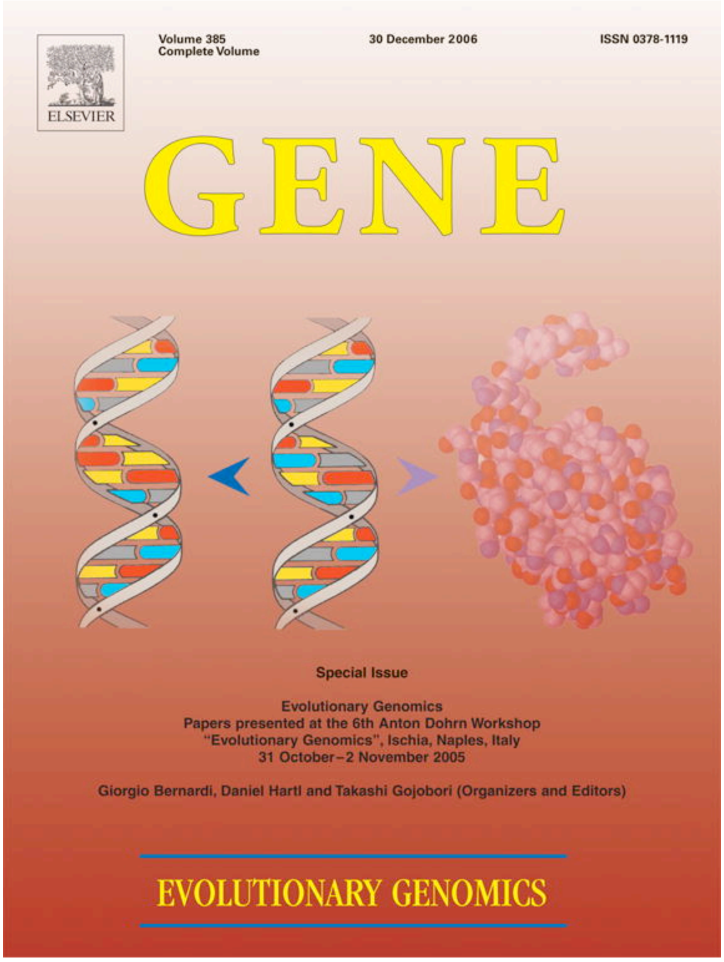
Provided for non-commercial research and educational use only.
Not for reproduction or distribution or commercial use.



Volume 385
Complete Volume
30 December 2006
ISSN 0378-1119

ELSEVIER

GENE

Special Issue

Evolutionary Genomics
Papers presented at the 6th Anton Dohrn Workshop
"Evolutionary Genomics", Ischia, Naples, Italy
31 October–2 November 2005

Giorgio Bernardi, Daniel Hartl and Takashi Gojobori (Organizers and Editors)

EVOLUTIONARY GENOMICS

This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

http://www.elsevier.com/locate/permissionusematerial

Available online at www.sciencedirect.com
ScienceDirect

Gene 385 (2006) 71 – 74

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

A new perspective on isochore evolution

Laurent Duret [a], Adam Eyre-Walker [b,c], Nicolas Galtier [d,*]

[a] CNRS UMR 5558, BBE, Université C. Bernard Lyon 1, France
[b] National Evolutionary Synthesis Center, Durham, NC, USA
[c] Centre for the Study of Evolution, University of Sussex, UK
[d] CNRS UMR 5171, GPIA, CC63, Université Montpellier 2, Place E. Bataillon, 34095 Montpellier, France

