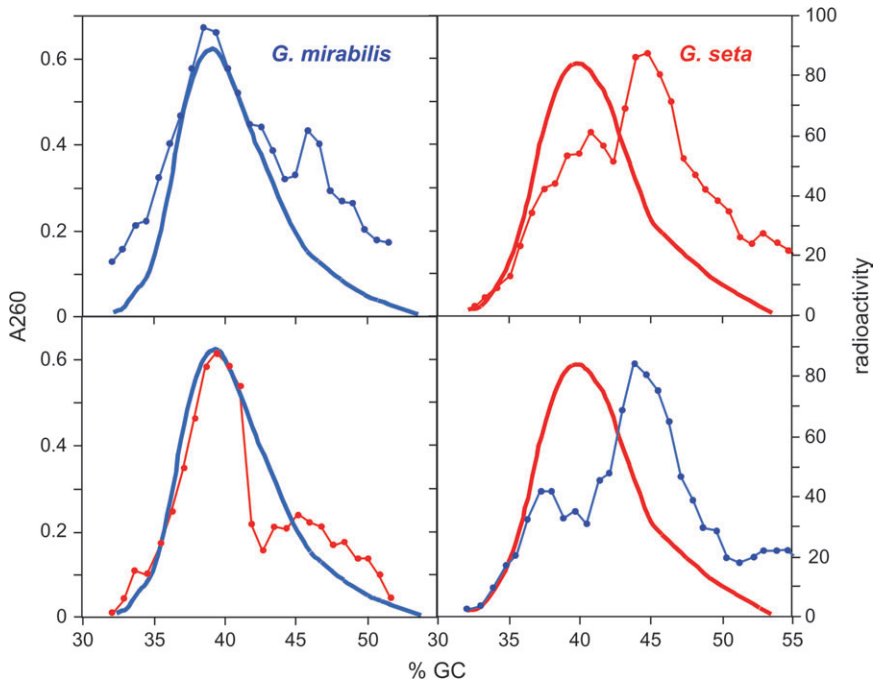

FIG. 3.—Hybridization of *Gillichthys mirabilis* and *Gillichthys seta* cDNAs on shallow gradient fractions of both genomes. The solid curves show the A260 profiles of the shallow gradient and the dotted curves the hybridization signals. Blue color corresponds to *G. mirabilis* and red color to *G. seta*.

(see supplementary fig. 3, Supplementary Material online), this trend being especially pronounced in those genes with high *Ka*:*Ks* ratios. Note that although the direction of changes cannot be determined for every single site (due to the lack of an appropriate outgroup), this excess is unambiguously indicated by the overall values which show a clear predominance of G and C in *G. seta* at evolutionarily variable sites. This contrasts with the normal AT bias found in vertebrate genomes (Eyre-Walker 1999; Smith and Eyre-Walker 2001; Alvarez-Valin et al. 2002; Bernardi 2004, 2007) Finally, an assessment of 5-methylC in the two nuclear genomes showed a significantly lower level (1.54%) in *G. seta* compared with *G. mirabilis* (2.20%).

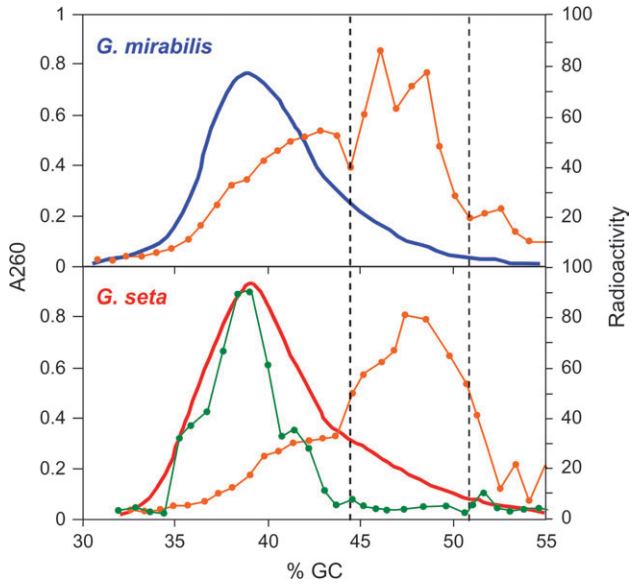

FIG. 4.—Hybridization of the *Gillichthys seta Alu*I sequences on shallow gradient fractions. The solid curves represent the shallow gradient A260 profile and the dotted curves the hybridization signals. The hybridization of the *G. seta* PstI sequence (another repeated sequence, whose primary structure is shown in supplementary fig. 2, Supplementary Material online) on shallow gradient fractions of *G. seta* DNA is represented by the green dotted curve. As in figures 2 and 3, blue color corresponds to *Gillichthys mirabilis* and red color to *G. seta*. Hybridization levels (radioactivity) are normalized to match the A260 profiles. The vertical lines delimit the hybridization peak used to estimate their percentage on the total hybridization.

