JOURNAL OF **MOLECULAR EVOLUTION**

The Decline of Isochores in Mammals: An Assessment of the GC Content Variation Along the Mammalian Phylogeny

Elise M.S. Belle,[1] Laurent Duret,[2] Nicolas Galtier,[3] Adam Eyre-Walker[1]

[1] Centre for the Study of Evolution—School of Life Sciences, University of Sussex, Brighton BN1 9QG, UK
[2] Laboratoire de Biométrie et Biologie évolutive, UMR CNRS 5558, Université Claude Bernard Lyon 1, 16 rue Raphael Dubois, 69622, Villeurbanne Cedex, France
[3] Laboratoire Génome, Populations, Interactions, Adaptations, CNRS Université Montpellier 2, IFREMER UMR 5000, CC63 Place Eugene Bataillon, 34095 Montpellier Cedex, France

Abstract. Whether isochores, the large-scale variation of the GC content in mammalian genomes, are being maintained has recently been questioned. It has been suggested that GC-rich isochores originated in the ancestral amniote genome but that whatever force gave rise to them is no longer effective and that isochores are now disappearing from mammalian genomes. Here we investigated the evolution of the GC content of 41 coding genes in 6 to 66 species of mammals by estimating the ancestral GC content using a method which allows for different rates of substitution between sites. We found a highly significant decrease in the GC content during early mammalian evolution, as well as a weaker but still significant decrease in the GC content of GC-rich genes later in at least three groups of mammals: primates, rodents, and carnivores. These results are of interest because they confirm the recently suggested disappearance of GC-rich isochores in some mammalian genomes, and more importantly, they suggest that this disappearance started very early in mammalian evolution.

Key words: Isochore — Genome — GC content — Noncoding DNA — Mammal — Primate

Introduction

The mammalian genome shows an unusual patterning of nucleotide composition: there are some large regions, over hundreds of kilobases, of relatively homogeneous base composition, some regions of the genome having a GC content as low as 35%, while others are at 60%; these regions have been called isochores (Bernardi et al. 1985). Although we now know that they are not as homogeneous as previously thought (IHGSC 2001), we will still use the term *isochores* to define these large regions of relatively homogeneous base composition (Eyre-Walker and Hurst 2001).

It has been shown that both mammal and bird genomes have an isochore structure, with some regions of very high GC content (Kadi et al. 1993; Bernardi 2000). Since both mammals and birds are warm-blooded, Bernardi (1993) proposed that the isochore structure could be an adaptation to homeothermy. However, the discoveries of an isochore structure in some reptiles, namely, crocodile and turtle (Hugues et al. 1999), demonstrated that the origin of isochores preceded the divergence between mammals and reptiles and, hence, the acquisition of homeothermy. Moreover, the absence of a relationship between temperature and GC content in vertebrates (Belle et al. 2002; Ream et al. 2003) also led to the rejection of this hypothesis.

Duret et al. (2002) recently found some evidence of a decline of GC-rich isochores in mammals. They

analyzed the synonymous substitution pattern in coding sequences from closely related species in primates, cetartiodactyles, and rodents and found an excess of GC to AT substitutions. The erosion of GC-rich isochores has been confirmed by the analysis of substitution patterns in noncoding regions in primates (Webster et al. 2003) and in repetitive retroelements in mammals (Arndt et al. 2003b).

However, Duret et al. (2002) and Lercher et al. (2002) reported, from the analysis of human sequence polymorphism data, some evidence of a GC-biased fixation bias, that is, of an evolutionary force opposing this decline in GC content. Duret et al. (2002) suggest that this force is not strong enough now to compensate for the strong mutation bias toward AT and therefore for the disappearance of GC-rich isochores from mammalian genomes.

Duret et al. (2002) used the method of maximum parsimony to infer the direction of the substitutions. Parsimony, however, can give biased results, even when levels of divergence are quite modest, especially if there is substantial variation in the rate of substitution between sites. Parsimony tends to incorrectly infer an excess of common to rare changes (Eyre-Walker 1998). Therefore, to avoid this problem with the use of parsimony, we used a maximum likelihood method to test whether isochores have been vanishing from mammalian genomes. We used the nonhomogeneous nucleotide substitution model of Galtier and Gouy (1998), which allows GC content to change across the tree and for sites to have different rates of substitution. Under this model, we can estimate the ancestral GC content and the GC content at each internal node of the tree. This method is much more reliable than maximum parsimony and can be used for relatively high levels of divergence between sequences, allowing us to approach isochore evolution along the whole mammalian tree.

