J Mol Evol (1990) 31:282–293

Journal of
Molecular Evolution

Compositional Transitions in the Nuclear Genomes of Cold-Blooded Vertebrates

Giacomo Bernardi and Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, Tour 43, 2 Place Jussieu, 75005 Paris, France

Summary. The compositional properties of DNAs from 122 species of fishes and from 18 other cold-blooded vertebrates (amphibians and reptiles) were compared with those from 10 warm-blooded vertebrates (mammals and birds) and found to be substantially different. Indeed, DNAs from cold-blooded vertebrates are characterized by much lower intermolecular compositional heterogeneities and CsCl band asymmetries, by a much wider spectrum of modal buoyant densities in CsCl, by generally lower amounts of satellites, as well as by the fact that in no case do buoyant densities reach the high values found in the GC-richest components of DNAs from warm-blooded vertebrates.

In the case of fish genomes, which were more extensively studied, different orders were generally characterized by modal buoyant densities that were different in average values as well as in their ranges. In contrast, different families within any given order were more often characterized by narrow ranges of modal buoyant densities, and no difference in modal buoyant density was found within any single genus (except for the genus *Aphyosemion*, which should be split into several genera).

The compositional differences that were found among species belonging to different orders and to different families within the same order are indicative of compositional transitions, which were shown to be essentially due to directional base substitutions. These transitions were found to be independent of geological time. Moreover, the rates of directional base substitutions were found to be very variable and to reach, in some cases, extremely high values, that were even higher than those of silent substitutions in primates. The taxonomic and evolutionary implications of these findings are discussed.

Key words: Genome — Isochores — Vertebrates — Directional mutations

Introduction

Earlier investigations from our laboratory revealed that the DNAs of cold-blooded and warm-blooded vertebrates are characterized by strikingly different compositional properties (Macaya et al. 1976; Thiery et al. 1976; Cortadas et al. 1979; Bernardi et al. 1985, 1988; Bernardi 1989). The former are generally characterized by only slightly asymmetrical CsCl bands and by low levels of intermolecular compositional heterogeneities. The latter exhibit strongly asymmetrical CsCl bands and high levels of intermolecular compositional heterogeneities; both differences are due to the existence in warm-blooded vertebrates of GC-rich DNA components (namely, families of DNA fragments in the 50–100-kb size range having a similar composition) that are absent or very scarcely represented in most cold-blooded vertebrates. These differences were initially established in investigations concerning 11 species of mammals, 2 species of birds, and 6 species of cold-blooded vertebrates.

Later expansions of the sample of cold-blooded vertebrates investigated (Hudson et al. 1980; Pizon 1983; Pizon et al. 1984; Bernardi and Bernardi 1990) provided results in agreement with the earlier data. Moreover, it was shown that modal buoyant densities of DNAs from cold-blooded vertebrates covered a much wider spectrum than those of DNAs from warm-blooded vertebrates and that satellite

Offprint requests to: Giorgio Bernardi

DNAs were generally present in lower amounts (Bernardi and Bernardi 1990). Other differences between the genomes of cold-blooded and warm-blooded vertebrates concerned coding sequences (Bernardi et al. 1985, 1988; Mouchiroud et al. 1987, 1988; Perrin and Bernardi 1987). These differences paralleled those found at the DNA level, in that coding sequences from cold-blooded vertebrates were generally characterized by low GC levels, whereas most coding sequences from warm-blooded vertebrates were characterized by high GC levels. Such differences were also found in comparisons of homologous genes.

In the present work, we attempted (1) better to define the features of the large compositional changes of the genome that accompanied the transition from cold- and warm-blooded vertebrates; and (2) to analyze the frequent compositional changes (indicated by the wide spectrum of modal buoyant densities) that took place in the genomes of cold-blooded vertebrates over evolutionary time.

It should be recalled here that two modes of evolution have been recognized in the vertebrate genome on the basis of comparisons of homologous coding sequences (Bernardi et al. 1988). In the conservative mode, which is predominant in the evolution of mammalian genomes (and was also found, at the DNA level, in avian genomes), enormous numbers of nucleotide substitutions accumulated over many millions of years, but the composition of coding sequences (and of their different codon positions, including third positions), of the associated introns, and of the intergenic noncoding sequences remained within very narrow limits. Such an invariance is accompanied by nucleotide divergences as high as 50% (without correction for multiple hits), and by differences in GC levels attaining 50% in the third codon positions of pairs of homologous coding sequences. The conclusion was then drawn that the conservative mode is not only due to some extent of compositional conservation in the nucleotide substitution process itself (namely to the tendency toward equal rates in the forward and backward changes from AT to GC base pairs), but also to negative selection acting at a regional (isochore) level to eliminate any strong deviation from a functionally optimal composition of isochores. If such were not the case, it would be impossible to explain, for instance, why extremely high GC values (in excess of 90%) are conserved in third codon positions of a number of homologous mammalian genes over very extended evolutionary times.

In the shifting (or transitional) mode, large compositional changes occur. In the case of the two independent major transitions that corresponded to the appearance of mammals and birds, respectively, relatively rapid, regional accumulations of nucleotide substitutions biased toward GC increases took place. As already mentioned, comparisons of the compositional properties of coding sequences from cold- and warm-blooded vertebrates (Bernardi et al. 1985, 1988; Mouchiroud et al. 1987, 1988; Perrin and Bernardi 1987) showed that the differences seen at the DNA level were paralleled by differences found at the gene sequence level. Such changes were ascribed to negative selection of isochores with decreased GC levels and positive selection of isochores with increased GC levels, under the influence of the functional advantages associated with such compositional shifts (Bernardi and Bernardi 1986; Bernardi et al. 1988).

The methods used in the present work were described in the preceding article (Bernardi and Bernardi 1990).

Results

Comparisons of DNAs from Cold-Blooded and Warm-Blooded DNAs

In Fig. 1 the modal buoyant densities, ρ_0, of DNAs from Osteichthyes and mammals are compared and it appears that values from mammals correspond to the lower values of Osteichthyes. This comparison is, however, misleading. Indeed, modal buoyant densities characterize DNAs from cold-blooded vertebrates well, since they exhibit low asymmetries and low intermolecular compositional heterogeneities, but modal buoyant densities are inappropriate for DNAs of warm-blooded vertebrates, where the opposite is true. In fact, as shown by Fig. 2, the CsCl band asymmetry, A, and the intermolecular compositional heterogeneity, H, of DNAs from warm-blooded vertebrates are so large as to fall practically outside the range exhibited by the DNAs from cold-blooded vertebrates. It should be stressed that the A and H values used in Fig. 2 and Table 1 (see

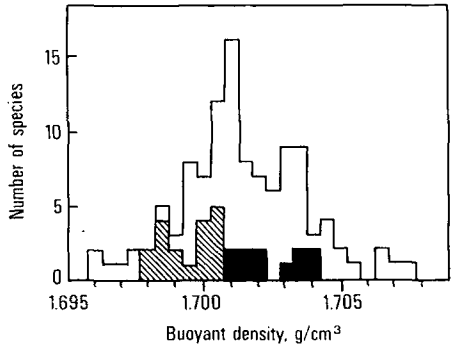

Fig. 1. Modal buoyant densities of DNAs from Osteichthyes (open histogram from Bernardi and Bernardi, 1990) are compared with modal and mean buoyant densities of DNAs from mammals (cross-hatched and solid histograms).

below) for warm-blooded vertebrates concern main bands, and do not include the contribution of satellite DNAs. The latter contribution can usually be estimated by studies involving preparative Cs_2SO_4/Ag^+ and Cs_2SO_4/BAMD density gradient centrifugation (Corneo et al. 1968; Filipski et al. 1973; Thiery et al. 1976; Bernardi and Bernardi 1990).

