J Mol Evol (1993) 37:93–108

Journal of Molecular Evolution

The Isochore Patterns of Mammalian Genomes and Their Phylogenetic Implications

Georgette Sabeur,* Gabriel Macaya,† Farida Kadi, and Giorgio Bernardi

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

Abstract. The compositional distributions of high molecular weight DNA fragments from 20 species belonging to 9 out of the 17 eutherian orders were investigated by analytical CsCl density gradient centrifugation and by preparative fractionation in Cs_2SO_4/BAMD density gradients followed by analysis of the fractions in CsCl. These compositional distributions reflect those of the isochores making up the corresponding genomes.

A "general distribution" was found in species belonging to eight mammalian orders. A "myomorph distribution" was found in Myomorpha, but not in the other rodent infraorders Sciuromorpha and Histricomorpha, which share the general distribution. Two other distributions were found in a megachiropteran (but not in microchiropteran, which, again, shares the general distribution) and in pangolin (a species from the only genus of the order Pholidota), respectively.

The main difference between the general distribution and all other distributions is that the former contains sizable amounts (6–10%) of GC-rich isochores (detected as DNA fragments equal to, or higher than, 1.710 g/cm^3 in modal buoyant density), which are scarce, or absent, in the other distributions. This difference is remarkable because gene concentrations in mammalian genomes are paralleled by GC levels, the highest gene concentrations being present in the GC-richest isochores.

The compositional distributions of mammalian genomes reported here shed light on mammalian phylogeny. Indeed, all orders investigated, with the exception of Pholidota, seem to share a common ancestor. The compositional patterns of the megachiropteran and of Myomorpha may be derived from the general pattern or have independent origins.

Key words: DNA — Base composition — Vertebrates — Eutheria — Evolution

Introduction

Twenty years ago we discovered, using Cs_2SO_4 preparative density gradient centrifugation in the presence of a sequence-specific DNA-ligand, Ag^+, (a method previously used in order to fractionate mammalian satellite DNAs; Corneo et al. 1968), that high molecular weight bovine DNA is characterized by a strong compositional heterogeneity due to the existence of a small number of discrete components covering a wide GC range (Filipski et al. 1973). This heterogeneity concerned the "main-band" DNA (as distinct from satellite and ribosomal DNAs) and was, therefore, different from the heterogeneity previously detected in analytical CsCl gradient (Sueoka 1959), which is essentially due (1) to the eight GC-rich satellite DNAs that form 23% of the bovine genome (Filipski et al. 1973; Cortadas et al. 1977; Macaya et al. 1978; Kopecka

Presented at the NATO Advanced Research Workshop on *Genome Organization and Evolution,* Spetsai, Greece, 16–22 September 1992

* *Present address:* Faculté de Médecine Nord Boulevard Pierre Dramard, 13326 Marseille Cedex 15

† *Permanent address:* Universidad de Costa Rica, Ciudad Universitaria Rodrigo Facio, San José, Costa Rica

Correspondence to: G. Bernardi

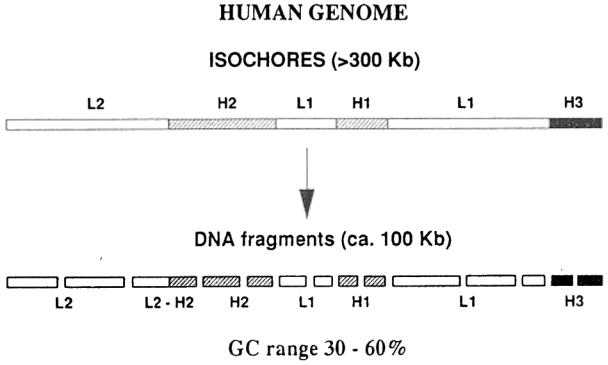


Fig. 1. Scheme of the compositional organization of the genomes from warm-blooded vertebrates. In the example shown, which concerns the human genome, the DNA consists of long (>300 kb, on the average) segments, the isochores, which are compositionally homogeneous (above an average size of 3 kb) and belong to a small number of families, GC-poor (L1 and L2), GC-rich (H1 and H2), and very GC-rich (H3). Physical and enzymatic degradation occurring during DNA preparation generates large DNA fragments, routinely in a size range of 50–100 kb. (Modified from Bernardi et al. 1985.)

et al. 1978), and, to a small extent, (2) to ribosomal DNA (Meunier-Rotival et al. 1979). Subsequent work (Thiery et al. 1976) showed that discrete "main-band" DNA components covering a broad GC range could be resolved in all other genomes of mammals and birds which were investigated, whereas the genomes of cold-blooded vertebrates were characterized by a much lower degree of compositional heterogeneity.

