J Mol Evol (1987) 26:301–310

Journal of Molecular Evolution

Directional Fixation of Mutations in Vertebrate Evolution

Pascale Perrin[1] and Giorgio Bernardi[2]

[1] Institut d'Evolution Moléculaire, Université Claude Bernard—Lyon I,
43, bd du 11 Novembre 1918, 69622 Villeurbanne, France
[2] Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2, Place Jussieu—75005 Paris, France

Summary. We have made pairwise comparisons between the coding sequences of 21 genes from cold-blooded vertebrates and 41 homologous sequences from warm-blooded vertebrates. In the case of 12 genes, GC levels were higher, especially in third codon positions, in warm-blooded vertebrates compared to cold-blooded vertebrates. Six genes showed no remarkable difference in GC level and three showed a lower level. In the first case, higher GC levels appear to be due to a directional fixation of mutations, presumably under the influence of body temperature (see Bernardi and Bernardi 1986b). These GC-richer genes of warm-blooded vertebrates were located, in all cases studied, in isochores higher in GC than those comprising the homologous genes of cold-blooded vertebrates. In the third case, increases appear to be due to a limited formation of GC-rich isochores which took place in some cold-blooded vertebrates after the divergence of warm-blooded vertebrates. The directional changes in the GC content of coding sequences and the evolutionary conservation of both increased and unchanged GC levels are in keeping with the existence of compositional constraints on the genome.

Key words: Genome — Isochores — Gene composition — Vertebrates — Neutral theory

Introduction

The genomes of cold-blooded and warm-blooded vertebrates exhibit a different organization of their nucleotide sequences, in that the latter show a strong compositional heterogeneity that is absent or much weaker in the former. More specifically, among the isochores (the long, compositionally fairly uniform regions that form the genomes of vertebrates), the GC-rich ones correspond to one-third of the genome of warm-blooded vertebrates, but are absent or scarce in the genomes of cold-blooded vertebrates (Thiery et al. 1976; Bernardi et al. 1985). This major difference is paralleled by a difference in Giemsa and Reverse chromosome banding. In fact, metaphase chromosomes from the vast majority of cold-blooded vertebrates do not exhibit a Giemsa banding, or exhibit a poor one (Cuny et al. 1981; Medrano et al. 1988).

The appearance of GC-rich isochores in the genomes of warm-blooded vertebrates can be explained (1) by de novo formation of sequences; and/or (2) by amplification of rare, pre-existing, interspersed GC-rich sequences; and/or (3) by regional increases in GC of pre-existing GC-poor sequences (Bernardi et al. 1985). The first mechanism, which is the one generally believed to result in the formation of highly repeated (satellite) DNA, can be ruled out as a satisfactory explanation for the phenomena observed in view of both the large amount and the evolutionary conservation of GC-rich isochores, as well as of the demonstrated high concentration of genes and interspersed repeats in them (Bernardi et al. 1985; Mouchiroud et al. 1987). The second mechanism can certainly play a role, as exemplified by the amplification of GC-rich interspersed repeats (like the sequences of the Alu family), and by their accumulation in some particular compartments of the human genome (Zerial et al. 1986). This mechanism is very far, however, from fully explaining the vast changes under consideration, be-

Offprint requests to: G. Bernardi

cause it concerns very small amounts of DNA. The third explanation was, therefore, considered to account for the appearance of GC-rich isochores. Very interestingly, the compositional changes leading to the formation of GC-rich isochores in the genomes of warm-blooded vertebrates also concern genes and, more specifically, coding sequences.

The GC enrichment of genes from warm-blooded vertebrates is of special interest in view of the following findings. (1) GC-rich genes are predominant over GC-poor genes in warm-blooded vertebrates, in spite of the fact that GC-rich isochores represent only one-third of nuclear DNA (Bernardi et al. 1985; Mouchiroud et al. 1987). It is likely that this predominance is even greater than estimated at present because of the current underrepresentation in data banks of housekeeping genes, which seem to be more frequently GC-rich than other genes (Mouchiroud et al. 1987). In contrast, GC-rich genes are poorly represented in the genome of cold-blooded vertebrates (Bernardi et al. 1985; Mouchiroud et al. 1987). The predominance of GC-rich genes in warm-blooded vertebrates and their scarcity in cold-blooded vertebrates stresses the point that a very large number of genes underwent GC increases during the transition between cold-blooded and warm-blooded vertebrates. (2) GC-rich genes are often associated with GC-rich promotors (Wolf and Migeon 1985; Bird 1986; Dynan 1986). (3) Proteins encoded by GC-rich genes show differences in amino acid compositions relative to proteins encoded by GC-poor genes (Bernardi and Bernardi 1986a,b). The last two points indicate that both structural and functional differences may accompany compositional differences.

In the present study, we have investigated the base changes found in the coding sequences from warm-blooded vertebrates relative to homologous coding sequences from cold-blooded vertebrates and we discuss the general meaning of these base changes.

Materials and Methods

The ACNUC retrieval system and ANALSEQ program (Gouy et al. 1984, 1985) were used to extract and analyze the available homologous coding sequences of cold-blooded and warm-blooded vertebrates from the latest GenBank release (Release 48, February 1987). In the present paper, we do not concern ourselves with the question of whether the homologous warm-blooded vertebrate genes that we are examining are orthologous or paralogous relative to the cold-blooded vertebrate genes. The word "isologous" will be used to indicate similarity in sequences.

