J Mol Evol (1995) 40:280–292

255

JOURNAL OF MOLECULAR EVOLUTION

Nonrandom Frequency Patterns of Synonymous Substitutions in Homologous Mammalian Genes

Simone Cacciò, Serguei Zoubak,* Giuseppe D'Onofrio, Giorgio Bernardi**

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France

Received: 15 July 1994

Abstract. All 69 homologous coding sequences that are currently available in four mammalian orders were aligned and the synonymous positions of quartet and duet (fourfold and twofold degenerate) codons were divided into three classes (that will be called conserved, intermediate, and variable) according to whether they show no change, one change, or more than one change, respectively. We observed (1) that the frequencies of conserved, intermediate, and variable positions of quartet and duet codons are different in different genes; (2) that the frequencies of the three classes are significantly different from expectations based on a random substitution process in the majority of genes (especially for GC-rich genes) for quartet codons and in a minority of genes for doublet codons; and (3) that the frequencies of the three classes of positions of quartet codons are correlated with those of duet codons, the conserved positions of quartet and duet codons being, in addition, correlated with the degree of amino acid conservation. Our main conclusions are that synonymous substitution frequencies: (1) are gene-specific; (2) are not simply the result of a stochastic process in which nucleotide substitutions accumulate at random, over time; and (3) are correlated in quartet and duet codons.

Key words: Synonymous substitutions – Homologous mammalian genes

Introduction

Recent investigations have shown that the frequencies of synonymous substitutions in mammals cover a wide range, are gene-specific, and are correlated with the frequencies of nonsynonymous substitutions (Mouchiroud et al. 1995). These conclusions were based on comparisons of frequencies (1) of synonymous substitutions, as determined on homologous genes from two different pairs of mammals; and (2) of synonymous and nonsynonymous substitutions, as determined on the same genes.

Here, we investigated the frequency patterns of synonymous substitutions in homologous mammalian genes. Basically, we aligned all 69 homologous coding sequences that are currently available in four mammalian orders and studied the frequencies of conserved, intermediate, and variable synonymous positions (as defined by the presence of no change, one change, or more than one change, respectively) from quartet and duet (fourfold and twofold degenerate) codons. We then compared the frequencies found in quartet and duet codons with those expected if the synonymous substitutions that took place between the (reconstructed) ancestral and the present-day sequences were random in their location in quartet or duet codons. Moreover, we studied the correlations between the frequencies of the three classes of positions of

* *Permanent address:* Institute of Molecular Biology and Genetics, Zabolotnogo str. 150, 252627 Kiev, Ukraine

** *Permanent address:* Stazione Zoologica, Villa Comunale 1, 80121 Naples, Italy

Correspondence to: G. Bernardi

quartet and duet codons, as well as those of the frequencies of conserved positions of quartet and duet codons with those of conserved amino acids.

Methods

The 69 genes studied in this work included all the complete orthologous coding sequences available in four mammalian orders, primates, artiodactyls, rodents (murids), and lagomorphs. However, twenty genes from carnivores and eight genes from perissodactyls were also used to make a four-order comparison possible, when some sequences were missing in the four orders mentioned above (this mainly concerned genes missing in lagomorphs). Only a very small number of genes (nine) were represented in more than four orders. Apart from a few genes initially chosen on the basis of common knowledge, the search for orthologous genes was done using the HOVERGEN program (Duret et al. 1994) and GenBank release 83 (June 1994). The mnemonics of the genes investigated are available upon request.

Table 1 lists the number of codons analyzed, the corresponding GC_3 values (the average GC levels of third codon positions from the analyzed codons) of homologous genes from four orders (one species per order; including murids) or from three orders (excluding murids), and the levels of amino acid conservation in the encoded proteins (as calculated from the analyzed codons). Codons excluded from the analysis comprised initiation and termination codons, codons showing deletions, and codons for methionine and tryptophan. In this and in the following paper (Zoubak et al. 1995) genes are always listed in the order of decreasing GC_3 from the gene set not including murids.

Several genes were not included in the analyses because of doubtful orthology (P450 IID, P450 A1, and interleukins) and/or problems in determining reliable sequence alignments (β- and κ-caseins, α-lactalbumin), because of their small sizes (colipase, insulin, interferon γ-3, phospholamban, protamine 1), or because sequences were incomplete (lactoferrin/transferrin, interphoto receptor-binding protein, IRBP).

Homologous genes from different orders exhibited close GC_3 values, as expected from previous work (Mouchiroud et al. 1987, 1988; Mouchiroud and Bernardi 1993), with the exception of genes from murids, in which case genes having extreme compositions were characterized by a compositional shift in third codon positions compared to genes from mammals exhibiting the general pattern (Salinas et al. 1986; Zerial et al. 1986; Mouchiroud et al. 1988; Bernardi et al. 1988; Mouchiroud and Gautier 1990; Sabeur et al. 1993; Mouchiroud and Bernardi 1993). This "minor shift" (so-called in order to distinguish it from the "major shift" that took place at the transition between cold- and warm-blooded vertebrates; see Bernardi et al. 1985; Bernardi and Bernardi 1990a,b, 1991) is responsible for the lower or higher GC_3 levels of murid genes in the high and low GC ranges, respectively, relative to their homologs from mammalian genomes exhibiting the general pattern. The effect of the minor shift is strong enough to be apparent even in the average values of Table 1.

Amino acid and nucleotide alignments were done by using the CLUSTAL program (Higgins and Sharp 1988). Nucleotide alignments of coding sequences were deduced from the alignments of the corresponding proteins.

The analysis of genes was performed using a program developed in order to distinguish quartet and duet codons (isoleucine codons being neglected and sextet codons being counted separately as quartet and duet codons), as well as conserved, intermediate, and variable positions (as defined below). The results of this analysis are exemplified by the alignments of Fig. 1, which concern the β-globin gene.

Although our analysis was centered on synonymous quartet codons, synonymous duet codons were also investigated. Interorder comparisons were routinely done using four orders and coding sequences from one species per order (the same species for different genes, whenever possible); in a few cases, all orders available were analyzed.

The synonymous divergence (or synonymous difference frequency, SDF; Mouchiroud and Gautier 1990) was calculated (as in Bernardi et al. 1993; and in Mouchiroud et al. 1995) as the percentage of synonymous codons that are different in third positions of aligned sequences. SDF_2 and SDF_4, the synonymous divergences occurring in duet and quartet codons, respectively, were also taken into consideration.

Synonymous positions from quartet and duet codons were classified as follows: (1) conserved positions, showing no change in the comparisons made; (2) intermediate positions, showing a single difference; and (3) variable positions, showing more than one difference.

In order to decide whether the synonymous substitution frequencies deviated significantly from those expected for a random process, the percentages of each class of positions actually found in the coding sequences studied were compared with expectations based on a random substitution process taking place between the "ancestral" (consensus) and the present-day sequence. The crucial point of this "randomization" is that it was done simply by reshuffling among synonymous positions from duet and quartet codons the nucleotides that were changed in the present day (actual) compared to the "ancestral" sequences (Zoubak et al. 1995).

Results

The Conserved, Intermediate, and Variable Synonymous Positions from Different Homologous Genes Show Different Frequencies

Table 2 presents the percentages of each class of positions for both quartet and duet codons of homologous genes from four orders (one species per order) including murids. Figure 2 displays the data for quartet and duet codons of Table 2 in the form of histograms. These results indicate that conserved, intermediate, and variable positions of quartet and duet codons show wide and different ranges (defined as the ratios of highest to lowest frequencies). In quartet and duet codons these ranges are 28–75%, 16–50%, 3–40%, and 45–82%, 14–50%, 2–18%, respectively. The threefold ranges of conserved and intermediate positions of quartet codons are remarkable if one considers that only four orders were compared, leading to generally high values of those positions (Fig. 2). More remarkable still was the 13-fold range of variable positions in quartet codons. Ranges in duet codons were less extended that those in quartet codons, except for intermediate positions.