The "major components" of the mammalian genomes, as the "main-band" DNA fragment families were called to contrast them with "satellite" components and with "minor" components (e.g., ribosomal DNA), did not vary in relative amounts or modal buoyant densities over an average molecular size range extending from 3 kb to over 300 kb (Macaya et al. 1976). This demonstrated that the DNA fragments making up the major DNA components derive (see Fig. 1), by the unavoidable mechanical and enzymatic degradation occurring during DNA preparation, from much longer DNA segments (>300 kb; see also Bettecken et al. 1992), homogeneous in composition, later called isochores (for similar regions; Cuny et al. 1981). These form the bulk of genomes of vertebrates and belong to a small number of families covering a broad compositional range (30–60% GC in the human genome). Incidentally, satellite and minor components may also be viewed as isochore families because of their compositional homogeneity (Bernardi 1989). The relative amounts and GC levels of isochore families define a compositional pattern which is characteristic of a genome (Bernardi et al. 1985, 1988) and represents a "genome phenotype" (Bernardi and Bernardi 1986). Other compositional patterns are those made up by the compositional distributions of coding sequences (or their different codon positions) and introns (Bernardi et al. 1985, 1988; Mouchiroud et al. 1987, 1988; Aïssani et al. 1991; Mouchiroud and Bernardi 1993).

The original observations (Thiery et al. 1976; Macaya et al. 1976) showing a large difference in the compositional patterns of cold- and warm-blooded vertebrates were later confirmed and extended by much more detailed studies, which indicated that the former never reach GC levels as high as those exhibited by the latter (Hudson et al. 1980; Bernardi and Bernardi 1990a,b, 1991). These studies also showed small differences between the compositional patterns of mammals and birds (Cortadas et al. 1979; Mouchiroud et al. 1987; Kadi et al. 1993) as well as between the compositional patterns exhibited by the mouse and the human genomes (Salinas et al. 1986; Zerial et al. 1986; see below). The significance of such differences was unequivocally confirmed by the fact that they were paralleled by compositional differences in homologous coding sequences (Mouchiroud et al. 1988; Bernardi et al. 1988; Mouchiroud and Gautier 1990; Bernardi 1993a; Mouchiroud and Bernardi 1993).

The first difference which was identified in the compositional patterns of mammalian genomes concerned the mouse and human genomes (Salinas et al. 1986; Zerial et al. 1986; Mouchiroud et al. 1987, 1988; Bernardi et al. 1988; Mouchiroud and Gautier 1990; Mouchiroud and Bernardi 1993). The mouse genome (like those of other Myomorpha, but not of Sciuromorpha or Caviomorpha) has a narrower compositional distribution, both at the DNA and at the coding sequence level, compared to the human genome (and to a number of other mammalian genomes). As a consequence, the GC-richest component, the isochore family H3 of the human genome, which is the richest one in genes (Mouchiroud et al. 1991), is simply absent in the mouse genome. The genes present in the H3 family of human isochores and their flanking sequences are found, however, in the GC-richest isochore family (H2) of mouse, and the compositional ranking order of third codon positions of homologous genes is preserved to a very large extent (Mouchiroud et al. 1988; Mouchiroud and Bernardi 1993). Moreover, the mouse genome is characterized by a smaller amount of the GC-poorest components and coding sequences compared to the human genome.

The results and observations mentioned above prompted a more detailed study aiming at further investigating mammalian genomes for differences and similarities in their compositional patterns. Indeed, these investigations should help in understanding how isochores arose in the evolution of vertebrates, how their composition was conserved

Table 1. A classification of mammals[a] and the species investigated in the present work