We applied this method to orthologous sequences of 41 mammalian genes for which we had at least one marsupial. We found some strong evidence of an early decline of the GC content shortly after the divergence of mammals and a later less striking but still significant decrease in the GC content of GC-rich isochores in several mammalian lineages, namely, primates, rodents, and carnivores.

Materials and Methods

Orthologous sequences of mammalian genes for which at least one marsupial was available were extracted using the HOVERGEN database (release 42, 2002 [Duret et al. 1994]). For each of the 48 genes found, protein sequences were aligned using the Clustal-W alignment program (Thompson et al. 1994) and the graphical sequence alignment editor Seaview (Galtier et al. 1996). Nucleotide alignments were then deduced from the aligned amino acid sequences. For each gene, we had between 6 and 66 species of

mammals, and for most of them we had at least one representative of the following eutherian taxa: rodents, lagomorphs, primates, carnivores, cetartiodactyles, and perissodactyles.

We estimated the ancestral GC content at the root of the tree and at each node using the program Eval_nh developed by Galtier and Gouy (1998). This program implements a model of DNA substitution which allows the evolutionary processes to vary among lineages: the assumptions of homogeneity and stationarity of traditional models are relaxed to account for variable base compositions among present-day sequences. The substitution process on any branch follows Tamura's (1992) model with unequal equilibrium GC contents among branches, so that the GC level can vary with time and between lineages. The program allows to model the rate of variation between sites according to a gamma distribution. The GC content at the root and the location of the root are two additional parameters of the model. A step-by-step optimization algorithm was designed to estimate the values of these parameters which would maximize the likelihood of the model. Parameters, including ancestral GC contents, are estimated using the maximum-likelihood method (Felstenstein 1981).

For each gene, we supplied the program with a tree, based on known phylogenies (Murphy et al. [2001] and Liu et al. [2001] for eutherians; Colgan [1999] for metatherians; NCBI and TREE of Life Web sites), the nucleotide sequences of the third codon position in each species, and a file specifying the options of the program (default options). We also constructed phylogenetic trees on the basis of the protein sequences using the neighbor-joining method implemented by Phylo_win and the K2 evolutionary distance of Kimura. In order to double-check for possible paralogy problems, in cases where we obtained a tree in striking discordance with the known phylogenies, we took the sequence which showed a discordance with the known phylogeny and performed a BLAST search against the human genome. We excluded two genes for which we suspected some paralogy, that is for which we did not find the exact same human match.

The output of the program provides a tree for each gene with the estimated GC content of the ancestor and at each node of the tree as well as the branch lengths. We calculated the GC content at the third codon positions (GC_3) and, as a control, at the third codon positions of the fourfold degenerated codons only (GC_4). The reason for focusing on the GC_3 of coding sequences is because it has been shown that there is a good correlation between the GC3 and the GC content of the genomic region in which the gene is located (Aissaini et al. 1991). The results with the GC_3 and the GC_4 were highly consistent (data not shown), suggesting that using all third codon positions did not bias our analyses.