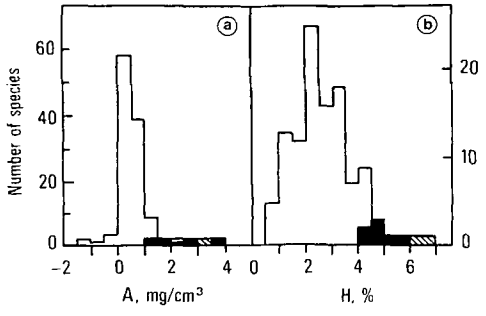

Fig. 2. Comparison of CsCl band asymmetry (**a**) and intermolecular compositional heterogeneity (**b**) of DNAs from Osteichthyes (open histograms; from Bernardi and Bernardi 1990) with DNAs from mammals (solid histogram) and birds (cross-hatched histogram)

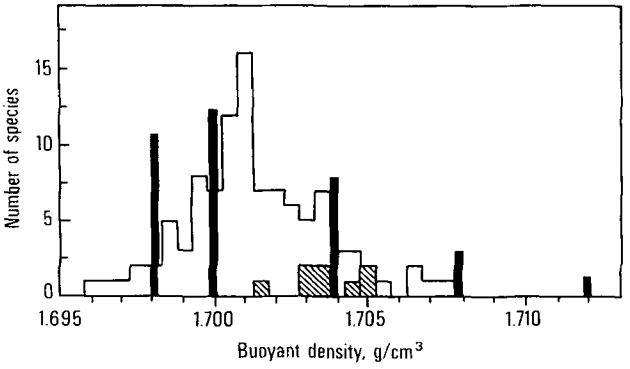

Fig. 3. The histogram of the modal buoyant densities of major DNA components (black bars) from the human genome (which is representative of the genomes from most mammals) is compared with the histogram of modal buoyant densities of the genomes from Osteichthyes (open histogram) and Chondrichthyes (cross-hatched histogram). Major DNA components from murids, cricetids, and spalacids did not reach values higher than 1.710 g/cm³, whereas those of birds reached higher values than those shown (unpublished).

The comparison under consideration can be improved by using (Fig. 1) the mean buoyant densities $\langle\rho\rangle$ of DNAs from mammals, whereas the modal buoyant densities of cold-blooded vertebrates can be used in the comparison, as they are very close to the mean buoyant densities. In this case, the values from warm-blooded vertebrates correspond to the higher values of Osteichthyes.

It should be pointed out, however, that the asymmetry, heterogeneity, and mean buoyant density are calculated by weight-averaging methods; and that the relative amounts of DNA fragments from warm-blooded vertebrates decrease as their GC level increases. As a consequence, the compositional differences between the genomes of cold-blooded and warm-blooded vertebrates, which have been discussed above, are underestimated by the approaches just described. Differences become much more evident when the modal buoyant densities and relative amounts of DNA components from warm-blooded vertebrates (namely, of the families of DNA fragments in the 30–100-kb size range, that are characterized by close GC levels) are taken into consideration. If this is done (Fig. 3), it is clear that DNA components exist in the genomes of warm-blooded vertebrates and they are higher in GC levels than any DNA from cold-blooded vertebrates.

Another way to compare DNAs from cold- and warm-blooded vertebrates is to take into consideration average $\bar{\rho}_0$, $\langle\bar{\rho}\rangle$, $\bar{A}$, and $\bar{H}$ values and their standard deviations. The results of Table 1 indicate that the average modal buoyant densities, $\bar{\rho}_0$, are lower for DNAs from warm-blooded vertebrates compared to DNAs from cold-blooded vertebrates, as expected from the histogram of Fig. 1. The average mean buoyant densities, $\langle\bar{\rho}\rangle$, are essentially

Table 1. Properties of DNAs from vertebrates

	Number of species[a]	$\bar{\rho}_0$ (g/cm³)	$\sigma_{\rho 0}$ (mg/cm³)	$\langle\bar{\rho}\rangle$ (g/cm³)	$\sigma_{\langle\rho\rangle}$ (mg/cm³)	$\bar{A}$ (mg/cm³)	σ_A (mg/cm³)	$\bar{H}$ %GC	σ_H %GC
Chondrichthyes	12 (9)	1.7035	1.1	1.7044	1.4	0.87	0.54	3.4	0.9
Osteichthyes	110 (98)	1.7014	2.2	1.7016	2.2	0.65	0.54	2.6	1.1
Amphibians	5 (3)	1.7020	1.7	1.7026	1.7	0.58	0.13	2.9	0.05
Reptiles	13 (6)	1.7019	1.8	1.7028	1.7	0.73	0.54	3.0	1.05
Mammals	10 (7)	1.7000	0.7	1.7025	1.1	2.5	0.9	4.7	0.4
Birds	2	1.6999	0.1	1.7031	0.0	3.1	0.2	6.7	0.2
Perciformes	30 (29)	1.7004	0.8	1.7009	0.9	0.42	0.57	2.4	0.8
Tetraodontiformes	11	1.7027	1.4	1.7033	1.6	0.56	0.40	2.8	1.1
Murids + cricetids	3	1.7004	0.3	1.7018	0.3	1.43	0.09	4.3	0.3
Other mammals	8 (4)	1.6997	0.8	1.7027	1.1	3.08	0.42	5.0	0.3

[a] Numbers concern the species compared in ρ_0, $\langle\rho\rangle$, and A values; numbers in parentheses concern the species compared in H values, if different. Data from Bernardi and Bernardi (1990) for fishes, amphibians, and reptiles; from Thiery et al. (1976) for mammals and birds (in this case H values were calculated in the present work)

Table 2. Standard deviations of modal buoyant densities of DNAs from different genera, families, and orders of fishes