A *t*-test showed that the average values for quartet codons were significantly different from those expected on the basis of a random nucleotide substitution process operating between the "ancestral" (consensus) and the present-day sequences, P values being lower than 0.001 for both the intermediate and variable classes, but only lower than 0.1 for the conserved class. In contrast in the case of duets, only average values for intermediate positions were significantly different ($P < 0.05$) from those expected. Average values for all different classes of both duet and quartet codons were very significantly different from each other.

Table 1. Number of codons analyzed, the corresponding GC_3 and conserved amino acids in homologous genes from four mammalian orders[a]

Gene	Codons (number)	GC_3 (no murids)	GC_3 (murids)	Conserved aa (%)
1 Apo E	290	90.6	87.3	51.7
2 Creatin kinase B	366	90.5	87.6	93.2
3 A1 adenosine	310	88.6	86.0	89.0
4 H,K ATPase β subunit	277	86.8	83.5	74.4
5 α-globin	139	86.2	81.6	72.7
6 Apo A1	252	86.1	82.7	53.6
7 Na-H exchange protein	784	85.7	83.8	89.8
8 Serine pyruvate aa transferase	378	85.7	81.8	67.2
9 Dipeptidase	389	85.5	81.2	66.6
10 GMP-phosphodiesterase-α	811	85.1	79.7	88.4
11 CD8 α chain	220	83.0	78.2	41.8
12 Glutathione peroxidase	190	82.7	79.9	78.4
13 H,K ATPase α subunit	994	82.2	79.2	96.8
14 Retinol-binding proein	185	82.2	79.2	80.5
15 Glucose Glut3	469	82.0	79.7	93.4
16 Prostaglandin E receptor	311	81.9	81.9	77.5
17 Prolyl-4-hydroxylase β	493	80.6	76.0	86.0
18 Growth hormone	205	80.2	79.3	61.0
19 TNFα	226	80.1	77.9	65.0
20 Ferritin L	169	79.8	77.7	77.5
21 Myoglobin	147	79.4	76.9	71.4
22 Apo CIII	89	79.2	76.3	38.2
23 TNF β	189	75.5	73.6	65.6
24 Hydrophobic-surfactant-associated factor	177	75.1	72.6	63.8
25 Phospholipase A2	141	74.5	75.0	64.5
26 Phenyl tRNA ligase	454	74.2	72.4	80.0
27 Tissue inhibitor of metalloproteinase	195	74.2	70.9	63.1
28 Guanine-nt-binding protein	383	73.0	73.8	99.2
29 Polymeric Ig receptor	723	71.7	68.6	39.1
30 Colony-stimulating factor	133	71.6	69.1	45.1
31 CD4 antigen	428	70.9	70.2	35.3
32 Erythropoietin	182	70.4	67.3	69.8
33 Gastrin	95	68.0	64.4	52.6
34 Na^+/nucleoside	567	68.0	68.6	78.8
35 D-amino-acid oxidase	328	67.1	66.2	68.3
36 Ferritin H	156	66.5	64.7	84.0
37 Protein kinase C	644	66.3	66.0	97.8
38 ANP	144	66.0	66.2	63.9
39 β-globin	141	65.3	65.4	71.6
40 Potassium channel	460	64.9	64.1	97.6
41 Cytochrome b5	127	64.7	63.4	77.2
42 Endothelin	193	62.7	62.7	54.4
43 Phagocytic glycoprotein I	341	62.3	61.2	72.1
44 Prolactin	168	62.3	59.5	51.8
45 Interleukin 2 receptor	250	60.1	59.3	38.8
46 Tissue factor	274	59.2	58.2	43.4
47 β2 microglobulin	114	59.1	57.6	52.6
48 Na-K ATPase β-1 subunit	292	57.1	59.4	86.0
49 CD3 ϵ antigen	180	55.3	57.2	50.0
50 Na-Ca exchange protein	935	54.2	54.7	93.7
51 Ca-ATPase	954	53.4	54.6	98.3
52 Urate oxidase	289	53.2	55.9	82.0
53 Selectin	460	52.7	53.5	54.1
54 Prolactin receptor	533	52.6	51.3	53.7
55 Link protein	344	51.2	50.4	92.7
56 SOD Cu/Zn	147	50.6	49.7	73.5
57 Flavin-containing monooxygenase	510	50.3	51.3	75.5
58 Pancreatic triglyceride lipase	446	50.1	51.0	66.1
59 Osteopontin	249	48.7	50.4	43.8
60 Casein kinase II α subunit	371	48.5	48.8	97.8
61 Apo H	335	43.9	46.7	61.2
62 Calpastatin	587	42.3	41.0	52.5

Table 1. Continued

Gene	Codons (number)	GC_3 (no murids)	GC_3 (murids)	Conserved aa (%)
63 Stem cell factor/Kit ligand	261	41.0	41.3	75.5
64 Serum albumin	598	40.8	44.9	58.7
65 HSP 108	773	40.4	41.5	95.7
66 Macrophage scavenger	433	38.7	39.3	55.2
67 Protein phosphatase X catalytic	284	38.2	39.5	100.0
68 Rab 2	205	36.9	39.3	98.0
69 Rab 1	210	35.8	36.9	100.0

[a] GC_3 values concern data from four orders including murids or from three orders without murids. Conserved amino acids were calculated as number of (QuS+DuS)/number of analyzed codons. QuS and DuS are synonymous quartet and duet codons, respectively (see also Methods and legend of Fig. 1)

The Frequencies of the Three Classes of Synonymous Positions in Quartet Codons of Different Genes are Correlated with Synonymous Divergence

Strong correlations were found between the frequencies of the conserved, intermediate, and variable classes of positions in quartet codons and: (1) SDF—namely, the divergence in all synonymous positions as judged from pairwise comparisons; and (2) SDF_2, the synonymous divergence in duet codons (not shown). These results stress the link of the three classes of positions in quartet codons with SDF (and, in turn, with the corresponding K_s values; see Mouchiroud et al. 1955) and with SDF_2; the (synonymous substitution rate) link with SDF_4 (the synonymous divergence in quartet codons) was also found, as expected.

The Frequencies of Each of the Three Classes of Positions Are Correlated in Quartet and Duet Codons

Figure 3 shows that a significant correlation ($R = 0.6$; $P = 10^{-4}$) exists between the percentages of conserved synonymous positions of quartet and duet codons from the same genes. The intercept and the slope of the least-square lines of Fig. 3 indicate, however, that the degree of conservation is higher in duet than in quartet synonymous positions for low conservation values, but reaches the same levels for high conservation values. Correlations similar to those just described for conserved positions hold for intermediate and variable positions, although with lower correlation coefficients (Fig. 3).

The Conserved Synonymous Positions in Different Genes Are Correlated with the Degree of Amino Acid Conservation

The percentages of conserved synonymous positions in quartet and duet codons are well correlated with those of conserved amino acids (excluding the amino acids encoded by the conserved quartet and duet codons, respectively; Fig. 4A and B).

This result expands the previously reported good correlation between synonymous and nonsynonymous positions (Mouchiroud et al. 1995) in that it shows that a correlation concerning homologous genes from four different mammalian orders also holds between the conserved quartet and duet codons and conserved amino acids. Incidentally, there is no significant correlation between GC_3 and amino acid or synonymous position conservation (not shown; however, see Fig. 2 for the latter point).

Conserved, Intermediate, and Variable Synonymous Positions of Quartet Codons from Homologous Genes Show Frequencies That Are Generally Different from Expectations Based on a Random Substitution Process

The difference histograms of Fig. 5 show that the quartet codons from actual sequences have higher frequencies of conserved and variable positions and much lower frequencies of intermediate positions, respectively, when compared to the simulated sequences. It may be worth pointing out that the sum of differences from the expected values for conserved + intermediate + variable positions for any single gene must be zero because of the nature of the randomization (Zoubak et al. 1995).