Nucleotide Substitutions in the Mitochondrial Genomes

We also sequenced the mitochondrial genomes from the two fishes and observed a NS:S ratio equal to 0.15, which is not indicative of positive selection. Selection was, however, also investigated by comparing the divergence in the protein-coding genes with the divergence at the control region (the most variable, presumed nearly neutral mitochondrial locus). Out of 13 protein-coding gene pairs, 12 exhibited a higher or a similar divergence compared with the control region, whereas a lower divergence was expectedly found for all five pairs of sister fish species whose complete mitochondrial genomes were available (fig. 5). Incidentally, preferential AT to GC changes were not observed, but an analysis of amino acid changes revealed a remarkable increase in *G. seta* of alanine, at the expense of serine and threonine (see supplementary table 4 [Supplementary Material online]; an assessment of amino acid changes could not be done in the case of proteins encoded by nuclear genes because of the small number of changes). The usually conserved ribosomal genes (12S rRNA and 16S rRNA) also showed a higher relative rate compared with the other sister species.

Table 1
List of the Orthologous CDS Investigated

	Gene	CDS Size	Sequenced	%	NS[a]	S[a]	NS/S[a]	*Ka*[b]	*Ks*[b]	*Ka*/*Ks*	tr[c]	tv[c]	tr/tv[c]	AT → GC[d]	GC → AT[d]	AT → GC:GC → AT[d]
1	E3	1,041	86		7	4	1.8	0.011	0.016	0.69	7	4	1.8	6	4	1.5
2	gygl	1,018	85		8	7	1.1	0.011	0.036	0.31	8	7	1.1	9	4	2.3
3	dio1	765	82		4	3	1.3	0.009	0.021	0.43	4	3	1.3	2	3	0.7
4	Qtrtd1	1,030	79		5	7	0.7	0.008	0.036	0.22	7	5	1.4	5	4	1.3
5	NDPK-B	450	94		1	2	0.5	0.003	0.014	0.21	2	1	2.0	0	2	GC → AT
6	RL27	411	94		2	4	0.5	0.008	0.032	0.25	5	1	5	4	2	2.0
7	RL11	537	95		2	4	0.5	0.006	0.022	0.27	6	0	tr	4	2	2.0
8	ampd	2,199	78		2	4	0.5	0.002	0.009	0.22	4	2	2.0	2	3	0.7
9	S23	432	94		3	7	0.4	0.012	0.071	0.17	8	2	4.0	8	1	8.0
10	eno1	1,299	93		3	11	0.3	0.003	0.039	0.08	8	6	1.3	8	6	1.3
11	eno1b	1,299	93		2	11	0.2	0.002	0.039	0.05	8	5	1.6	7	6	1.2
12	sdhb	882	82		1	3	0.3	0.002	0.018	0.11	2	2	1.0	2	1	2.0
13	atp5b	1,554	98		2	8	0.3	0.002	0.019	0.11	7	3	2.3	1	8	0.1
14	GDI	621	78		1	4	0.3	0.002	0.032	0.06	3	2	1.5	2	2	1.0
15	rplp1	342	92		1	6	0.2	0.004	0.050	0.08	5	1	5.0	4	2	2.0
16	aco2	2,428	76		4	22	0.2	0.003	0.041	0.07	16	10	1.6	15	9	1.7
17	TPI	747	95		1	6	0.2	0.002	0.030	0.07	5	2	2.5	5	2	2.5
18	GADPH	1,002	94		1	6	0.2	0.001	0.019	0.05	6	1	6.0	4	3	1.3
19	pgam2	768	85		1	11	0.1	0.002	0.059	0.03	10	2	5.0	11	1	11.0
20	S19	444	93		0	2	S				1	1	1.0	1	1	1.0
21	GABA	369	93		0	1	S				1	0	tr	0	1	GC → AT
22	RL23	423	94		0	1	S				1	0	tr	1	0	AT → GC
23	S13	456	94		0	2	S				1	1	1.0	1	1	1
24	RL24	474	95		0	1	S				0	1	tv	0	0	—
25	NRGN	198	87		0	2	S				2	0	tr	1	1	1
26	LDH-A	999	100		0	3	S				2	1	2.0	1	2	0.5
27	pvalb1	330	93		0	1	S				0	1	tv	1	0	AT → GC
28	RXR	1,114	60		0	3	S				2	1	2	2	0	AT → GC
29	S5	612	94		0	4	S				4	0	tr	1	3	0.3
30	S18	459	94		0	4	S				6	1	6.0	2	5	0.4
31	znf207	1,443	63		0	4	S				2	2	1.0	2	2	1.0
32	btub	1,338	92		0	14	S				11	3	3.7	5	8	0.6
33	S8	627	96		0	0	S									
34	HSPG	520	93		0	0										
Total		28,631	85[e]		51	172					154	71		117	89	
Sequenced[f]		49,331														1.3
Average ratio							0.5						2.2			