Genus		Family		Order		
		Carcharinidae (3)	0.60	Lamniformes (5)	0.90	1
				Squaliformes (2)	0.00	2
Torpedo (2)	0.05			Rajiformes (5)	0.98	3
				Osteoglossiformes (3)	0.57	4
Anguilla (2)	0.35					5
		Clupeidae (2)	0.05			6
		Cyprinidae (4)	0.64	Cypriniformes (5)	0.63	7
Salmo (3)	0.25					
Oncorhynchus (3)	0.20	Salmonidae (7)	0.44			8
Synodus (2)	0.20	Synodontidae (3)	0.54			9
Urophycis (2)	0.25			Gadiformes (3)	1.26	10
		Batrachoidae (2)	0.10			11
Aphyosemion (10)	1.83	Aplocheilidae (15)	2.00			
Cyprinodon (4)	0.43	Cyprinodontidae (6)	0.45			
		Poecilidae (2)	0.10	Cyprinodontiformes (23)	2.15	12
				Scorpaeniformes (4)	0.98	13
Scorpaena (2)	0.35					
Epinephelus (2)	0.55	Serranidae (3)	1.42			
Oreochromis (5)	0.27	Cichlidae (8)	0.24			
Sphyraena (2)	0.45					
Trematomus (4)	0.44	Notothenidae (5)	0.42	Perciformes (10)	1.95	14
Limanda (2)	0.85			Pleuronectiformes (3)	1.19	15
		Balistidae (4)	1.39			
Arothron (2)	0.45	Tetraodontidae (4)	1.10	Tetraodontiformes (10)	1.50	16

This table is constructed so as to match exactly Fig. 5. Numbers on the rightmost column refer to the fish orders. Standard deviations equal to, or in excess of, unity are underlined. They correspond to significant differences in the modal buoyant densities of DNAs from the species belonging to the group under consideration. Values in parentheses indicate the numbers of species analyzed. In the case of Perciformes, *Synchiropus splendidus* was not taken into account for reasons given in the text; moreover, if all values (30) for Perciformes were taken into consideration, the standard deviation would decrease to 0.78, because of the large number of very close values

the same for the DNAs of amphibians, reptiles, mammals, and birds, whereas a lower value is found for Osteichthyes and a higher value for Chondrichthyes. When average CsCl band asymmetries and intermolecular compositional heterogeneities are compared, however, warm-blooded vertebrates are characterized by much higher values relative to cold-blooded vertebrates. Moreover, when the standard deviations of $\bar{\rho}_0$, $\langle\bar{\rho}\rangle$, $\bar{A}$, and $\bar{H}$ are compared, DNAs from cold-blooded vertebrates are generally characterized by higher values relative to warm-blooded vertebrates; the low σ values of $\bar{A}$ and $\bar{H}$ for DNAs of amphibians are, in all likelihood, associated with the particular small sample used.

Two remarks concerning these findings are (1) that the 10 species of mammals (belonging to six orders) used in the comparison are those previously studied by Thiery et al. (1976) except for *Glis glis*, which was not taken into account because its CsCl profile was unsatisfactory for heterogeneity analysis; and (2) that even if the number of birds in Table 1 is limited to two, preliminary results from four other species belonging to different orders are in agreement with those results (F. Kadi, G. Sabeur, and G. Bernardi, unpublished). The conclusion should, therefore, be drawn that the comparison made reflects real differences between cold- and warm-blooded vertebrates.

The comparisons of Table 1 reveal two additional points: (1) they show, in yet another way, the existence of differences (Salinas et al. 1986; Zerial et al. 1986; Mouchiroud et al. 1987, 1988; Bernardi et al. 1988) in the compositional patterns of murids and cricetids on the one hand, and other mammals, on the other hand; the former exhibit a much lower asymmetry (1.43 mg/cm^3) compared with the latter (3.08 mg/cm^3), but not a much lower heterogeneity (4.3% GC versus 5.0% GC); (2) they show that different orders of fishes (exemplified by Perciformes and Tetraodontiformes) exhibit very different spectra of modal buoyant densities; this latter point was investigated in more detail (see the following section).

Although the differences displayed by the data of Table 1 are quite evident, they may become more meaningful if comparisons are also made with external references, like bacterial DNAs, and individual DNA components from the genome of warm-blooded vertebrates. Such comparisons were already made for the genomes of 33 fish species (Hudson et al. 1980; Cuny et al. 1981) and showed that the latter exhibited DNA heterogeneities that fell between those of the DNAs from *Haemophilus influenzae* (1.5% GC) and *Escherichia coli* (2.8% GC) and between those of light (1.7–2.9% GC) and heavy components (3.1–3.8% GC) of chicken,

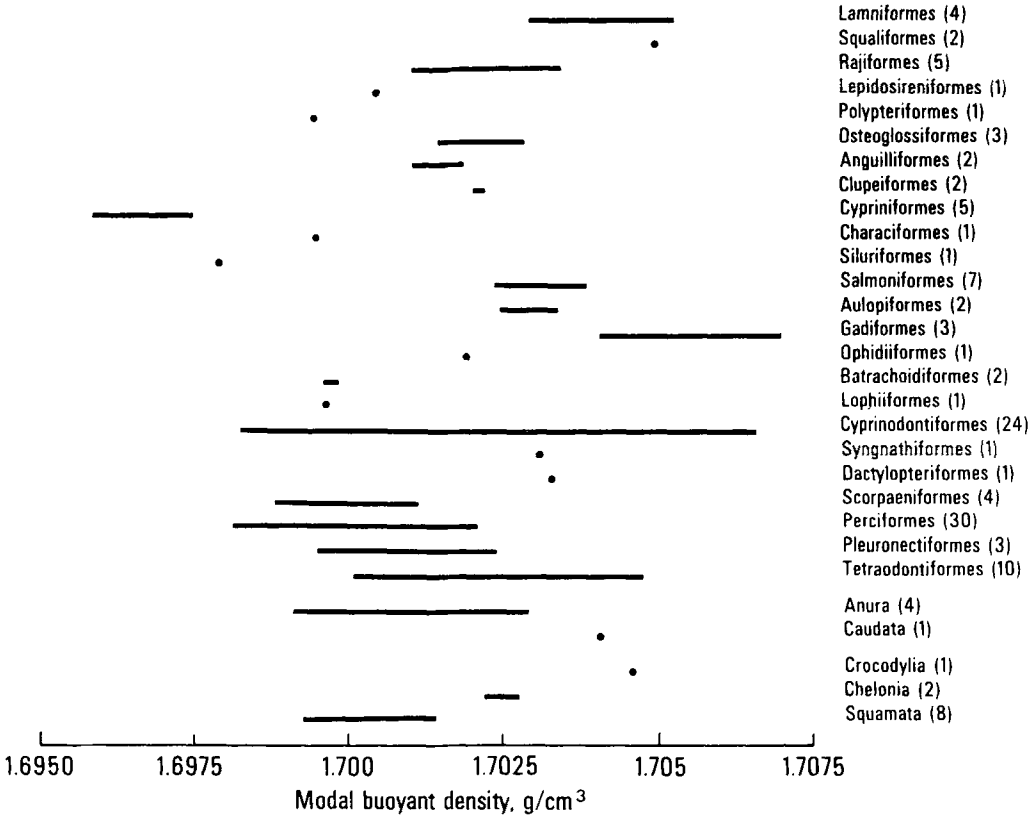

Fig. 4. Comparison of the ranges of modal buoyant densities exhibited by DNAs from species of cold-blooded vertebrates belonging to the same order. Figures in parentheses indicate the numbers of species analyzed. In the case of Perciformes, *Synchiropus splendidus* was not taken into account for reasons given in the text.

mouse, and human DNAs. In this respect, the much larger fish DNA sample studied here, as well as DNAs from amphibians and reptiles, show the same properties as those found previously.