An assessment of the significance of the deviation was carried out using a χ^2 test (Table 3). This revealed that, when a four-order strategy (including murids) was used and the three classes of positions were combined for each gene, 57% of the genes showed P values lower than 0.05. If only genes including more than 50 or 100 synonymous quartets were used, P values lower than 0.05 were obtained for 60% and 69% of the genes, respectively. These results indicate that, after correction for size, the majority (two-thirds) of the genes tested show significant deviations from statistical expectations.

Significant deviations were more frequent in GC-rich genes than in GC-poor genes. For example, in the 23 genes having the highest GC_3 level, 17 genes (74%) showed significant differences, whereas in the two 23 gene sets having the lowest GC_3 level, only 11 (48%) did so. These trends are of interest in connection with findings presented in the following paper (Zoubak et al. 1995).

```
HUMAN  ATG GTG CAC CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC CTG TGG GGC AAG GTG AAC GTG GAT GAA GTT GGT
CALF   ATG --- --- CTG ACT GCT GAG GAG AAG GCT GCC GTC ACC GCC TTT TGG GGC AAG GTG AAA GTG GAT GAA GTT GGT
RABBIT ATG GTG CAT CTG TCC AGT GAG GAG AAG TCT GCG GTC ACT GCC CTG TGG GGC AAG GTG AAT GTG GAA GAA GTT GGT
RAT    ATG GTG CAC CTA ACT GAT GCT GAG AAG GCT GCT GTT AAT GCC CTG TGG GGA AAG GTG AAC CCT GAT GAT GTT GGT
Symbol --- $$$ $$$ --+ +-+ **- -++ --- --- *-- --* --* -++ --- +-+ --- --+ --- --- --* +++ --+ --+ --- ---
Codon  $$$ $$$ $$$ QuS QuN QuN DuN DuS DuS QuN QuS QuS QuN QuS QuN $$$ QuS DuS QuS DuN QuN DuN DuN QuS QuS

HUMAN  GGT GAG GCC CTG GGC AGG CTG CTG GTG GTC TAC CCT TGG ACC CAG AGG TTC TTT GAG TCC TTT GGG GAT CTG TCC
CALF   GGT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCC TGG ACT CAG AGG TTC TTT GAG TCC TTT GGG GAC TTG TCC
RABBIT GGT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCA TGG ACC CAG AGG TTC TTC GAG TCC TTT GGG GAC CTG TCC
RAT    GGC GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCT TGG ACC CAG AGG TAC TTT GAT AGC TTT GGG GAC CTG TCC
Symbol --+ --- --- --- --- --- --- --- --+ --- --- --* --- --+ --- --- -+- --+ --+ ++- --- --- --+ +-- ---
Codon  QuS DuS QuS QuS QuS DuS QuS QuS QuS QuS DuS QuS $$$ QuS DuS DuS DuN DuS DuN QuN DuS QuS DuN QuN QuS

HUMAN  ACT CCT GAT GCT GTT ATG GGC AAC CCT AAG GTG AAG GCT CAT GGC AAG AAA GTG CTC GGT GCC TTT AGT GAT GGC
CALF   ACT GCT GAT GCT GTT ATG AAC AAC CCT AAG GTG AAG GCC CAT GGC AAG AAG GTG CTA GAT TCC TTT AGT AAT GGC
RABBIT TCT GCA AAT GCT GTT ATG AAC AAT CCT AAG GTG AAG GCT CAT GGC AAG AAG GTG CTG GCT GCC TTC AGT GAG GGT
RAT    TCT GCC TCT GCT ATC ATG GGT AAC CCT AAG GTG AAG GCC CAT GGC AAG AAG GTG ATA AAC GCC TTC AAT GAT GGC
Symbol *-- +-* *+- --- +-+ --- **+ --+ --- --- --- --- --* --- --- --- --+ --- +-* +*+ +-- --* -+- +-+ --+
Codon  QuN QuN DuN QuS QuN $$$ QuN DuS QuS DuS QuS DuS QuS DuS QuS DuS DuS QuS QuN DuN QuN DuS DuN DuN QuS

HUMAN  CTG GCT CAC CTG GAC AAC CTC AAG GGC ACC TTT GCC ACA CTG AGT GAG CTG CAC TGT GAC AAG CTG CAC GTG GAT
CALF   ATG AAG CAT CTC GAT GAC CTC AAG GGC ACC TTT GCT GCG CTG AGT GAG CTG CAC TGT GAT AAG CTG CAT GTG GAT
RABBIT CTG AGT CAC CTG GAC AAC CTC AAA GGC ACC TTT GCT AAG CTG AGT GAA CTG CAC TGT GAC AAG CTG CAC GTG GAT
RAT    CTG AAA CAC TTG GAC AAC CTC AAG GGC ACC TTT GCT CAT CTG AGT GAA CTC CAC TGT GAC AAG CTG CAT GTG GAT
Symbol +-- +** --+ +-+ --+ +-- --- --+ --- --- --- --+ *** --- --- --* --+ --- --- --+ --- --- --* --- ---
Codon  QuN DuN DuS QuN DuN DuN QuS DuS QuS QuS DuS QuS DuN QuS DuS DuS QuS DuS DuS DuN DuS QuS DuS QuS DuS

HUMAN  CCT GAG AAC TTC AGG CTC CTG GGC AAC GTG CTG GTC TGT GTG CTG GCC CAT CAC TTT GGC AAA GAA TTC ACC CCA
CALF   CCT GAG AAC TTC AAG CTC CTG GGC AAC GTG CTA GTG GTT GTG CTG GCT CGC AAT TTT GGC AAG GAA TTC ACC CCG
RABBIT CCT GAG AAC TTC AGG CTC CTG GGC AAC GTG CTG GTT ATT GTG CTG TCT CAT CAT TTT GGC AAA GAA TTC ACT CCT
RAT    CCT GAG AAC TTC AGG CTC CTG GGC AAT ATG ATT GTG ATT GTG TTG GGC CAC CAC CTG GGC AAG GAA TTC ACC CCC
Symbol --- --- --- --- -+- --- --- --- --+ +-- +-* --* *+- --- +-- ++* -+* +-* +-+ --- --* --- --- --+ --*
Codon  QuS DuS DuS DuS DuN QuS QuS QuS DuS QuN QuN QuS OdN QuS QuN QuN DuN DuN DuN QuS DuS DuS DuS QuS QuS

HUMAN  CCA GTG CAG GCT GCC TAT CAG AAA GTG GTG GCT GGT GTG GCT AAT GCC CTG GCC CAC AAG TAT CAC TAA
CALF   GTG CTG CAG GCT GAC TTT CAG AAG GTG GTG GCT GGT GTG GCC AAT GCC CTG GCC CAC AGA TAT CAT TAA
RABBIT CAG GTG CAG GCT GCC TAT CAG AAG GTG GTG GCT GGT GTG GCC AAT GCC CTG GCT CAC AAA TAC CAC TGA
RAT    TGT GCA CAG GCT GCC TTC CAG AAG GTG GTG GCT GGA GTG GCC AGT GCC CTG GCT CAC AAG TAC CAC TAA
Symbol *** +++ --- --- -+- -*+ --- --+ --- --- --- --+ --- --+ -+- --- --- --* --- -+* --* --+ -+-
Codon  QuN QuN DuS QuS QuN DuN DuS DuS QuS QuS QuS QuS QuS QuS DuN QuS QuS QuS DuS DuN DuS DuS $$$
```

Fig. 1. Alignment of β-globin coding sequences from four species belonging to four mammalian orders. *Symbols* –, +, and * indicate conserved, intermediate, and variable positions (as defined in Methods). $ indicates codons that were excluded from analysis; these comprised initiation and termination codons, codons for methionine and tryptophan, and codons showing deletions. *QuS, QuN, DuS* and *DuN* refer to synonymous and nonsynonymous quartet and duet codons, respectively.