A likelihood-ratio test was devised to check the adequacy of the model to the data and the robustness of ancestral GC estimates. For each gene we ran the program excluding the marsupials and recorded the log-likelihood value ln_{L1}. We then ran the program on the same file again (without marsupials) but this time forcing the ancestral eutherian GC content at the value found with the main analysis including the marsupials, and we again recorded the log-likelihood value ln_{L2}. We calculated twice the difference $\Delta = 2(ln_{L1} - ln_{L2})$ and we excluded any gene for which this difference was greater than a chi-square value with one degree of freedom (3.84, or ≈ 4). A large difference between ln_{L1} and ln_{L2} indicates that the estimate of the ancestral eutherian GC content is not robust, since it is strongly affected by the inclusion/exclusion of marsupials. We excluded three genes on this basis; we also excluded from the analyses two other genes for which we had fewer than 100 codons, after gaps were removed. The following results therefore only concern the remaining 41 genes. We subsequently divided our 41 genes into three approximately equal categories according to their GC_3 content: GC-poor (13 genes; GC < 0.550), GC-medium (15 genes; 0.550 < GC < 0.655) and GC-rich (13 genes; GC > 0.655). Since the GC_3 content is not independent of the variable of interest (the difference in GC_3), we also repeated all our analyses using the
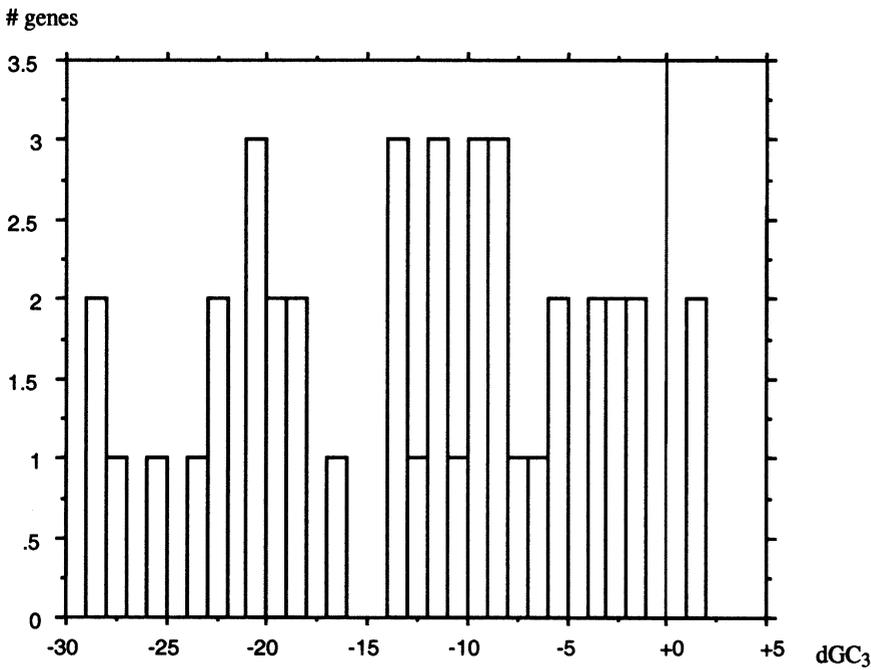
genes

Fig. 1. Histogram of the mean differences in GC_3 across genes between the estimated GC_3 of the common ancestor of mammals and the mean observed GC_3 across orders, for each of the 41 genes surveyed.

GC content at the first and second codon positions (GC_{12}) instead of the GC_3 to divide our genes into different categories. We found no substantial difference in the results; we therefore only present the results obtained using the GC_3 content.

Simulation Studies

Sequences were simulated under the model of Tamura (1992), except that we allowed higher transition rates in CpG dinucleotides. We used the following parameters: a transition to transversion ratio of 2.75; and an increase in the transition rate in CpG of 10-fold. Simulations were driven as follows. To generate ancestral sequences, we started from a random sequence and we let it evolve until each site experienced an average of 10 substitutions (i.e., reached equilibrium). We generated a set of ancestral sequences of different GC content (from 40 to 80%), using different values of the GC-bias parameter. We then made these sequences evolve, with a GC bias corresponding to an equilibrium GC content of about 40%, following a tree topology similar to the real phylogeny in our data: ((((Taxon1:0.1, Taxon 2:0.1):0.05, Taxon 3:0.15):0.1, ((Taxon4:0.1, Taxon5:0.1):0.05, Taxon6:0.15):0.1) :0.0, ((Taxon7:0.15, (Taxon8:0.1, Taxon9:0.1): 0.05):0.1, ((Taxon10:0.1, Taxon11:0.1):0.05, Taxon12:0.15):0.1): 0.0). The alignment of the 12 simulated sequences was then used to estimate their ancestral GC content with Eval_nh.

Results

We have applied a nonhomogeneous model of base substitution which allows different rates of substitution between sites on 41 genes for which we had at least one marsupial species. Using this model, we estimated the GC content at each node of the tree including the ancestral GC content. We calculated the change in GC content along each branch of the tree. The program also estimates the GC content at the terminal branches (expected GC), which is often slightly different from the real current GC content (observed GC). Generally the difference is not great and the results are qualita-

tively unchanged whether the observed or the estimated terminal values are used. We therefore only present the results with the observed values.

For each gene, we calculated the mean change in GC_3 from the common ancestor of mammals to the current eutherian orders; these values are represented in Fig. 1. There is strong evidence of an overall very significant GC content decrease across all genes (-12.7%) between the estimated GC_3 of the common ancestor of mammals and the mean observed GC_3 across orders (t-test, $p < 0.0001$). Note here that throughout the paper the use of percentage refers to the value of the decrease in GC content, and not to the percentage of the decrease; for example, if we have an ancestral value of 75% and a terminal value of 50%, we talk about a decrease in GC content of 25% (and not 33.33%).