Received 9 January 2006; accepted 7 April 2006
Available online 5 August 2006

Abstract

The genomes of mammals and birds show dramatic variation in base composition over large scales, the so called isochore structure of the genome. The origin of isochores is still controversial: various neutral and selectionist models have been proposed – and criticized – since the discovery of isochores in the 1970s. The availability of complete mammalian genomes has yielded new opportunities for addressing this question. In particular, it was recently proposed that (i) the isochore structure is declining in many mammalian groups, and that (ii) GC-content is influenced by local recombination rate, possibly via the mechanism of GC-biased gene conversion. In this article we review the existing support for these two hypotheses, and discuss how they can be combined to provide a new perspective on isochore evolution.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The genomes of mammals and birds show dramatic variation in base composition over large scales, the so called isochore structure of the genome (Bernardi et al., 1985). However, although isochores were discovered some 30 years ago, there is still no consensus as to why the genomes of birds and mammals show such conspicuous compositional heterogeneity (Eyre-Walker and Hurst, 2001). Unfortunately the question may have just got harder to answer, since there is an increasingly good evidence that the isochore structure of the genome is not being actively maintained and that the genome is regressing to a state of relative homogeneity, at least in some groups of mammals. If this is indeed the case, then the question of why isochores exist is difficult to answer since it is a question about a process which occurred several hundred million years ago. Recent findings about the influence of recombination and biased gene conversion on GC-content evolution might help to make progress with this difficult issue.

2. Declining isochores

Prior to the late 1990s it had been generally accepted that the isochore structure was fairly stable in mammals, except in murids (Mouchiroud et al., 1987). Murids show a substantial reduction in compositional heterogeneity, both in genomic DNA and protein coding sequences; this was termed the minor shift. However, the evidence for the stability of the isochore structure in other mammals was based on limited data, and an analysis which would not have detected most compositional changes, even if they existed. The evidence for stability was two-fold. First, different mammals show similar GC-content distributions when their genomic DNA is fractionated, and second, there are strong correlations between the GC contents of protein coding sequences from different mammalian groups (Bernardi, 2000). However, these analyses ignore common ancestry – most mammals are quite closely related to one another – and they ignore the fact that differences will only become apparent between two groups if the process of evolution is different. For example, GC-content could be declining in two species but not be apparent if the rate of decline is similar in the two lineages.

Recently data has become available which allows one to determine the direction of substitutions. The data are of two forms; orthologous gene copies in three or more species and

---

[*] Corresponding author. Tel.: +33 467 14 46 84; fax: +33 467 14 45 54.
E-mail address: galtier@univ-montp2.fr (N. Galtier).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.04.030

repetitive DNA elements from within the same genome. The results are the same from both sources of data — there has been a dramatic decrease in GC-content both in protein coding and non-coding DNA, at least in the some mammalian lineages, including murids, primates and carnivores (IHGSC, 2001; Duret et al., 2002; Smith and Eyre-Walker, 2002; Arndt et al., 2003; Webster et al., 2003; Belle et al., 2004; Arndt et al., 2005). Analysis of ancient repetitive DNA suggests that this decline started ~250 MYR ago, but accelerated ~90 MYR ago, when mammals diversified (Arndt et al., 2003).

Alvarez-Valin et al. (2004) have argued that some of these results could be an artifact of parsimony, which was used to determine the direction of substitution in several of these analyses. However, while problems with parsimony may explain a couple of the results they do not explain them all. First, the analysis of Belle et al. (2004) used the method of Galtier and Gouy (1998), which has been shown to be accurate at inferring ancestral GC contents in equilibrium and non-equilibrium conditions, the task it was designed for. Second, the results of Duret et al. (2002) and Webster et al. (2003), who did use parsimony, are not likely to be biased for the primate datasets they considered, since the divergences between the species considered were low. Alvarez-Valin et al. (2004) argue that the primate results could have been due to other problems, such as alignment errors, CpG effects and paralogy, but they present no evidence that these are actually problems in the data analysed by Duret et al. (2002) and Webster et al. (2003). Moreover, the analysis of a larger set of genomic alignments with careful consideration of CpG effects, again confirmed the erosion of GC-rich isochores in primates (Meunier and Duret, 2004). Antezana (2005) recently suggested that mammalian GC-content was very close to equilibrium. However, we showed that this assertion was incorrect, because of a methodological error in the counting of substitutions (Duret, 2006). Overall, there is clear evidence that the GC-content has declined in at least some mammalian lineages. Webster et al. (2003) estimate that the human genome is likely to homogenise to an average GC-content of around 42%.