Some genes comprised a relatively small number of positions, especially in the variable class. If only genes for which the actual number of positions is at least 30 are taken into consideration, the sample comprises only 50 genes for the conserved class, 39 genes for the intermediate class, and 18 genes for the variable class. In this case, the percentage of genes showing significant differences becomes 28% for the conserved class, 46% for the intermediate class, and 78% for the variable class of positions, i.e., values higher than those obtained for each class before any selection: 23% of conserved, 38% of intermediate, and 52% of variable positions.

In contrast with the results mentioned above for homologous genes from different mammalian orders, homologous genes from different species belonging to the same mammalian order did not display significant differences compared to statistical expectations (see Table 3, end; the same genes show significant differences in the four-order comparison level).

Figure 6A shows a comparison of percentages of each class of codons, as observed in actual sequences, with the distributions expected from a random process for the H,K ATPase (α) gene as represented in five different orders (including murids). In this case, the actual values fall outside the distributions of simulated values. In contrast, this was not true (Fig. 6B) for the β-globin gene from four species belonging to the same order (primates). Results similar to those of Fig. 6B were obtained for the two other sets of genes (α-globin genes from primates and growth hormone genes from artiodactyls) that are available from four species belonging to the same order (data not shown).

Table 2. Number of synonymous quartet and duet codons and percentage of conserved, intermediate, and variable classes of positions in homologous genes from four mammalian orders[a]

Gene	Quartets	% QuC	% QuI	% QuV	Duets	% DuC	% DuI	% DuV
1 Apo E	75				73	74.0	21.9	4.1
2 Creatin kinase B	177				148	77.0	19.6	3.4
3 A1 adenosine	150				120	69.0	22.0	9.0
4 H,K ATPase β subunit	88				107	64.0	24.0	12.0
5 α-globin	53				44	56.8	40.9	2.3
6 Apo A1	61				68	63.0	32.0	5.0
7 Na-H exchange protein	374				317	71.0	22.4	6.6
8 Serine pyruvate aa transferase	141				99	55.6	34.3	9.1
9 Dipeptidase	135				110	49.1	46.4	4.5
10 GMP-phosphodiesterase α	297				399	53.1	33.3	13.6
11 CD8 α chain	51				29	69.0	27.6	3.4
12 Glutathione peroxidase	82				63	54.0	28.6	17.4
13 H,K ATPase α subunit	497				424	63.0	28.0	9.0
14 Retinol-binding protein	61				80	72.5	22.5	5.0
15 Glucose Glut3	253				176	69.3	25.0	5.7
16 Prostaglandin E receptor	140				73	71.2	23.3	5.5
17 Prolyl-4-hydroxylase-β	179				233	52.4	33.9	13.7
18 Growth hormone	57				64	50.0	42.0	8.0
19 TNFα	79				61	64.0	33.0	3.0
20 Ferritin L	63				60	60.0	23.0	17.0
21 Myoglobin	47				54	67.0	28.0	5.0
22 Apo CIII	20				14	78.6	14.3	7.1
23 TNF β	77				45	62.2	28.9	8.9
24 Hydrophobic-surfactant-associated factor	67	41.8	34.3	23.9	45	66.6	26.7	6.7
25 Phospholipase A2	31	38.7	32.3	29.0	55	60.0	30.9	9.1
26 Phenyl tRNA ligase	148	44.6	29.1	26.4	206	62.6	25.2	12.2
27 Tissue inhibitor of metallo-proteinase	53	43.4	41.5	15.1	63	52.4	33.3	14.3
28 Guanine-nt-binding protein	159	73.0	18.2	8.8	221	76.9	16.7	6.4
29 Polymeric Ig receptor	141	45.4	36.2	18.4	124	54.8	33.1	12.1
30 Colony-stimulating factor	37	48.6	40.5	10.8	20	45.0	50.0	5.0
31 CD4 antigen	86	48.8	29.1	22.1	51	54.9	29.4	15.7
32 Erythropoietin	70	44.3	44.3	11.4	49	57.1	32.7	10.2
33 Gastrin	28	42.9	50.0	7.1	22	45.5	36.4	18.2
34 Na^+/nucleoside	223	41.7	36.8	21.5	201	53.7	36.3	10.0
35 D-amino-acid oxidase	126	47.8	36.5	16.7	94	61.7	30.8	7.5
36 Ferritin H	50	54.0	36.0	10.0	77	64.0	31.0	5.0
37 Protein kinase C	262	50.0	32.1	17.9	358	66.2	26.0	7.8
38 ANP	42	57.0	19.0	24.0	41	61.0	34.0	5.0
39 β-globin	54	67.0	20.0	13.0	43	63.0	26.0	12.0
40 Potassium channel	192	56.2	34.4	94	247	68.4	22.3	9.3
41 Cytochrome b5	40	27.5	32.5	40.0	46	60.9	23.9	15.2
42 Endothelin	39	46.2	30.8	23.1	56	69.6	25.0	5.4
43 Phagocytic glycoprotein I	116	33.6	40.5	25.9	119	59.7	31.9	8.4
44 Prolactin	36	31.0	36.0	33.0	49	69.0	29.0	2.0
45 Interleukin 2 receptor	43	44.2	30.2	25.6	53	45.3	43.4	11.3
46 Tissue factor	49	36.7	30.6	32.7	67	50.7	35.8	13.5
47 β2 microglobulin	25	36.0	40.0	24.0	28	50.0	39.3	10.7
48 Na-K ATPase β-1 subunit	100	55.0	29.0	16.0	131	64.1	28.2	7.6
49 CD3 ϵ antigen	42	61.9	33.3	4.8	37	73.0	18.9	8.1
50 Na-Ca exchange protein	403	53.6	30.0	16.4	456	60.1	32.2	7.7
51 Ca-ATPase	471	51.0	32.0	17.0	446	58.0	32.0	10.0
52 Urate oxidase	99	44.4	37.4	18.2	130	66.9	26.9	6.2
53 Selectin	117	42.0	42.0	16.0	127	54.0	37.0	9.0
54 Prolactin receptor	126	50.8	32.5	16.7	151	60.9	30.5	8.6
55 Link protein	150	62.7	26.0	11.3	147	58.5	32.0	9.5
56 SOD Cu/Zn	57	42.1	40.4	17.5	49	65.3	22.4	12.3
57 Flavin-containing monooxy-genase	184	52.2	31.5	16.3	192	66.7	26.0	7.3
58 Pancreatic triglyceride lipase	139	42.4	39.6	18.0	150	57.3	34.0	8.7
59 Osteopontin	47	42.6	44.7	12.8	62	56.5	35.5	8.0

Table 2. Continued

Gene	Quartets	% QuC	% QuI	% QuV	Duets	% DuC	% DuI	% DuV
60 Casein kinase II alpha subunit	150	66.7	24.0	9.3	186	71.0	26.3	2.7
61 Apo H	96	42.7	36.5	20.8	104	49.0	39.4	11.6
62 Calpastatin	123	42.3	34.1	23.6	173	51.4	37.0	11.6
63 Stem cell factor/Kit ligand	77	75.3	15.6	9.1	116	81.9	13.8	4.3
64 Serum albumin	133	42.1	44.4	13.5	203	52.2	37.4	10.3
65 HSP 108	284	51.4	33.8	14.8	435	63.9	29.0	7.1
66 Macrophage scavenger	108	46.3	38.9	14.8	121	45.6	41.3	14.1
67 Protein phosphatase X catalytic	134	58.2	26.9	14.9	147	62.6	31.3	6.1
68 Rab 2	82	60.0	32.0	8.0	117	77.0	20.0	3.0
69 Rab 1	92	72.0	25.0	3.0	112	78.0	16.0	5.0
Average	123.0 ± 101.2	49.0 ± 9.7	32.7 ± 6.8	18.3 ± 7.2	130.3 ± 109.7	61.7 ± 9.0	29.7 ± 7.4	8.5 ± 3.9
Average of simulated sequences		46.5 ± 9.2	39.4 ± 4.1	14.1 ± 5.4		61.1 ± 8.5	32.3 ± 6	6.5 ± 2.7

