J Mol Evol (1993) 37:583–589

Journal of Molecular Evolution

Silent Substitutions in Mammalian Genomes and Their Evolutionary Implications

Giorgio Bernardi,[1] Dominique Mouchiroud,[2] Christian Gautier[2]

[1] Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France
[2] Laboratoire de Biométrie, Génétique et Biologie des Populations, URA 243, Université Claude Bernard, 69600 Villeurbanne, France

Received: 20 July 1993 / Accepted: 15 September 1993

Abstract. An analysis of silent substitutions in pairwise comparisons of homologous genes from different mammals has shown that, in spite of individual fluctuations, their frequencies (which are very strongly correlated with the frequency of substitutions per synonymous site calculated according to Li et al. 1985) do not vary, on the average, with the GC levels of silent positions. This holds in the general case, in which silent positions of pairs of homologous genes share the same composition, namely in the human/other primates, human/artiodactyls, and in the mouse/rat pairs, as well as in the special cases in which the composition of silent positions are different, namely in the human/rabbit and the human/rat (or human/mouse) pairs. A slightly lower frequency found for low GC values in the human/bovine and human/pig pairs seems to be due to the specific gene samples used. These results contradict the previously claimed existence of differences in mutation rates and of mutational biases in third codon positions of coding sequences located in different isochores of mammalian genomes. They also imply that the variations in nucleotide precursor pools through the cell cycle and the differences in replication timing, or in repair efficiency, which were reported for different isochores, do not lead, as claimed, to differences in mutation rates, not in mutational biases in mammals. The differences claimed appear to be due to using small gene samples when individual fluctuations from gene to gene are relatively large.

Key words: Isochores — Mutation rates — Mammals — DNA replication — DNA repair

Introduction

The mammalian genome is a mosaic of isochores, long DNA segments (>300 Kb on the average) that are remarkably homogeneous in base composition and that can be subdivided into a small number of families characterized by different GC levels (Bernardi et al 1985; Bernardi 1989, 1993a,b). In the human genome, which is representative of the majority of mammalian genomes (Sabeur et al. 1993, Mouchiroud et al. 1993), isochores cover a broad GC range, 30–60% (Mouchiroud et al. 1991). The silent codon positions of human genes are compositionally correlated with the isochores in which the corresponding genes are located (Bernardi et al. 1985, Aïssani et al. 1991), but they cover a much broader GC range, 25–97.5% (Mouchiroud and Bernardi 1993). Under these circumstances, a pertinent question is whether substitution rates, and particularly silent substitutions rates, are the same or vary over different isochore families and over the genes contained in them. This question, which has important implications for genome evolution, was already investigated in the past with conflicting results.

Indeed, previous reports indicated (i) that synonymous substitutions are approximately uniform in

Correspondence to: G. Bernardi

rates for different pairs of mammalian genes (Miyata et al. 1982); (ii) that silent substitution rates are independent of the composition of third codon positions in primate genes (Filipski 1988; this also appears to be the case for the human/Old World monkeys comparisons made by Wolfe et al. 1989); (iii) that silent positions in GC-poor rodent sequences mutate four times faster than their human equivalents or than GC-rich rodent sequences (Filipski 1988); (iv) that homologous human/rat genes exhibit increasing silent substitution rates with increasing A and T (Ticher and Graur 1989); (v) that the rate of silent substitutions varies among murid genes, that it is correlated with the base composition of genes and their flanking DNA, and that it shows a peak at 50% GC (Wolfe et al. 1989); (vi) that homologous mammalian genes do not exhibit significant silent substitution rate differences in the absence of differences in third codon position GC, whereas they appear to show higher rates in the presence of such differences (Saccone et al. 1989).

Here we have examined silent substitutions in pairwise comparisons of homologous genes from different mammals using much larger samples than those previously studied and we have unambiguously shown that their frequencies do not vary, on the average, with the GC levels of the corresponding codon positions. This holds both in the general case, in which silent positions of pairs of homologous genes share the same composition, as well as in the special case in which the composition of silent positions is different in pairs of homologous genes. A slightly lower substitution frequency observed in the GC-poor range of the human/bovine and human/pig comparison appears to be due to the particular gene sample investigated. The fact that the average mutation rate and mutational bias do not depend upon GC has important implications which are discussed.