The decrease in GC content observed across genes appears to be largely due to GC-rich genes, which exhibit the most significant decrease (-14.9%; $p < 0.0001$) before GC-medium genes (-14.4%; $p < 0.01$) and GC-poor genes (-9.4%; $p < 0.01$). There is also a very significant proportion of genes showing a decrease in GC content since the last common ancestor of mammals to the average observed GC_3 across present orders (39/41, or 95%; $p < 0.0001$, binomial test).

We therefore have evidence of a very significant decrease in GC content, especially in GC-rich genes. The question now is, When in mammalian evolution did this change occur? To answer this question, we divided mammalian evolution up into several phases: between the common ancestor (CA) of mammals (CA_{Mam}) and the CA of eutherians (CA_{Euth}), between CA_{Mam} and marsupials, and between CA_{Euth} and current species from several eutherian orders:
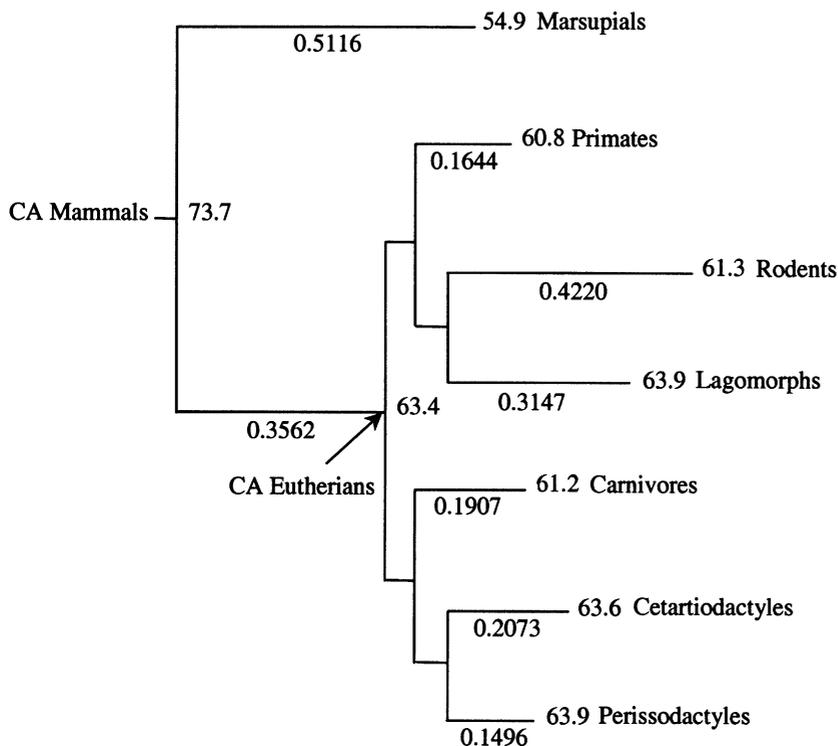
Fig. 2. Schematic mammalian tree showing the mammalian orders used in the analyses with the average estimated GC_3 content for the common ancestor (CA) of mammals and for the CA of eutherians, the average current GC_3 content for present-day mammalian orders, and the average branch lengths.

rodents, lagomorphs, primates, carnivores, perissodactyles, and cetartiodactyles (Fig. 2).

We found a very strong and significant decrease in the average GC content between CA_{Mam} and CA_{Euth} (-10.3%; $p < 0.0001$) and between CA_{Mam} and marsupials (-18.8%; $p < 0.000l$) for all genes and also, surprisingly, for each GC content category (Table 1), including GC-poor genes.

The picture is less clear from CA_{Euth} to each branch leading to each of the present-day eutherian lineages (rodents, lagomorphs, primates, carnivores, perissodactyles, and cetartiodactyles). In the rodent, primate, and carnivore lineages, we found some significant evidence of an ongoing disappearance of GC-rich isochores: in GC-rich genes of rodents the GC content decrease was -10.9%, in primates -4.1%, and in carnivores -3.7%. The GC content of GC-rich genes of perissodactyles and cetartiodactyles also decreased, although not significantly. In GC-medium genes, the decrease in GC content was significant only in primates (-2.4%). Regarding GC-poor genes, we found a significant increase in the GC content in rodents ($+5.2\%$). Interestingly, the GC content of lagomorphs and perissodactyles is also increasing, although not significantly.

The Decrease in GC Content: A Reflection of the Branch Lengths?

It can be argued that the contrast between the great decrease observed in marsupials and later mammalian evolution could reflect the fact that the average branch length within eutherians (0.241) is shorter than the average branch length from CA_{Mam} to CA_{Euth} (0.356) or to marsupials (0.512). There is indeed a significant correlation between the difference in GC_3 and the length of the branches in about three-quarters of the genes surveyed (31/41) (Supplementary Table).

