FEBS Letters 581 (2007) 5819–5824

Different functional classes of genes are characterized by different compositional properties

Giuseppe D'Onofrio[a,*], Tapash Chandra Ghosh[b], Salvatore Saccone[c]

[a] *Laboratorio di Evoluzione Molecolare, Stazione Zoologica A. Dohrn, Villa Comunale, 80121 Napoli, Italy*
[b] *Bioinformatics Centre, Bose Institute, P 1/12, C.I.T. Scheme VII M, Kolkata 700 054, India*
[c] *Dept. Biologia Animale "M. La Greca" University of Catania, Via Androne, 81 – 95124 Catania, Italy*

Received 13 August 2007; revised 14 November 2007; accepted 16 November 2007

Available online 26 November 2007

Edited by Takashi Gojobori

Abstract A compositional analysis on a set of human genes classified in several functional classes was performed. We found out that the GC3, i.e. the GC level at the third codon positions, of the genes involved in cellular metabolism was significantly higher than those involved in information storage and processing. Analyses of human/Xenopus ortologous genes showed that: (i) the GC3 increment of the genes involved in cellular metabolism was significantly higher than those involved in information storage and processing; and (ii) a strong correlation between the GC3 and the corresponding GCi, i.e. the GC level of introns, was found in each functional class. The non-randomness of the GC increments favours the selective hypothesis of gene/genome evolution.

Keywords: Base composition (GC); Biased gene conversion (BGC); Chromosomal band; Codon usage; Intron; Isochore

1. Introduction

From the time of availability in public databases of several hundred coding sequences, it was evident that human genes covered a broad compositional range, especially at the third codon positions [1,2]. The same features were observed in all mammals and birds, so far analyzed, but not in amphibians and fish [1,3,4] and, for a review, [5].

Comparative analyses of coding sequences from cold- and warm-blooded vertebrates defined two classes of genes, i.e. the GC-poor and the GC-rich. The former were essentially those that showed close base composition among all vertebrates, whereas the latter were essentially those that underwent a compositional transition, becoming CG-richer in warm-blooded vertebrates [6], reviewed in [7]. Interestingly, comparative analyses of the GC3, i.e. the GC levels at the third codon positions, confirmed the results obtained from comparing the genomes of cold- and warm-blooded vertebrates [8]. Significant correlations found between the GC3 of the genes and the corresponding GC of introns and flanking regions, further supported the role of GC3 as parameter to understand not only gene, but also genome evolution [9–11].

The compositional properties of GC-poor and GC-rich human genes were paralleled by several observations concerning structural and functional properties, such as: gene density [10], gene structure [12,13], gene length [14], amino acid frequencies and codon usage [2,15,16], expression breadth [17–19], expression levels [20–22], chromosomal band localization and chromatin structure [23–25], substitution rates, at both gene and protein levels [26,27]. However, in spite of the large number of studies, the adaptive or the stochastic nature of the forces driving compositional changes, observed at coding and non-coding levels, is still a matter of debate in the field of molecular evolution [5,7,28–30].

Here, in this study, we have performed a compositional analysis on a set of human genes classified in several functional classes. We found out that the GC3 was significantly higher in the genes involved in cellular metabolism than in those involved in information storage and processing. The comparison of human/Xenopus orthologous genes showed that the above result was due to a non-uniform GC3 increment among functional classes. The same trend was found comparing the non-coding sequences of human genes. Moreover, within each functional class, a strong correlation between GC3 and GCi, i.e. the GC levels of the introns, was found. According to the above results, the selective hypotheses of gene/genome evolution were favored.