Materials and Methods

All analyses presented here concern pairwise comparisons of homologous sequences. These sequences were retrieved from GenBank (Bilofsky et al. 1986) Release 76 (March 1993) using the data base management system ACNUC (Gouy et al. 1985). The procedure used to select the homologous pairs was described by Mouchiroud and Gautier (1990). Comparisons comprised the mammalian genome pairs for which a large enough number of homologous sequences was available, namely the human/other primates, the human/artiodactyls, the human/rabbit, the mouse/rat and the human/rat pairs (Mouchiroud and Bernardi 1993). The features exhibited by human/mouse homologous gene pairs were similar to those of the human/rat pairs and will not be presented here.

To quantify dissimilarity between homologous sequences, the silent difference frequency (SDF) was used. SDF is the percentage divergence in third codon positions and does not rely on any hypothesis concerning the nature of the substitution process in contrast with Ks, the frequency of substitutions per synonymous site, as calculated according to Li et al. (1985). In every case, the mean SDF values were estimated not only for all points, but also for those characterized by low (<50%), intermediate (50–77%), and high (>77%) GC levels, in order to detect possible differences. These GC values roughly correspond to the borders between genes located in L1, L2-H2, and H3 human isochore families, respectively (see Mouchiroud et al. 1991).

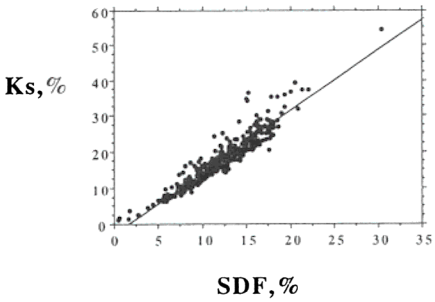

Fig. 1. Relationship between the frequency of substitutions per synonymous site (calculated according to Li et al. 1985) and the silent divergence frequency, SDF (the percentage divergence of third codon positions), for homologous gene pairs from the rat and mouse genomes.

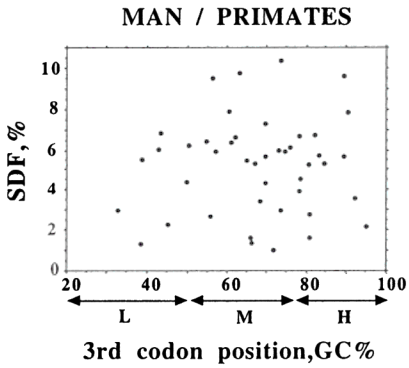

Fig. 2. SDF between homologous genes of man and other primates is plotted against GC of third codon positions of human genes. For correlation coefficients and average SDF values for the L (<50% GC) M (50–77% GC) and H (>77% GC) sections, see Table 1.

Results

Before presenting our results on silent difference frequencies (SDF), it should be stressed that a strong linear correlation exists between SDF and classical estimations (for instance, Li et al. 1985) of silent substitutions per synonymous site (Fig. 1). Identical conclusions can be drawn from SDF and K4 (the frequency of substitutions per fourfold degenerate site) in the case of the rat/mouse comparison (see Fig. 5).

Plots of frequency of substitution in third codon positions (SDF) against GC levels of those positions show no significant correlation in the comparisons man/other primates (Fig. 2), man/sheep (Fig. 3), man/rabbit (Fig. 4), rat/mouse (Fig. 5), and man/rat (Fig. 6). Such lack of correlation was therefore