The algorithm of Smith and Waterman (1981) was used to align nucleotide or amino acid sequences. In most cases, alignments were possible over the whole sequences. When alignments revealed codon insertions, these were eliminated in order to compare equal-length sequences. In two cases (fibrinogen and rabbit immunoglobulin variable segment), only the most similar regions in the alignments were used; these alignments concerned 76% and 70% of the whole amino acid sequences, respectively. Finally, partial protein sequences were used for bovine, rat, and human enkephalin (82% of the sequence), frog alpha crystallin (81%), and human proopiomelanocortin (50%).

Results and Discussion

The Gene Comparisons

Pairwise comparisons were made between the coding sequences of 21 genes from cold-blooded vertebrates and 41 homologous (see Materials and Methods) sequences from warm-blooded vertebrates. These comparisons, which correspond to the totality of those that can be done using the latest GenBank release, deserve some comments.

First of all, cold-blooded vertebrates examined included one reptile (caiman), two amphibians (*Xenopus*, frog), four bony fishes (anglerfish, catfish, trout, salmon), two cartilaginous fishes (the torpedo species *T. marmorata* and *T. californica*), and one cyclostome (lamprey). Different phylogenetic distances were, therefore, involved in the comparisons. Second, comparisons were made between one gene from a cold-blooded vertebrate and one or more homologous genes from warm-blooded vertebrates. Some comparisons concerned, therefore, homologous genes from two or more warm-blooded vertebrates; moreover, one comparison concerned two homologous genes from two fishes. Third, the genes analyzed comprised housekeeping genes (histone, creatine kinase, and calmodulin genes), and tissue-specific genes expressed in different tissues under different developmental and hormonal controls. In other words, comparisons were not limited to a single family or a single class of genes. Finally, because of both the different nature of the genes and the different evolutionary distances, a spectrum of isology levels in both nucleotide and amino acid sequences was explored.

The Changes in Homologous Nucleotide and Amino Acid Sequences from Warm-Blooded and Cold-Blooded Vertebrates

Table 1 shows the number of nucleotide changes found in the three codon positions and in the complete coding sequences; also indicated are the ratios of transitions over transversions, the numbers of amino acid changes, and the nucleotide and amino acid isology levels.

The data of Table 1 indicate that amino acid isology levels were extremely high for certain proteins, ranging from 95–100% for calmodulins, histone H4, and histone H2B, to 75–85% for histone H1, protamine, creatine kinase, acetylcholine re-

ceptor alpha subunit, and alpha crystallin. All other proteins were in an amino acid isology range between 44% and 70%.

In the case of the most conserved proteins, nucleotide isologies ranged from 75% to 90%, with an exceptionally low value, 49%, for the protamine gene. In all other cases, nucleotide isologies were in the 54–68% range; in contrast with the previous genes, nucleotide isology was equal or, more often, higher than amino acid isology for the less conserved proteins.

The data of Table 1 also show the expected decrease in the number of changes for any given gene when going from the third to the first and to the second position; on the average, isology levels were about 50%, 30%, and 20% for these positions. In the most conserved genes, histone H4, histone H2B, and calmodulins A and B, no (or almost no) changes were found in second positions. Again as expected, variations in the extent of nucleotide changes decreased from the second to the first to the third position; in the latter case, most of the values were in the 50–60% range, with only a few lower (30–40%) and higher (80%) values.

Finally, the nucleotide changes under consideration were characterized by levels of transversions which in most cases were higher than those of transitions. This finding suggests a "multiple hit" situation which is not surprising in view of the large evolutionary distances under consideration. An extremely low level of the transitions/transversions ratio, 0.17, was found for the protamine gene, whereas the highest values, 1.78 and 2.25, were found for calmodulins A and B.

The GC Changes in Coding Sequences from Warm-Blooded and Cold-Blooded Vertebrates

In most cases, the nucleotide changes in coding sequences just discussed were accompanied by GC changes. Table 2 shows, for the three coding positions and for the entire coding sequences, the number of nucleotide changes causing increases (columns +), no alteration (columns =), and decreases in GC levels (columns −), as found in warm-blooded vertebrates relative to isologous sequences from cold-blooded vertebrates.

The data of Table 2 are further elaborated in Table 3, which presents the R ratios of GC increases over GC decreases (column + values over column − values of Table 2) for the coding sequences and for the three codon positions. Table 3 also presents the GC levels of the coding sequences, their differences, and their ratios.

The data of Table 3 indicate that the first 12 genes listed present GC increases in their coding sequences; for each gene, increases range from 4% to 13%; this corresponds to 10–30% increases in the GC levels. As already mentioned, these increases were also estimated as the ratios (R1, R2, R3, and Rt) of GC increases over GC decreases for each codon position and for the total coding sequences. Expectedly, the largest changes in these ratios were those of R3, which ranged from about two to about six.

The following set of six genes showed variations in GC within the extreme limits of ±10%; in most cases, however, GC levels did not change. R3 values were between 0.8 and 1.6, with two exceptionally lower values, 0.3.