2. Materials and methods

Human genes were retrieved from NCBI/CoreNucleotide database. CleanUp 2.0 [31] was used to remove redundant entries. Sequences without AUG and/or stop codons, as well as with internal stop codons, were detected by CodonW 1.4.2 (J. Peden; http://www.molbiol.ox.ac.uk/Win95.codonW.zip) and removed. The remaining 15112 CDS were classified in 25 functional classes. As described in COG database [32,33], functional classes, denoted by capital letter, were grouped in three broad categories, namely: (i) information storage and processing; (ii) cellular processes and signaling; and (iii) metabolism. The base composition of sequences was calculated by CodonW 1.4.2. The average GC3 level was compared among functional classes by pairwise comparisons. Moreover, the average GC3 of each functional class was compared with the average GC3 of the whole dataset. The significance of the differences found in the pairwise comparisons was assessed by the *t*-Student's test, with Bonferroni's correction for multiple-comparisons. Indeed, in order to avoid a lot of spurious positives, the threshold (alpha level) of significance was lowered to account for the number of comparisons performed. The human genes of four functional classes were positioned in the chromosomal bands, according to published data [34]. The comparative analysis was extended to a specific subset of human/Xenopus orthologous genes

*Corresponding author.
E-mail address: donofrio@szn.it (G. D'Onofrio).

0014-5793/$32.00
doi:10.1016/j.febslet.2007.11.052

retrieved from HOVERGEN (release 48, May 2007) [35]. For those genes, the corresponding intronic regions, localized between the AUG and the stop codon of the human CDS, were retrieved from UCSC database (http://genome.ucsc.edu/), and the interspersed repeats removed by Repeat Masker (http://ftp.genome.washington.edu/).

3. Results

The classification of the human genes in 25 functional classes, denoted by capital letter, and the corresponding number of genes was reported in Table 1. Functional classes were grouped in three main categories: information storage and processing (in blue), cellular processes and signaling (in black) and metabolism (in red). Each category respectively represented, about 21%, 33% and 15% of the all dataset. About 30% of the all entries were not classified in any of the above-mentioned categories because the gene function was unknown or predicted only (in grey). Four functional classes, namely [H], [M], [N] and [Y], were represented by less than a hundred sequences, and therefore removed from subsequent statistical analyses.

The GC3 values, i.e. average and variance of each functional class, and those of the whole dataset of human genes, were reported in Table 2. In the same table the statistically significant pairwise comparisons (alpha level 2.1×10^{-4}; $P < 0.05$) were also reported. The average GC3 of the functional classes ranged from 53% to 63%, while that of the whole dataset was 58%. Sorting functional classes by GC3, it was observed: (i) the clustering of the blue and red classes at the opposite extremes of the GC3 range, whereas the black classes did not show any biased distribution (for color code see Table 1); and (ii) the clustering of the statistically significant pairwise comparisons towards highest GC3 values (Table 2). The average GC3 of each functional class was compared with the average GC3 of the whole dataset of human genes. Nine classes turned out to be statistically significant (Table 2, last line). The average GC3 of four classes, namely [A], [L], [J] and [O], were significantly lower than the average GC3 of the whole data set, whereas the average GC3 of five classes, namely [K], [T], [G], [E] and [P], were significantly higher. It is worth stressing that three out of four classes of the first group, namely [A], [L] and [J], belonged to the category of information storage and processing, while three out of five of the second group, namely [G], [E] and [P], belonged to the category of metabolism (see Table 1). Two classes, namely [O] and [T], confirmed the absence of any specific trend in the category of cellular pro-

