Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 317 (2003) 161–163
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

GC_3 heterogeneity and body temperature in vertebrates

Kamel Jabbari[a], Oliver Clay[b], Giorgio Bernardi[a,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 3 March 2003; received in revised form 25 April 2003; accepted 12 May 2003

Abstract

A recent paper by Belle et al. (J. Mol. Evol. 55 (2002) 356) reported an analysis of mean GC_3 (the GC level of third codon positions) and standard deviations of GC_3 of vertebrate genomes as related to body temperature, and concluded that "the thermal stability hypothesis does not appear to explain the general patterns of composition", apparently contradicting a previous working hypothesis from our laboratory. We have analyzed the data of Belle et al. and find that their data not only do not contradict the thermal stability hypothesis, but if anything support it.

Keywords: Body temperature; Base composition

1. Introduction

The gene-dense compartments (the "genome core") of the ancestors of mammals and birds underwent two independent "major shifts". The first explanation (Bernardi and Bernardi, 1986) proposed for these GC increases and for the maintenance of the new patterns was that they were due to natural selection acting on the "genome phenotype", namely on the compositional pattern of the genomes. As a working hypothesis, we suggested that the selective advantages provided by the compositional genome transitions that accompanied the emergence of warm- from cold-blooded vertebrates were the higher thermodynamic stabilities of DNA, of RNA, and of the proteins encoded by the newly formed GC-rich coding sequences, all these advantages being achieved simultaneously (Bernardi and Bernardi, 1986; see Bernardi, 2003, for a general review).

Three main reasons for this thermodynamic stability hypothesis were that (i) vertebrates are a very small taxon sharing most genetic and genomic properties, i.e. the vast majority of inputs influencing genome composition are, in all likelihood, very similar; (ii) a major difference between cold- and warm-blooded vertebrates is body temperature; (iii) the "major shifts" were never observed in fishes and amphibians, involved both coding and non-coding sequences, and only concerned the gene-richest part of these genomes (which represents 10–15% of the human genome).

This last point has been recently clarified by Saccone et al. (2002), who showed that the gene-richest and GC-richest isochores of warm-blooded vertebrates (the "genome core") consistently have an open chromatin structure in the interphase nucleus. Since there is every reason to believe that in cold-blooded vertebrates the gene-dense regions of the ancestral genome core are also characterized by an open chromatin structure (see Bernardi, 2003), and are likewise centrally located in the nucleus, whereas the gene-poor regions of the empty quarter are also packed at the periphery, then a plausible explanation for the differential compositional transition of the ancestral genome core is that, as body temperature increased with the appearance of homeothermy, the DNA of the open chromatin of the genome core needed to be stabilized by an increasing GC level. In contrast, this was not needed by the gene-poor regions because their stabilization was assured by its dense chromatin structure. This stabilization also was not necessary in cold-blooded vertebrates, because of their lower body temperature and higher DNA methylation (Jabbari et al., 1997), a well known factor for stabilizing DNA.

Based on an analysis of mean and standard deviation of GC_3 values (GC level of third codon positions) of vertebrate genes, and using the method of orthogonal contrasts (see

Abbreviations: GC, molar fraction of guanine and cytosine; GC_3, GC in third codon positions of protein-coding genes; T_{max}, T_{mean}, maximum, mean body temperatures.

* Corresponding author. Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, Villa Comunale, I-80121 Naples, Italy. Tel.: +39-81-5833300; fax: +39-081-2455807.
E-mail address: bernardi@szn.it (G. Bernardi).

0378-1119/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0378-1119(03)00699-1

Felsenstein, 1985), Belle et al. (2002) concluded that "the thermal stability hypothesis does not appear to explain general patterns of composition". Indeed, the absence of a correlation was reported between either the mean GC_3, or the standard deviation in GC_3 (two ways of evaluating the spreading of the GC_3 histogram towards higher values), and body temperature of 18 cold-blooded and 3 warm-blooded vertebrates. Another recent paper, which again appears to contradict a relation between temperature and GC, is discussed elsewhere (Clay et al., 2003).