NOTE.—Full and sequenced sizes of CDS (in base pairs), and the number of nonsynonymous and synonymous nucleotide changes, transitions and transversions, and AT → GC and GC → AT changes from *Gillichthys mirabilis* to *Gillichthys seta* (see supplementary tables 1–4 [Supplementary Material online] for additional information) are presented.

[a] NS and S indicate nonsynonymous or synonymous changes, respectively. In the ratio column, S indicates the presence of only synonymous changes.

[b] *Ka* and *Ks* indicate the rates of nonsynonymous or synonymous substitutions, respectively.

[c] Transitions and transversions are indicated by tr and tv. In the ratio column, tr and tv indicate the presence of only transitions and only transversions.

[d] AT → GC and GC → AT changes and their ratios are shown. Here, AT → GC means that either T and A is observed in *G. mirabilis* and G or C in *G. seta*. Likewise, GC → AT means that G or C is observed in *G. mirabilis* and A or T in *G. seta*.

[e] Weight average.

[f] This size concerns the orthologous genes from both fishes.

Reassessment of Divergence Time between *G. mirabilis* and *G. seta*

Based on mitochondrial cytochrome b sequences, the divergence between *G. mirabilis* and *G. seta* was originally dated at 4.6–11.6 Mya (Huang and Bernardi 2001). Because the mutation rate in protein-coding genes is now known to be much faster than originally thought, this divergence time certainly was a vast overestimate. A divergence assessment based on the control region, 10.4%, is likely to provide a more realistic estimate for the time of divergence between these species. On the basis of divergence of the control regions, which are assumed to be under less selective pressures, and using fish calibrated control region molecular clocks (Domingues et al. 2005), the divergence time may now be estimated at less than 0.66–0.75 Mya. Incidentally, the short divergence time is confirmed by the lack of AT → GC changes in the stems of (nuclear) 18S ribosomal RNA, a molecule which undergoes such changes in vertebrates living at high temperature (Varriale et al. 2008).

Discussion

The striking observations made on the nuclear and mitochondrial genomes of the two fishes can be summarized and commented upon as follows.

(i) The mitochondrial genomes of *G. seta* and *G. mirabilis* exhibit faster substitution rates in all protein CDSs (except for COX-1) compared with D-loops, as