Table 1
Classification of the human genes

	INFORMATION STORAGE AND PROCESSING	#Genes
[A]	RNA processing and modification	497
[B]	Chromatin structure and dynamics	233
[J]	Translation, ribosomal structure and biogenesis	1203
[K]	Transcription	1046
[L]	Replication, recombination and repair	278
	CELLULAR PROCESSES AND SIGNALING	
[D]	Cell cycle control, cell division, chromosome partitioning	245
[M]	Cell wall/membrane/envelope biogenesis	62
[N]	Cell motility	28
[O]	Post-translational modification, protein turnover, chaperones	1427
[T]	Signal transduction mechanisms	1723
[U]	Intracellular trafficking, secretion, and vesicular transport	639
[V]	Defense mechanisms	164
[W]	Extracellular structures	272
[Y]	Nuclear structures	18
[Z]	Cytoskeleton	531
	METABOLISM	
[C]	Energy production and conversion	336
[E]	Amino acid transport and metabolism	367
[F]	Nucleotide transport and metabolism	167
[G]	Carbohydrate transport and metabolism	386
[H]	Coenzyme transport and metabolism	93
[I]	Lipid transport and metabolism	379
[P]	Inorganic ion transport and metabolism	343
[Q]	Secondary metabolites biosynthesis, transport and catabolism	217
	POORLY CHARACTERIZED	
[R]	General function prediction only	3271
[S]	Unknown function	1187

Table 2
Average and variance of GC3 levels

Classes[b]	GC3 Average	GC3 σ^2	Pairwise comparison[a] [A][L][J][O][D][U][R][C][S][F][W][Q][I][B][Z][V][K][T][G][E][P]
[A]	0.534	0.024	-
[L]	0.538	0.023	-
[J]	0.543	0.013	-
[O]	0.551	0.022	-
[D]	0.556	0.024	-
[U]	0.569	0.026	S S -
[R]	0.572	0.029	S S S -
[C]	0.574	0.025	-
[S]	0.575	0.023	S S S -
[F]	0.589	0.024	S -
[W]	0.591	0.022	S S S S -
[Q]	0.595	0.024	S S S S -
[I]	0.601	0.023	S S S S -
[B]	0.607	0.029	S S S S -
[Z]	0.610	0.024	S S S S S S S S S -
[V]	0.612	0.029	S S S S -
[K]	0.613	0.028	S S S S S S S S S -
[T]	0.614	0.025	S S S S S S S S S -
[G]	0.616	0.019	S S S S S S S S S -
[E]	0.632	0.021	S S S S S S S S S -
[P]	0.637	0.027	S S S S S S S S S -
All dataset	0.582	0.025	S S S S S S S S S

S = significant pairwise comparisons.
[a] *t*-Student's test with Bonferroni's correction [$n^*(n-1)/2 + n$]. Alpha level 2.1×10^{-4}; $P < 0.05$.
[b] Functional classes sorted according to the average GC3 levels. Color code as in Table 1.

cesses and signaling. Since the former was significantly lower, and the latter significantly higher than the average GC3 of the whole data set of human genes. Interestingly, only one class of the information storage and processing category, i.e. the functional class [K] grouping genes involved in the transcription process, was significantly higher than the average GC3 of the whole dataset, stressing a link between transcription and metabolism rate.

The human genes of the classes [A], [L], [E] and [P] were localized in the four compositionally different subsets of chromosomal bands, namely H3+, H3−, L1−, and L1+ bands (Fig. 1). Two kinds of distributions were found: that of the classes [A] and [L], and that of the classes [E] and [P]. The two distributions were significantly different (χ^2 test, $P < 1.7 \times 10^{-2}$), because genes were differently distributed in H3+ and H3− bands (Fisher's exact test, $P = 0.0034$). Actually,

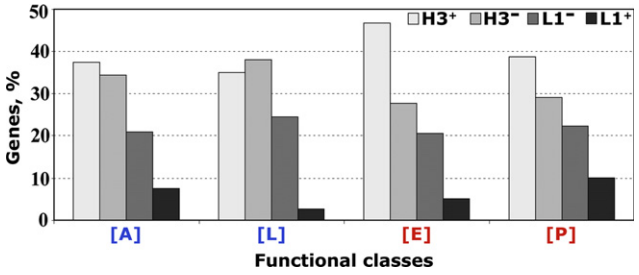

Fig. 1. Histogram of the gene distribution in chromosomal bands.