However, although the overall pattern seems to be one of decline in mammals, this may not be true of all mammalian groups. Belle et al. (2004) found evidence of a strong decline in GC-content early in mammalian evolution, a decline which appears to have continued in the primate, murid and carnivore lineages. However, there was no evidence of a decline in lagomorphs and little evidence in perissodactyls and cetartiodactyls. It is unclear whether this was due to a lack of data, since relatively few genes were analysed, or due to genuine differences in the way these groups evolve.

3. GC-biased gene conversion

In parallel to these findings about the history of isochores, our knowledge of the evolutionary forces influencing GC-content evolution in mammals has progressed. The two models traditionally competing to explain isochore origin and evolution, namely variable mutation bias versus variable selective optimum, lack any empirical evidence (Eyre-Walker and Hurst, 2001). A third hypothesis, the biased gene conversion model (BGC), was therefore considered. Allelic gene conversion, i.e. the copy/paste of one allele onto the other one at heterozygous loci, occurs during meiotic recombination as a consequence of the repair of mismatched heteroduplexes (Marais, 2003). We proposed that this process could be biased toward GC, so that an AT/GC heterozygote would produce more GC than AT gametes (Eyre-Walker, 1993; Galtier et al., 2001), leading to an advantage of G and C over A and T alleles, and an increase of GC-content in highly recombining regions. Various pieces of evidence supporting this view have been reported during the past 5 years.

There is, in the first place, a significant correlation between recombination rate and GC-content in mammalian genomes, observed at various levels. The non-recombining Y chromosome, for instance, has a significantly lower GC-content than the female-recombining X chromosome, and the X has a lower GC-content than autosomes (Galtier et al., 2001). Among autosomes, short chromosome arms show a higher GC-content than long ones (IHGSC, 2001), which is a prediction of the BGC model: because one obligate crossing-over per meiosis occurs on every chromosome arm, the per kilobase density of recombination events increases as the length of chromosome arm decreases. The relationship is even stronger in birds, in which microchromosomes are by far GC-richer than macrochromosomes (ICGSC, 2004).

At a smaller scale, the recombination/GC relationship appears weaker. Plotting the GC-content against the estimated recombination rate of 1 MB-long fragments in the human genome yielded only a marginally significant correlation (Kong et al., 2002). However, when the predicted equilibrium GC-content (i.e., the GC-content towards which the focal fragment is evolving), estimated thanks to a ((Human, Chimpanzee), Baboon) triplet, was used, a remarkably linear relationship appeared (Meunier and Duret, 2004). This result strongly suggests that recombination drives GC-content, not the reverse, since recombination predicts future base composition more accurately than current base composition. Another line of evidence came with the report of a GC-biased substitution process in Alu repeats located in highly recombining regions of the human genome (Webster et al., 2003, 2005). Eisenbarth et al. (2000) also showed that the NF1 isochore boundary in human corresponds to a break-point in the recombination map, between a highly recombining section which is GC-rich, and a low-recombining section which is GC poor.

The influence of recombination on GC-content, finally, was spectacularly illustrated at the local scale by the peculiar history of the Fxy gene in mouse. The 3′ part of this gene recently moved from the X chromosome, a relatively low-recombination region, to the pseudoautosomal region (PAR) in the house mouse *Mus musculus* (Perry and Ashworth, 1999). The PAR is a ~600 kb-long region of homology between the X and Y chromosomes. Behaving like a very short autosome at meiosis, the PAR is intrinsically highly recombining. Subsequent to this move, the 3′ Fxy evolved a very high GC-content (GC3 increased from 50% to ~85%), a reduced intron size (from 5 to 10 kb to less than 1 kb), an increased difference between GC3 and intron GC-content, and several new minisatellite loci — all of these features are well-known (or newly discovered) characteristics of GC-rich isochores (Montoya-Burgos et al., 2003). The 5′ part of Fxy, also translocated but still X-linked, has not undergone any of these changes, indicating that the increased recombination rate was the single determinant of the novel, GC-rich–isochore-like features of the PAR-linked fragment.

