Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 317 (2003) 157–160
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

LDH-A and α-actin as tools to assess the effects of temperature on the vertebrate genome: some problems

Oliver Clay, Stilianos Arhondakis, Giuseppe D'Onofrio, Giorgio Bernardi*

Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121, Naples, Italy

Received 15 March 2003; received in revised form 10 May 2003; accepted 12 May 2003

Abstract

In a recent paper written with the purpose of shedding light on the question of whether genomic GC levels are related to temperature in vertebrates, Ream et al. [Mol. Biol. Evol. 20 (2003) 105] offered an analysis of two sets of homologous genes: those coding for α-actin and lactate dehydrogenase-A (LDH-A). The conclusion was that "there is no consistent relationship between adaptation temperature and the percentage of thermal stability-enhancing G + C base pairs in protein-coding genes". We argue here that the data presented neither prove nor suggest such a conclusion because of conceptual and methodological errors.

Keywords: Temperature adaptation; Base composition; Isochores

1. Introduction

Almost 30 years ago, we reported (Thiery et al., 1976) a striking difference between the genomes of cold- and warm-blooded vertebrates. The former DNAs were found to be characterized by low degrees of asymmetry and heterogeneity of CsCl profiles. In contrast, the latter displayed remarkable compositional asymmetries and heterogeneities, essentially due to the presence of GC-rich components, which were absent or scarce in the former. A series of subsequent investigations (summarized by Bernardi, in press) showed that vertebrate genomes exhibit an essentially bimodal gene distribution. Gene density is low in the large GC-poor majority of the genome, whereas it is high in the small GC-rich minority of the genome. The latter, the "genome core", is very GC-rich in warm-blooded vertebrates, whereas the corresponding "ancestral genome core" of cold-blooded vertebrates is only modestly GC-rich. Fig. 1 indicates that a compositional transition led to the formation of the GC-rich genome core at the emergence of warm-blooded vertebrates.

This transition can be observed not only at the DNA level but also at the coding sequence level, as shown by the plot of Fig. 2, which displays the correlation between GC_3 levels of orthologous genes of *Xenopus* and human, and shows that while GC-poor genes from both *Xenopus* and human are close to the diagonal (because they are similar in GC_3 levels), human GC_3-rich genes exhibit increasing higher deviations from the diagonal. Another way to display the results is to construct histograms of GC-poor ($GC_3 < 60\%$) and GC-rich ($GC_3 > 60\%$) human codons and to compare their average values with those from orthologous genes of *Xenopus* (Cruveiller et al., 2000). This way of presenting data stresses the lack of differences between average codons from GC_3-poor genes and contrasts it with the large differences found for GC_3-rich genes. It should be pointed out that these histograms show no difference for either GC-poor or GC-rich genes in the case of human and calf (two species belonging to mammalian orders that have evolved separately for about 100 million years), or in the case of human and chicken (whose common ancestor goes back in time to about 300 millions years ago). Very recent results (Romero et al., in press) have also shown no difference between *Cyprinids* and *Xenopus* (the common ancestor being at about 450 million years ago).

We have previously suggested (Bernardi and Bernardi, 1986) that the compositional changes discussed so far were due to natural selection, the advantage being related to an

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

* Corresponding author. Tel.: +39-81-5833300; fax: +39-81-2455807.
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)00698-X

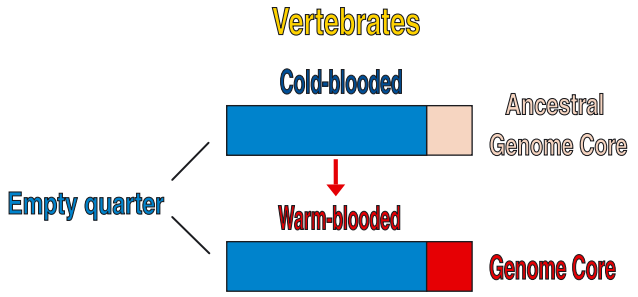

Fig. 1. Scheme of the compositional transition from cold- to warm-blooded vertebrates. The "empty quarter" of their genomes is GC-poor and gene poor (blue box) and essentially did not undergo any compositional change. The gene-dense, moderately GC-rich "ancestral genome core" (pink box) underwent a compositional change into a gene-dense, GC-rich "genome core" (red box).