These findings are of interest not only because they provide additional comparisons, but also because they show comparable values of heterogeneity for DNAs from cold-blooded vertebrates and DNAs from bacteria. Because the latter have genome sizes that are smaller by two to three orders of magnitude, this observation can only be understood (Hudson et al. 1980; Cuny et al. 1981) if the genomes of cold-blooded vertebrates are made of compositionally homogeneous regions, the isochores (Bernardi et al. 1985). Likewise, DNAs of cold-blooded vertebrates show asymmetries and heterogeneities that are comparable with those of individual major components from mammalian genomes, in spite of the fact that some of the latter (like the H2 and H3 components) have kinetic complexities that are much smaller (H2 represents about 8% of the genome and H3 about 5% of the mammalian genome).

Comparisons of DNAs from Cold-Blooded Vertebrates

A comparison of the modal buoyant densities of DNAs from cold-blooded vertebrates indicates that different orders are generally characterized by different average values and different ranges of values (see Table 2 and Fig. 4). This point was studied in more detail in Fig. 5, which displays histograms of modal buoyant densities exhibited by DNAs of fish species as analyzed by genus, family, and/or order. Expectedly, the scatter of values generally increased when moving from the genus to the family and to the order, but large scatters were found not only in some orders, but also, although less frequently, in families and even within one genus (which might, however, be considered as several genera; see Discussion).

Table 2 presents the standard deviations of modal buoyant densities corresponding to these groups of DNAs. As already mentioned, because of the generally low asymmetries of fish DNA bands, results concerning mean buoyant densities are similar to those of modal buoyant densities (see Table 3 of Bernardi and Bernardi 1990). In fact, because asymmetries are, in part at least, due to GC-rich cryptic or poorly resolved satellites (Bernardi and Bernardi 1990), modal buoyant density data are likely to be more reliable in the case of fish DNAs. Standard deviations, σ, of values higher than unity correspond to differences that were very significant under our experimental conditions, and definitely indicated, therefore, compositional transitions. Such high val-

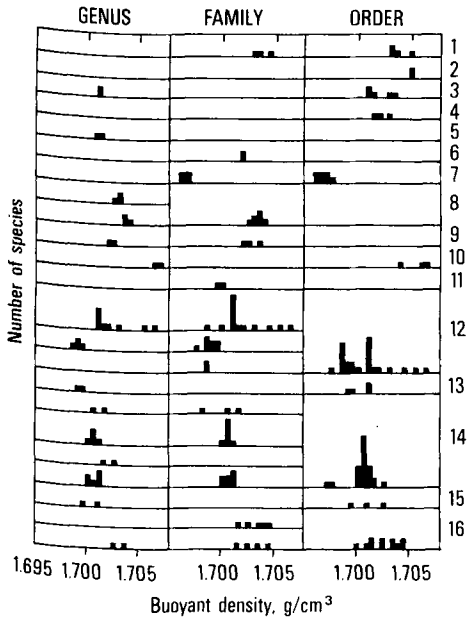

Fig. 5. Histograms showing modal buoyant densities of DNA from fish species belonging to the same genus, family, and/or order. The numbering of the right side of the figure refers to the fish orders (see Table 2). Data are from Table 3 of Bernardi and Bernardi (1990).

ues were found in 1 genus (*Aphyosemion*; see Discussion) out of 16, in 4 families (Aplocheilidae, Serranidae, Balistidae, and Tetraodontidae) out of 14, and in 7 orders (Rajiformes, Gadiformes, Cyprinodontiformes, Scorpaeniformes, Perciformes, Pleuronectiformes, Tetraodontiformes) out of 11.

Discussion

Compositional Transitions in the Genomes of Vertebrates: The Cold- to Warm-Blooded Vertebrate Transition

The differences found between the DNAs from cold- and warm-blooded vertebrates (see the first section of the Results) definitely establish the existence of a major discontinuity between these DNAs. The accompanying changes in coding sequences essentially parallel the DNA changes, in that genes located in the GC-rich isochores of warm-blooded vertebrates show increased GC levels in their coding sequences and introns (Bernardi et al. 1985, 1988; Mouchiroud et al. 1987, 1988; Perrin and Bernardi 1987). Such changes were demonstrated in homologous coding sequences as well, thus unequivocally showing that directional mutations had been fixed in those genes.

It should be stressed that the compositional changes just discussed did not concern the totality of the genome of cold-blooded vertebrates. Indeed, the genome of warm-blooded vertebrates consists (Bernardi 1989) of two main compartments. (1) The paleogenome is similar to what it presumably was, and still is, in cold-blooded vertebrates; this compartment, which corresponds to two-thirds of the genome, is made up of late-replicating, compositionally homogeneous, GC-poor isochores containing GC-poor genes. (2) The neogenome, which corresponds to the remaining third, is characterized by compositional features that are different from what they were in the genomes of cold-blooded vertebrates; in the neogenome, the ancestral, early-replicating, GC-poor isochores were changed into compositionally heterogeneous GC-rich isochores that contain abundant genes. Whether this early-replicating genome compartment, which is not (or is much less) compositionally distinct in cold-blooded vertebrates, already contained (and still contains) a much higher gene concentration than the late-replicating compartment is not known but is possible. Otherwise, the compositional transition from cold- to warm-blooded vertebrates should have been accompanied by massive genome rearrangements, leading to the very high gene concentrations that are found at present in the GC-richest regions of the genome of warm-blooded vertebrates. Some rearrangements definitely took place, however. Suffice it to mention here that the α- and β-globin genes, which are contiguous and GC poor in the genome of *Xenopus*, are on different chromosomes in both mammals and birds. Moreover, in the case of mammals, the α-globin gene cluster became GC rich, whereas the β-globin gene cluster remained GC poor; in the case of birds both the α- and the β-globin gene clusters became GC rich.

Compositional Transitions in the Genomes of Cold-Blooded Vertebrates

The data of Table 2 show significant compositional differences among the DNAs of fishes belonging to the same order and/or family. In fact, compositional transitions are not only frequent in the genomes of fishes, but also in those of other cold-blooded vertebrates, amphibians and reptiles. For instance, the order Anura among amphibians, and the order Sauria among reptiles definitely exhibit compositional transitions in their families ($\sigma > 1$). Here we will concentrate, however, on the compositional transitions of fishes, because of the much larger number of DNAs investigated and because an extensive paleontological record is available to date the time of appearance of the fish groups under consideration (Nelson 1984; Carroll 1988; and references quoted therein).

In the case of fish genomes compositional transitions (as indicated by standard deviations of ρ_O values equal to, or higher than, unity) were found

in 6% of the genera, 29% of the families, and 64% of the orders examined. If account is taken of the possibility that *Aphyosemion* should be split into several genera, no significant deviation would exist in any genus investigated by us, whereas this would be the case for 36% of the families. This result is particularly striking if one considers that the standard deviation of modal buoyant densities of DNAs from a whole vertebrate class, mammals, is only 0.7, a value below the significance threshold, considered to be equal to one (see Table 1).