The above arguments stress the link between recombination and GC-content, but do not demonstrate that BGC is the underlying mechanism. This issue has been addressed by analysing loci undergoing ectopic gene conversion, i.e. gene conversion between paralogous copies from the same haploid genome. We first noticed that ribosomal RNAs, ribosomal spacers, transfer RNAs, specific regions of the MHC, and replication-dependent histones, for which frequent ectopic gene conversion between paralogues is notoriously occurring, are all GC-enriched (Galtier et al., 2001). A consistent result was recently reported for the HINTW duplicated gene in birds (Backstrom et al., 2005). A detailed analysis of histone genes in human and mouse revealed that paralogues involved in clusters of nearly-identical copies, presumably evolving through gene conversion, show a significantly higher GC3 than single copies, physically isolated, and distantly related to any other paralogue in the genome (Galtier, 2003). Similarly, only those regions of the mammalian HSP70 genes showing evidence for homogenisation through gene conversion are GC-enriched (Kudla et al., 2004). Finally, experimental evidence for a GC-biased Base-Excision Repair process has been reported in mammals (Bill et al., 1998), providing a plausible mechanistic basis to the BGC model.

4. Isochore origin and evolution

To summarize, we now know that GC-rich isochores are declining in most mammalian orders, and we have identified a probable mechanism controlling GC-content evolution in mammals. The next question is whether these two discoveries, taken together, can help in providing a reasonable scenario of isochore evolution, accounting for the various sources of data available.

A number of problems have to be addressed. It should first be noticed that the two findings we report could appear contradictory: we show that GC-rich isochores are declining, but exhibit a process that tends to increase GC-content. The decline of GC-rich isochores implies that the evolutionary forces that contributed to their creation are weaker now than they used to be in the past, or have been compensated by new, opposing forces. We also need to understand how the local recombination rate could control the (relatively stable) isochores given the rapid evolutionary rate of recombination maps. We should, finally, explain why only amniotes have “true” isochores (whatever this means, Cohen et al., 2005), although BGC appears to occur in many different organisms (Birdsell, 2002).

These apparent paradoxes can perhaps be solved by considering the time/space dynamics of GC-content in mammals, and the asymmetric nature of the underlying evolutionary forces. Consider a model in which local GC-content in the mammalian genome reflects a balance between an AT-biased mutation process and BGC, a GC-biased fixation process. The major part of the genome is not affected by BGC, and is at the mutational equilibrium, close to 35–40% GC. At specific time periods and genomic locations, when/where the recombination rate is high enough for BGC to be effective, GC-content increases rapidly, as a consequence of the “advantage” of G and C over A and T alleles, which makes their fixation probability higher than the neutral one. Now if BGC stops, the local GC-content will decrease slowly, approaching the new, AT-rich equilibrium at the relatively low mutation rate. This asymmetric process (quick GC-increase but slow GC-decrease, Galtier, 2004) could explain why the global pattern is a decrease of GC-content in GC-rich isochores, although BGC is still active in the most highly recombining regions of the genome.

Under this model, the present GC-content of a given genomic region would be basically determined by the integral of its past recombination rate (times the effective population size) over the last hundreds of million years. A region having hosted many recombination hotspots during the last 100 My should be GC-rich, whereas a consistently (in time) low-recombining region should be GC-poor. A recent genome-wide analysis of linkage disequilibrium patterns in humans reveals that although the location of recombination hotspots evolves quickly, the large-scale recombination rates appear quite stable in time (Myers et al., 2005). This finding is consistent with the fact that the strongest GC/recombination correlations are observed at large scales. The two-level recombination map dynamics, i.e. a rapid birth/death of hotspots within a framework of stable hotspot-containing regions (Myers et al., 2005), might also explain why GC-rich isochores are less homogeneous than the GC-poor ones (Cohen et al., 2005; Clay and Bernardi, 2005).