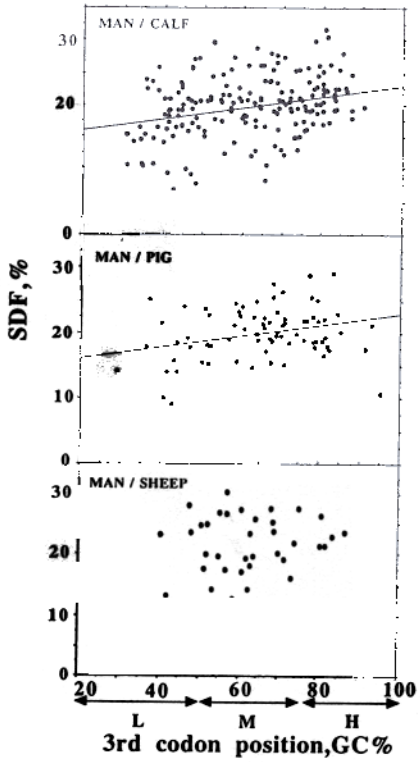

Fig. 3. SDF between homologous genes of man and artiodactyls (calf, pig, and sheep) is plotted against GC of third codon positions of human genes. Other indications as in Fig. 2.

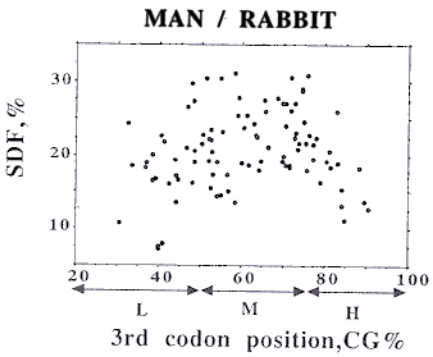

Fig. 4. SDF between homologous genes of man and rabbit is plotted against GC of third codon positions of human genes. Other indications as in Fig. 2.

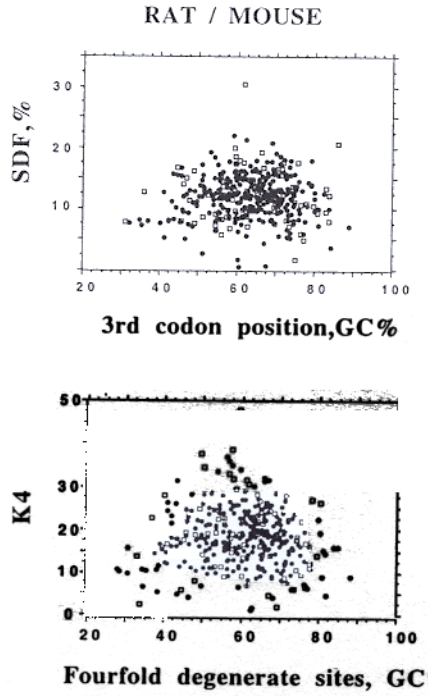

Fig. 5. **(Top)** SDF between homologous genes of mouse and rat is plotted against the average GC level of third codon positions. Black points correspond to coding sequences longer than 180 codons. Other indications as in Fig. 2. **(Bottom)** K4, the frequency of substitutions per fourfold degenerate sites, is plotted against the average GC level of those sites. Black points correspond to coding sequences longer than 180 codons.

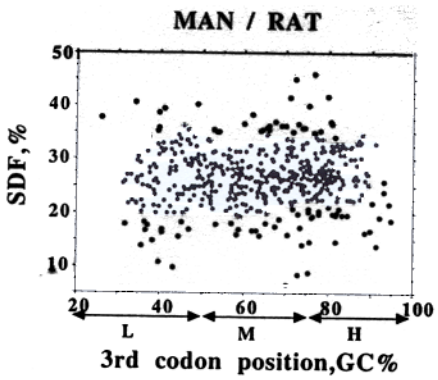

Fig. 6. SDF between homologous genes of man and rat is plotted against GC of third codon position of human genes. Other indications as in Fig. 2.

found both in the cases characterized by a conserved base composition in third codon positions and in those characterized by a different base composition. A correlation which is weak, yet significant (Fig. 3), was found in the comparisons man/calf and man/pig, in which SDF showed a slight increase with increasing GC. This increase appears, however, to be due to slightly lower SDF values for low GC third codon positions (see Table 1). Since this phenomenon is not found in any other case, and in particular in the two comparisons involving the largest number of genes, the mouse/rat and the man/rat comparisons, the only explanation which can be offered at present is that it is due to the particular, small sample of genes present in the low GC section of the plot. This explanation is supported by the finding that, when 58 homologous genes from man, calf, and rat were studied, a result similar to that of Fig. 3 was also obtained for the man/rat pair (not shown). A similar, yet weaker, phenomenon is probably responsible for the higher SDF value of the middle GC section of the man/rabbit comparison (see Table 1).