In fact, the spread of the transitional mode of evolution in fish genomes must be severely underestimated on the basis of the data of Table 2 for the following reasons. (1) The small size of most samples raises the possibility that some higher standard deviations were missed because certain species were not studied. (2) Significant changes that occurred in one direction (GC increases, for instance) could have subsequently been counterbalanced by changes in the opposite direction (GC decreases). (3) Compositional changes could occur without being accompanied by easily detectable differences in modal buoyant densities. Indeed, changes in CsCl band asymmetry and in DNA heterogeneity can occur without practically affecting modal buoyant densities (compare murids + cricetids with other mammals in Table 1). (4) The presence of identical values of modal buoyant densities within a group, due to the overrepresentation of a genus or a family, may cause a decrease in the standard deviations. These were, therefore, also calculated by taking into account only one value out of groups of identical ones from the same genus or family. The increases in standard deviations so obtained were, however, negligible, except for Perciformes. In this case, the standard deviation became higher than unity and, therefore, significant. Interestingly, the existence of this problem was pointed out by the fact that one family of Perciformes, Serranidae, showed a standard deviation higher than unity, whereas the order as a whole did not.

The Compositional Transitions of Fish Genomes: The Nature of the Changes

Two mechanisms, not exclusive of each other, may be responsible for compositional transitions. The first mechanism concerns expansion–contraction phenomena affecting repeated sequences.

1) Changes in highly repeated, clustered sequences (satellite DNAs) can be neglected here for two different reasons. First, fish satellite DNAs are systematically GC rich and are almost always present in small relative amounts (Bernardi and Bernardi 1990); and second, calculating GC levels from $\langle\rho\rangle$ values (without taking into account satellites; as done in Table 3 of Bernardi and Bernardi 1990) or from ρ_0 values very largely or completely eliminates compositional changes associated with the presence of different amounts of satellite DNAs.

2) Changes in interspersed repeated sequences (and in noncoding intergenic sequences, in general) need more serious consideration. Indeed, we have shown (Bernardi and Bernardi 1990) that if polyploid genomes are neglected (see below), GC levels decrease or increase when *c* values increase or decrease, respectively. Such changes are apparently due to the expansion or contraction of noncoding intergenic sequences (and of the interspersed repeat families contained in them), which are lower in GC than the coding sequences that they surround. Changes due to this cause are small, however, as an increase of 1 pg in *c* value is required to obtain, on the average, a 5% GC change. It is reasonable to conclude that changes in genome size due to regional expansion–contraction phenomena are not conducive to large GC changes and are, therefore, in general, not responsible for compositional transitions. In fact, in the species under consideration a change of 1 pg is an enormous change, and may more often be the consequence of diploidization (which does not cause by itself any significant change in composition) than of regional compositional changes (Bernardi and Bernardi 1990).

The above considerations suggest that compositional transitions, as investigated here, are essentially due to an alternative possibility, namely to directional nucleotide changes. Ideally, this alternative explanation should be supported by the finding that homologous coding sequences show GC changes (especially in third codon positions) that

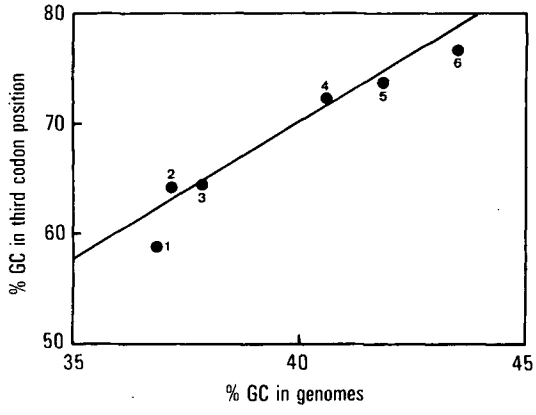

Fig. 6. Average GC levels of third codon positions of genes from six fish species are plotted against the GC levels of the corresponding genome. Fish species are 1, *Brachydanio rerio* (2); 2, *Cyprinus carpio* (6); 3, *Carassius auratus* (2); 4, *Lophius americanus* (3); 5, *Torpedo marmorata* (2); and 6, *Salmo gairdneri* (4). Values in parentheses indicate the number of coding sequences studied; genes for histones, protamines, and additional copies of genes from multigene families were disregarded.

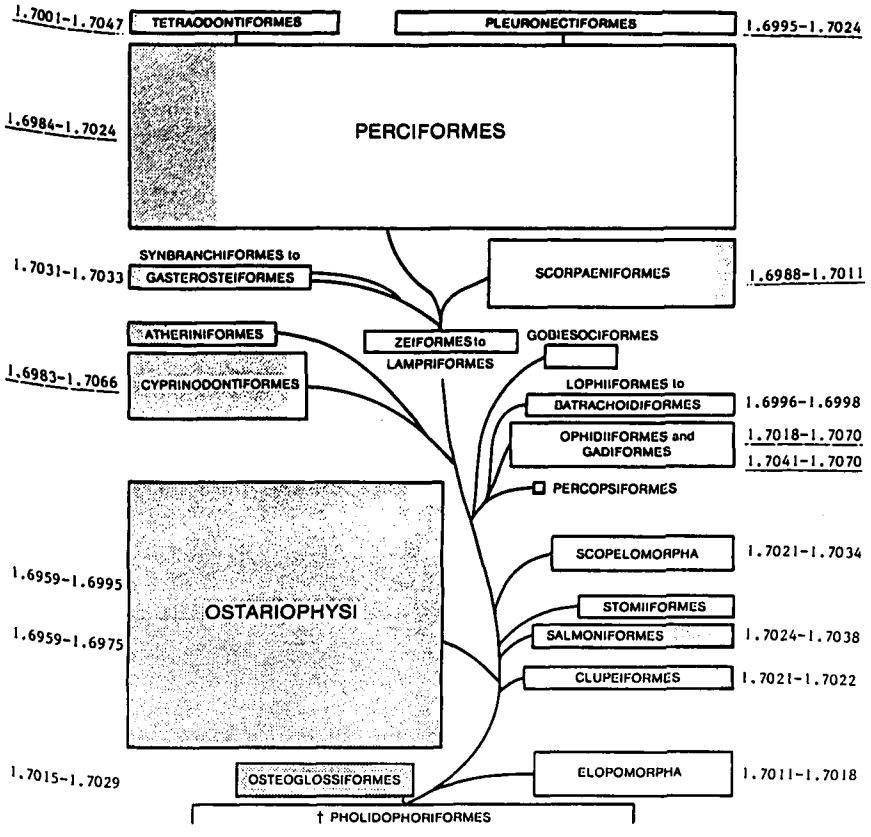

Fig. 7. Ranges of modal buoyant densities from teleostean fish species belonging to different orders (data from Bernardi and Bernardi 1990) are presented on a generalized diagram of relationships of the major groups of extant teleosts (Nelson 1984; all lines indicating postulated relationships are shown as the same, regardless of whether the affinity is considered highly speculative or relatively certain; the area of the blocks is roughly proportional to the number of species recognized in the indicated group; the gray area represents the percentage of species normally confined to fresh water). In the case of Perciformes, *Synchiropus splendidus* was not taken into account, for reasons given in the text. In the case of Ostariophysi, the first value refers to the superorder and the second to the order Cypriniformes. In the case of Ophidiiformes and Gadiformes, the first value refers to both orders and the second to Gadiformes only. Underlined values correspond to significant compositional transitions within the designated order ($\sigma > 1$).

correlate with the overall compositional transition of the genome. Although this information is not available so far, we know, at least, that the compositional transitions under consideration here are accompanied by only minor changes in *c* values; and that GC levels of third codon positions averaging genes from the same genomes are correlated with the GC levels of the genomes (Fig. 6). It should be stressed that, although the number of genes investigated is necessarily small because sequence data for fish genes are still scanty, the results of Fig. 6 are credible because they fit with the genome hypothesis (Grantham et al. 1980), which applies to compositionally homogeneous genomes (Bernardi and Bernardi 1986). Obviously, the data of Fig. 6 support the explanation of directional nucleotide changes, because they indicate that changes in genome composition are paralleled by changes in gene compositions.