Finally, the last three genes listed showed GC decreases that were moderate for calmodulin B and glucagon, but quite strong for fibrinogen. In the latter case, the GC level decreased by 30% and R3 was as low as 0.1. It should be recalled, however, that the fibrinogen comparison involved the largest phylogenetic distance among those under consideration, namely that between lamprey and man, and that only 70% of the sequence could be aligned and compared. Interestingly, one of the exceptionally low R3 values of the previous class of genes also concerned a gene (immunoglobulin variable segment) that could only be aligned over 76% of the sequence.

Table 4 summarizes the results of Table 3 by providing average $\bar{R}$ values for total coding sequences and third codon positions. If the log $\bar{R}3$ values of Table 4 are taken into consideration, the isologous genes investigated here fall very neatly into the three different classes mentioned above, the limits of the central class being taken as ±0.25.

The most abundant class showed higher GC levels in all warm-blooded vertebrates tested, compared to cold-blooded vertebrates. The $\bar{R}3$ ratios of these genes ranged from 6.2 to 1.8. $\bar{R}t$ ratios ranged from 3.8 to 1.3 (with the exception of histone H2B). A second class showed $\bar{R}3$ values close to unity, in the 0.6–1.3 range. In this case, $\bar{R}t$ ratios ranged from 0.8 to 1.2. The third class showed $\bar{R}3$ values lower than 0.6 and $\bar{R}t$ ratios in the 0.3–0.7 range.

If the data of Tables 2, 3, or 4 are compared with those of Table 1, it appears that there is no correlation between the percentage changes of genes and their belonging to any one of the three classes just discussed, except that the most conserved genes were concentrated in the second class.

The compositional changes associated with the transition from cold-blooded to warm-blooded vertebrates were also compared with those occurring in the same genes within the warm-blooded vertebrates (or, in one case, within cold-blooded vertebrates). In this case (Table 5), since the evolutionary directionality that exists between cold-blooded and warm-blooded vertebrates is absent, all possible pairwise comparisons were made. $\bar{R}t$, $\bar{R}3$, and log

Table 1. Changes in homologous nucleotide and amino acid sequences from warm-blooded and cold-blooded vertebrates

Genes and species[a]	Changes[b]					
	Codon positions				Total sequences	
	1st		3rd	Tt/Tv	Nucleotides	Amino acids
Acetylcholine receptor alpha subunit						
Ray/bovine	68 (15)	39 (9)	268 (60)	1.16	375 (72)	88 (80)
Ray/human	70 (16)	37 (8)	250 (56)	1.04	357 (74)	90 (80)
Histone H2B						
Trout/chicken	9 (7)	5 (4)	41 (33)	0.62	55 (85)	6 (95)
Alpha-globin						
Xenopus/chicken	46 (32)	42 (29)	78 (55)	0.68	166 (61)	65 (55)
Xenopus/mouse	48 (44)	41 (29)	79 (55)	0.77	168 (61)	55 (62)
Xenopus/rabbit	47 (33)	38 (27)	81 (57)	0.87	166 (61)	59 (59)
Xenopus/human	48 (44)	35 (24)	83 (58)	0.87	166 (61)	53 (63)
Crystallin gamma 2						
Frog/mouse	46 (27)	32 (19)	88 (52)	1.27	166 (67)	66 (63)
Frog/rat	46 (27)	32 (19)	86 (51)	1.22	164 (68)	62 (63)
Frog/bovine	39 (33)	33 (19)	93 (55)	1.32	165 (68)	61 (64)
Frog/human	45 (27)	36 (21)	85 (50)	1.13	166 (67)	66 (61)
Acetylcholine receptor gamma subunit						
Ray/chicken	147 (29)	95 (19)	292 (58)	0.94	534 (65)	190 (62)
Enkephalin						
Xenopus/rat	43 (20)	40 (19)	123 (58)	0.98	206 (68)	71 (67)
Xenopus/bovine	53 (25)	39 (18)	123 (57)	0.90	215 (67)	75 (65)
Xenopus/human	53 (25)	39 (18)	116 (54)	0.89	208 (68)	71 (67)
Homeo box						
Xenopus/mouse	20 (28)	15 (21)	38 (54)	0.52	73 (66)	25 (65)
Proopiomelanocortin						
Salmon/human	46 (46)	33 (26)	66 (52)	0.79	145 (62)	66 (48)
Histone H1						
Trout/chicken	38 (19)	26 (13)	79 (39)	0.47	143 (76)	45 (78)
Creatine kinase						
Torpedo/chicken	51 (13)	34 (9)	171 (45)	0.77	256 (78)	63 (84)
Torpedo/rat	45 (12)	35 (9)	151 (40)	0.70	231 (80)	58 (85)
Torpedo/dog	63 (16)	49 (13)	160 (42)	0.72	272 (76)	86 (77)
Beta-globin						
Xenopus/mouse	66 (45)	50 (34)	89 (61)	0.71	205 (54)	82 (44)
Xenopus/rabbit	66 (45)	43 (29)	80 (54)	0.75	189 (57)	80 (46)
Xenopus/bovine	58 (40)	42 (29)	85 (58)	0.68	185 (58)	74 (49)
Xenopus/human	64 (44)	46 (31)	90 (61)	0.77	200 (55)	80 (46)
Vitellogenin						
Xenopus/chicken	26 (37)	19 (27)	40 (56)	0.70	85 (60)	37 (48)
Histone H4						
Trout/chicken	0 (0)	0 (0)	38 (37)	0.65	38 (88)	0 (100)
Trout/mouse	2 (2)	0 (0)	48 (46)	0.92	50 (84)	0 (100)
Trout/human	8 (8)	0 (0)	52 (50)	0.94	60 (81)	4 (96)
Alpha crystallin						
Frog/mouse	28 (19)	13 (9)	68 (45)	0.88	109 (76)	32 (79)
Frog/rat	22 (18)	12 (10)	57 (47)	0.94	91 (75)	23 (81)
Protamine						
Trout/mouse	16 (59)	4 (15)	21 (78)	0.17	41 (49)	7 (74)
Calmodulin A						
Xenopus/chicken	5 (3)	0 (0)	45 (30)	1.78	50 (89)	0 (100)