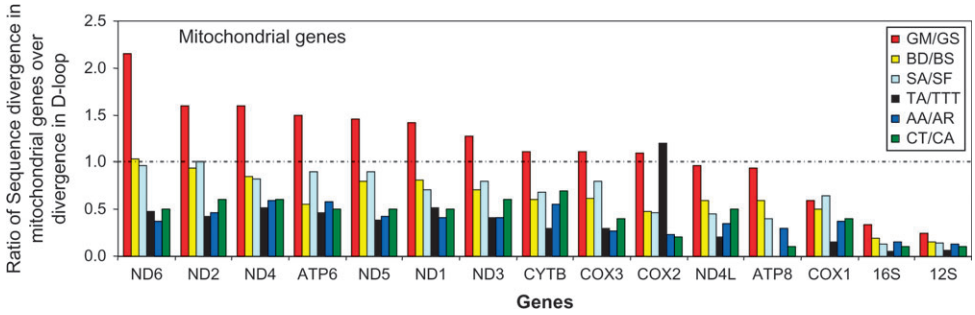

FIG. 5.—Histograms representing the ratios of sequence divergence between mitochondrial genes and the control region of congeneric fish species. The horizontal broken line indicates a ratio of 1. The cytochrome oxydase 2 (COX2) gene of the tuna pair was an exception, possibly related to the unusual thermal control of these species. Species pairs are *Gillichthys mirabilis–Gillichthys seta* (red), *Beryx decadactylus–Beryx splendens* (pale green), *Salvelinus alpinus–Salvelinus fontinalis* (pale blue), *Thunnus alalunga–Thunnus thynnus thynnus* (black), *Anguilla anguilla–Anguilla rostrata* (royal blue), and *Chaunax tosaensis–Chaunax abei* (dark green).

opposed to the other pairs of congeneric fishes investigated, which exhibit the expected trend of faster D-loops. This is the first report of such an acceleration in substitution rates in mitochondrial genes. This finding raises two questions concerning the cause of such high rates and the functional consequences of the changes, respectively. These questions are of great interest because of their potential connection with the environmental changes (temperature and the accompanying hypoxia) but require further work to be answered.

(ii) The strong decrease of nuclear DNA methylation in *G. seta* compared with *G. mirabilis* fits with previous observations on the decrease of mC (5-methylcytosine) with increasing body temperature in fishes (Varriale and Bernardi 2006). This may open up genome regions that were locked by methylation.

(iii) As far as the nucleotide substitutions in nuclear genes are concerned, it should be stressed that the direction of changes is *G. mirabilis* → *G. seta* as indicated by *G. seta* being a paedomorphic variant of *G. mirabilis* (Bois and Jeffreys 1999). The excess of AT → GC over GC → AT changes (a feature particularly evident in genes with high *Ka*:*Ks* ratios) observed in *G. seta* is an indication that CDSs in this species are also undergoing a GC increase. Understandably, this increase takes place at a lower pace than the overall GC increase of the gene-rich regions. These different rates can be readily explained by the evolutionary intrinsic rates responsible for them because changes in the copy number of a repetitive sequence (such as the GC-rich *Alu*I repeat; see below) are more rapid than point mutations (Bois and Jeffreys 1999; Richard and Paques 2000). Concerning the low transition:transversion ratio, it should be recalled that very low ratios were found when comparing orthologous gene from cold- and warm-blooded vertebrates (Perrin and Bernardi 1987).

(iv) The proportion of genes that are very likely under positive selection in *G. seta* (possibly an underestimate due to methodological limitations) appears to be high. It should be recalled that positive selection is classically considered to be very rare. In agreement with this view, only four genes of cichlid fishes (prime examples of adaptive radiation) out of 12,000 tested were found to be under positive selection (Salzburger et al. 2008). Examples of positive selection are, however, increasing in the literature (see, e.g., Endo et al. 1996; Bustamante et al. 2005; Voight et al. 2006; Sabeti et al. 2007).

(v) The expansion of the GC-rich *Alu*I repeats in the *G. seta* genome clearly is the main factor responsible for the gene shift because, as already mentioned, the compositional changes in CDSs are too small to account for it. At the same time, this phenomenon contributes to the stabilization of the gene-rich regions.

It is interesting to compare this expansion of the minisatellites with what happens in the transition from cold- to warm-blooded vertebrates. In the latter case, the GC increase in the gene-rich regions of the genome involved both coding and non-CDSs and took place in a process which extended over very long evolutionary times (see Bernardi 2004; Costantini and Bernardi 2008). In the case of the compositional genome transition from *G. mirabilis* to *G. seta*, CDSs were only at the beginning of the GC increase process. In contrast, the instability and proneness to expansion of minisatellites (Bois and Jeffreys 1999; Richard and Paques 2000) were favored by the temperature increase, the major environmental change in the case under consideration, and were fast. Whether the *Alu*I repeats influence the expression of contiguous genes, a possibility indicated by recent work (Arhondakis et al. 2008; Mahmud et al. 2009), needs further investigations.

At this point, we will consider other factors and mechanisms that might have played a role in the genome changes observed in *G. mirabilis* and *G. seta* as well as in the formation of GC-rich isochores in the genomes of mammals and birds. 1) Major changes in population size are certainly present, *G. seta* having a population size orders of magnitude smaller than that of *G. mirabilis*, but the acceleration in mutation rate in *G. seta* does not account for observations, such as the amplification of minisatellite and the high ratio of AT → GC/GC → AT. 2) As far as life history strategies are concerned, little changes have been observed (Barlow 1961). 3) Biased gene conversion or changes in the pattern of mutational biases that have been proposed to explain the formation of GC-rich isochores can hardly be invoked in the *Gillichthys* case where the nucleotide changes, although

very clear, are limited, and the major quantitative phenomenon is the amplification of a GC-rich minisatellite.