Separate analyses of coding sequences charac-

Table 1. Statistical analysis of silent difference frequencies in mammalian genes

Plots	Man/primate	Man/calf	Man/sheep	Man/pig	Man/rabbit	Rat/mouse	Man/rat
Number of genes	44	194	39	79	97	386	523
SDF vs. GCIII(Man)[a]	R = 0.07	R = 0.29**[b]	R = 0.16	R = 0.29**	R = 0.16	R = 0.045	R = 0.03
Average SD							
All sequences	5.2 (2.3)[c]	19.7 (4.9)	21.4 (4.8)	19.9 (4.3)	20.8 (5.3)	12.5 (5.3)	26.8 (5.4)
L (<50% GC)	4.2 (2.1)	17.2 (5.3)	19.7 (6.3)	16.5 (5.1)	18.2 (5.8)		26.7 (6.1)
M (50–77% GC)	5.5 (2.5)	20.2 (4.5)	21.4 (4.8)	20.7 (3.6)	22.7 (4.6)		26.7 (5.3)
H (>77% GC)	5.2 (2.3)	21.8 (4.1)	22.6 (3.7)	20.6 (4.4)	17.8 (4.1)		27.1 (4.9)[d]

[a] Third codon position GC of human sequences. In the mouse/rat comparison, the average between the third codon position GC of the two sequence pairs was used.
[b] Asterisks refer to statistical significancy (*5%, **1%).
[c] Values in parentheses are standard deviations.

terized by high or low degrees of conservation did not show any significant difference (not shown).

Discussion

This discussion will be divided in three parts. First, we discuss the previous reports in the light of our present findings. It should be stressed that the comparison between our data, which are percentages of third codon position differences, with the data of other authors, which concern silent substitution rates, is a valid one in view of the strong correlation between the frequency of silent substitutions per synonymous site and SDF (Fig. 1) and of the comparison of SDF and K4 plots (Fig. 5). Then we reanalyze a series of interpretations claimed to account for previous reports. Finally, we discuss the general issues.

The Frequencies of Silent Substitutions Do Not Differ Among Regions of the Mammalian Genome

The results of Figs. 2–6, concerning all pairwise comparisons involving a large number of genes, unambiguously show that, while the numbers of silent differences exhibit relatively large fluctuations in different genes, they do not show any significant trend over the very extended range of third codon position GC under consideration. The above results raise the question why discrepant results on the dependence of silent nucleotide substitution rates upon silent position GC were previously found. Five conflicting sets of results reported in the literature will be commented upon in the following paragraphs.

(i) The report by Smithies et al. (1981), that base substitutions in the 5 Kb duplication that led to the human Gγ and Aγ fetal globin genes correlate positively with the local GC level, really concerns short-segment variations. This is a problem different from that under consideration here and will, therefore, not be discussed.

(ii) A second set of data by Miyata et al. (1982), showing a lack of differences in silent substitution rates, is in apparent agreement with the present results. This failed, however, to prove the point made in this work because of the small size of the samples studied. Indeed, the data only concern a total of 17 comparisons of genes from man, rat, rabbit, and monkey. Although some of the coding sequences investigated did show differences in silent position GC levels, the lack of differences in silent mutation rates may well be due to the small size of the sample studied, the largest pairwise comparison, man/rat, comprising only nine genes. No definite conclusion can, therefore, be drawn from those data (and none was drawn by the authors, who were, in fact, not concerned with this problem). Indeed, if a systematic variation existed, it would have been missed. The same conclusion applies to the 12 comparisons of primate genes made by Filipski (1988), in which no correlation was found, and also to the 13 human/Old World monkeys comparisons made by Wolfe et al. (1989).