Compositional Transitions in Fish Genomes: Independence from Evolutionary Time

The compositional transitions of fish genomes raise an important question about how the underlying nucleotide changes are related to evolutionary time.

This point can be approached in different ways. First of all, no correlation appears to exist between modal buoyant densities of DNAs from species belonging to different orders of teleosts and the evolutionary relationships among these orders. As shown in Fig. 7, there are no visible trends in the compositional properties of the genome from the older groups to the more recent ones. Moreover, groups comprising very large numbers of species, like Ostariophysi, may exhibit a narrow range of modal buoyant densities, whereas groups comprising a much smaller number of species, like Cyprinodontiformes and Tetraodontiformes, show a very large spectrum. Second, if the spread of modal buoyant densities (as indicated by their standard deviations) exhibited by a fish order or family (showing no large differences in *c* values and, therefore, no major role of expansion–contraction phenomena affecting intergenic sequences) is plotted against the time of their respective appearance (Fig. 8), no significant correlation is found. In other words, the average compositional divergence that took place among the species under consideration since the appearance of the corresponding groups (and that led to the differences in modal buoyant density) is not related to evolutionary time. (One should consider here that all small recent compositional divergences accumulate in the lower right-end corner of Fig. 8, and that all mammals investigated exhibit a standard deviation of only 0.7.) In fact, the averaging method used to produce the data of Fig. 8 hides the fact that changes much larger than the average took

place within a time that obviously was shorter than the time elapsed since the appearance of the group. Indeed, even within a given group, pairwise comparisons of modal buoyant densities give different values (see below).

The independence of compositional transitions upon evolutionary time can be explored in more detail. If we take into consideration the two groups of Cyprinodontiformes, the aplocheiloids and the other Cyprinodontiformes (Poecilidae, samples 69, 70 of Table 3 of Bernardi and Bernardi 1990; fundulines, sample 65; and New World cyprinodontines, samples 64–68), for which the time of separation is established at 200 million years (Myr) ago (Parenti 1981), it is obvious (Fig. 8 and Table 2) that the former group underwent a much larger compositional divergence compared to the latter. In both cases, the number of genera (6 and 5, respectively) and species (17 and 8, respectively) investigated appears to be satisfactory for the present purpose. Pairwise comparisons of modal buoyant densities indicate in addition that the maximal compositional divergence was much larger in the first group than in the second. A GC level difference of 8.6% was found between two aplocheiloids, *Aphyosemion australe* (sample 56) and *Rivulus holmiae* (sample 61). In contrast a maximal difference of only 1.3% GC was found between two species from the other Cyprinodontiformes, *Cyprinodon salinus* (sample 64) and *Jordanella floridae* (sample 68).

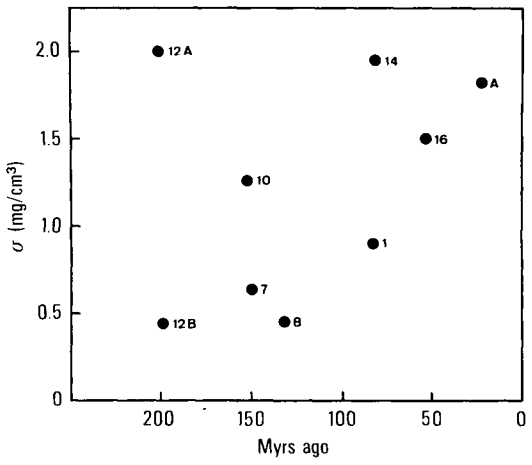

Fig. 8. Standard deviations of average modal buoyant densities, σ, of DNAs from fish species belonging to the same family, or order, are plotted against the time of the appearance of these groups. Groups are: Lamniformes, 1; Cypriniformes, 7; Salmonidae, 8; Gadiformes, 10; Aplocheilidae, 12A; Cyprinodontidae, 12B; *Aphyosemion*, A; Perciformes, 14; Tetraodontiformes, 16. Numbers correspond to those of Table 2 and Fig. 5.

Compositional Transitions in Fish Genomes: The Rates of Directional Nucleotide Substitutions

The GC differences characterizing compositional transitions allow one to calculate directional substitution rates, provided that an estimate of the divergence time is available and that GC changes are only or essentially due to directional nucleotide substitutions. If these conditions are satisfied, the rates of directional mutations (AT $\rightleftharpoons$ GC) per site and per year can be calculated in pairwise comparisons as $(GC_1 - GC_2)/2N$, where GC_1 and GC_2 are the GC percentages of the genomes having the higher and lower GC level, respectively, and N is the number of years since the appearance of the group (order or family) comprising the two species.

A case that lends itself to such an analysis and that is particularly relevant here is that of three species from the genus (or genera) *Aphyosemion*. Indeed, *A. punctatum* (sample 46 of Table 3 of Bernardi and Bernardi 1990) has a genome characterized by a GC level of 41.9% and a *c* value of 0.6 pg, whereas the genomes of *A. australe* (sample 56) and *A. striatum* (sample 55) have GC levels of 46.7% and 48.1%, respectively, and the corresponding *c* values are 0.73 and 0.65 pg, respectively. These slightly higher *c* values correspond to differences that are small enough not to cause more than a 0.5% GC difference, whereas the GC differences under consideration are between 5 and 6%. Moreover, these *c* value differences, if due to amplification of interspersed repeats, should lead to lower GC levels, whereas the opposite is found. In fact, at least in the case of *A. australe*, the increased *c* value may be due to the influence of a GC-rich satellite (see Fig. 1 of Bernardi and Bernardi 1990).

The time of divergence for the group *Aphyosemion* was estimated using Nei's genetic distances as calculated by Douchement et al. (1984). Divergence times are based on Sarich's (1977) correlations of genetic distance and divergence times for vertebrates, as modified by Carlson et al. (1978; see also Graves and Somero 1982). This leads to an estimate of 23.2 ± 4.3 Myr for the separation of *Archiaphyosemion* (*A. punctatum*), *Mesoaphyosemion* (*A. australe* and *A. striatum*), and *Paraphyosemion*. Using this time, one can calculate a compositional substitution rate equal to $1.55.10^{-9}$ per site per year. This value, if corrected for multiple hits, would be higher than the rate of synonymous substitutions as measured in primates (2.10^{-9}, corrected for multiple hits; Li et al. 1987). Needless to stress that this conclusion is most striking, as the comparison concerns the numbers of AT to GC changes (4 changes) on the one hand and the numbers of all (12) changes on the other hand.

