Body temperature and evolutionary genomics of vertebrates: a lesson from the genomes of *Takifugu rubripes* and *Tetraodon nigroviridis*

Kamel Jabbaria,1, Giorgio Bernardia,b,*

*a* Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, F-75005 Paris, France

*b* Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, Villa Comunale, I-80121, Naples, Italy

Received 15 December 2003; accepted 5 February 2004

Available online 23 April 2004

Abstract

In this paper, we provide evidence for the body temperature effect on the formation of GC-rich isochores, by analysing genomic sequences from two puffer fishes living at different temperatures. The higher body temperature of *Tetraodon nigroviridis* compared to *Takifugu rubripes* (DTf15°C) appears to be the cause of a higher compositional heterogeneity of the former due to the formation of GC-rich regions. Such an effect does not only concern large DNA segments but also coding sequences.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Genomics; *Takifugu rubripes*; *Tetraodon nigroviridis*

1. Introduction

The compositional approach in the analysis of vertebrate genomes (see Bernardi, 2004 for a review) led to some important discoveries concerning the evolution of those genomes: (1) the compositional compartmentalization into mosaics of isochores; (2) the difference in compositional patterns exhibited by different vertebrate classes; (3) the existence of compositional correlations between coding and flanking noncoding sequences; (4) the bimodal distribution of gene density, which is low in the abundant GC-poor isochores (the genome desert) and high in the scarce GC-rich isochores (the genome core); (5) the bimodal expression, replication and recombination, which are all related to gene distribution.

The major difference in compositional patterns (or genome phenotypes) concerns cold- and warm-blooded vertebrates. The genomes of the former show a narrow compositional spectrum compared to the latter, the difference being essentially due to the presence of GC-rich isochores in the genomes of warm-blooded vertebrates.

The sequencing of human and *Takifugu* genomes confirmed the compositional properties just mentioned (Lander et al., 2001; Aparicio et al., 2002). The fact that the major transition from cold- to warm-blooded vertebrates was accompanied by a change in body temperature led to the thermodynamic stability hypothesis (Bernardi and Bernardi, 1986), which proposed that the formation of GC-rich isochores was linked to an increase in body temperature at the transition from cold- to warm-blooded vertebrates. In line with this, a recent analysis of mean GC3 (the GC level of third codon positions) and standard deviations of GC3 of vertebrate genomes showed a positive correlation between body temperature and GC3 level (Jabbari et al., 2003). Moreover, if families of prokaryotic genomes are examined, correlations between optimal growth temperature and GC levels are found (Naya et al., 2004; see also Nishio et al., 2003).

Until recently, no large DNA sequences from fish genomes were available in data banks. During the last 5 years, the genome projects of *Takifugu* and *Tetraodon* have yielded a wealth of sequences, which allowed us to conduct a large-scale compositional analysis on coding and noncoding sequences.
2. Material and methods

Genomic sequences from Takifugu draft assembly were downloaded from the Joint Genome Institute (JGI) ftp site (Takifugu assembly release 2–17th May 2002). Tetraodon genomic sequences were downloaded from the Genoscope ftp site (Release date: 6th May 2002). Coding sequences from Takifugu were extracted from the Ensembl (www.ensembl.org) database (Clamp et al., 2003) using the data mining tool EnsMart that allows retrieving customised data sets from annotated genomes. Tetraodon CDSs were provided by the French Sequencing Center Genoscope (www.genoscope.fr).

3. Results and discussion

3.1. Compositional distribution of DNA and coding sequences

As shown in Fig. 1, GC histograms of 20, 10 and 5 kb DNA sequences from the two fishes, indicate that the genome of Takifugu is more homogeneous (sd20 kb = 2.5 vs. 3.2) and GC-poorer (45.2% vs. 49.3%) than that of Tetraodon. The two genomes exhibit positive skewness values (0.43 for Takifugu and 0.61 for Tetraodon) with more than 35% of the Tetraodon genome made up of large sequences having a GC level higher than 50%. When the sample size was increased by using smaller window sizes, the results remained fundamentally the same. Yet, the average GC level of the Takifugu sequences appeared to be rather stable whatever the window size, whereas the Tetraodon showed slight changes. These differences are explained by the fact that 5-kb segments represent a higher percentage of the genomes compared to 10- and 20-kb segments, but much more so in Tetraodon compared to Takifugu. Thus, the 5-kb sample is more representative of the whole genome than 10- and 20-kb samples.

As already mentioned, the compositional distribution of high molecular weight DNA molecules from different vertebrates (Thiery et al., 1976; Hudson et al., 1980; Bernardi and Bernardi 1990a,b; Bucciarelli et al., 2002) had shown that, as a general rule, DNAs from cold-blooded vertebrates are characterised by lower intermolecular compositional heterogeneities. The latter, expressed here as standard deviations, basically reflect the spread of GC levels among different fragments in a given DNA preparation. It is interesting to note that the GC richness of Tetraodon is also accompanied by a slight decrease in...
genome size, 385 Mb (Neafsey and Palumbi, 2003) compared to 414 Mb for Fugu (Brenner et al. 1993).

The difference in GC level between orthologous coding sequences showed (Fig. 2) that about 65% of Tetraodon sequences are GC-richer ($\Delta G C > 0$) compared to those from Takifugu. This is in agreement with the difference in GC levels of the respective genomes.

These results can be explained by the difference ($\sim 15 ^\circ C$) in body temperature between the two fishes. Indeed, Tetraodon lives in tropical fresh water, Takifugu in the sea of Japan. The higher heterogeneity of DNA molecules from Tetraodon is what we would expect from the thermodynamic stability hypothesis (Bernardi and Bernardi, 1986), where an increased thermal stability of DNA was made necessary at the emergence of homeothermy.

The reason why only a small part, the gene-rich part of the genome, underwent the change can now be understood. Indeed, the genome core in the interphase nucleus is in an open chromatin structure, whereas the gene-poor compartment (the genome desert) is in a packed chromatin configuration (Saccone et al., 2002). While the latter DNA is stabilised by its dense chromatin structure, the former required to be stabilised at the higher body temperature of warm-blooded vertebrates by a GC increase.

The higher body temperature of warm-blooded vertebrates (birds and mammals) has been accompanied by the formation of GC-rich isochores that are thermally resistant. In line with this result, we recently noticed that GC$_3$ heterogeneity tends also to increase with body temperature of vertebrates (Jabbari et al., 2003).

References