(iii) The third set of results concerns the rat-mouse comparison in which variations of mutation rates were reported, even if the results were different in two series of data. Indeed, in one case a strong increase with decreasing GC was found (Filipski 1988), whereas in another one a peak at 50% GC was reported (Wolfe et al. 1989). It is clear that these discrepant results were both due to the fact that the sizes of the gene samples used (30 genes in the work by Filipski 1988; 23 large genes in the work of Wolfe et al. 1989) were still too small. Indeed, the variation in silent divergence being relatively large even for genes having exactly the same GC levels in silent codon positions, any correlation may be found when using a small sequence sample.

(iv) Ticher and Graur (1989) reported a correlation between silent substitutions rate and the percentage of different nucleotides at silent positions. This correlation was positive for A and T, negative for C and nonsignificant for G. It concerned 42 homologous genes from man and rat having GC levels

in third codon positions higher than 45%. This correlation could not be confirmed in the present work.

(v) The last set of data (Saccone et al. 1989) indicated no significant rate difference for 17 human/artiodactyl gene pairs that showed no silent position GC differences. This is, however, again a small sample from which no general rule can be drawn (see (iii) above). Human or artiodactyl/rodent comparisons showed that some 9 pairs of genes, with a small or no difference in silent position GC, exhibited lower rates compared to another 9 or so pairs of genes, which showed large differences. The significance of these differences in rates is, however, doubtful in view of the present results on the human versus rat (or mouse) comparisons. Indeed, since in such a case differences in silent position GC exist for both GC-poor and GC-rich genes (Mouchiroud et al. 1988, Mouchiroud and Gautier 1991, Mouchiroud and Bernardi 1993), one should notice higher numbers of substitutions for those "extreme" genes compared to genes having a more balanced composition, which is not the case.

Differences in Repair Efficiency Do Not Affect the Rates nor the Biases of Silent Mutations

The higher rate of accumulation of mutation in GC-poor sequences in rodents compared to primates, as well as their compositional bias (claimed by Filipski 1988), was explained as due to less efficient DNA repair in these regions of the genome. Obviously, the inexistence of such higher rate does not question the existence of a less efficient repair of DNA lesions in rodent compared to human cells (Hart and Setlow 1974), nor the evidence for between-gene differences in efficiency of DNA repair (Bohr et al. 1987). The lack of rate difference indicates, therefore, that such less efficient or differential repairs do not influence, on the average, the silent substitutions in rates nor in biases, even if between gene differences can result from fluctuations. This is an important conclusion, because DNA repair has been repeatedly considered to be a cause for changes and biases in the mutation process (Filipski 1987, Sueoka 1988, 1992, 1993).

Differences in Process of Mutation Associated with Replication Timing Do Not Affect Rates nor Biases of Silent Substitutions and Are Not Responsible for the Origin of Isochores

The main conclusions drawn by Wolfe et al. (1989) was "that much of the intragenomic variation in silent substitution rate and base composition in mammals results from variation in the process of mutation, rather than from natural selection (Bernardi and Bernardi 1986, Gillespie 1986)." According to Wolfe et al. (1989), "the variation in both silent substitution rate and base composition" (which they consider as "two facets of the same phenomenon") "is due to systematic differences in the rate and pattern of mutation over regions of the genome, the differences arising because mutation patterns vary with the timing of replication of different chromosomal regions in the genome." Indeed, "most germline mutations are thought to arise from misincorporation errors made by the DNA replication apparatus (Friedberg 1985, Topal and Fresco 1976). It has been demonstrated that different genes replicate at different stages of the cell cycle in different cells (Holmquist 1987) and this is presumably also true in the germline. The number and type of replication errors are likely to vary during the cell cycle if the chemical environment in the nucleus changes. In fact, the abundances (both relative and absolute) of free dNTPs in the nucleus change with time (Leeds et al. 1985), as do the activities of the DNA polymerase enzymes and their accessory proteins (Kelly and Stillman 1988)" . . . "as S-phase progresses, the G + C content of both dNTP pools and the replicating DNA decreases (Holmquist et al. 1982, 1987, Fersht and Kwill-Jones 1981)." Wolfe et al. (1989), therefore, "propose that isochores arise as a result of the synchronous replication of megabase stretches of DNA under varying dNTP pool conditions."