increased thermal stability of DNA, which 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 empty quarter) is in a packed chromatin configuration (Saccone et al., 2002). While the latter DNA was stabilized by its dense chromatin structure, the former needed to be stabilized by a GC increase at the higher body temperature of warm-blooded vertebrates. Such a need did not exist in cold-blooded vertebrates, e.g. because of their lower body temperature and higher DNA methylation level (Jabbari et al., 1997).

2. Discussion

Ream et al. (2003) analyzed two sets of homologous genes, those coding for α-actin and lactate dehydrogenase-A (LDH-A), in 51 vertebrates in order to detect any possible correlation between the base composition of these genes and the body temperature of the source vertebrates. These authors concluded that "there is no consistent relationship between adaptation temperature and the percentage of thermal stability-enhancing G+C base pairs in protein-coding genes".

Unfortunately, the approach of Ream et al. has several problems or errors that render it invalid or unreliable. We will discuss first the methodological problems (which concern paralogy, the contrast method and temperature assessment), since they may reappear in the literature, and then an apparent conceptual misunderstanding.

2.1. Paralogy

Interspecific comparisons of widely sequenced protein-coding genes from a single gene family might, in principle, allow a glimpse into the evolution of genic GC levels. Indeed, the alignment of corresponding coding sequences from different species might, in principle, allow one to infer, at the nucleotide level, many of the substitutions that were responsible for the compositional changes, and to investigate their possible causes.

For such a programme, there is a necessary condition: orthologous genes, rather than simply homologous (i.e., possibly paralogous) genes, must exist and be easily recognizable in the selected gene families, for all taxa that are to be included in the study. Unless this condition is met, one will risk comparing GC levels of non-corresponding genes, and have the problem of paralog noise. Unfortunately, the LDH and especially the actin genes analyzed exemplify this problem. For example, LDH-B (a paralog of LDH-A that has 52.2% GC_3) was surprisingly chosen for *Xenopus*, instead of the available LDH-A sequence (which has 70.6% GC_3; see Mannen and Li, 1999). The α-actin gene set used by the authors is even more problematic; several of the species represented by only one or two sequences in the authors' list are in fact also represented by other, distinct coding sequences that cluster with the α-actins at the amino acid level and are labelled as such (see, e.g., HOVERGEN/Query, http://pbil.univ-lyon1.fr; Duret et al., 1994). Although these paralogs have sometimes very different GC_3 levels, they are not mentioned or included by the authors.

In the case of α-actins, paralogies are so frequent that it is difficult to obtain reliable trees, and not even the pair shown in Fig. 2 is certain to represent true orthologs. The actin gene family has long been known for its remarkable

