

Structural and Evolutionary Fish Genomics

Giorgio Bernardi

Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, 80121 Napoli, Italy

Abstract: Investigations from our laboratory have shown that the genomes of fish (and amphibians) exhibit properties that are common to all vertebrates (essentially a strikingly nonuniform gene distribution and a biphasic replication timing) and properties that are remarkably different from those of mammals and birds. The different base compositional heterogeneity, DNA methylation, and CpG levels exhibited by cold- and warm-blooded vertebrates imply a transition in these properties took place. The causes of such massive changes were investigated and led to a better understanding of the evolution of vertebrate genomes and of the relative role played by selection and random drift in shaping those genomes.

Key words: genome organization, genome evolution, vertebrates.

Investigations from our laboratory have shown that the genomes of fish (and cold-blooded vertebrates in general) exhibit properties that are common to all vertebrates and properties that are remarkably different from those of mammals and birds.

Common properties include (1) gene density, which is nonuniform, with a very high gene density in the small “genome core” (the genome compartments characterized by an open chromatin structure) and a very low gene density in the large “genome desert,” which is endowed with a closed chromatin structure (see Figure 1); (2) GC (guanine + cytosine) levels, which are higher in the “genome core” than in the “genome desert”; and (3) replication timing, which is biphasic in the cell cycle, the “genome core,” and the “genome desert” showing an early and a late replication, respectively.

Different properties include (1) the genomes of fish (and cold-blooded vertebrates in general) are characterized by a much smaller compositional heterogeneity (i.e., by a narrower compositional distribution of DNA molecules and coding sequences) compared with the genomes of warm-blooded vertebrates. In general, the highest GC levels attained by fish genomes are much lower than in mammals and birds. (2) Chromosomes of fish (and cold-blooded vertebrates in general) do not show R(everse) bands, but only poor G(iemsa) bands, in sharp contrast with chromosomes from warm-blooded vertebrates which exhibit striking R and G bands. (3) DNA methylation and CpG levels are twice as large in fish (and amphibians) compared with mammals and birds. (4) The genomes of fish (and cold-blooded vertebrates in general) show overall compositional changes that are much larger than those shown by the genomes of warm-blooded vertebrates. (5) Genome contraction–expansion phenomena (which are responsible, for instance, for the very compact genomes of *Tetraodontiformes*) are wide-ranging and common in fish but not in warm-blooded vertebrates.

These data were presented at the Marine Biotechnology Conference 2003 in Chiba, Japan, September 21–27.

■ Corresponding author: Giorgio Bernardi; e-mail: ■

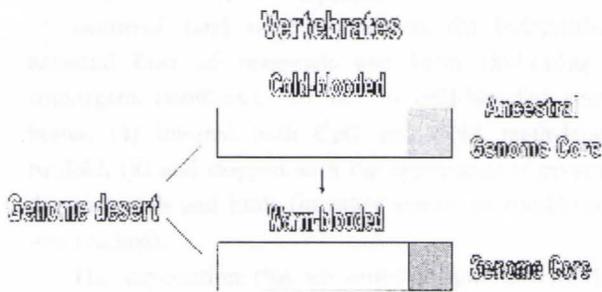


Figure 1. Scheme of the compositional transition shown by the genomes of cold- to warm-blooded vertebrates. The "genome desert" is GC-poor and gene-poor and did not undergo any compositional change. The gene-dense, moderately GC-rich "ancestral genome core" underwent a compositional change into a gene-dense, GC-rich "genome core" (from Bernardi, 2004).

The first three points (the only ones that are discussed here) raise the problem of the existence of a compositional transition (which is depicted in Figure 1) between cold- and warm-blooded vertebrates and of the causes of such transition. Before and after this transition, the genomes of fish/amphibians and of mammals/birds were in a state of compositional equilibrium, a point that can be demonstrated in a very rigorous way by comparing the GC levels of third codon positions (the most variable positions of codons) of orthologous genes from different mammalian orders.

Indeed, in spite of the fact that different mammalian orders diverged from each other about 100 million years ago according to a star-like phylogeny, the GC levels of third codon positions of orthologous mammalian genes are very close to each other (see Figure 2). This compositional equilibrium (as well as the equilibrium of DNA methylation and CpG frequencies) is maintained in the conservative mode of evolution in spite of a strong mutational AT bias, namely, the fact that GC \rightarrow AT changes are more frequent than AT \rightarrow GC changes. Because of the strong linear correlations between GC₃ and GC of intra- or intergenic noncoding DNA, the conservation of GC₃ implies a compositional conservation of isochores (the long, compositionally fairly homogeneous regions forming the vertebrate genome), which can be achieved only by eliminating by negative selection whole regions of the genome, specifically the regions that underwent a strong GC decrease because of the AT bias.