In order to check that the decrease in GC content was genuinely more important early in mammalian evolution and was not an artifact of branch lengths, we compiled the following differences in GC content *divided* by the branch lengths (Table 2): between CA_{Mam} and CA_{Euth}, between CA_{Mam} and marsupials, and between CA_{Euth} and any of the six eutherian orders: rodents, lagomorphs, primates, perissodactyles, cetartiodactyles, and carnivores (we took the average branch length from CA_{Euth} to any eutherian order when several species of the same eutherian order were available).

The decrease in GC content divided by the branch length is significantly different between CA_{Mam}–CA_{Euth} and any CA_{Euth}–eutherian (Wilcoxon rank signed test, $p < 0.05$). It is also significantly different between CA_{Mam}–marsupials and CA_{Euth}–eutherian ($p < 0.05$). But the differences between CA_{Mam}–marsupials and CA_{Mam}–CA_{Euth} are not significantly different. This result suggests that the decrease in GC content has indeed been stronger at the beginning of mammalian evolution from the common ancestor of mammals to the common ancestor of eutherians or marsupials. After applying the Bonferroni correction

Table 1. Mean variation in GC_3 content by gene category according to GC_3 richness (high, medium, poor) along each major branch of mammalian lineages for the 41 genes surveyed

Lineage/dGC_3	All genes	Low GC_3	Medium GC_3	High GC_3
CA mammals/CA eutherians	−10.32***	**−10.40****	**−12.25*****	**−8.29*****
CA mammals/marsupials	−18.80***	**−11.39****	**−19.53*****	**−26.64*****
CA eutherians/rodents	−2.11	**5.19***	−1.76	**−10.87*****
CA eutherians/lagomorphes	1.79	2.01	3.12	0.30
CA eutherians/primates	−2.56***	−1.42	**−2.36***	**−4.09****
CA eutherians/perissodactyles	−0.94	1.93	−3.00	−2.02
CA eutherians/cetartiodactyles	−0.31	0.71	0.11	−1.73
CA eutherians/carnivores	−1.76*	0.09	−1.89	**−3.68****

Note. CA, common ancestor; considering the average if several values were available in a lineage. *t*-test: ***$p < 0.0001$; **$p < 0.01$; *$p < 0.05$.

for multiple tests, during the evolution of eutherians, only the differences between CA_{Euth}–primates and CA_{Euth}–cetartiodactyles were significantly different. This suggests that maybe there has been a stronger or faster decrease in primates compared to other eutherian lineages.

The Disappearance of GC-Rich Isochores: A Methodological Artifact?

The use of parsimony to infer the synonymous substitution pattern can lead to incorrect results: parsimony can be biased and the bias can be exacerbated by a variation in the rate of substitution between sites (Eyre-Walker 1998). We therefore used the method of Galtier and Gouy (1998), which allows variation in the substitution rate between sites and should not suffer from this problem. However, one potential problem with this method is that it assumes that the probability of substitution of a given base does not depend on the nature of its flanking bases (i.e., no neighboring effect), whereas it is known that in vertebrates, cytosines located in a CpG dinucleotide mutate at a very high rate.

To check the validity of the model, we performed a simulation study, using a model of evolution mimicking the real situation in mammals, i.e., including transition and transversion rates, GC bias, CpG hypermutability, and nonstationarity of GC content.

We generated ancestral sequences of different GC content (poor, 0.40; medium, 0.48; rich, 0.62; and very rich, 0.75). We then let these sequences evolve toward an equilibrium GC content of about 40%, following a tree topology similar to that of eutherian mammals (see Materials and Methods). We then calculated the differences (Δ) between the real ancestral GC content and the GC content estimated with the Eval_nh model.

We found that these mean differences increased with the GC content and the branch lengths (Table 3). Interestingly, Δ was always found to be positive, suggesting that Eval_nh has a tendency to

Table 2. Average differences in GC_3 content (dGC3) *divided* by the branch lengths (BL) along each major mammalian branch for the 41 genes surveyed

Lineage	dGC_3	dGC_3/BL
CA mammals/CA eutherians	−10.3	−39.6
CA mammals/marsupials	−17.7	−49.5
CA eutherians/rodents	−2.9	−8.5
CA eutherians/lagomorphes	0.25	5.5
CA eutherians/primates	−2.3	−17.1
CA eutherians/perissodactyles	−1.0	−1.5
CA eutherians/cetartiodactyles	−0.9	−1.7
CA eutherians/carnivores	−2.1	−11.5

Note. See Wilcoxon signed rank tests in Results.

slightly underestimate the ancestral GC content. These differences ranged from +0.4% per 0.1 substitution in GC-poor genes to +0.9% in "very GC-rich" genes. Thus the decrease in GC content in the branches from the common ancestor of eutherians to present-day eutherian orders is probably about 3% stronger than reported in Table 1.