In conclusion, our results indicate that the environment can not only affect the gene expression pattern (Gracey et al. 2001; Buckley et al. 2006) but also mold the genome through natural selection. As such, the changes observed in *G. mirabilis* and *G. seta* provide a model for the genome changes that accompany the emergence of mammals and birds. It has not escaped our notice that such molding of the genome may also affect the tempo and mode of speciation of *Gillichthys*.

Supplementary Material

Supplementary figures 1–3 and supplementary tables 1–5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We wish to thank our colleague Annalisa Varriale for a number of experiments including the determination of 5 mC that were carried out in the early phase of this work. We also thank the Molecular Biology Service of the Stazione Zoologica Anton Dohrn for the library screening and the clone sequences of the present paper, in particular Dr Elio Biffali, Dr Marco Borra, and Mrs Elvira Mauriello.

Literature Cited

Akihito, Fumihito A, Ikeda Y, et al. (11 co-authors). 2008. Evolution of pacific ocean and the sea of Japan populations of the gobiid species, Pterogobius elapoides and Pterogobius zonoleucus, based on molecular and morphological analyses. Gene. 427:7–18.

Akihito, Iwata A, Kobayashi T, et al. (14 co-authors). 2000. Evolutionary aspects of gobioid fishes based upon a phylogenetic analysis of mitochondrial cytochrome b genes. Gene. 259:5–15.

Alvarez-Valin F, Lamolle G, Bernardi G. 2002. Isochores, GC3 and mutation biases in the human genome. Gene. 300:161–168.

Arhondakis S, Clay O, Bernardi G. 2008. GC level and expression of human coding sequences. Biochem Biophys Res Commun. 367:542–545.

Barlow GW. 1961. Gobies of the genus Gillichthys with comments on the sensory canals as a taxonomic tool. Copeia. 1961:423–437.

Bernardi G. 2004. Structural and evolutionary genomics. Natural selection in genome evolution. Amsterdam: Elsevier. (reprinted in 2005)

Bernardi G. 2007. The neoselectionist theory of genome evolution. Proc Natl Acad Sci USA. 104:8385–8390.

Bernardi G, Bernardi G. 1986. Compositional constraints and genome evolution. J Mol Evol. 24:1–11.

Bernardi G, Bernardi G. 1990. Compositional patterns in the nuclear genome of cold-blooded vertebrates. J Mol Evol. 31:282–293.

Bois P, Jeffreys AJ. 1999. Minisatellites instability and germline mutation. Cell Mol Life Sci. 55:1636–1648.

Bucciarelli G, Bernardi G, Bernardi G. 2002. An ultracentrifugation analysis of 200 fish genomes. Gene. 295:153–162.

Buckley BA, Gracey AY, Somero GN. 2006. The cellular response to heat stress in the goby Gillichthys mirabilis: a cDNA microarray and protein-level analysis. J Exp Biol. 209:2660–2677.

Bustamante CD, Fledel-Alon A, Williamson S, et al. 2005. Natural selection on protein coding genes in the human genome. Nature. 437:1153–1157.

Cossins AR, Crawford DL. 2005. Fish as models for environmental genomics. Nat Rev Genet. 6:324–333.

Costantini M, Bernardi G. 2008. The short-sequence designs of isochores from the human genome. Proc Natl Acad Sci USA. 10:13971–13976.

De Sario A, Geigl EM, Bernardi G. 1995. A rapid procedure for the compositional analysis of yeast artificial chromosomes. Nucleic Acids Res. 23:4013–4014.

De Vlaming V. 1971. Thermal selection behaviour in the estuarine goby Gillichthys mirabilis Cooper. J Fish Biol. 3:277–286.

Di Filippo M, Bernardi G. 2008. Mapping Dnase-I hypersensitive sites on human isochores. Gene. 419:62–65.

Domingues VS, Bucciarelli G, Almada VC, Bernardi G. 2005. Historical colonization and demography of the Mediterranean damselfish, Chromis chromis. Mol Ecol. 14:4051–4063.