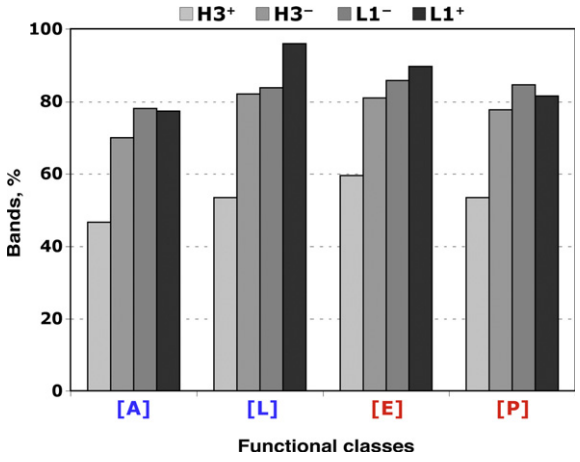

Fig. 2. Histogram of the chromosomal "empty-bands".

the classes [E] and [P] were more represented in the H3+ bands than the classes [A] and [L].

All four functional classes showed a significant clustering of genes in the H3+ bands. The percent of "empty-bands" was very high, not only as expected in the genepoor L1+ and L1− bands, but also in the gene-rich H3+ bands. In fact, the percent of H3+ "empty-bands" ranged from 46%, in [A], up to 60%, in [E] (Fig. 2). Moreover, all four functional classes showed that the percent of genes as a "single-genes/band" (black bars, Fig. 3) was lower than that of a " ⩾2genes/band" (grey bars, Fig. 3), especially in the H3+ bands.

In order to check if compositional differences among functional classes were just a human genome peculiarity or a more general facet of the gene/genome evolution in the transition from cold- to warm-blooded vertebrate, a set of human/Xenopus orthologous genes was retrieved. The analysis presented in this paper was restricted to the GC3-poor classes, namely [A] and [L], and to the GC3-rich classes, namely [E] and [P].

The number of orthologous genes found in each class was of the same order of magnitude for [A], [L], [E] and [P], respectively ~24%, ~22%, ~18% and ~25%. For the sake of simplicity, human/Xenopus orthologous classes of genes from now on will be denoted as [Ah], [Lh], [Eh] and [Ph] for the human genes, and [Ax], [Lx], [Ex] and [Px] for the Xenopus ones. The corresponding average and variance of GC3 were reported in Table 3. From a compositional point of view, the [A], [L], [E] and [P] classes were well represented by the corresponding [Ah], [Lh], [Eh] and [Ph] classes, since the GC3 values were not significantly different between corresponding functional classes. Needless to say, the average GC3 of the classes [Ah] and [Lh] were both significantly lower than those of the classes [Eh] and [Ph] (alpha level 6.2×10^{-3}; $P < 0.05$). Among the classes [Ax], [Lx], [Ex] and [Px], significant differences of the average GC3 values were not found. In human, however, the corresponding GC3 values were significantly higher (alpha level 6.2×10^{-3}; $P < 0.05$). More precisely, the GC3 increments were higher in the classes of genes belonging to the category of metabolism, i.e. [Eh] and [Ph], than those of the genes belonging to the category of information storage and processing, i.e. [Ah] and [Lh] (Table 3).

Last but not least, the average and variance of GCi, i.e. the GC levels of the introns, were analyzed (Table 3). Between the classes [Ah] and [Lh], no significant differences were observed. The same result was found comparing the classes [Eh] and [Ph]. However, the average GCi levels of the classes [Ah] and [Lh] were both significantly lower than those of the classes [Eh] and [Ph] (P-values ranging from $P < 4.6 \times 10^{-3}$ to $P <$