Table 1. Continued

Genes and species[a]	Changes[b]					
	Codon positions				Total sequences	
	1st		3rd	Tt/Tv	Nucleotides	Amino acids
Immunoglobulin						
Caiman/mouse	51 (44)	39 (33)	64 (55)	0.54	154 (56)	67 (43)
Caiman/rabbit	29 (38)	27 (35)	30 (39)	0.95	86 (63)	41 (47)
Caiman/human	39 (33)	34 (29)	54 (46)	0.70	127 (64)	56 (52)
Somatostatin						
Anglerfish/human	43 (37)	41 (36)	72 (63)	0.66	156 (55)	64 (44)
Catfish/human	44 (39)	33 (29)	71 (62)	0.49	148 (57)	64 (44)
Calmodulin B						
Xenopus/chicken	5 (3)	0 (0)	60 (40)	2.25	65 (86)	0 (100)
Glucagon						
Anglerfish/human	50 (41)	34 (28)	77 (63)	0.77	161 (56)	61 (50)
Fibrinogen						
Lamprey/human	117 (46)	94 (29)	212 (65)	0.75	423 (57)	148 (55)

[a] GenBank mnemonics for the sequences used are: Acetylcholine receptor alpha subunit: FSCACHRA.PE1/BOVACHRA.PE1; HUMACHRA1.PE1. Histone H2B: FSBHIS42B.PE2/CHKH2BA.PE1. Alpha-globin: XELHBAT3X.PE1/CHKHBAM.PE1; MUSHBA.PE1; RABHBA.PE1; HUMHBA4.PE1. Crystallin gamma 2: FRGCRYG2.PE1/MUSCRYG2.PE1; RATCRYG.PE1; BOVCRYG.PE1; HUMCRYGX3.PE1. Acetylcholine receptor gamma subunit: FSCACHRGS.PE1/CHKACHR2.PE1. Enkephalin: XELENKA1.PE1/RATENK2.PE1; BOVENKEPH.PE1; HUMENK1.PE1. Homeo box: XELHOMMM3.PE1/MUSHOM. Proopiomelanocortin: FSBPOMC.PE1/HUMPOMC.PE1. Histone H1: FSBHIS1.PE1/CHKH1G.PE1. Creatine kinase: FSCCK.PE1/CHKCKMX.PE1; RATCKM.PE1; DOGCKB.PE1. Beta-globin: XELHBBC.PE1/MUSHBBMAJ.PE1; RABHBB1.PE1; BOVHBB.PE1; HUMHBB.PE5. Vitellogenin: XELB1VIT1.PE1/CHKVITII2.PE1. Histone H4: FSBHIS42B.PE1/CHKH234G.PE5; MUSHIST4.PE1; HUMH4.PE1. Alpha crystallin: RANCRYA2.PE1/MUSCRYAA.PE1; RATCRYA.PE1. Protamine: FSBPRCPC2.PE1/MUSPRO1S.PE1. Calmodulin A: XELCAMA.PE1/CHKCAM1.PE1. Immunoglobulin: CRCIGHV.PE1/MUSIGHAP.PE1; RABHAB.PE1; HUMIGHVA.PE1. Somatostatin: FSBSOMI.PE1/HUMSOMI.PE1; FSBSOM14.PE1/HUMSOMI.PE1. Calmodulin B: XELCAMB.PE1/CHKCAM1.PE1. Glucagon: FSBAFGI.PE1/HUMGG.PE1. Fibrinogen: FSBFBRG.PE1/HUMFBRG.PE1

[b] For each comparison, the numbers of nucleotide and amino acid changes are indicated. Values in parentheses are percentage values, or isology percentages for total nucleotide and amino acid changes. The Tt/Tv ratio is the ratio of transitions over transversions

$\bar{R}3$ values indicate that GC changes within warm-blooded vertebrates do not show the directionality of those occurring between cold-blooded and warm-blooded vertebrates. This result is not surprising in view of the unidirectional compositional changes exhibited by different warm-blooded vertebrate genes relative to isologous cold-blooded reference genes. The strong standard deviation exhibited by the alpha-globin gene is correlated with the different localization (in H2 or H3 compartments) of this gene in different warm-blooded vertebrates (Bernardi et al. 1985); in contrast, that shown by the immunoglobulin gene is likely to be correlated with the alignment problems presented by this gene.