[a] QuC, QuI, QuV, DuC, DuI, and DuV refer to conserved, intermediate, and variable quartet and duet codons, respectively

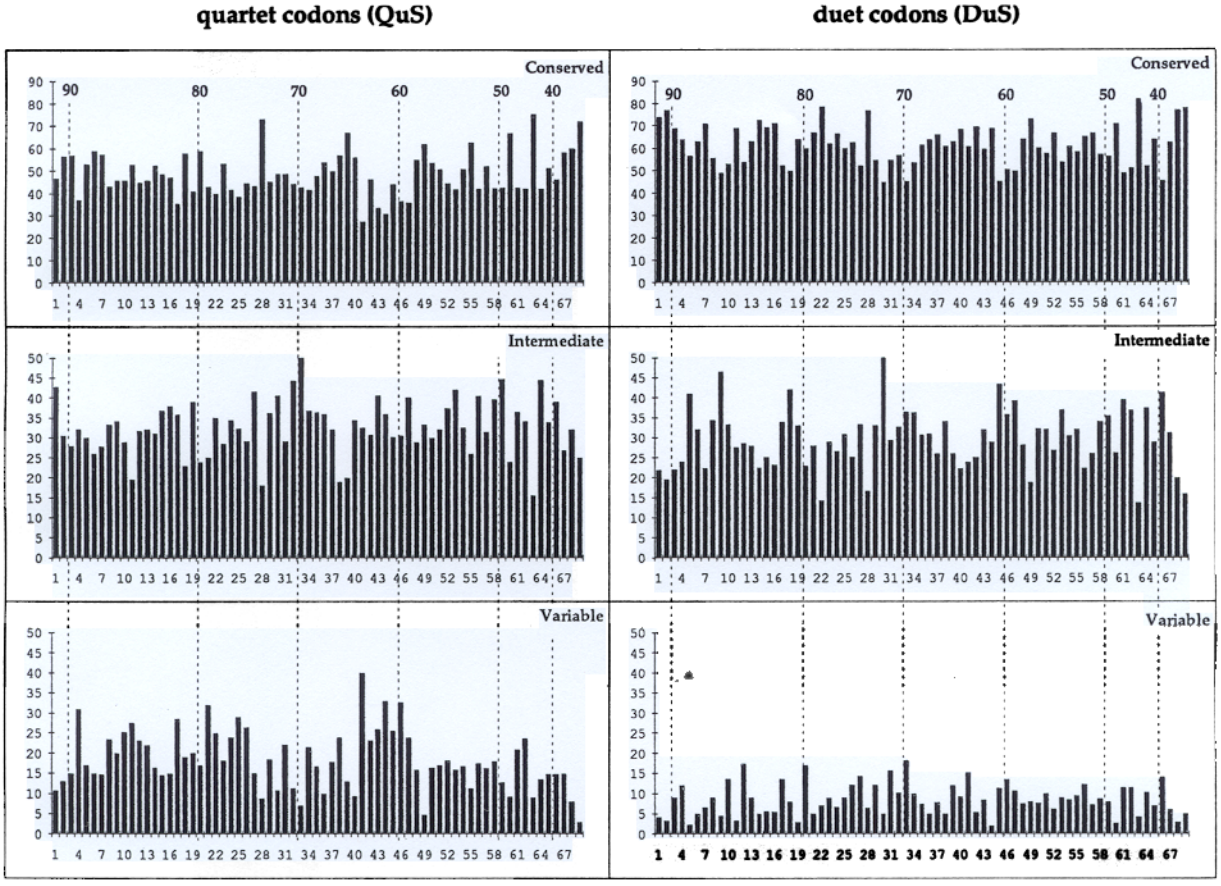

Fig. 2. Histograms displaying (ordinates) the percentages of each class of synonymous positions (conserved, intermediate, and variable) for quartet and duet codons of the genes investigated (Table 2). Data refer to four-order comparisons (including murids). Genes are arranged in order of decreasing GC_3 and numbered as in Table 2 (figures on the horizontal bottom line). Vertical dashed lines and figures on the top horizontal line refer to GC_3.

The Three Classes of Duet Codons Show Frequencies That Are Sometimes Different from Expectations Based on a Random Substitution Process

An analysis similar to that of Table 3 was carried out for duet codons (Table 4). While the general trends are the same as those described above for quartet codons, the percentages of significant deviations from statistical expectations were definitively lower. Indeed, when the χ^2 values for the three classes were combined for each gene, only 27.5% of the genes showed significant differences (against a 57% value for quartet codons). This value

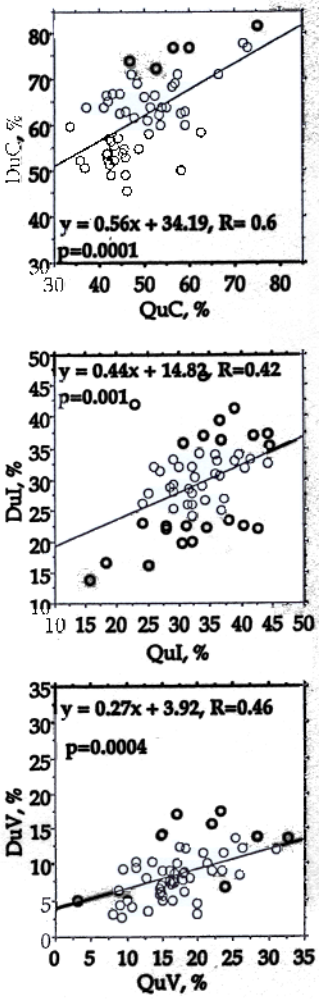

Fig. 3. The percentages of conserved, intermediate, and variable positions in duet codons are plotted against those in quartet codons of the same genes; 15 genes comprising less than 150 synonymous codons were omitted in this plot.

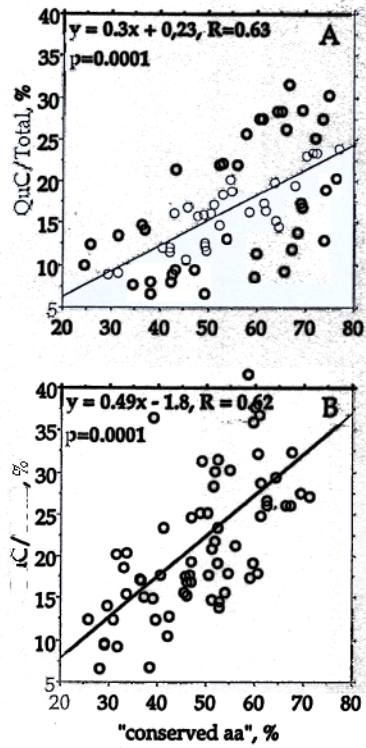

Fig. 4. The percentages of conserved third codon positions in synonymous quartet **(A)** and duet **(B)** codons of homologous genes from four mammalian orders are plotted against the percentages of conserved amino acids in the corresponding encoded proteins. Conserved amino acids (as obtained from Table 1) corresponding to conserved quartet and conserved duet were omitted in **A** and **B**, respectively.

increased to 34% when neglecting sequences comprising less than 50 synonymous duets (the corresponding value being 60% for quartet codons).