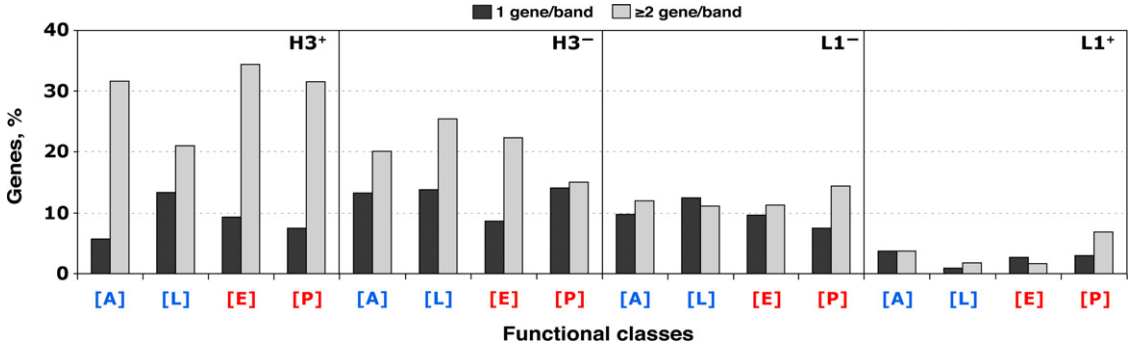

Fig. 3. Histograms of the chromosomal band distribution of "single-genes/band" and " ⩾2 genes/band" in each functional class.

Table 3
Average and σ^2 of the GC3 and GCi levels in ortologous genes

	GC3							
	[A]*		[L]*		[E]*		[P]*	
Human	0.540	0.0239	0.504	0.0210	0.636	0.0232	0.640	0.0281
Xenopus	0.472	0.0082	0.444	0.0056	0.462	0.0066	0.505	0.0098
Human GCi	0.437	0.0075	0.424	0.0059	0.485	0.0081	0.481	0.0095

[a]See Table 1 for description.

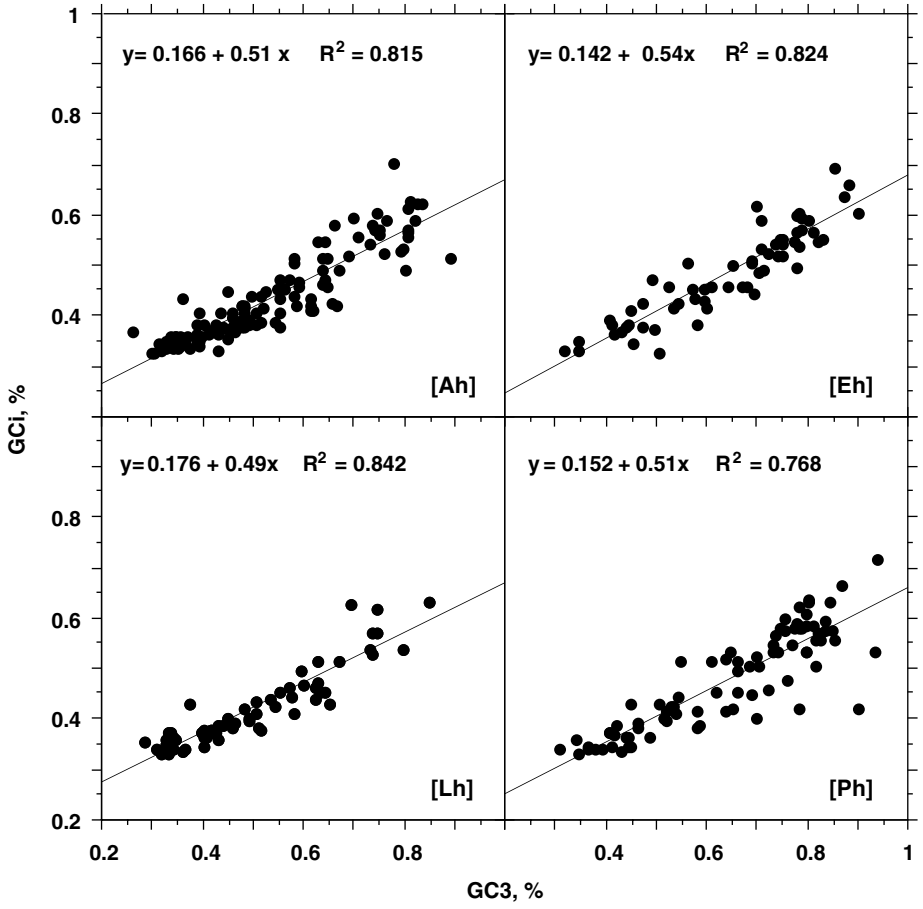

Fig. 4. Correlation within each functional class between the GC levels at the third codon positions (GC3) and the GC levels of the corresponding introns (GCi). Equations of linear correlation are shown.