The Meaning of GC Changes in Isologous Coding Sequences from Warm-Blooded and Cold-Blooded Vertebrates

The GC increases in most of the coding sequences studied can be understood in the light of previous work (Bernardi et al. 1985; Bernardi and Bernardi 1986a,b; Mouchiroud et al. 1987). Indeed, we know (1) that regional GC increases in the genome have accompanied the transition from cold-blooded to warm-blooded vertebrates; (2) that genes located in the GC-rich isochores of warm-blooded vertebrates have higher GC levels compared to genes located in GC-poor isochores; (3) that GC-rich genes are predominant over GC-poor genes in the genomes of warm-blooded vertebrates.

On the basis of these points, and of specific examples (those of alpha- and beta-globin genes from *Xenopus*, chicken, and three mammals), it was postulated that the GC-poor genes from cold-blooded vertebrates, which were located in GC-rich isochores in warm-blooded vertebrates, had changed the base composition of their coding sequences by point mutation events (Bernardi et al. 1985).

The present results provide a definitive demonstration for this conclusion. Indeed, in the most abundant class of genes investigated here, GC-poor coding sequences from cold-blooded vertebrates did increase their GC levels by point mutations at the time of the transition between cold-blooded and warm-blooded vertebrates, since isologous genes

Table 2. Compositional (GC) changes in homologous coding sequences from warm-blooded and cold-blooded vertebrates

Gene and species[a]	Changes											
	1st position			2nd position			3rd position			Total[b]		
	+	=	−	+	=	−	+	=	−	+	=	−
Acetylcholine receptor alpha subunit												
Ray/bovine	35	15	18	15	10	14	188	49	31	238 (63)	74 (20)	63 (17)
Ray/human	33	18	19	15	9	13	174	48	28	222 (62)	75 (21)	60 (17)
Histone H2B												
Trout/chicken	4	2	3	—	1	4	20	17	4	24 (44)	20 (36)	31 (20)
Alpha-globin												
Xenopus/chicken	17	9	20	13	15	14	44	22	12	74 (44)	46 (28)	46 (28)
Xenopus/mouse	19	14	15	15	17	9	36	24	19	70 (42)	55 (33)	43 (25)
Xenopus/rabbit	18	9	20	11	16	11	52	20	9	81 (49)	45 (27)	40 (24)
Xenopus/human	21	12	15	13	12	10	54	22	7	88 (53)	46 (28)	32 (19)
Crystallin gamma 2												
Frog/mouse	25	7	14	18	6	8	61	17	10	104 (63)	30 (18)	32 (19)
Frog/rat	25	7	14	18	6	8	56	19	11	99 (60)	32 (19)	33 (20)
Frog/bovine	21	5	13	19	7	7	57	13	23	97 (59)	25 (15)	43 (26)
Frog/human	27	9	9	18	11	7	57	16	12	102 (61)	36 (22)	28 (17)
Acetylcholine receptor gamma subunit												
Ray/chicken	74	39	34	42	28	25	180	70	42	296 (55)	137 (26)	101 (19)
Enkephalin												
Xenopus/rat	24	8	11	12	9	19	77	30	16	113 (55)	47 (23)	46 (22)
Xenopus/bovine	30	11	12	11	12	16	78	27	18	119 (55)	50 (23)	46 (21)
Xenopus/human	28	12	13	12	8	19	64	29	23	104 (50)	49 (23)	55 (26)
Homeo box												
Xenopus/mouse	10	7	3	5	4	6	15	18	5	30 (41)	29 (40)	14 (19)
Proopiomelanocortin												
Salmon/human	24	11	11	14	6	13	26	30	10	64 (44)	47 (32)	34 (23)
Histone H1												
Trout/chicken	12	15	11	12	9	5	26	42	11	50 (35)	66 (46)	27 (19)
Creatine kinase												
Torpedo/chicken	24	12	15	9	13	12	83	57	31	116 (45)	82 (32)	58 (23)
Torpedo/rat	17	17	11	7	17	11	59	53	39	83 (36)	87 (38)	61 (26)
Torpedo/dog	36	10	17	18	21	10	71	63	26	125 (46)	94 (35)	53 (19)
Beta-globin												
Xenopus/mouse	25	23	18	18	12	20	45	23	21	88 (43)	58 (28)	59 (29)
Xenopus/rabbit	29	20	17	12	12	19	39	23	18	80 (42)	55 (29)	54 (29)
Xenopus/bovine	25	17	16	10	13	19	42	23	20	77 (42)	53 (29)	55 (29)
Xenopus/human	29	21	14	15	12	19	44	25	21	88 (44)	58 (29)	54 (27)
Vitellogenin												
Xenopus/chicken	8	10	8	7	5	7	18	12	10	33 (39)	27 (32)	25 (29)
Histone H4												
Trout/chicken	—	—	—	—	—	—	23	15	0	23 (61)	15 (39)	0 (0)
Trout/mouse	1	—	1	—	—	—	19	17	12	20 (40)	17 (34)	13 (26)
Trout/human	1	1	6	—	—	—	19	14	19	20 (33)	15 (25)	25 (42)
Alpha crystallin												
Frog/mouse	14	6	8	2	8	3	25	18	25	41 (38)	32 (29)	36 (33)
Frog/rat	11	6	5	2	7	3	23	15	19	36 (39)	28 (31)	27 (30)
Protamine												
Trout/mouse	6	1	9	2	2	—	5	10	6	13 (32)	13 (32)	15 (36)
Calmodulin A												
Xenopus/chicken	3	—	2	—	—	—	16	10	19	19 (38)	10 (20)	21 (42)