As already indicated, the conservative mode is not the only mode of genome evolution. Indeed, there is a transitional mode which comprises the compositional

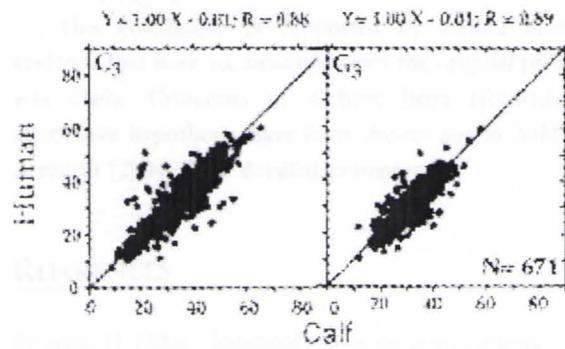


Figure 2. Correlation between G₃ and C₃ values of orthologous genes from human and calf. The orthogonal regression lines coincide with the diagonals. The equations of the regression lines and the correlation coefficients are shown. *N* is the number of gene pairs explored (from Bernardi, 2004).

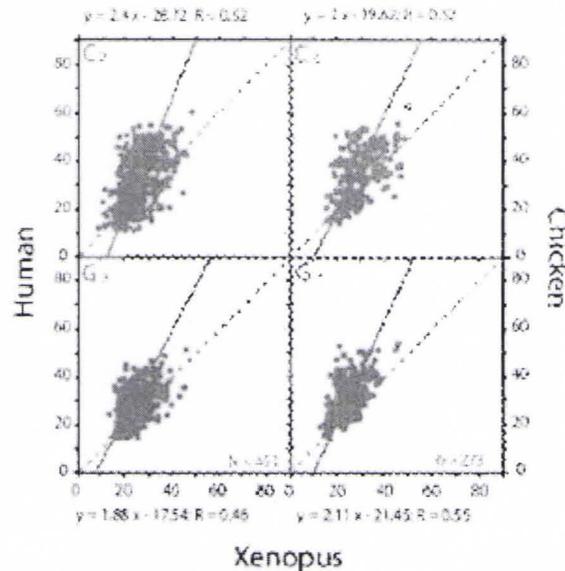


Figure 3. Correlation between G₃ and C₃ values of orthologous genes from human, or chicken, and *Xenopus*. The orthogonal regression lines are shown together with the diagonals (dashed lines). The equations of the regression lines and the correlation coefficients are indicated. *N* is the number of gene pairs explored (from Bernardi, 2004).

changes occurring at the transition between cold- and warm-blooded vertebrates (Figure 1). As in the case of the conservative mode, GC₃ plots of human (or chicken) genes against *Xenopus* (or fish) orthologous genes provide the best evidence for the transitional mode of evolution (see Figure 3). These changes (GC increases) (1) concerned the ancestral genome core and not the genome

desert; (2) affected both coding and noncoding sequences of the genome core at comparable and correlated levels; (3) occurred (and were similar) in the independent ancestral lines of mammals and birds (indicating a convergent evolution), but not in cold-blooded vertebrates; (4) lowered both CpG and DNA methylation twofold; (5) and stopped with the appearance of present-day mammals and birds (in other words, an equilibrium was reached).

The explanation that we originally put forward for the formation and maintenance of GC-rich isochores was natural selection, more precisely negative selection. We suggested that the selective advantages provided by the GC increases were the increased thermodynamic stability of DNA, RNA, and proteins, all these advan-

tages being achieved simultaneously (Bernardi and Bernardi, 1986).

This conclusion is supported by several lines of evidence that have accumulated since the original proposal was made. Criticisms of it have been rebutted and alternative hypotheses have been shown not to hold [see Bernardi (2004) for a detailed presentation].

REFERENCES

- Bernardi, G. (2004). *Structural and Evolutionary Genomics. Natural Selection in Genome Evolution*. Amsterdam: Elsevier.
- Bernardi, G., and Bernardi, G. (1986). Compositional constraints and genome evolution. *J Mol Evol* 24:1–11.