Endo T, Ikeo K, Gojobori T. 1996. Large-scale search for genes on which positive selection may operate. Mol Biol Evol. 13:685–690.

Eyre-Walker A. 1999. Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. Genetics. 152:675–683.

Fields PA, Kim YS, Carpenter JF, Somero GN. 2002. Temperature adaptation in *Gillichthys* (Teleost: Gobiidae) A4-lactate dehydrogenase: identical primary structures produce subtly different conformations. J Exp Biol. 205:1293–1303.

Gracey AY, Troll JV, Somero GN. 2001. Hypoxia-induced gene expression profiling in the euryoxic fish Gillichthys mirabilis. Proc Natl Acad Sci USA. 98:1993–1998.

Hochachka PW, Somero GN. 2002. Biochemical adaptation: mechanism and process in physiological evolution. Oxford: Oxford University Press. p. 290–312.

Huang D, Bernardi G. 2001. Disjunct Sea of Cortez—Pacific Ocean *Gillichthys mirabilis* populations and the evolutionary origin of their paedomorphic relative, *Gillichthys seta*. Mar Biol. 138:421–428.

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116:153–159.

Kay ERM, Simmons NS, Dounce NL. 1952. An improved preparation of sodium desoxyribonucleate. J Am Chem Soc. 74:1724–1726.

Li WH. 1993. Unbiased estimation of the rates of synonymous and non-synonymous substitution. J Mol Evol. 36:96–99.

Macaya G, Thiery JP, Bernardi G. 1976. An approach to the organization of eukaryotic genomes at a macromolecular level. J Mol Biol. 108:237–254.

Mahmud AA, Amore G, Bernardi G. 2009. Compositional genome contexts affect gene expression control in sea urchin embryo. PLoS ONE. 3:e4025.

Musto H, Naya H, Zavala A, Romero H, Alvarez-Valín F, Bernardi G. 2004. Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. FEBS Lett. 573:73–7.

Musto H, Naya H, Zavala A, Romero H, Alvarez-Valín F, Bernardi G. 2006. Genomic GC level, optimal growth temperature, and genome size in prokaryotes. Biochem Biophys Res Commun. 347:1–3.

Nelson JS. 2006. Fishes of the world. New York: John Wiley and Sons, Inc.

Pamilo P, Bianchi NO. 1993. Evolution of the Zfx and Zfy, genes: rates and interdependence between the genes. Mol Biol Evol. 10:271–281.

Perrin P, Bernardi G. 1987. Directional fixation of mutations in vertebrate evolution. J Mol Evol. 26:301–310.
Richard G, Paques F. 2000. Mini- and microsatellite expansions: the recombination connection. EMBO Rep. 1:122–126.
Sabeti PC, Varilly P, Fry B, et al. (263 co-authors). 2007. Genome-wide detection and characterization of positive selection in human populations. Nature. 449:913–918.
Saccone S, Federico C, Bernardi G. 2002. Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. Gene. 300:169–178.
Salzburger W, Renn S, Steinke D, Braasch I, Hofmann HA, Meyer A. 2008. Annotation of expressed sequence tags for the East African cichlid fish Astatotilapia burtoni and evolutionary analyses of cichlid ORFs. BMC Genomics. 9:96.
Schildkraut CL, Marmur J, Doty P. 1962. Determination of the base composition of DNA from its buoyant density in CsCl. J Mol Biol. 4:430–443.
Smith NG, Eyre-Walker A. 2001. Nucleotide substitution rate estimation in enterobacteria: approximate and maximum-likelihood methods lead to similar conclusions. Mol Biol Evol. 18:2124–2126.
Tamura K, Dudley J, Nei M, Kumar S. 2004. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.
Tang H, Wu C. 2006. A new method for estimating non-synonymous substitutions and its applications to detecting positive selection. Mol Biol Evol. 23:372–379.
Thiery JP, Macaya G, Bernardi G. 1976. An analysis of eukaryotic genomes by density gradient centrifugation. J Mol Biol. 108:219–235.
Varriale A, Bernardi G. 2006. DNA methylation and body temperature in fishes. Gene. 385:111–112.
Varriale A, Torelli G, Bernardi G. 2008. Compositional properties and thermal adaptation of 18S rRNA in vertebrates. RNA. 14:1492–1500.
Voight F, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4:e72.

Norihiro Okada, Associate Editor

Accepted February 6, 2009