Table 2. Continued

Gene and species[a]	Changes											
	1st position			2nd position						Total[b]		
							±			+	=	
Immunoglobulin												
Caiman/mouse	10	18	23	11	11	17	9	26	29	30 (19)	55 (36)	69 (45)
Caiman/rabbit	8	9	12	8	8	11	7	13	10	23 (27)	30 (35)	33 (38)
Caiman/human	11	9	19	14	7	13	15	23	16	40 (31)	39 (31)	48 (38)
Somatostatin												
Anglerfish/human	14	13	16	13	12	16	10	29	33	37 (24)	54 (34)	65 (42)
Catfish/human	14	19	11	12	6	15	17	33	21	43 (29)	58 (39)	47 (32)
Calmodulin B												
Xenopus/chicken	3		2				16	9	35	19 (29)	9 (14)	37 (57)
Glucagon												
Anglerfish/human	18	13	19	15	8	11	18	18	41	51 (32)	39 (24)	71 (44)
Fibrinogen												
Lamprey/human	28	33	56	30	34	30	19	40	153	77 (18)	107 (25)	239 (57)

[a] For mnemonics see footnote a of Table 1
[b] Values in parentheses are the percentage values

from warm-blooded vertebrates showed higher GC levels in *all* species considered (see below for a discussion of orthologous or paralogous relationships between these genes and those of cold-blooded vertebrates). Interestingly, this GC increase affects, as a rule, all three codon positions; a "compensatory" effect can, however, be detected in some cases at second codon positions.

The localization of mammalian beta-globin genes in L2 compartments and of mammalian alpha-globin and chicken alpha- and beta-globin genes in H2 or H3 compartments, whereas both genes are located in the L1 compartment of the *Xenopus* genome (Bernardi et al. 1985), fits with the known linear relationship between GC levels of coding sequences and GC levels of the isochores containing them.

The other two classes of genes reflect two alternative situations. The second class did not undergo any important compositional change. As already stressed, this class of genes underwent the same extent of third position changes as the other two classes of genes; whereas first and second positions showed either high levels of change or low levels in the case of the most conserved genes (histone H4 and calmodulin). Interestingly, another highly conserved gene, histone H2B, does belong to the first class, confirming the lack of relationship between GC increase and extent of nucleotide change. The genes belonging to this class are located, in all likelihood, in isochores having the same composition as the isochores harboring the isologous genes in cold-blooded vertebrates.

Finally, the third class can be interpreted as formed by genes that have undergone GC increases in cold-blooded vertebrates *after* the divergence of warm-blooded vertebrates. It is known that a small minority of cold-blooded vertebrates have also undergone, although to a much lesser extent, compositional changes similar to those mentioned above for warm-blooded vertebrates. Some reptiles, some fishes, and some amphibians exhibit this phenomenon, which is, apparently, due to the same cause underlying the formation of GC-rich isochores in warm-blooded vertebrates, namely increased environmental temperature (Bernardi and Bernardi 1986b). This explanation would be supported by the finding that the particular genes showing higher GC levels in the cold-blooded vertebrates investigated show lower levels in other cold-blooded vertebrates. Unfortunately, only an extremely small number of such comparisons are possible at the present time because of lack of data. The insulin-coding sequences fulfill this expectation, being low in GC in the carp (51% in third position), but high in anglerfish (63%) and in salmon (65%); incidentally, these genes were not analyzed here because of alignment problems. It should be stressed that the third class of genes definitely does not represent a "compensation" of the GC increases found in the first class, since GC-poor genes represent the vast majority of cold-blooded vertebrate genes, but only a minority of those from warm-blooded vertebrates (Bernardi et al. 1985; Mouchiroud et al. 1987).

A final question concerns the validity of the sequence comparisons made in the present work. So far we have refrained from deciding whether the comparisons made among homologous genes were

Table 3. Compositional (GC) ratios in homologous coding sequences from warm-blooded and cold-blooded vertebrates[a]