1×10^{-4}, namely for [Ph] vs. [Ah] and for [Eh] vs. [Lh]). Within each functional class, a strong correlation between GC3 and GCi was found. Correlation coefficients (R^2) ranged from 0.77 to 0.84, respectively for [Ph] and [Lh] functional classes (Fig. 4).

4. Discussion

The evolution of the vertebrate genomes is still a matter of intense selectionist/neutralist debate devoted to understand the adaptive or stochastic nature of forces driving the compositional heterogeneity in the genomes of mammals and birds. Several hypotheses have been proposed, such as: (i) new thermodynamic conditions [5], or different nucleosome formation potential [30], supporting the adaptive hypotheses; and (ii) mutational bias [29] or recombination events, coupled with bias gene conversion (BGC) and bias mutation repair [28], supporting the stochastic hypotheses.

With the aim to produce new evidences supporting the selectionist or the neutralist hypothesis, we analyzed the base composition of human genes classified in 25 functional classes (Table 1). The results can be summarized in the following points:

(i) The average GC3 was significantly different among functional categories (see Table 2). More precisely, the GC3 of metabolic category was significantly higher than that of the whole dataset, whereas the information storage

and processing one was significantly lower. Functional classes of genes involved in cellular processing and signaling did not show any specific trend. Indeed, the average GC3 of the genes belonging to this category was lower, higher or not significantly different from that of the whole dataset.

(ii) Genes from the metabolic category and from the information storage and processing one were positioned in the chromosomal bands. The two categories showed a different chromosomal band distribution (Fig. 1), that of metabolism being biased towards the H3+ bands. A significant clustering of genes, especially in those bands, was observed in all functional classes (Fig. 3).

(iii) For the same categories, a set of human/Xenopus orthologous genes was analyzed. In human, the average GC3 of the genes classified in the metabolic category were significantly higher than those classified in the information storage and processing one. In Xenopus, no significant differences were observed. Pairwise comparison of human/Xenopus orthologous genes showed that human genes belonging to both categories underwent a GC3 increment, that, however, was higher in the metabolic one (Table 3).

(iv) In each functional class a strong correlation between GC3 and GCi was found (Fig. 4). The average GC3 was always higher than the corresponding average GCi (Table 3).

The first point raised some questions on the randomness of the factors affecting the base composition of coding and non-coding regions. Indeed, from a stochastic point of view, it would not be expected that a few functional classes were GC-richer than others. However, the different chromosomal band localization of the two categories of genes, point (ii), could give support to the mutational mechanism or the BGC hypothesis [29,28]. Indeed, the hypothesis of a mutational mechanism that acts unlike in different gene regions [29] could be easily extended to different genome regions. According to the BGC hypothesis [28], it could be argued that in the human genome metabolic genes preferentially cluster close to recombination hotspots [36]. Regarding the mutational mechanism producing a mutational bias toward GC, it was already shown that a mutational bias was found, indeed, in human genome, but was towards AT [27]. Regarding the BGC, point (iii) stressed the fact that compositional differences were not found among functional classes in Xenopus. Two pictures can be drawn out according to the absence or the presence of recombination hotspots in Xenopus genome. In the first case, it should be assumed a non-random appearance of recombination hotspots in the human genome, biased, indeed, towards metabolic genes. In the second case, it should be assumed a biased reshuffling producing a clustering of both metabolic genes and recombination hotspots in the human genome. In the two cases, due to the low probabilities of each event and to the report that recombination hotspot sites were poorly conserved among genomes [37], the organization and gene distribution in the human genome should be unique. However, comparative analyses among mammals and birds showed that base compositions of orthologous genes were significantly correlated [5].