2. Results and discussion

The assertion of Belle et al. hinges on *p* values that would not be significant, especially if warm-blooded vertebrates are excluded. In the opinion of the authors, such *p* values permit them to reject the temperature hypothesis. We offer here three comments on this reasoning.

The first is that the method of independent contrasts, which the authors used to calculate the *p* value, emphasizes (i.e., rewards) changes, but neglects constraints or selection pressures maintaining a status quo, which it mistakenly interprets as phylogenetic inertia (for critiques of the method's general applicability, see Westoby et al., 1995, and Martins, 2000). In other words, no points are given for maintaining similar GC_3 distributions at similar temperatures, e.g., by combatting tendencies that might homogenize GC_3 (e.g., a genome-wide mutational bias) or cause GC-rich DNA to drift. Points are given only for changing the GC_3 distributions when temperature changes. Furthermore, when one calculates *R* and *p* values via the independent contrast method, one implicitly assumes a Brownian motion model, i.e., a molecular clock for GC (or for its frequency distribution in third codon positions). GC does not satisfy this condition, since its changes do not correlate well with time, even among fish taxa (Bernardi and Bernardi, 1990a,b). In other words, GC evolution in vertebrates is apparently not characterized by gradual diffusion, but rather by rare, relatively rapid shifts. The hypothesis tested and rejected by the method used by Belle et al. (if the temperature and standard deviation estimates are precise; see, however, below) is that essentially any temperature change, either upward or downward, either large or small, should be expected to yield a linearly corresponding change in the GC_3 distribution's standard deviation (regardless of other environmental or physiological conditions). This very special hypothesis does not seem close to any temperature hypothesis that we proposed in the past.

The second comment is that the error or variability of the temperature estimates was not quantified by the authors, and is likely to be large. Various Internet sources are given as the sources of the temperatures. It is well known (Precht et al., 1955; Elliott, 1994) that "body temperature" of a vertebrate species differs widely according to the method, conditions or criteria (optimal, mean, maximum, minimum, growth or lethal temperature) used to measure it. Neglecting the large shifts between mammals or birds and cold-blooded vertebrates, some of the temperature contrasts among species are likely to be well within the measurement or error range for a single species. Experimental noise of this kind will obviously increase scatter and lead to underestimates of correlation coefficients. Inaccurate estimates of GC_3 distribution, e.g., where few genes were available, will contribute further scatter.

The third comment is that one would not expect a larger, indiscriminate sampling of species to increase the statistical significance of contrast correlations. Indeed, only few large compositional shifts are known to have occurred during vertebrate evolution. Instead, one can invoke the functional