Gene and species[b]	GC%	GC	RGC	Rt	R1	R2	R3
Acetylcholine receptor alpha subunit							
Ray/bovine	39/52	13	1.3	3.8	1.9	1.1	6.1
Ray/human	39/51	12	1.3	3.7	1.7	1.1	6.2
Histone H2B							
Trout/chicken	59/62	3		2.2	1.3		5.0
Alpha-globin							
Xenopus/chicken	52/58	6	1.1	1.6	0.8	0.9	3.7
Xenopus/mouse	52/58	6	1.1	1.6	1.3	1.7	1.9
Xenopus/rabbit	52/61	9	1.2	2.0	0.9	1.0	5.8
Xenopus/human	52/65	13	1.2	2.8	1.4	1.3	7.7
Crystallin gamma 2							
Frog/mouse	46/60	14	1.3	3.3	1.8	2.2	6.1
Frog/rat	46/59	13	1.3	3.0	1.8	2.2	5.1
Frog/bovine	46/56	10	1.2	2.3	1.6	2.7	2.5
Frog/human	46/60	14	1.3	3.6	3.0	2.6	4.7
Acetylcholine receptor gamma subunit							
Ray/chicken	44/57	13	1.3	2.9	2.2	.7	4.3
Enkephalin							
Xenopus/rat	41/52	11	1.3	2.5	2.2	0.6	4.8
Xenopus/bovine	41/53	12	1.3	2.6	2.5	0.7	4.3
Xenopus/human	41/49	8	1.2	1.9	2.1	0.6	2.8
Homeo box							
Xenopus/mouse	55/62	7	1.1	2.1	3.3	0.8	3.0
Proopiomelanocortin							
Salmon/human	59/67	8	1.1	1.9	2.2	1.1	2.6
Histone H1							
Trout/chicken	64/68	4	1.1	1.8	1.1	2.4	2.4
Creatine kinase							
Torpedo/chicken	56/61	5	1.1	2.0	1.6	0.8	2.7
Torpedo/rat	56/58	2	1.0	1.4	1.5	0.6	1.5
Torpedo/dog	56/63	7	1.1	2.4	2.1	1.8	2.7
Beta-globin							
Xenopus/mouse	49/55	6	1.1	1.5	1.4	0.9	2.1
Xenopus/rabbit	49/54	5	1.1	1.5	1.7	0.6	2.2
Xenopus/bovine	49/54	5	1.1	1.4	1.6	0.5	2.1
Xenopus/human	49/56	7	1.1	1.6	2.1	0.8	2.1
Vitellogenin							
Xenopus/chicken	44/48	4	1.1	1.3	.0	1.0	1.8
Histone H4							
Trout/chicken	61/68	7	1.1	–	–	–	–
Trout/mouse	61/63	2	1.0	1.5	1.0	–	1.6
Trout/human	61/59	−2	1.0	0.8	0.2	–	1.0
Alpha crystallin							
Frog/mouse	56/57	1	1.0	1.1	1.7	0.7	1.0
Frog/rat	55/57	2	1.0	1.3	2.2	0.7	1.2
Protamine							
Trout/mouse	74/72	−2	1.0	0.9	0.7		0.8
Calmodulin A							
Xenopus/chicken	41/41	0	1.0	0.9	1.5		0.8
Immunoglobulin							
Caiman/mouse	59/48	−11	0.8	0.4	0.4	0.6	0.3
Caiman/rabbit	61/56	−5	0.9	0.7	0.7	0.7	0.7
Caiman/human	59/57	−2	1.0	0.8	0.6	1.1	0.9

Table 3. Continued

Gene and species[b]	GC%	GC	RGC	Rt	R1	R2	R3
Somatostatin							
Anglerfish/human	67/59	−8	0.9	0.6	0.9	0.8	0.3
Catfish/human	64/59	−5	0.9	0.9	1.3	0.8	0.8
Calmodulin B							
Xenopus/chicken	45/41	−4	0.9	0.5	1.5		0.5
Glucagon							
Anglerfish/human	53/47	−6	0.9	0.7	0.9	1.4	0.4
Fibrinogen							
Lamprey/human	59/42	−17	0.7	0.3	0.5	1.0	0.1

[a] The GC column gives the GC values of the coding sequence pair under consideration. Other columns concern warm-blooded/cold-blooded GC differences (ΔGC), GC ratios for overall GC levels (RGC), GC increase/GC decrease ratios (columns +/− from Table 2) for complete coding sequences (Rt), and for codon positions 1, 2, and 3 (R1, R2, R3)

[b] For mnemonics see footnote a of Table 1

orthologous or paralogous. It should be pointed out, however, that, in this particular case, both orthologous and paralogous comparisons would be valid. Indeed, (1) the phylogenetic distances between cold-blooded and warm-blooded vertebrates are very large compared to those between duplicated genes and their "master copies" in warm-blooded vertebrates (essentially, gene duplications need to be considered for the large genomes of warm-blooded vertebrates rather than for the small ones of cold-blooded vertebrates); and (2) alignments were excellent for most genes.

Table 4. Summary of compositional comparisons between homologous coding sequences from cold-blooded and warm-blooded vertebrates

Gene[a]	R̄t	R̄3	log R̄3[b]
Acetylcholine receptor alpha subunit	3.8 ± 0.1	6.2 ± 0.1	0.79
Histone H2B	2.2	5.0	0.70
Alpha-globin	2.0 ± 0.5	4.8 ± 2.2	0.68
Crystallin gamma 2	3.1 ± 0.5	4.6 ± 1.3	0.66
Acetylcholine receptor gamma subunit	2.9	4.3	0.63
Enkephalin	2.3 ± 0.3	4.0 ± 0.9	0.60
Homeo box	2.1	3.0	0.48
Proopiomelanocortin	1.9	2.6	0.41
Histone H1	1.8	2.4	0.38
Creatine kinase	1.9 ± 0.4	2.3 ± 0.6	0.36
Beta-globin	1.5 ± 0.1	2.1 ± 0	0.32
Vitellogenin	1.3	1.8	0.26
Histone H4	1.0	1.3 ± 0.3	0.11
Alpha crystallin	1.2 ± 0.1	1.1 ± 0.1	0.04
Protamine	0.9	0.8	−0.10
Calmodulin A	0.9	0.8	−0.10
Immunoglobulin	0.6 ± 0.2	0.6 ± 0.3	−0.22
Somatostatin	0.8 ± 0.2	0.6 ± 0.3	−0.22
Calmodulin B	0.5	0.5	−0.30
Glucagon	0.7	0.4	−0.40
Fibrinogen	0.3	0.1	−1