Discussion

The Frequency Patterns of the Three Classes of Synonymous Positions

The results of Table 2 and Fig. 2 show that both quartet and duet codons derived from different homologous genes as present in four mammalian orders (including murids) exhibit frequencies of the three classes of positions which are significantly different from each other and from the averages of simulated sequences for the intermediate and variable classes of quartets and for the intermediate class of duets. Different frequencies in different genes suggest a gene-specific phenomenon. (See following section.)

The parallel behavior of the three classes of synonymous positions in quartet and duet codons (Fig. 3) suggests common features in the nucleotide substitution process, as it occurs throughout the genes in those two sets of synonymous codons.

The implications of the good correlations between conserved synonymous positions in quartet and duet codons and conserved amino acids (Fig. 4), which is in agreement with the previous data of Mouchiroud et al. (1995), will be discussed elsewhere.

Finally, the existence of specific frequency patterns suggests that the synonymous substitution process is nonrandom, some codons being conserved in the mammalian genes studied, while other ones accumulate changes. This point will be discussed in more detail in the following section.

The Frequency Pattern of Synonymous Substitutions is Nonrandom

As shown in Fig. 5, the frequencies of the three classes in the actual sequences are quite distinct from those found in the simulated sequences, in that they show an

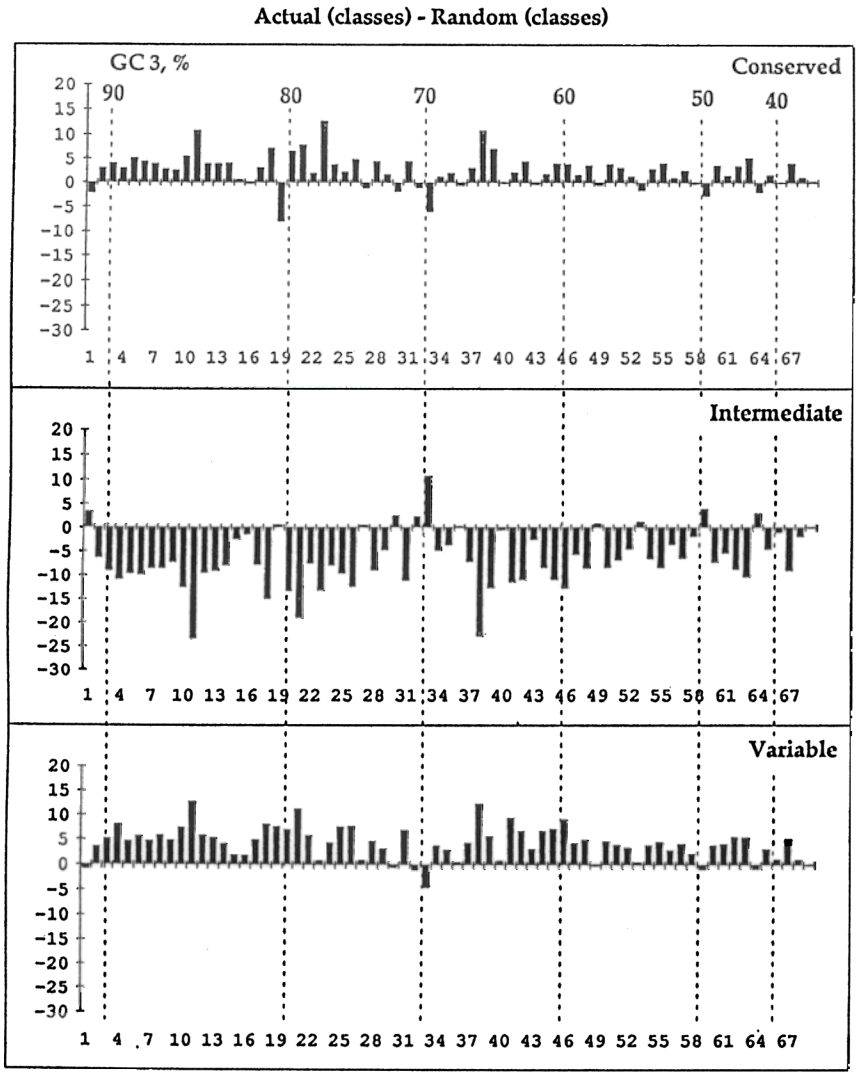

Fig. 5. Difference histograms of the frequencies of conserved, intermediate, and variable positions as found in quartet codons from actual and simulated sequences. For other indications, see legend of Fig. 2.

excess of conserved and of variable positions and a scarcity of intermediate positions. This indicates that a number of synonymous positions of homologous genes seem to have been spared by the nucleotide substitution process, whereas other ones have accumulated substitutions. Obviously, in the simulated sequences synonymous substitutions tend to be scattered in a more uniform way over the gene, as shown by the higher percentage of intermediate positions generated by the random synonymous substitution process compared to the actual sequences.

The statistical analysis of the frequencies of synonymous substitutions in quartets (Table 3) shows that the majority of the genes tested (especially the GC-rich genes) exhibit a significant difference in the frequency of different classes compared to expectations based on a process in which nucleotide substitutions accumulate at random in synonymous quartet positions. This leads to the conclusions that the synonymous substitution process is nonrandom and that the three classes of positions were not simply the result of a stochastic process in which, with time, nucleotide substitutions accumulate at random in the genes under consideration.

The nonrandomness of the process will be discussed further in the following paper (Zoubak et al. 1995).

Since, in most genes, the frequencies of the conserved synonymous positions found in different genes are not simply due to the fluctuations in a stochastic substitution process, one should also draw the conclusion from the present results that the substitution process is largely gene-specific, in agreement with the conclusion of Mouchiroud et al. (1995).

The reason why only a two-thirds majority of the genes show a significant difference may be that the present analysis was done at a four-order level. It is expected, indeed, that comparisons at levels higher than four-order would show significant differences for genes which do not so at the four-order level. This suggestion

Table 3. χ^2 values obtained comparing the actual frequencies of conserved, intermediate, and variable positions from quartet codons with expectations based on a random substitution process