The fact that the conclusion of Wolfe et al. (1989) "that the substitution rate and the base composition of silent sites vary together in a systematic way" is wrong has two important consequences. First, if changes in the nucleotide pools in the germline do exist (as assumed on the basis of what happens in somatic cells), the fact that mutation patterns do not vary with the timing of replication (because mutation rates are the same on the average for genes which replicate early or late) means that changes in nucleotide pools do not cause biases in mutation patterns. In fact, it was already pointed out that in the somatic tissues of mammals, late replicating DNA, like satellites and the inactive X chromosome, may be both GC-rich and GC-poor (Bernardi et al. 1988), and it has been recently shown that early and late replicating genes may be both GC-poor and GC-rich (Eyre-Walker 1992). This finding can be understood because of the presence of GC-poor isochores in R-bands which replicate early and of GC-rich isochores in G-bands which replicate late (Gardiner et al. 1990, Pilia et al. 1993; Saccone et al. 1992, 1993). Moreover, such early and late replication patterns also exist in cold-blooded vertebrates (see Bernardi 1989, and papers quoted therein), which never developed strong compositional differences in their genomes (Bernardi and Bernardi 1990a,b, 1991).

Second, if mutations patterns do not vary with the timing of replication of different chromosomal regions in the germline, the explanation of Wolfe et al. (1989) "for both the origin of isochores in mammalian genomes and the observation that silent nucleotide substitutions in different mammalian genes (Li et al. 1987) do not have the same molecular clock" does not hold anymore.

In conclusion, the differences in mutation rates and in mutational biases in different isochore families of mammalian genomes appear to be due to two reasons: (i) the existence of relatively large individual fluctuations from gene to gene; and (ii) the use of small, non-representative gene samples. Under these circumstances, explanations other than differences in mutation rates and in mutational biases have to be taken into consideration in order to account for the large differences in GC levels of third codon positions of genes located in different isochore families from mammalian genomes (see Bernardi and Bernardi 1986, Bernardi et al. 1988, Bernardi 1993a, and paper in preparation).

References

Aïssani 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

Aota S, Ikemura T (1986) Diversity in G + C content at the third position of codons in vertebrate genes and its cause. Nucleic Acids Res 14:6345–6355

Bernardi G, Oloffson 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

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

Bernardi G (1989) The isochore organization of the human genome. Ann Rev Genet 23:637–661

Bernardi G (1993a) The vertebrate genome: Isochores and evolution. Mol Biol Evol 10:186–204

Bernardi G (1993b) The isochore organization of the human genome and its evolutionary history: A review. Gene (in press)

Bilofsky HS, Burks C, Fickett JW, Goad WB, Lewitter FI, Rindone WP, Swindell LD, Tung CS (1986) The Genbank (R) genetic sequence database. Nucleic Acids Res 14:1–4

Bohr VA, Philips DH, Hanawalt PC (1987) Heterogeneous DNA damage and repair in the mammalian genome. Cancer Res 47:6426–6436

Efstratiadis A, Posakony JW, Maniatis T, Lawn RL, O'Connell C, Spritz RA, Deriel JK, Forget B, Weissman SM, Slightom JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot NJ (1980) The structure and evolution of the human β-globin gene family. Cell 21:653–668

Eyre-Walker A (1992) Evidence that both G + C-rich and G + C-poor isochores are replicated early and late during the cell cycle. Nucleic Acids Res 20:1497–1501

Fersht AR, Knill-Jones JW (1981) DNA polymerase accuracy and spontaneous mutation rates: Frequencies of purine-purine, purine-pyrimidine, pyrimidine-pyrimidine mismatches during DNA replication. Proc Natl Acad Sci USA 78:4251–4255

Filipski J (1987) Correlation between molecular clock ticking, codon usage, fidelity of DNA repair, chromosome banding and chromatin compactness in germline cells. FEBS Lett 217:184–186

Filipski J (1988) Why the rate of silent codon substitutions is variable within a vertebrate's genome. J Theor Biol 134:159–164

Friedberg EC (1985) DNA repair. Freeman, New York

Gardiner K, Aissani B, Bernardi G (1990) A compositional map of human chromosome 21. EMBO J 9:1853–1858