The criticism that the selective pressure would be confined solely to the codon usage, a selection that, indeed, may act in human genes [38], was countered by point (iv). The strong correlations found between GC3 and GCi in each functional class implied not only selection on codon usage, but also a selection on a larger genomic scale, i.e. at isochore levels [5,30]. The fact that among functional classes GC3 was higher than GCi do not contradict the above hypotheses, but simply supported the expectation of a selective pressure stronger at coding than non-coding regions.

Therefore, the favoured answer to the question "Adaptation or biased gene conversion?" [28], it is "adaptation." Indeed, the different base composition of genes in different functional classes can be modulated by selective pressure which is not a conclusion in contrast with the neutral theory of molecular evolution [39] as reviewed in [7].

Acknowledgements: Thanks are due to Dieter Hafner for the kind statistical assistance, to Philip Sommese for the valuable proofreading, and to the anonymous referees for constructive criticisms.

References

[1] Mouchiroud, D., Fichant, G. and Bernardi, G. (1987) Compositional compartmentalization and gene composition in the genome of vertebrates. J. Mol. Evol. 26, 198–204.
[2] D'Onofrio, G., Mouchiroud, D., Äissani, B., Gautier, C. and Bernardi, G. (1991) Correlations between the compositional properties of human genes, codon usage, and amino acid composition of proteins. J. Mol. Evol. 32, 504–510.
[3] Bernardi, G. (1990) Compositional patterns in the nuclear genome of coldblooded vertebrates. J. Mol. Evol. 31, 265–281.
[4] D'Onofrio, G., Jabbari, K., Musto, H., Alvarez-Valin, F., Cruveiller, S. and Bernardi, G. (1999) Evolutionary genomics of vertebrates and its implications. Ann. NY Acad. Sci. 870, 81–94.
[5] Bernardi, G. (2004). Structural and Evolutionary GenomicsNatural Selection in Genome Evolution, Elsevier, Amsterdam, pp. 434.
[6] Bernardi, G., Mouchiroud, D. and Gautier, C. (1988) Compositional patterns in vertebrate genomes: conservation and change in evolution. J. Mol. Evol. 28, 7–18.
[7] Bernardi, G. (2007) The neoselectionist theory of genome evolution. Proc. Natl. Acad. Sci. USA 104, 8385–8390.
[8] Bernardi, G., Olofson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F. (1985) The mosaic genome of warmblooded vertebrates. Science 228, 953–958.
[9] Aissani, B., D'Onofrio, G., Mouchiroud, D., Gardiner, K., Gautier, C. and Bernardi, G. (1991) The compositional properties of human genes. J. Mol. Evol. 32, 493–503.
[10] Mouchiroud, D., D'Onofrio, G., Aissani, B., Macaya, G., Gautier, C. and Bernardi, G. (1991) The distribution of genes in the human genome. Gene 100, 181–187.
[11] Clay, O., Caccio, S., Zoubak, S., Mouchiroud, D. and Bernardi, G. (1996) Human coding and noncoding DNA: compositional correlations. Mol. Phylogenet. Evol. 5, 2–12.
[12] Vinogradov, A.E. (2001) Bendable genes of warm-blooded vertebrates. Mol. Biol. Evol. 18, 2195–2200.
[13] Vinogradov, A.E. (2003) DNA helix: the importance of being GC-rich. Nucleic Acids Res. 31, 1838–1844.
[14] Duret, L., Mouchiroud, D. and Gauthier, C. (1995) Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J. Mol. Evol. 40, 308–317.
[15] Eyre-Walker, A.C. (1991) An analysis of codon usage in mammals: selection or mutation bias? J. Mol. Evol. 33, 442–449.
[16] Kotlar, D. and Lavner, Y. (2006) The action of selection on codon bias in the human genome is related to frequency, complexity, and chronology of amino acids. BMC Genom. 7, 67.
[17] Lercher, M.J., Urrutia, A.O. and Hurst, L.D. (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat. Genet. 31, 180–183.
[18] Vinogradov, A.E. (2003) Isochores and tissue-specificity. Nucleic Acids Res. 31, 5212–5220.