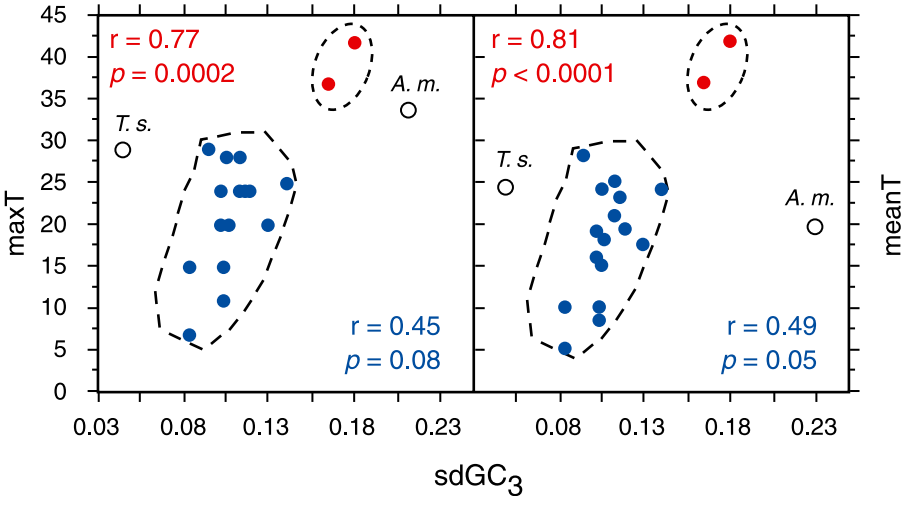

Fig. 1. The maximal and mean body temperatures of the species studied by Belle et al. (2002) are plotted against the standard deviations of GC_3 of genes from cold- (blue points) and warm-blooded vertebrates (human and chicken; red points).Data are from Belle et al. (2002). The standard deviations of GC_3 for human, chicken and *Xenopus* coincide with our previous data (see Bernardi, 2003). Correlation coefficients and *p* values (in red figures for all points, in blue figures for cold-blooded vertebrate points only) are shown neglecting the two outliers (*Alligator mississipiens*, *A.m.*, and *Trachemys scripta*, *T.s.*).

or structural properties, correlated with rising body temperature, rising GC and rising GC_3, that accompanied the evolution of two well-studied, independent lineages, those leading to eutherian mammals and to birds (see Section 1, and Bernardi, 2003). The concordance among these different properties in the two lineages gives strong support to an explanation based on temperature.

Surprisingly, in view of the author's conclusion, the maximal (and mean) body temperatures of the vertebrates investigated by Belle et al. suggest a general correlation with the standard deviation if data from *Alligator* and *Trachemys* (a turtle), two outliers, are neglected (Fig. 1). This elimination is justified by the fact that these species had not only a very low number of genes (17 and 16, respectively), but also extreme (highest and lowest, respectively) standard deviations, inspite of no striking difference in the distribution of DNA molecules (Hughes et al., 2002). The only plausible explanation for the anomalous standard deviations must, therefore, be associated with the particular gene samples used.

If only the data of Belle et al. for cold-blooded vertebrates are considered (still neglecting the two outliers), the standard (not contrast) correlation coefficients are $r=0.45$, $p=0.08$ for T_{max} and $r=0.49$, $p=0.05$ for T_{min} (Fig. 1). Obviously, these values greatly improve if warm-blooded vertebrates are included in the analysis.

We conclude, therefore, that the formation of GC-rich isochores in cold-blooded vertebrates be correlated with body temperature. While not being the final proof for the thermodynamic stability hypothesis, the results of Fig. 1 certainly are in favour of it. Other arguments (Bernardi, 2003) also point in the same direction.

References

Belle, E.M., Smith, N., Eyre-Walker, A., 2002. Analysis of the phylogenetic distribution of isochores in vertebrates and a test of the thermal stability hypothesis. J. Mol. Evol. 55, 356–363.

Bernardi, G., 2003. Structural and evolutionary genomics. Natural Selection in Genome Evolution. Elsevier, Amsterdam.

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

Bernardi, G., Bernardi, G., 1990a. Compositional patterns in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. 31, 265–281.

Bernardi, G., Bernardi, G., 1990b. Compositional transitions in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. 31, 282–293.

Clay, O., Arhondakis, S., D'Onofrio, G., Bernardi, G., 2003. LDH-A and α-actin as tools to assess the effects of temperature on the vertebrate genome: some problems. Gene 317, 157–160 (this issue).

Elliott, J.M., 1994. Quantitative Ecology and the Brown Trout. Oxford Univ. Press, Oxford.

Felsenstein, J., 1985. Phylogenies and the comparative method. Am. Nat. 125, 1–15.

Jabbari, K., Cacciò, S., Pais de Barros, J.P., Desgres, J., Bernardi, G., 1997. Evolutionary changes in CpG and methylation levels in the genome of vertebrates. Gene 205, 109–118.

Hughes, S., Clay, O., Bernardi, G., 2002. Compositional patterns in reptilian genomes. Gene 295, 323–329.

Martins, E.P., 2000. Adaptation and the comparative method. Trends Ecol. Evol. 15, 296–299.

Precht, H., Christopherson, J., Hensel, H., 1955. Temperatur und Leben. Springer-Verlag, Berlin.

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.

Westoby, M., Leishman, M.R., Lord, J.M., 1995. On misinterpreting the "phylogenetic correction". J. Ecol. 83, 531–534.