Order, sub-, and infraorder	Family	Species investigated	
Monotremata			
Marsupialia			
Insectivora	Erinaceidae	1. Hedgehog	*Erinaceus europeus*
	Soricidae	2. Shrew	*Crocidura russula*
	Talpidae	3. Mole	*Talpa europea*
Dermoptera			
Chiroptera			
Megachiroptera	Pteropodidae	4. Fruit bat	*Pteropus sp.*
Microchiroptera	Vespertilionidae	5. Bat	*Myotis myotis*
Primates	Hominidae	6. Man	*Homo sapiens*
Edentata			
Pholidota	Manidae	7. Pangolin	*Manis sp.*
Lagomorpha	Leporidae	8. Rabbit	*Oryctolagus cuniculus*
Rodentia			
Sciurognathi			
Sciuromorpha	Sciuridae	9. Squirrel	*Sciurus vulgaris*
		10. Woodchuck	*Marmota monax*
Myomorpha	Muridae	11. Rat	*Rattus norvegicus*
		12. Mouse	*Mus musculus*
	Cricetidae	13. Hamster	*Cricetus norvegicus*
	Spalacidae	14. Mole rat	*Spalax sp.*
	Gliridae	15. Dormouse	*Glis glis*
Histricognathi			
Caviomorpha	Caviidae	16. Guinea pig	*Cavia porcellus*
Cetacea			
Carnivora	Canidae	17. Dog	*Canis familiaris*
	Felidae	18. Cat	*Felis domesticus*
Pinnipedia			
Tubulidentata			
Proboscidea			
Hyracoidea			
Sirenia			
Perissodactyla	Equidae	19. Horse	*Equus caballus*
Artiodactyla	Bovidae	20. Calf	*Bos taurus*

[a] From Nowak and Paradiso (1983). Rodents are classified according to Colbert and Morales (1991)

in some cases, and how it was changed in others, a subject on which contrasting views exist (see Bernardi 1993a). Moreover, they also might shed light on mammalian phylogeny.

In the present work, we report on the compositional patterns, as investigated at the DNA level, of the genomes of 20 species belonging to nine out of the 17 eutherian orders. Our efforts concentrated on studying possible differences in the highest GC range of DNA fragments because of the previously found differences between the genomes of Myomorpha and those of most other mammals investigated, and because the GC-richest isochores are characterized by the highest gene concentrations. In the following paper, the compositional distributions of the genomes of a smaller number of mammalian species will be examined at the level of coding sequences (Mouchiroud and Bernardi 1993).

Materials and Methods

Sources of Animals and Tissues. A garden dormouse, a hedgehog, a mole, and bats were captured in Normandy, in the Massif Central, in West Germany (Thiery et al. 1976), and in the south of France, respectively. Shrews were obtained from Dr. R. Fons, Laboratoire Arago, Banyuls-sur-Mer, France. A fruit bat and a pangolin were obtained in Ho Chi Minh Ville, Vietnam. A human placenta was from a Parisian Hospital. Rats, mice, hamsters, guinea pigs, and rabbits were from the animal house of our institute; a squirrel came from a local pet store. Mole rats were provided by Prof. E. Nevo (Tel Aviv University, Tel Aviv, Israel). Cat and dog livers were from the Laboratory of Physiology of Paris VI University; horse liver, calf liver, and calf thymus were from local slaughterhouses.

DNA Preparations. DNA was prepared according to Kay et al. (1952), in most cases from either freshly excised or frozen livers; in other cases, from fresh thymus (hedgehog, calf) or placenta (man). The average size of DNA molecules, as determined by gel electrophoresis, was over 30 kb. Woodchuck DNA was obtained from Dr. G. Fourel, Pasteur Institute, Paris.

DNA Centrifugation and Analysis. Analytical centrifugation in CsCl density gradient was carried out as described previously (Thiery et al. 1976; see this paper, or Bernardi and Bernardi 1990a, for other details and for the definitions of modal, ρ_o, and mean, $\langle\rho\rangle$, buoyant densities). Preparative centrifugations of DNA in Cs_2SO_4/BAMD density gradients were done at 20°C and 40,000 rpm. BAMD is 3,6-bis (acetato-mercuri-methyl) dioxane