[19] Vinogradov, A.E. (2005) Dualism of gene GC content and CpG pattern in regard to expression in the human genome: magnitude versus breadth. Trends Genet. 21, 639–643.
[20] Goncalves, I., Duret, L. and Mouchiroud, D. (2000) Nature and structure of human genes that generate retropseudogenes. Genome Res. 10, 672–678.
[21] Arhondakis, S., Auletta, F., Torelli, G. and D'Onofrio, G. (2004) Base composition and expression level of human genes. Gene 325, 165–169.
[22] Semon, M. and Duret, L. (2004) Evidence that functional transcription units cover at least half of the human genome. Trends Genet. 20, 229–232.
[23] Federico, C., Andreozzi, L., Saccone, S. and Bernardi, G. (2000) Gene density in the Giemsa bands of human chromosomes. Chromosome Res. 8, 737–746.
[24] Saccone, S., Federico, C. and Bernardi, G. (2002) Localization of the generichest and the gene-poorest isochores in the interphase nuclei of mammals and birds. Gene 300, 169–178.
[25] Federico, C., Scavo, C., Cantarella, C.D., Motta, S., Saccone, S. and Bernardi, G. (2006) Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates. Chromosoma 115, 123–128.
[26] Chiusano, M.L., D'Onofrio, G., Alvarez-Valin, F., Jabbari, K., Colonna, G. and Bernardi, G. (1999) Correlations of nucleotide substitution rates and base composition of mammalian coding sequences with protein structure [In Process Citation]. Gene 238, 23–31.
[27] Alvarez-Valin, F., Lamolle, G. and Bernardi, G. (2002) Isochores, GC3 and mutation biases in the human genome. Gene 300, 161–168.
[28] Galtier, N. and Duret, L. (2007) Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. 23, 273–277.
[29] Press, W.H. and Robins, H. (2006) Isochores exhibit evidence of genes interacting with the large-scale genomic environment. Genetics 174, 1029–1040.
[30] Vinogradov, A.E. (2005) Noncoding DNA, isochores and gene expression: nucleosome formation potential. Nucleic Acids Res. 33, 559–563.
[31] Grillo, G., Attimonelli, M., Liuni, S. and Pesole, G. (1996) CLEANUP: a fast computer program for removing redundancies from nucleotide sequence databases. Comput. Appl. Biosci. 12, 1–8.
[32] Tatusov, R.L. et al. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29, 22–28.
[33] Tatusov, R.L. et al. (2003) The COG database: an updated version includes eukaryotes. BMC Bioinf. 4, 41.
[34] Costantini, M., Clay, O., Federico, C., Saccone, S., Auletta, F. and Bernardi, G. (2007) Human chromosomal bands: nested structure, high-definition map and molecular basis. Chromosoma 116, 29–40.
[35] Duret, L., Mouchiroud, D. and Gouy, M. (1994) HOVERGEN: a database of homologous vertebrate genes. Nucleic Acids Res. 22, 2360–2365.
[36] Galtier, N., Piganeau, G., Mouchiroud, D. and Duret, L. (2001) GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159, 907–911.
[37] Ptak, S.E. et al. (2005) Fine-scale recombination patterns differ between chimpanzees and humans. Nat. Genet. 37, 429–434.
[38] Lavner, Y. and Kotlar, D. (2005) Codon bias as a factor in regulating expression via translation rate in the human genome. Gene 345, 127–138.
[39] Kimura, M. (1983) The Neutral Theory of Molecular Evolution, Cambridge Univ. Press, UK, Cambridge.