This underestimation of the decline in GC content is likely to be due to the presence of CpG dinucleotides. Indeed, Eval_nh does not take into account the high GC to AT mutation rate at CpG dinucleotides (also present in the simulated sequences): Eval_nh is able to take into account variations of substitution rates between sites, but not a specific GC bias at rapidly evolving sites. Note that some methods have been developed to take CpGs into account (Arndt et al. 2003a), but that they require a very large number of sequences and hence could not be used in these analyses.

Discussion

We found evidence of a significant decrease in the GC content during mammalian evolution; this appears to be particularly evident in early mammalian evolution, where there is a dramatic decrease in the GC content from the common ancestor of mammals to the com-

Table 3. Summary table of the results of the simulation studies: GC category, observed ancestral GC content, ancestral GC content estimated with the Eval_nh program, difference Δ (Obs − Est) = observed − estimated ancestral GC content, average branch length (BL), and error expected for a branch length of 0.1 ($\Delta/0.1$BL)

GC category	Observed GC	Estimated GC	Δ (Obs − Est)	Average BL	$\Delta/0.1$BL
Poor	0.400	0.389	+ 0.011	0.2707	+ 0.004
Medium	0.479	0.462	+ 0.017	0.2524	+ 0.006
High	0.617	0.591	+ 0.025	0.2927	+ 0.008
Very high	0.751	0.722	+ 0.029	0.3193	+ 0.009

mon ancestor of eutherians and to marsupials in all gene categories (GC-poor, -medium, or -rich). However, later in mammalian evolution (from the common ancestor of eutherians to current mammalian orders), this decrease is less evident or absent; it is nevertheless still significant for GC-rich genes in at least three lineages—primates, rodents, and carnivores—and marginal in perissodactyles and cetartiodactyles. But note that these decreases are probably slightly underestimated by our method (See Results).

We found evidence of a change in the rate of decay of GC content during mammalian evolution in our analysis where we corrected for the branch lengths (see Results). The rate of decline early in mammalian evolution (−10.3%) appears to be over four times faster than late in mammalian evolution (−2.4%). One simple explanation for this difference would be that as the GC content approaches the equilibrium GC content, the rate of decrease is declining, even if the underlying pressure toward low GC is constant with time, this process being nonlinear and following a negative exponential curve, i.e., the slope of the curve of the GC decay is expected to decrease in time to reach zero at the equilibrium GC content.

To examine this possibility, consider the following simple model: let μ be the rate of change from GC to AT per generation and v the rate of change from AT to GC; then the change in GC content per generation is $\Delta f = -f\mu + (1-f)v$. At equilibrium $\Delta f = 0 = -f_e\mu + (1 - f_e)v$ and $f_e = v/(\mu + v)$. Without loss of generality let $\mu = 1-v$; then $f_e = v$ and $\Delta f = -f(1-f_e) + (1-f)f_e = f_e - f$. If we assume an equilibrium GC content, f_e, of 40%, then the rate of decay at the common ancestor of mammals is $\Delta f = 0.40 - 0.74 = -0.34$ and the rate of decay for present-day mammals is $\Delta f = 0.40 - 0.60 = -0.20$. This difference, which is only 1.7-fold, is much lower than the reduction in the rate of decay we actually observed (4.3). Therefore, this simple model cannot explain the decline in the rate of decay.

An alternative explanation is that the strong erosion could be due to a rapid decay of CpG dinucleotides. Indeed, Arndt et al. (2003b) found an abrupt increase (four- to eight-fold) in the rate of substitution at CpG dinucleotides (C→T and G→A substitutions), that occurred around 90 MYA. This shift in CpG substitution rate apparently coincides with the rapid erosion of GC-rich isochores that we observed between CA_{Mam} and CA_{Euth}. To investigate the potential impact of the CpG decay on the decrease in GC content, we tried to estimate the level of CpG in the common ancestor of mammals. For this purpose, we calculated the ratio of observed over expected CpG dinucleotides ($CpG_{o/e}$) in several chordate species for which we had sufficient amount of genomic sequence. We found a progressive decline in this ratio (Table 4) from Urochordata (where there is no CpG depletion) to Amniotes ($CpG_{o/e} = 0.2$ to 0.3). The fact that there is a strong depletion in amphibians, birds, and mammals indicates that the genome of CA_{Mam} was already strongly depleted ($CpG_{o/e} \leq 0.3$). Thus, although CpG substitutions contribute to the decline in GC content, the strong erosion observed between CA_{Mam} and CA_{Euth} cannot be simply attributed to the decay of CpGs since the common ancestor of mammals was already strongly depleted in CpG dinucleotides.