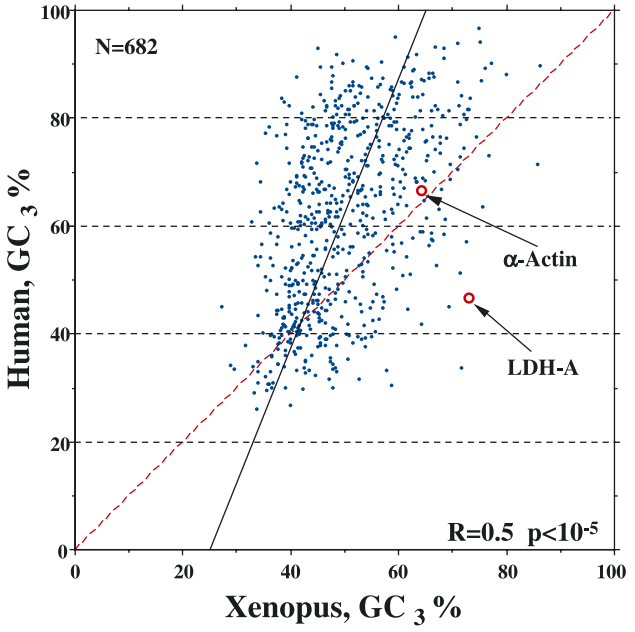

Fig. 2. Scatterplot of the GC_3 levels of putative orthologous genes from *Xenopus* and human. Each point represents a gene "family" as defined by HOVERGEN (Duret et al., 1994), and shows the GC_3 levels of the sequenced human and mouse genes in that gene family that are most similar (as assessed at the amino acid level). In many, but not all, cases, the genes will be orthologs; although the criterion extracts only potential orthologs, it may be the strictest automatic criterion that can be applied in the absence of complete sequence information for all species. The sequence pairs for LDH-A and α-actin are indicated by red circles.

chromosomal dispersion and compositional diversity within a single genome, which is also reflected in the large differences observed in third codon positions in spite of sometimes very high amino sequence conservation (Dodemont et al., 1982; Soriano et al., 1982; Bernardi et al., 1985; Britten, 1993). The α-actin subfamily is no exception. For example, two human α-actin members differ by only 4/377 amino acids, yet one has 89.2% GC_3 (HSSAACT/HSACTASK) and the other has 66.4% GC_3 (BC009978/HSACTCA4); these sequences were not included by Ream et al. (2003). Similarly, paralogs (not mentioned by these authors) differing by about 20% in GC_3 have been sequenced in cow and pig. In view of the many α-actin sequences that they omitted, including a *Xenopus* paralog that has 64.0% GC_3, it cannot be said that their analyses of the α-actin genes "consistently showed" that "*Xenopus* spp. tend to be outliers that differ from other ectothermic species". It is this gene (or, rather, this multigene family) that is an outlier, not *Xenopus*. That *Xenopus* is not an outlier was independently shown by Romero et al. (in press), who found an identical average codon usage for orthologous genes from *Xenopus* and *Cyprinids*. Similarly, in the LDH-A comparisons, *Xenopus* would not have been an outlier if LDH-A, rather than LDH-B, had been used to represent this species.

Where recent duplication events have apparently caused a compositional bifurcation into two paralogous genes (as in *Coryphaenoides*, where the reported α-actin GC_3 levels differ by >11% between paralogs but only by <0.4% between orthologs), it is difficult to know if the GC_3 of one of them, or their mean (as Ream et al. assumed), can be used to represent the species in an interspecific comparison.

The point is that the variations in genic GC levels, third positions and third positions of quartet codons presented by Ream et al. do not correspond to a simple monitoring of two genes during vertebrate evolution. One cannot identify which of the multiple α-actin genes in one species corresponds to which of the multiple α-actin genes in another; as explained above, α-actin trees based on protein alignments alone are unreliable, being based on very few amino acid changes (some of which, such as Glu/Asp in the 5th amino acid, might even represent single nucleotide polymorphisms, SNPs in human or *Xenopus*). It is, therefore, not surprising that the "phylogeny based on the first two positions of the α-actin gene was poorly supported for the 13 different vertebrate orders represented" (Ream et al., 2003). By contrast, genes exhibiting low paralog noise tend to yield well-supported and consistent phylogenies, at least where species span all placental orders (Madsen et al., 2001; Murphy et al., 2001).