Two remarks are appropriate at this point. The first is that the separation time of the species under consideration is certainly overestimated. Indeed, the

extremely high compositional substitution rate, as calculated for the whole genomes (and therefore mainly for intergenic sequences that represent the vast majority of the genome) should be accompanied by higher first and second codon position substitutions, which are responsible for the amino acid changes used to determine separation time. This obviously leads to overestimating separation time. The second remark is that, although our data on *Aphyosemion* correspond to the highest compositional substitution rate found by us so far, other data are likely to approach them; this should be true of rates concerning at least some Tetraodontiformes.

Compositional Transitions in Fish Genomes: Evolutionary Implications

The results discussed in the preceding section have important evolutionary implications that will be elaborated further elsewhere. Suffice it to mention here that the existence of extremely high rates of nucleotide substitutions (as exemplified by the case of *Aphyosemion*) indicates that the molecular clock (Zuckerkandl and Pauling 1962; Zuckerkandl 1987) may be disrupted during at least some compositional transitions. Effects along the same line have been detected between homologous coding sequences having undergone compositional changes in mammals (Mouchiroud and Gautier 1988, 1990; Saccone et al. 1989).

More generally, the independence of compositional changes from evolutionary time indicates that the genome phenotype (namely, the compositional pattern of the genome; Bernardi and Bernardi 1986), like the classical phenotype (corresponding to the gene products) does not undergo changes according to a molecular clock. Yet the genome phenotype and the classical phenotype do not behave symmetrically relative to each other.

In the conservative mode of genome evolution there is no change in the genome phenotype, and yet the classical phenotype may be altered because of a very small number of changes in critical genome sites, which do not show up in the compositional pattern of the genome.

In the transitional or shifting mode of evolution, the genome phenotype undergoes compositional transitions, and these transitions are always accompanied by changes in the classical phenotype. The reason is that the GC increases or decreases occurring in third codon positions, in introns, and in noncoding intergenic sequences, which change the genome phenotype, are accompanied by GC increases or decreases, respectively, in first and second codon positions. (This is exemplified by the transition from cold-blooded to warm-blooded vertebrates.) Changes are, therefore, simultaneously introduced in the genome phenotype and in the classical phenotype.

In conclusion, although changes in the genome phenotype are always accompanied by changes in the classical phenotype, the reverse is not necessarily true. In particular, the classical phenotype changes associated with speciation may be accompanied by a spectrum of compositional changes ranging from undetectable to very large ones. Large and rapid compositional changes, like those exemplified by *Aphyosemion,* may be taken as molecular evidence of punctuated equilibria (Eldredge and Gould 1972), a point that will be discussed further elsewhere.

Compositional Transitions in Fish Genomes: Taxonomic Implications

As just mentioned, changes in the classical phenotype, and in particular those associated with speciation, may be accompanied or not by changes in the genome phenotype. On the other hand, changes in the genome phenotype, at least those large enough to be detected by differences in modal buoyant densities, are inescapably accompanied by large changes in the classical phenotype. It is conceivable, therefore, that they always are associated with speciation. Although this is obvious for genomes, like the bacterial genomes, that consist almost solely of genes, it is not so for genomes predominantly made up of noncoding intergenic sequences, like those of vertebrates. In this case, the association with speciation is due to the fact that compositional changes in the predominant noncoding intergenic sequences are accompanied by similar although slighter compositional changes in the first and second codon positions of coding sequences. Accordingly, modal buoyant density, a parameter related to the genome phenotype, should be of important taxonomic value not only in the case of prokaryotic genomes (a point already recognized by microbiologists; see Marmur et al. 1963; Mandel 1969), but also in that of eukaryotic genomes.

The following are some examples in which modal buoyant densities helped to solve taxonomic problems in ichthyology by supporting one among divergent taxonomic proposals.

1) Tetraodontiformes, an order that is problematic as far as taxonomy is concerned (Rosen 1982; Zehren 1987), exhibit a wide range of modal buoyant densities, large differences being shown even by species belonging to the same family. In particular, Balistidae (which are heterogeneous in ρ_0) were split (Matsuura 1979) into two families, Balistidae and Monacanthidae. At least in partial agreement with this proposal, Monacanthidae are homogeneous in terms of ρ_0 (see Table 3 of Bernardi and Bernardi 1990). This suggests that additional investigations

on the compositional patterns of Tetraodontiformes could clarify their complex taxonomic relationships.

2) The DNA of *Synchiropus splendidus* exhibits a modal buoyant density, 1.7073 g/cm^3, which is the highest among those observed in the present work. This density is much closer to those found in Gadiformes than to those of Perciformes and favors the placement, already proposed by Gosline (1970) of Callionymidae (the family to which *S. splendidus* belongs), in the superorder Paracanthopterygii (see Table 2 of Bernardi and Bernardi 1990) together with the order Gadiformes (as accepted by Nelson 1976) rather than in the order Perciformes (as proposed more recently by Nelson 1984).

3) The families Pomacentridae, Cichlidae, Labridae, and Scaridae were put together into a group called Pharyngognathi (Liem and Greenwood 1981; see, however, Stiassny and Jensen 1987), but anatomical features of Scaridae placed them apart from the other families (Monod 1986). In agreement with the latter proposal, the modal buoyant density of Scaridae is different from those of Pomacentridae, Cichlidae, and Labridae.

4) Species from the genus *Aphyosemion* are characterized by DNAs ranging in modal buoyant density from 1.7008 to 1.7066 g/cm^3. This genus has been split, however, into four genera on the basis of morphology (Clausen 1967) and of electrophoretic mobility of proteins (Douchement et al. 1984). Interestingly, species belonging to three of these genera (*Archiaphyosemion, Mesoaphyosemion,* and *Paraphyosemion*; these comprise, respectively, species 47–51 and 57–58, species 55–56, and species 52–54 of Table 3 of Bernardi and Bernardi 1990) were analyzed and found to fall into three much more homogeneous classes of modal buoyant densities (Bernardi and Bernardi 1990).

Compositional Transitions in Fish Genomes: The Causes

As far as the causes of compositional transitions are concerned, two possibilities should be considered, namely mutational pressure (Sueoka 1988; Wolfe et al. 1989) or selection. Although the problems associated with the idea that mutational pressure may be the cause (and not simply the mechanism) of compositional transitions have already been discussed (Bernardi et al. 1988) and will be further commented upon elsewhere, here we make some remarks about selection as an explanation.

In general, the selective advantages leading to compositional transitions in the genome are difficult to define, presumably because they are associated with a number of different biological requirements. The observed compositional transitions probably are the result of an equilibrium between different selective pressures. For example, it is impossible to define at present the selective advantages associated with the compositional transitions of the *Aphyosemion* genomes. However, in the case of the transition from cold-blooded to warm-blooded vertebrates, one major environmental change appears to have been involved, namely an increase in body temperature. In such a case, one could expect to identify specific advantages of certain compositional transitions. Indeed, these advantages have been detected (Bernardi and Bernardi 1986) in that the regional GC increases accompanying the transition lead to a higher thermostability of chromosomal regions, transcripts, and proteins.