Our results are consistent with recent studies suggesting the disappearance of GC-rich isochores in different mammalian orders (Duret et al. 2002; Webster et al. 2003; Arndt et al. 2003b). Duret and colleagues analyzed the evolution of isochores by looking at the synonymous substitution pattern in orthologous coding sequences of rodents, primates, and cetartiodactyles. Using parsimony, they found a significant excess of GC to AT over AT to GC changes for GC-rich genes in the three orders. In this study, we found a significant decrease in the GC content of GC-rich genes in primates, rodents, and carnivores. We also observed the same trend in cetartiodactyles and perissodactyles, although the decrease was not statistically significant. Note, however, that the method we used in this study tends to underestimate the ancestral GC content (Table 4) and, hence, underestimate the decrease. The only exception to this general pattern is observed in lagomorphs. Thus it appears that the erosion of GC-rich isochores occurs in most mammalian orders.

Arndt et al. (2003b) used repetitive elements to infer the evolution of the substitutional patterns in humans using a maximum likelihood method. They suggest that the pattern of substitution has changed

Table 4. Levels of CpG depletion in several chordate species: number of sequences, GC content, and $CpG_{o/e}$ ratio

Taxon	Species	No. seqs.	GC %	$CpG_{o/e}$
Urochordata	*Ciona savignyi*	49	37.9	0.96
Cephalochordata	*Branchiostoma lanceolatum*	39	41.6	0.74
Teleostei	*Fugu rubripes*	159	45.8	0.60
	Danio rerio	4831	36.4	0.52
Amphibia	*Silurana tropicalis*	128	41.2	0.40
	Xenopus laevis	13	41.3	0.38
Aves	*Gallus gallus*	235	43.4	0.28
Mammalia	*Bos taurus*	294	43.9	0.26
	Sus scrofa	302	46.0	0.32
	Canis familiaris	221	42.8	0.25
	Felis catus	253	43.0	0.31
	Oryctolagus cuniculus	40	47.5	0.30
	Papio anubis	264	43.1	0.25
	Mus musculus	a	40.9	0.22
	Homo sapiens	a	41.7	0.19

a, Complete genome.

from one dependent on the GC content (where there were more transitions toward GC in regions of high GC content and more transitions toward AT in regions of high AT content) to one independent of the GC content. They suggest that this transition occurred early in mammalian evolution, about 90 MYA. Our data would suggest an earlier decline of isochores, starting as early as the divergence between eutherians and marsupials, estimated to have occurred around 130 MYA (Marshall and Westerman 2002). However, these estimates are not very accurate, therefore we can consider our findings to be roughly consistent with the Arndt et al. (2003b) study.

The interesting contribution of this result to the very much debated isochore topic is that it shows that the disappearance of isochores has not been constant with time. It has been very fast at the start of mammalian radiation and has then slowed down on the branches leading to present-day mammals. We can therefore question why the evolution of GC content is not stable in time. There are three evolutionary processes which could be responsible for the evolution of isochores and this variation and thus be unstable through time: mutation bias, selection, and biased gene conversion (BGC).

According to the mutation bias hypothesis, the rate of change from AT to GC would be greater than the rate of change from GC to AT. This process would depend on the strength of the bias and of the cellular repair system involved.

If, instead, selection is responsible for this variation in the rate of decline of the GC content, we would expect this variation to be correlated with the evolution of a selected trait. Homeostasis, the ability to regulate body temperature, could be related to variation in the GC content since C and G nucleotides are more thermally stable. However, the evolution of homeostasis occurred at the beginning of mammalian evolution and probably did not change until present-day mammals. Selection for thermal stability is therefore again unlikely to explain this variation. This process would depend on the strength of the selective pressure and of the effective population size (N_e).