2.2. The contrast method

Where phylogenies are reliable, the independent contrast method can be used together with compositional data to suggest when abrupt GC shifts occurred during the evolution of certain lineages. The "expected" variation in Felsenstein's (1985) method, which is used by the authors, is a Brownian drift, or random walk. The observed variation in vertebrates turns out, instead, to be more abrupt, and does not correlate with time (Bernardi and Bernardi, 1990a,b; Bernardi, 2000a,b), suggesting that the Brownian assumption is far from satisfied. This means, for example, that one cannot simply calculate the expected value at an internal node by taking the mean of the values calculated from the two branches that lead to it. Furthermore, the correlations that the authors obtain would, if the method could be used, indicate to what extent small or large changes in body temperature (in either direction) are accompanied by small or large changes in GC, respectively. This is a rather special hypothesis. It differs, for example, from a less demanding hypothesis that only relatively large increases in temperature (not decreases or marginal increases in mean temperature) are correlated with observable shifts in GC.

Some minor noise was apparently introduced in the data of Ream et al. by typing errors, such as the GC_{total} and GC_3 levels for *Sphyraena lucasana* LDH-A which are slightly higher (59.5% and 84.7%; accession number U80002), not lower, than the levels for the colder-blooded (−5.5°) species *Sphyraena argentea* of the same genus. Anecdotal GC increases that accompany temperature rises could be of particular interest when the taxa are closely related, since other factors affecting GC should be similar.

2.3. Temperature

Temperature reports for individual vertebrates persistently show a large scatter. Systematic series of precise measurements using state-of-the-art techniques (see, e.g., Klaus et al., 1998), obtained under similar conditions, might provide a way out of this impasse if accompanied by a curated database of reported environmental or body temperatures. Unfortunately, the currently available data are often not directly comparable, and some of the contrasts presented by Ream et al. would need a high precision. Furthermore, the temperature table of Ream et al. contains some values that are very far from those given in other publications, and are likely to be incorrect. For example, the sources for their table report of 20° as the mean body temperature for brown trout (*Salmo trutta*), yet a detailed study (Elliott, 1994) reports an optimal growth temperature range around 7–17° for the former, and even lower temperatures for the egg stage. Such differences among reported values would correspond to intermediate temperature contrasts in the data of Ream et al.

Another large discrepancy between body temperatures was found for the T_{max} of *Carassius auratus* (goldfish), which was reported to be 20° by Belle et al. (2002) and 33° by Ream et al. (2003). Both T_{max} values are correct, but the former refers to the temperature of goldfish from its native regions (Asia and Eastern Europe), the latter to that of a non-indigenous region (North America), in which goldfish was established in the XVII century (Page and Burr, 1991). The temperature of 33° should be disregarded, since the

very short adaptation time could not have affected the GC level of genes in the goldfish genome.

The paper by Ream et al. is affected not only by the methodological problems just discussed but also by an apparent conceptual misunderstanding. Indeed, the points made in the Introduction indicate that looking for a correlation between GC_3 increase and body temperature requires (i) investigating genes belonging to the genome compartment that underwent the compositional change, since no change occurred in about 50% of the genes, the genes from the GC-poor "empty quarter"; and (ii) not being misled by the few rare outlier genes that exhibited anomalous compositional behaviour at the transition between cold- and warm-blooded vertebrates.

Most unfortunately, of the two genes chosen by Ream et al., LDH-A is an outlier, and the putative α-actin orthologs have a GC_3 value falling on the diagonal of the plot (see Fig. 2). In other words, α-actin has the same composition in *Xenopus* and human. Neither of the two genes are, therefore, acceptable candidates for the purpose of checking the GC_3 vs. temperature relationship. A better choice would have been two genes located on or near the major axis of the correlation of Fig. 2 and showing a high GC_3 value in mammals or birds. However, the inherent difficulty of ensuring functional orthology for any single gene, over the huge taxonomic range analyzed by the authors, suggests that their approach will not yield reliable results until many orthologous genes have been sequenced in diverse vertebrate taxa characterized by different body temperatures.