How can we explain that the genomes of amniotes show a strong heterogeneity in GC-content, with GC-poor and GC-rich isochores, whereas other vertebrate genomes are much more homogeneous in base composition ? According to the BGC model, the evolution of GC-content depends on the rate of recombination, on the effective population size and on the strength of the GC-bias in the repair of DNA mismatches. Any one of these factors might vary along the genome. In mammals and in birds, there is a strong heterogeneity in the rate of recombination within the genome. Notably recombination rate is strongly negatively correlated to the size of chromosome arms (Meunier and Duret, 2004; ICGSC, 2004). We therefore proposed the hypothesis that GC-rich isochores in mammals and birds might derive from microchromosomes in the karyotype of the ancestor of amniotes (Duret et al., 2002). We presently do not have data on genome-wide variation in recombination rate in non-amniote vertebrates. However, it seems *a priori* unlikely that the rate of recombination would have been generally more heterogenous in the genome of amniotes than in other vertebrates. It also seems unlikely that systematic differences in effective population sizes could account for differences in isochore organization between the genome of amniotes and other vertebrates. Hence, we have to suppose that GC-rich isochores emerged in the genome of the cenancestor of amniotes as a consequence of an increase of the GC-bias in the repair of DNA mismatches. It has been proposed that this GC-bias in mismatch repair reflects an adaptation to the hypermutability of cytosines in methylated genomes (Brown and Jiricny, 1987). Interestingly, the CpG deficiency is stronger in amniotes than in amphibia or teleost fishes (Belle et al., 2004), suggesting a higher level of CpG methylation. Thus, we hypothesize that the GC-bias in mismatch repair increased in the ancestor of amniotes, as a consequence of an increase in the level of CpG methylation. This increase would have led to an increase in GC-content in regions of high recombination.

Presently, this scenario remains speculative. However, thanks to the progress in the sequencing of the genomes from different vertebrates, it should soon be possible to reconstruct the karyotype of the last common ancestor of amniotes. It will therefore be possible to test our hypothesis that GC-rich isochores in mammalian genomes derive from microchromosomes in the genome of the amniote cenancestor. It also would be interesting to investigate the relationship between recombination rate and substitution patterns in non-amniote vertebrates, to try to understand the parameters that affect the evolution of GC-content in those taxa. Finally, thanks to the accumulation of polymorphism data, it will also be possible to investigate the relationship between the rate of recombination and the strength of the fixation bias in favor of GC-alleles.

References

Alvarez-Valin, F., Clay, O., Cruveiller, S., Bernardi, G., 2004. Inaccurate reconstruction of ancestral GC levels creates a “vanishing isochores” effect. Mol. Phylogenet. Evol. 31, 788–793.

Antezana, M., 2005. Mammalian GC content is very close to mutational equilibrium. J. Mol. Evol. 61, 834–836.

Arndt, P.F., Petrov, D.A., Hwa, T., 2003. Distinct changes of genomic biases in nucleotide substitution at the time of mammalian radiation. Mol. Biol. Evol. 20, 1887–1896.

Arndt, P.F., Hwa, T., Petrov, D.A., 2005. Substantial regional variation in substitution rates in the human genome: importance of GC content, gene density, and telomere-specific effects. J. Mol. Evol. 60, 748–763.

Backstrom, N., Ceplitis, H., Berlin, S., Ellegren, H., 2005. Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. Mol. Biol. Evol. 22, 1992–1999.

Belle, E.M., Duret, L., Galtier, N., Eyre-Walker, A., 2004. The decline of isochores in mammals: an assessment of the GC content variation along the mammalian phylogeny. J. Mol. Evol. 58, 653–660.

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

Bernardi, G., et al., 1985. The mosaic genome of warm-blooded vertebrates. Science 228, 953–958.