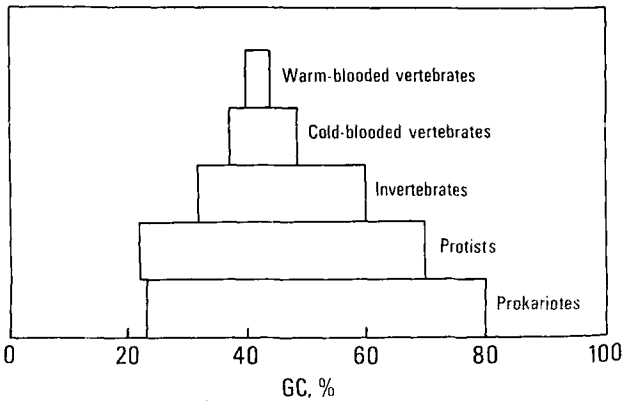

Fig. 9. GC level ranges of DNAs from warm- and cold-blooded vertebrates, invertebrates, protists, and prokaryotes. Data are from Normore and Brown (1973), Mandel (1973), and Thiery et al. (1976).

It should be noted that the observation that compositional transitions are so much more frequent in cold-blooded than in warm-blooded vertebrates is in line with the explanation given above. A higher degree of homeostasis and/or of environmental stability appears to lead to more stable compositional patterns of the genome, as indicated by the increasing spread of modal buoyant densities (or GC levels) of DNAs, when going from warm-blooded to cold-blooded vertebrates, to invertebrates, to unicellular eukaryotes, and to bacteria (see Fig. 9).

References

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

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

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

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

Carlson SS, Wilson AC, Maxson RD (1978) Do albumin clocks run on time? A reply. Science 200:1183–1185

Carroll RL (1988) Vertebrate paleontology and evolution. Freeman, New York

Clausen HS (1967) Tropical Old World cyprinodonts. Akademisk Forlag, Copenhagen

Corneo G, Ginelli E, Soave C, Bernardi G (1968) Isolation and characterization of mouse and guinea pig satellite DNAs. Biochemistry 7:4373–4379

Cortadas J, Olofsson B, Meunier-Rotival M, Macaya G, Bernardi G (1979) The DNA components of the chicken genome. Eur J Biochem 99:179–186

Cuny G, Soriano P, Macaya G, Bernardi G (1981) The major components of the mouse and human genomes: preparation, basic properties, and compositional heterogeneity. Eur J Biochem 111:227–233

Douchement J, Romand R, Pasteur N (1984) Biochemical differentiation in West-Africa cyprinodontoid fish of the genus *Aphyosemion*. Biochem Syst Ecol 12:325–333

Eldredge N, Gould SJ (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: Shopf JM (ed) Models in paleobiology. Freedman, Cooper & co, San Francisco, pp 82–115

Filipski J, Thiery JP, Bernardi G (1973) An analysis of the bovine genome by Cs_2SO_4/Ag^+ density gradient centrifugation. J Mol Biol 80:177–197

Gosline WA (1970) A reinterpretation of the fish order Gobiesociformes. Proc Calif Acad Sci, Ser 4 37:363–382

Grantham R, Gautier C, Gouy M, Mercier R, Paré A (1980) Codon catalog usage and the genome hypothesis. Nucleic Acids Res 8:r49–r62

Graves JE, Somero GN (1982) Electrophoretic and functional enzyme evolution in four species of eastern pacific barracudas from different thermal environments. Evolution 36:97–106

Hudson AP, Cuny G, Cortadas J, Haschemeyer AEV, Bernardi G (1980) An analysis of fish genomes by density gradient centrifugation. Eur J Biochem 112:203–210

Li W-H, Tanimura M, Sharp PM (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J Mol Evol 25:330–342

Liem KS, Greenwood PH (1981) A functional approach to the phylogeny of the pharyngognath teleosts. Am Zool 21:83–101

Macaya G, Thiery JP, Bernardi G (1976) An approach to the organization of eukaryotic genomes at a macromolecular level. J Mol Biol 108:237–254

Mandel M (1969) New approaches to bacterial taxonomy: perspective and prospects. Annu Rev Microbiol 23:239–274

Mandel M (1973) DNA base compositions of eucaryotic protists. In: Sober HA (ed) Handbook of biochemistry, ed 2. CRC Press, Boca Raton FL

Marmur J, Falkow S, Mandel M (1963) New approaches to bacterial taxonomy. Annu Rev Microbiol 17:329–372

Matsuura K (1979) Phylogeny of the superfamily Balistoidea (Pisces, Tetradontiformes). Mem Fac Fish Hokkaido University 26:49–169

Monod T (1986) Les scarides: quelques observations sur le crane et le squelette pharyngien. Oceania 12:339

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

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

Mouchiroud D, Gautier C (1990) Codon usage changes and evolutionary rate between man and rat. J Mol Evol (in press)

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

Nelson JS (1976) Fishes of the world, ed 1. Wiley, New York

Nelson JS (1984) Fishes of the world, ed 2. Wiley, New York

Normore WM, Brown JR (1973) Guanine + cytosine (G+C) composition of bacteria. In: Sober HA (ed) Handbook of biochemistry, ed 2. CRC Press, Boca Raton FL, pp H24–74

Parenti LR (1981) A phylogenetic and biogeographic analysis of cyprinodontiformes fishes (Teleostei, Atherinomorpha). Bull Am Mus Nat Hist 168:335–557

Perrin P, Bernardi G (1987) Directional fixation of mutations in vertebrate evolution. J Mol Evol 26:301–310

Pizon V (1983) Organisation des séquences nucléotidiques dans les génomes de poissons. Thèse de 3è cycle, Université Paris VI

Pizon V, Cuny G, Bernardi G (1984) Nucleotide sequence organization in the very small genome of a tetraodontid fish, *Arothron diadematus*. Eur J Biochem 140:25–30

Rosen DE (1982) Teleostean interrelationships, morphological function and evolutionary inference. Am Zool 22:261–273

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

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

Sarich VM (1977) Rates, sample sizes, and the neutrality hypothesis for electrophoresis in evolutionary studies. Nature 265:24–28

Stiassny MLJ, Jensen JS (1987) Labroid intrarelationships resisted: morphological complexity, key innovations, and the study of comparative diversity. Bull Mus Comp Zool 151: 269–319

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

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

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

Zehren SJ (1987) Osteology and evolutionary relationships of the boarfish genus *Antigonia* (Teleostei: Caproidae). Copeia 3:564–592

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

Zuckerkandl E (1987) On the molecular evolutionary clock. J Mol Evol 26:34–46

Zuckerkandl E, Pauling L (1962) Molecular disease, evolution and genetic heterogeneity. In: Kasha M, Pullman B (eds) Horizons in biochemistry. Academic Press, New York, pp 189–225

Received November 16, 1989/Revised and accepted May 1, 1990