Finally, gene conversion is a process which occurs at meiosis during recombination when heteroduplexes are found between maternal and paternal homologous chromosomes. If a heterozygous site is involved, a DNA mismatch occurs and is repaired by one of the cellular DNA repair system. If this process is biased, i.e., if AT alleles are converted to GC alleles more frequently than the opposite, we talk about BGC. It has been proposed that GC-rich isochores could have evolved from this process (Holmquist 1992; Eyre-Walker 1993, Galtier et al. 2001). In addition to the length of the heteroduplex, the strength of bias of the repair system, and the effective population size (N_e), this process also depends on the rate of formation of heteroduplexes and hence on the rate of recombination. In humans there is a significant correlation between the local rates of recombination and the GC contents (Kong et al. 2002). A change in the rate of recombination could occur if some major chromosomal rearrangements, such as fission or fusion, take place.

To conclude, we found some strong evidence supporting the disappearance of isochores very early in mammalian evolution, with an ongoing decrease in some lineages (primates, rodents, and carnivores). This variation in the rate of decline implies that there has been a change in the rate of the possible causes of isochores, in the strength of this evolutionary force, or in one of the variables it depends on, possibly the effective population size or the rate of recombination.

Since birds also exhibit very rich isochores, it would be interesting to assess the date of origin of

their isochores and to see whether their GC-rich isochores are also now vanishing. Bird genomes include very small chromosomes with an extremely high GC content (Kadi et al. 1993) which probably arose by fission. It would therefore be very informative to see if there is a correlation between the occurrence of these events and changes in the GC content of their genomes.

References

Aissaini B, D'Onofrio G, Mouchiroud D, Gardiner K, Gautier C, Bernardi G (1991) The compositional properties of human genes. J Mol Evol 32:493–503

Arndt PF, Burge CB, Hwa T (2003a) Modelling of neighbor-dependent mutation in DNA sequence evolution. J Comput Biol 10:313–322

Arndt PF, Petrov DA, Hwa T (2003b) Distinct changes of genomic biases in nucleotide substitution at the time of mammalian radiation. Mol Biol Evol 20:1887–1896

Belle EMS, 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 (2000) Isochores and the evolutionary genomics of vertebrates. Gene 241:3–17

Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J (1985) The mosaic genome of warm-blooded vertebrates. Science 228:953–958

Colgan DJ (1999) Phylogenetic studies of marsupials based on phosphoglycerate kinase DNA sequences. Mol Phylogenet Evol 11:13–26

Duret L, Mouchiroud D, Gouy M (1994) Hovergen: A database of homologous vertebrate genes. Nucleic Acids Res 22:2360–2365

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

Eyre-Walker A (1993) Recombination and mammalian genome evolution. Proc R Soc Lond B Biol Sci 252:237–243

Eyre-Walker A (1998) Problems with parsimony in sequences of biase base composition. J Mol Evol 47:686–690

Eyre-Walker A, Hurst LD (2001) The evolution of isochores. Nat Rev Genet 2:549–555

Festenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17:368–376

Galtier N, Mouchiroud D (1998) Isochore evolution in mammals: A human-like ancestral structure. Genetics 150:1577–1584

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

Galtier N, Gouy M, Gautier C (1996) SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput Applic Biosci 12:543–548

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

Holmquist GP (1992) Chromosome bands, their chromatine flavours and their functional features. Am J Hum Genet 51:17–37

Hughes S, Zelus D, Mouchiroud D (1999) Warm-blooded isochore structure in Nile crocodile and turtle. Mol Biol Evol 16:1521–1527

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

Kadi F, Mouchiroud D, Sabeur G, Bernardi G (1993) The compositional patterns of the avian genomes and their evolutionary implications. J Mol Evol 37:544–551

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Gene 31:241–247

Lercher MJ, Smith NG, Eyre-Walker A, Hurst LD (2002) The evolution of isochores. Evidence from SNP frequency distributions. Genetics 162:1805–1810

Liu F-GR, Miyamoto MM, Freire NP, Ong PQ, Tennant MR, Young TS, Gugel KF (2001) Molecular and morphological supertress for eutherian (placental) mammals. Science 291:1786–1789

Marshall JA, Westerman M (2002) Marsupial genetics and genomics. Trends Genet 18:517–521

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ (2001) Molecular phylogenetics and the origins of placental mammals. Nature 409:614–618

Ream RA, Johns GC, Somero GN (2003) Base compositions of genes encoding α-actine and lactate deshydrogenase-A from differently adapted vertebrates show no temperature-adaptive variation in G + C content. Mol Biol Evol 20:105–110

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. Mol Biol Evol 9:678–687

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Webster MT, Smith NG, Ellegren H (2003) Compositional evolution of noncoding DNA in the human and chimpanzee genomes. Mol Biol Evol 20:278–286