Bill, C.A., Duran, W.A., Miselis, N.R., Nickoloff, J.A., 1998. Efficient repair of all types of single-base mismatches in recombination intermediates in Chinese hamster ovary cells. Competition between long-patch and G–T glycosylase-mediated repair of G–T mismatches. Genetics 149, 1935–1943.

Birdsell, J.A., 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Mol. Biol. Evol. 19, 1181–1197.

Brown, T.C., Jiricny, J., 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. Cell 50, 945–950.

Clay, O., Bernardi, G., 2005. How not to search for isochores: a reply to Cohen et al. Mol. Biol. Evol. 22, 2315–2317.

Cohen, N., Dagan, T., Stone, L., Graur, D., 2005. GC composition of the human genome: in search of isochores. Mol. Biol. Evol. 22, 1260–1272.

Duret, L., 2006. The GC-content of primates and rodents genomes is not at equilibrium: a reply to Antezana. J. Mol. Evol. 62, 803–806.

Duret, L., Semon, M., Piganeau, G., Mouchiroud, D., Galtier, N., 2002. Vanishing GC-rich isochores in mammalian genomes. Genetics 162, 1837–1847.

Eisenbarth, I., Vogel, G., Krone, W., Vogel, W., Assum, G., 2000. An isochore transition in the NF1 gene region coincides with a switch in the extent of linkage disequilibrium. Am. J. Hum. Genet. 67, 873–880.

Eyre-Walker, A., 1993. Recombination and mammalian genome evolution. Proc. Biol. Sci. 252, 237–243.

Eyre-Walker, A., Hurst, L.D., 2001. The evolution of isochores. Nat. Rev. Genet. 2, 549–555.

Galtier, N., 2003. Gene conversion drives GC-content evolution in mammalian histones. Trends Genet. 19, 65–68.

Galtier, N., 2004. Recombination, GC-content, and the human pseudoautosomal boundary paradox. Trends Genet. 20, 347–349.

Galtier, N., Gouy, M., 1998. Inferring pattern and process : maximum likelihood implementation of a non-homogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15, 871–879.

Galtier, N., Piganeau, G., Mouchiroud, D., Duret, L., 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159, 907–911.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716.

Kong, X., et al., 2002. A high-resolution recombination map of the human genome. Nat. Genet. 31, 241–247.

Kudla, G., Helwak, A., Lipinski, L., 2004. Gene conversion and GC-content evolution in mammalian Hsp70. Mol. Biol. Evol. 21, 1438–1444.

Marais, G., 2003. Biased gene conversion: implications for genome and sex evolution. Trends Genet. 19, 330–338.

Meunier, J., Duret, L., 2004. Recombination drives the evolution of GC-content in the human genome. Mol. Biol. Evol. 21, 984–990.

Montoya-Burgos, J.I., Boursot, P., Galtier, N., 2003. Recombination explains isochores in mammalian genomes. Trends Genet. 19, 128–130.

Mouchiroud, D., Fichant, G., Bernardi, G., 1987. Compositional compartmentalization and gene composition in the genome of vertebrates. J. Mol. Evol. 26, 198–204.

Myers, S., Bottolo, L., Freeman, C., McVean, G., Donnelly, P., 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science 310, 321–324.

Perry, J., Ashworth, A., 1999. Evolutionary rate of a gene affected by chromosomal position. Curr. Biol. 9, 987–989.

Smith, N.G., Eyre-Walker, A., 2002. The compositional evolution of the murid genome. J. Mol. Evol. 55, 197–201.

Webster, M.T., Smith, N.G., Ellegren, H., 2003. Compositional evolution of noncoding DNA in the human and chimpanzee genomes. Mol. Biol. Evol. 20, 278–286.

Webster, M.T., Smith, N.G., Hultin-Rosenberg, L., Arndt, P.F., Ellegren, H., 2005. Male-driven biased gene conversion governs the evolution of base composition in human alu repeats. Mol. Biol. Evol. 22, 1468–1474.