We completely agree (although for a different reason) with the authors' title statement, namely that coding sequences for α-actin and LDH-A "from differently adapted vertebrates show no temperature-adaptive variation in G+C content." We do not agree, however, that their study of two such genes "indicates that... there is no consistent relationship between adaptation temperature and the percentage of thermal stability-enhancing G+C base pairs in protein-coding genes." For the reasons that we have discussed above, the two genes analyzed by Ream et al. do not show, indicate or suggest any such general conclusion regarding GC and temperature. In fact, the opposite seems to be true (see Bernardi, 2003; Jabbari et al., 2003).

Acknowledgements

We wish to thank Giuseppe Torelli for his help with computer analysis.

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., 2000a. Isochores and the evolutionary genomics of vertebrates. Gene 241, 3–17.

Bernardi, G., 2000b. The compositional evolution of vertebrate genomes. Gene 259, 31–43.

Bernardi, G., 2003. Structural and Evolutionary Genomics. Natural Selection in Genome Evolution. Elsevier, Amsterdam. In press.

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

Bernardi, G., Bernardi, G., 1990a. Compositional patterns in the nuclear genome 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.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., Rodier, F., 1985. The mosaic genome of warm-blooded vertebrates. Science 228, 953–958.

Britten, R., 1993. Forbidden synonymous substitutions in coding regions. Mol. Biol. Evol. 10, 205–220.

Cruveiller, S., D'Onofrio, G., Bernardi, G., 2000. The compositional transition between the genomes of cold- and warm-blooded vertebrates: codon frequencies in orthologous genes. Gene 261, 71–83.

Dodemont, H.-J., Soriano, P., Quax, W.J., Ramaekers, F., Lenstra, J.A., Groenen, M.A.M., Bernardi, G., Bloemendal, H., 1982. The genes coding for the cytoskeletal proteins actin and vimentin in warm-blooded vertebrates. EMBO J. 1, 161–171.

Duret, L., Mouchiroud, D., Gouy, M., 1994. HOVERGEN: a database of homologous vertebrate genes. Nucleic Acids Res. 22, 2360–2365.

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, 1–12.

Jabbari, K., Cacciò, S., Païs de Barros, J.-P., Desgrès, J., Bernardi, G., 1997. Evolutionary changes CpG and methylation levels in vertebrate genomes. Gene 205, 109–118.

Jabbari, K., Clay, O., Bernardi, G., 2003. GC_3 heterogeneity and body temperature in vertebrates. Gene 317, 161–163 (this issue).

Klaus, S., Münzberg, H., Trüloff, C., Heldmaier, G., 1998. Physiology of transgenic mice with brown fat ablation: obesity is due to lowered body temperature. Am. J. Physiol. 274, R287–R293.

Madsen, O., Scally, M., Douady, C.J., Kao, D.J., Debry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., De Jong, W.W., Springer, M.S., 2001. Parallel adaptive radiation in two major clades of placental mammals. Nature 409, 610–614.

Mannen, H., Li, S.S.-L., 1999. Molecular evidence for a clade of turtles. Mol. Phylogenet. Evol. 13, 144–148.

Murphy, M.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryderk, O.A., O'Brien, S.J., 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614–618.

Page, L.M., Burr, B.M., 1991. A Field Guide to Freshwater Fishes from North America North of Mexico. Peterson Field Guide Series. Houghton Mifflin Co., Boston.

Ream, R.A., Johns, G.C., Somero, G.N., 2003. Base compositions of genes encoding α-actin and lactate dehydrogenase-A from differently adapted vertebrates show no temperature-adaptive variation in G+C content. Mol. Biol. Evol. 20, 105–110.

Romero, H., Zavala, A., Musto, H., Bernardi, G., 2003. The influence of translational selection on codon usage in fishes from the family *Cyprinidae*. Gene (in press).

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.

Soriano, P., Szabo, P., Bernardi, G., 1982. The scattered distribution of actin genes in the mouse and human genomes. EMBO J. 1, 579–583.

Thiery, J.P., Macaya, G., Bernardi, G., 1976. An analysis of eukaryotic genomes by density gradient centrifugation. J. Mol. Biol. 108, 219–235.