Gene		χ^2 QuC		χ^2 QuV	$\Sigma\chi^2$ [a]	$P<$ [b]
1*	Apo E	0.66	0.40	0.15	1.21	0.700
2	Creatin kinase B	2.89	4.04	4.96	**11.89**	**0.005**
3	A1 adenosine	5.37	6.92	7.88	**20.17**	**0.001**
4*	H,K ATPase β subunit	1.31	4.90	8.78	**14.99**	**0.001**
5*	α-globin	2.14	2.42	2.25	**6.81**	**0.050**
6*	Apo A1	2.34	3.40	4.25	**9.99**	**0.010**
7	Na-H exchange protein	10.96	15.7	19.33	**45.99**	**0.001**
8	Serine pyruvate aa transferase	1.51	4.72	8.67	**14.9**	**0.001**
9	Dipeptidase	1.33	3.32	5.48	**10.13**	**0.010**
10	GMP-phosphodiesterase α	11.58	21.09	27.92	**60.59**	**0.001**
11*	CD8 α chain	9.55	13.64	15.93	**39.12**	**0.001**
12*	Glutathione peroxidase	1.53	3.07	4.49	**9.09**	**0.020**
13	H,K ATPase α subunit	12.38	19.36	23.74	**55.48**	**0.001**
14*	Retinol-binding protein	1.56	2.01	2.17	5.74	0.100
15	Glucose Glut3	0.17	0.84	1.90	2.91	0.250
16	Prostaglandin E receptor	0.01	0.17	0.84	1.02	0.700
17	Prolyl-4-hydroxylase β	2.12	4.64	6.61	**13.37**	**0.005**
18*	Growth hormone	4.83	6.85	8.35	**20.03**	**0.001**
19*	TNFα	2.05	3.57	5.03	**10.65**	**0.005**
20*	Ferritin L	4.68	6.45	7.83	**18.96**	**0.001**
21**	Myoglobin	3.40	6.80	9.56	**19.76**	**0.001**
22**	Apo CIII	0.08	0.44	0.96	1.48	0.500
23*	TNF β	0.00	0.26	1.09	1.35	0.700
24*	Hydrophobic-surfactant-associated factor	1.19	1.86	2.16	5.21	0.100
25**	Phospholipase A2	0.22	1.39	2.73	4.34	0.200
26	Phenyl tRNA ligase	4.38	9.26	13.82	**27.46**	**0.001**
27*	Tissue inhibitor of metalloproteinase	0.14	0.01	0.07	0.22	0.900
28	Guanine-nt-binding protein	10.18	12.05	13.25	**35.48**	**0.001**
29	Polymeric Ig receptor	0.54	1.54	2.80	4.88	0.100
30**	Colony-stimulating factor	0.25	0.13	0.03	0.41	0.900
31*	CD4 antigen	2.44	4.81	7.11	**14.36**	**0.001**
32*	Erythropoietin	0.17	0.20	0.19	0.56	0.800
33**	Gastrin	1.66	1.54	1.20	4.4	0.200
34	Na^+/nucleoside	0.35	2.34	5.82	**8.51**	**0.020**
35	D-amino-acid oxidase	0.15	0.83	1.99	2.97	0.250
36**	Ferritin H	0.04	0.00	0.02	0.06	0.975
37	Protein kinase C	3.63	6.64	9.38	**19.65**	**0.001**
38**	ANP	9.01	11.73	13.53	**34.27**	**0.001**
39*	β-globin	5.49	5.39	4.69	**15.57**	**0.001**
40	Potassium channel	0.00	0.06	0.31	0.37	0.900
41**	Cytochrome b5	0.22	2.33	5.32	**7.87**	**0.020**
42**	Endothelin	0.96	1.96	3.07	**5.99**	**0.050**
43	Phagocytic glycoprotein I	0.03	0.33	1.81	2.17	0.500
44**	Prolactin	0.07	1.03	2.70	3.8	0.200
45**	Interleukin 2 receptor	0.9	2.19	3.47	**6.56**	**0.050**
46**	Tissue factor	0.90	3.44	6.21	**10.55**	**0.010**
47**	β2 microglobulin	0.08	0.37	0.81	1.26	0.700
48*	Na-K ATPase β-1 subunit	2.22	4.05	6.05	**12.32**	**0.005**
49**	CD3 ϵ antigen	0.05	0.03	0.01	0.09	0.975
50	Na-Ca exchange protein	10.2	14.4	17.03	**41.63**	**0.001**
51	Ca-ATPase	5.81	10.63	15.47	**31.91**	**0.001**
52*	Urate oxidase	0.23	0.96	2.05	3.24	0.200
53	Selectin	0.6	0.07	0.11	0.78	0.700
54	Prolactin receptor	1.64	2.85	3.89	**8.38**	**0.020**
55	Link protein	5.63	7.16	8.37	**21.16**	**0.001**
56*	SOD Cu/Zn	0.07	0.37	0.87	1.31	0.700
57	Flavin-containing monooxygenase	1.48	3.45	6.01	**10.94**	**0.005**
58	Pancreatic triglyceride lipase	0.01	0.25	1.25	1.51	0.500
59**	Osteopontin	0.63	0.35	0.10	1.08	0.700
60	Casein kinase II α subunit	5.06	6.27	7.23	**18.56**	**0.001**
61*	Apo H	0.27	1.31	2.91	4.49	0.200

Table 3. Continued

Gene		χ^2 QuC	QuI	QuV	$\Sigma\chi^{2\,a}$	$P<^b$
62	Calpastatin	1.91	4.06			**0.005**
63*	Stem cell factor/Kit ligand	7.41	8.54			**0.001**
64	Serum albumin	0.92	0.56			0.500
65	HSP 108	1.15	3.12			**0.010**
66	Macrophage scavenger	0.00	0.05			0.900
67	Protein phosphatase X catalytic	3.71	5.74			**0.001**
68*	Rab 2	0.12	0.27			0.700
69*	Rab 1	0.00	0.00			0.995
	α-globin (primates)[c]	0.10	0.14			0.800
	β-globin (primates)	0.61	0.77			0.300
	Growth hormone (artiodactyls)	1.60	0.53			0.300

[a] $\Sigma\chi^2$ is the sum of the three χ^2 values for the three classes of quartet codons, using as a reference the sequences randomized according to Zoubak et al. (1995)

[b] *P* values were estimated using two degrees of freedom. *P* values lower than 0.05 are in underlined bold type. Asterisk and double asterisks refer to sequences with less than 100 or less than 50 quartet codons, respectively

[c] These three genes were compared for four species within the orders indicated

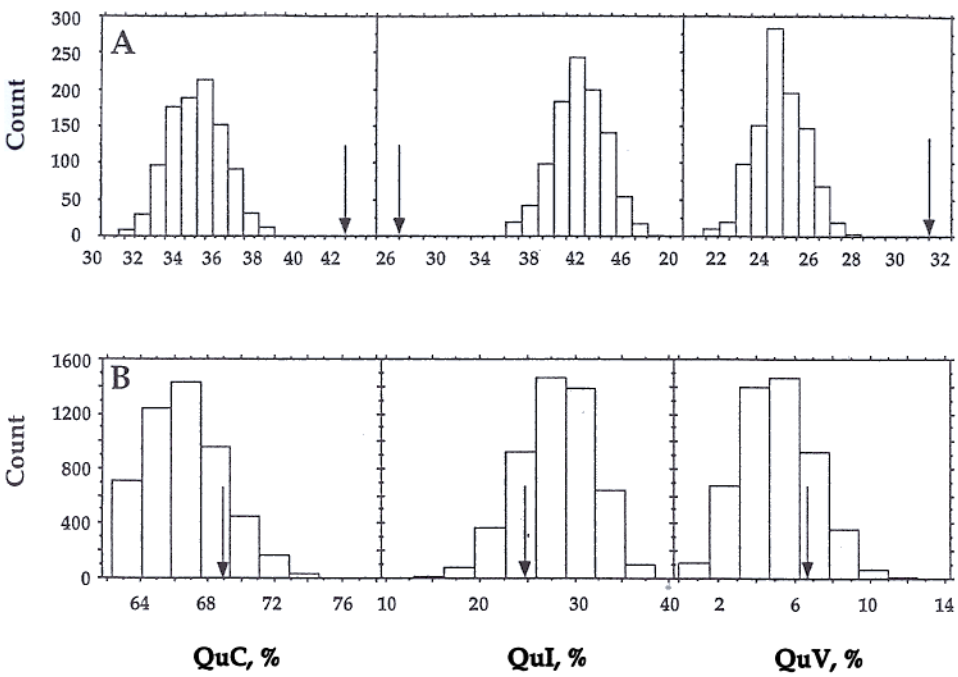

Fig. 6. The distribution of the percentages of each class of positions (conserved, intermediate, and variable) in fourfold degenerate codons in multi-alignments of "simulated" sequences (derived from the "ancestral" sequence by a random substitution process) are displayed along with the percentage of each class of position (arrows) as found in the actual sequence for the H,K ATPase (α) genes from five mammalian orders (1,000 simulations; **A**) and for the β-globin genes from four species of primates (5,000 simulations; **B**) For the construction of the "ancestral" (consensus) sequence and of the derived simulated present-day sequences, see Zoubak et al. (1995).

is supported by the few analyses which could be done at five-order level and which showed an increase of the χ^2 values in most cases and no decrease in the others, and also by the finding that the difference between inter- and intra-order comparison fits the expectations based on the widely different divergence times under consideration. The small extent or absence of deviation from randomness shown by some genes, especially GC-poor genes, may, however, also be due to other factors (Zoubak et al. 1995).