Gillespie JH (1986) Variability of evolutionary rates of DNA. Genetics 113:1077–1091

Gouy M, Gautier C, Attimonelli M, Lanave C, di Paola G (1985) ACNUC—A portable retrieval system for nucleic acid sequence databases: Logical and physical designs and usage. CABIOS 1:167–172

Hart RW, Setlow RB (1974) Correlation between deoxyribonucleic acid excision-repair and life-span in a number of mammalian species. Science 71:2169–2173

Holmquist G, Gray M, Porter T, Jordan J (1982) Characterization of Giemsa dark- and light-band DNA. Cell 31:121–129

Holmquist G (1992) Review article: Chromosome bands, their chromatin flavors and their functional features. Am J Hum Genet 51:17–37

Kelly T, Stillman B (eds) (1988) Eukaryotic DNA replication. Cancer Cell 6: Cold Spring Harbor Laboratory, New York

Leeds JM, Slabaugh MB, Mathews CK (1985) DNA precursor pools and ribonucleotide reductase activity: Distribution between the nucleus and cytoplasm of mammalian cells. Mol Cell Biol 5:3443–3450

Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Miyata T, Hayashida H, Kikuno R, Hasegawa M, Kobayashi M, Koike K (1982) Molecular clock of silent substitution: At least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J Mol Evol 19:28–35

Mouchiroud D, Gautier C (1988) High codon-usage changes in mammalian genes. Mol Biol Evol 5:192–194

Mouchiroud D, Gautier C, Bernardi G (1988) The compositional distribution of coding sequences and DNA molecules in humans and murids. J Mol Evol 27:311–320

Mouchiroud D, Gautier C (1990) Codon usage changes and sequence dissimilarity between human and rat. J Mol Evol 31: 81–91

Mouchiroud D, D'Onofrio G, Aïssani B, Macaya G, Gautier C, Bernardi G (1991) The distribution of genes in the human genome. Gene 100:181–187

Mouchiroud D, Bernardi G (1993) Compositional properties of coding sequences and mammalian phylogeny. J Mol Evol 37: 109–116

Nghiem Y, Cabrera M, Cupples CG, Miller JH (1988) The *mutY* gene: A mutator locus in *Escherichia coli* that generates G.C-T.A transversions. Proc Natl Acad Sci USA 85:2709–2713

Pilia G, Little RD, Aïssani B, Bernardi G, Schlessinger D (1993) Isochores and CpG islands in YAC contigs in human Xq26.1-qter. Genomics 17:456–462

Sabeur G, Macaya G, Kadi F, Bernardi G (1993) The isochore patterns of mammalian genomes and their phylogenetic implications. J Mol Evol 37:93–108

Saccone C, Pesole G, Preparata G (1989) DNA microenvironments and the molecular clock. J Mol Evol 29:407–411

Saccone S, De Sario A, Della Valle G, Bernardi G (1992) The highest gene concentrations in the human genome are in T-bands of metaphase chromosomes. Proc Natl Acad Sci USA 89:4913–4917

Saccone S, De Sario A, Wiegant J, Rap AK, Della Valle G, Bernardi G (1993) Correlations between isochores and chromosomal bands in the human genome. Proc Natl Acad Sci USA, in press

Smithies O, Engels WR, Devereux JR, Slightom JL, Shen S-H (1981) Base substitutions, length differences and DNA strand asymmetries in the human Gγ and Aγ fetal globin gene region. Cell 26:345–353

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA 85:2653–2657

Sueoka N (1992) Directional mutation pressure and molecular evolution: Equilibria and asymmetric phylogenetic branching. J Mol Evol 34:95–114

Sueoka N (1993) Directional mutation pressure and dynamics of molecular evolution. J Mol Evol 37:137–153

Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1:269–285

Ticher A, Graur D (1989) Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. J Mol Evol 28:286–298

Topal MD, Fresco JR (1976) Complementary base pairing and the origin of substitution mutations. Nature 263:285–289

Wolfe KH, Sharp PM, Li, WH (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285

Wu T-H, Clarke CH, Marinus MG (1990) Specificity of *Escherichia coli mutD* and *mutL* mutator strains. Gene 87:1–5