[a] For mnemonics see footnote a of Table 1

[b] These values have been used to group the genes into three classes, the central class being between +0.25 and −0.25 (see text)

Conclusions

The majority of the coding sequences from warm-blooded vertebrates studied in the present work show a very striking compositional directionality in the mutations undergone since their divergence from their cold-blooded ancestors. Very interestingly, this directionality is evident, with only relatively minor variations, in all comparisons between cold-blooded reference genes and isologous genes from different warm-blooded vertebrates. In other words, compositional changes undergone since the divergence of different warm-blooded vertebrates appear to be relatively negligible in extent, and, more so, in directionality, compared to those that occurred between them and their cold-blooded ancestors. This points to a change that should be traced back to the transition between cold-blooded and warm-blooded vertebrates and that was followed by a composi-

Table 5. Summary of compositional comparisons between homologous genes from warm-blooded vertebrates (A) and from cold-blooded vertebrates (B)

Gene-reference species	R̄t	R̄3	log R̄3
A. Acetylcholine receptor alpha subunit	1.1 ± 0.4	1.1 ± 0.4	0.04
Alpha-globin	1.3 ± 0.8	1.6 ± 1.6	0.20
Crystallin gamma 2	1.1 ± 0.5	1.3 ± 0.7	0.11
Creatine kinase	1.1 ± 0.5	1.2 ± 0.6	0.08
Beta-globin	1.1 ± 0.3	1.0 ± 0.1	0
Immunoglobulin	1.2 ± 0.8	1.7 ± 1.9	0.23
B. Somatostatin	2.6	1.4	0.15

tional conservation. Such compositional conservation was also followed in the coding sequences that do not show any significant change between warm-blooded and cold-blooded vertebrates. The explanation provided above for the case of GC decreases in coding sequences from cold-blooded vertebrates points in the same direction as the GC increases, in that the same presumable cause, body temperature increase (Bernardi and Bernardi 1986a,b), yielded the same consequences.

As already stressed elsewhere (Bernardi and Bernardi 1986a,b), the directionality of the compositional changes and their evolutionary conservation strongly point to compositional constraints on the genome and on its compartments in vertebrates. In turn, these constraints do not appear to be compatible with the neutral theory of evolution (Kimura 1968, 1983, 1986). Indeed, if most third position changes were neutral, as claimed, one would not expect compositional constraints which depend upon environmental factors. We will not reiterate here the arguments already made elsewhere, but simply state that the detailed analyses presented in this work provide definitive support for them.

Acknowledgments. One of us (P.P.) thanks the Fondation pour la Recherche Médicale for its financial support. We thank Prof. R. Grantham for his interest, encouragement, and suggestions, D. Mouchiroud and Giacomo Bernardi for discussions and help in the preparation of this manuscript, and Martine Brient for typing it.

References

Bernardi G, Bernardi G (1986a) The human genome and its evolutionary context. Cold Spring Harbor Symp Quant Biol 51:479–487

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

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 118:953–958

Bird A (1986) CpG-rich islands and the function of DNA methylation. Nature 321:209–213

Cuny G, Soriano P, Macaya G, Bernardi G (1981) The major components of the mouse and human genomes. I. Preparation, basic properties and compositional heterogeneity. Eur J Biochem 115:227–233

Dynan WS (1986) Promoters for housekeeping genes. Trends Genet 2:196–197

Gouy M, Milleret F, Mugnier C, Jacobzone M, Gautier C (1984) ACNUC: a nucleic acid sequence data base and analysis system. Nucleic Acids Res 12:121–127

Gouy M, Gautier C, Milleret F (1985) System analysis and nucleic acid sequence banks. Biochimie 67:433–436

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626

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

Kimura M (1986) DNA and the neutral theory. Philos Trans R Soc Lond (Biol) 312:343–354

Medrano L, Bernardi G, Couturier J, Dutrillaux B, Bernardi G (1988) Chromosome banding and genome compartmentalization in fishes. Chromosoma (in press)

Mouchiroud D, Fichant G, Bernardi G (1987) Compositional compartmentalization and gene composition in the genome of vertebrates. J Mol Evol (in press)

Salinas J, Zerial M, Filipski J, Bernardi G (1986) Gene distribution and nucleotide sequence organization in the mouse genome. Eur J Biochem 160:469–478

Smith TF, Waterman MS (1981) Comparison of biosequences. Adv Appl Math 2:482–489

Thiery JP, Macaya G, Bernardi G (1976) An analysis of eukaryotic genomes by density gradient centrifugation. J Mol Biol 108:219–235

Wolf SF, Migeon BR (1985) Clusters of CpG dinucleotides implicated by nuclease hypersensitivity as control elements of housekeeping genes. Nature 314:467–469

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

Received July 31, 1987/Revised and accepted October 1, 1987