Acknowledgments. We thank most warmly Laurent Duret for having provided us with the sequence alignments used in the present work and for very useful discussions, and Adam Eyre-Walker, Takashi Gojobori, Wen-Hsiung Li, Tomoko Ohta, and Ken Wolfe for critical reading of this paper.

Table 4. χ^2 values obtained comparing the actual frequencies of conserved, intermediate, and variable positions from duet degenerate codons with expectations based on a random substitution process

Gene	χ^2 DuC	DuI	DuV	$\Sigma\chi^{2\,a}$	$P<^b$
1* Apo E	1.39	1.48	1.53		0.100
2 Creatin kinase B	0.22	0.28	0.33		0.700
3 A1 adenosine	1.51	3.76	4.70		**0.010**
4 H,K ATPase β subunit	0.63	1.74	2.45		0.100
5** α-globin	1.20	1.12	0.88		0.200
6* Apo A1	0.75	0.51	0.18		0.500
7 Na-H exchange protein	0.62	1.65	2.62		0.100
8* Serine pyruvate aa transferase	0.23	0.02	0.04		0.900
9 Dipeptidase	**9.39**	**6.54**	2.03		**0.001**
10 GMP-phosphodiesterase α	0.78	**4.95**	**9.12**		**0.001**
11** CD8 α chain	0.24	0.29	0.33		0.700
12* Glutathione peroxidase	0.68	3.15	**4.39**		**0.020**
13 H,K ATPase α subunit	0.44	0.08	1.37		0.400
14* Retinol-binding protein	0.17	0.36	0.53		0.600
15 Glucose Glut3	0.04	0.29	0.67		0.600
16* Prostaglandin E receptor	0.01	0.26	0.65		0.700
17 Prolyl-4-hydroxylase β	2.08	**4.66**	**6.22**		**0.002**
18* Growth hormone	1.05	0.62	0.13		0.400
19* TNFα	0.81	0.66	0.42		0.400
20* Ferritin L	**5.25**	**9.08**	**9.43**		**0.001**
21* Myoglobin	0.06	0.19	0.32		0.800
22** Apo CIII	0.91	1.30	1.48		0.150
23** TNF β	0.09	0.08	0.47		0.750
24** Hydrophobic-surfactant-associated factor	0.03	0.15	0.29		0.800
25* Phospholipase A2	0.07	0.31	0.61		0.600
26 Phenyl tRNA ligase	**7.48**	**13.53**	**14.91**		**0.001**
27* Tissue inhibitor of metalloproteinase	0.06	0.62	1.34		0.400
28 Guanine-nt-binding protein	1.66	3.11	**3.91**		**0.015**
29 Polymeric Ig receptor	0.30	1.45	2.75		0.100
30** Colony-stimulating factor	1.11	1.13	1.05		0.200
31* CD4 antigen	0.63	2.07	2.81		0.070
32** Erythropoietin	0.07	0.35	0.69		0.400
33** Gastrin	0.79	1.41	1.57		0.150
34 Na^+/nucleoside	0.10	0.03	0.38		0.800
35* D-amino-acid oxidase	0.12	0.60	1.18		0.400
36* Ferritin H	0.10	0.23	0.38		0.700
37 Protein kinase C	2.00	**4.31**	**6.29**		**0.001**
38** ANP	0.07	0.01	0.00		1.000
39** β-globin	1.66	2.23	2.52		**0.050**
40 Potassium channel	**8.01**	**10.54**	**11.60**		**0.001**
41** Cytochrome b5	2.69	**4.16**	4.60		**0.001**
42* Endothelin	0.14	0.22	0.28		0.750
43 Phagocytic glycoprotein I	0.37	0.94	1.47		0.300
44** Prolactin	0.00	0.01	0.01		1.000
45* Interleukin 2 receptor	0.95	0.39	0.01		0.500
46* Tissue factor	0.23	0.02	0.50		0.700
47** β2 microglobulin	0.20	0.30	0.38		0.700
48 Na-K ATPase β-1 subunit	0.50	1.10	1.56		0.200
49** CD3 ϵ antigen	0.60	1.12	1.34		0.200
50 Na-Ca exchange protein	0.35	0.02	0.10		0.800
51 Ca-ATPase	2.13	**5.87**	**9.11**		**0.001**
52 Urate oxidase	1.75	2.46	2.83		**0.040**
53 Selectin	0.19	0.00	0.22		0.800
54 Prolactin receptor	0.21	0.91	1.77		0.250
55 Link protein	0.19	0.95	1.92		0.250
56** SOD Cu/Zn	4.50	**5.68**	**5.90**		**0.001**
57 Flavin-containing monooxygenase	1.64	2.89	**3.91**		**0.015**
58 Pancreatic triglyceride lipase	0.03	0.04	0.31		0.850
59* Osteopontin	0.12	0.04	0.00		0.900
60 Casein kinase II α subunit	0.44	0.29	0.12		0.700
61 Apo H	0.01	0.05	0.27		0.850

Table 4. Continued

Gene		χ^2 QuC	QuI	QuV	$\Sigma\chi^{2\,a}$	$P<^b$
62	Calpastatin	0.00	0.32			0.500
63	Stem cell factor/Kit ligand	3.12	**3.86**			**0.001**
64	Serum albumin	0.01	0.58			0.300
65	HSP 108	1.09	2.90			**0.015**
66	Macrophage scavenger	0.98	0.13			0.500
67	Protein phosphatase X catalytic	0.01	0.05			0.850
68	Rab 2	0.03	0.10			0.850
69	Rab 1	**6.74**	**8.09**			**0.001**

[a] $\Sigma\chi^2$ is the sum of the three χ^2 values for the three classes of duet codons, using as a reference the sequences randomized according to Zoubak et al. (1995). Boldface χ^2 values correspond to P values lower than 0.05

[b] P values were estimated using two degrees of freedom. P values lower than 0.05 are in underlined bold type. Asterisk and double asterisks refer to sequences with less than 100 or less than 50 duet codons, respectively

References

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

Bernardi G, Mouchiroud D, Gautier C, Bernardi G (1988) Compositional patterns in vertebrate genomes: conservation and change in evolution. J Mol Evol 28:7–18

Bernardi G, Bernardi G (1990a) Compositional patterns in the nuclear genome of cold-blooded vertebrates. J Mol Evol 31:265–281

Bernardi G, Bernardi G (1990b) Compositional transitions in the nuclear genomes of cold-blooded vertebrates. J Mol Evol 31:282–293

Bernardi G, Bernardi G (1991) Compositional properties of nuclear genes from cold-blooded vertebrates. J Mol Evol 33:57–67

Bernardi G, Mouchiroud D, Gautier C (1993) Silent substitutions in mammalian genomes and their evolutionary implications. J Mol Evol 37:583–589

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

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene 73:237–244

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

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, Bernardi G (1993) Compositional properties of coding sequences and mammalian phylogeny. J Mol Evol 37:441–456

Mouchiroud D, Gautier C, Bernardi G (1995) Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of non-synonymous substitutions. J Mol Evol 40:107–113

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

Salinas J, Zerial M, Filipski J, Bernardi G (1986) Gene distribution and nucleotide sequence organization in the mouse genome. Eur J Biochem 16:469–478

Zerial M, Salinas J, Filipski J, Bernardi G (1986) Gene distribution and nucleotide sequence organization in the human genome. Eur J Biochem 160:479–485

Zoubak S, D'Onofrio G, Cacciò S, Bernardi G, Bernardi G (1995) Specific compositional patterns of synonymous positions in homologous mammalian genes. J Mol Evol 40:293–307