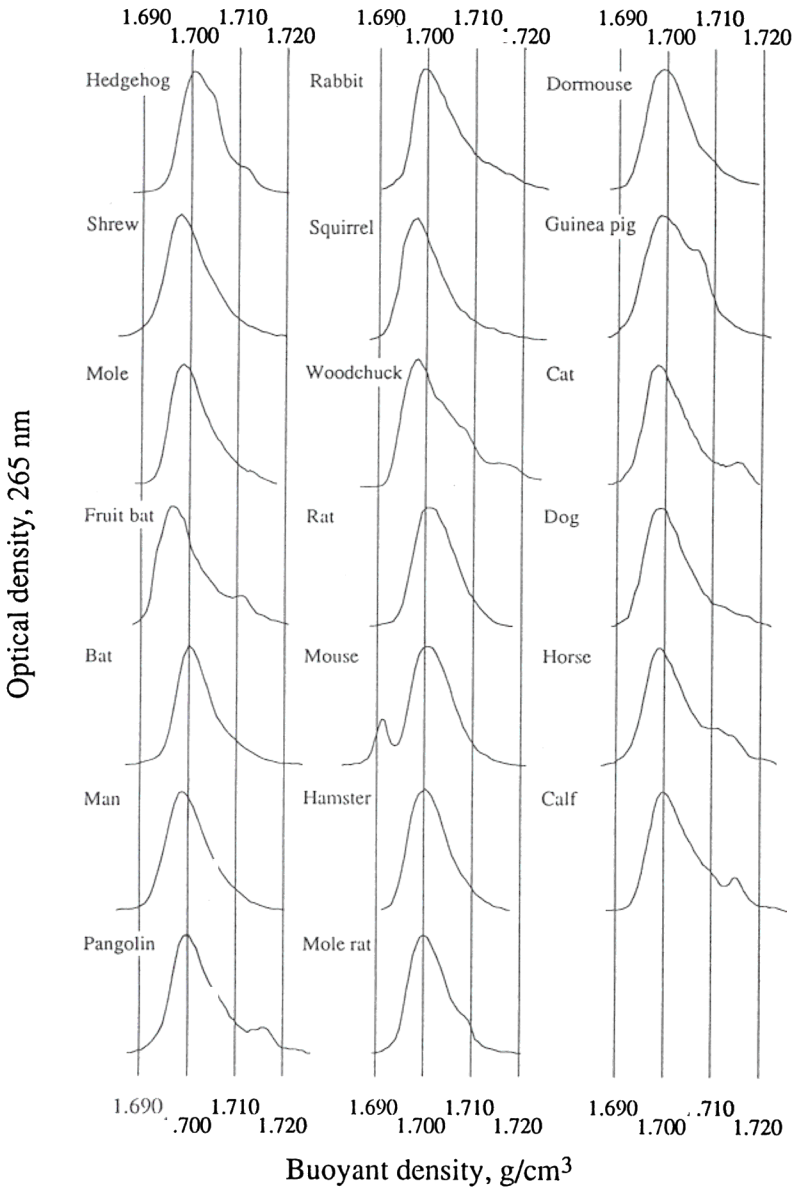


Fig. 2. Analytical CsCl density profiles of unfractionated DNA preparations. The hedgehog and calf profiles are from Thiery et al. (1976), the woodchuck profile from F. Kadi (personal communication). Profiles were normalized to the same height.

(Cortadas et al. 1977, 1979; Macaya et al. 1978). The BAMD/nucleotide ratio, r_f, was 0.10 in the case of shrew, fruit bat, pangolin, mole rat, 0.14 in recentrifugation experiments and in the experiment on the mouse genome (from Salinas et al. 1986) reported here, and 0.13 in all other cases.

Results

Table 1 presents a simple classification of mammals and a list of the species studied. This list includes species which had been investigated previously, like calf (Thiery et al. 1976) and mouse (Salinas et al. 1986), as well as three species only studied by analytical CsCl centrifugation: hedgehog, garden dormouse (Thiery et al. 1976), and woodchuck. Overall, 20 species belonging to nine out of the 17 eutherian orders were investigated. One species per order was studied, except for insectivores (3 species from 3 families), chiropters (2 species from 2 suborders), carnivores (2 species from 2 families), and rodents (8 species from two suborders, 3 infraorders, and 5 families).

Figure 2 displays analytical CsCl density profiles of unfractionated DNA preparations which represent, to a first approximation (i.e., neglecting molecular weight effects), distributions of buoyant densities and (neglecting DNA methylation) of GC levels (see Discussion). As shown in Table 2 and Fig. 3, mammalian main-band DNAs exhibit modal buoyant densities comprised between 1.6965–1.7008 g/cm^3, which correspond to a GC range of 4%. Modal buoyant densities belong in two main groups, centered on 1.698–1.699 and 1.700 g/cm^3, respectively, with one lower value (fruit bat, 1.6965

Table 2. Buoyant density properties of the DNAs investigated

Order	Species	ρ_o (g/cm^3)	$\langle\rho\rangle$ (g/cm^3)	$\langle\rho\rangle$-ρ_o (mg/cm^3)	>1.710 g/cm^3 (%)[a]
Insectivora		1.7004		3.1	
		1.6976		2.5	**7**
		1.6987		2.4	**9**
Chiroptera		1.6965		3.7	<1
		1.7000		2.4	**10**
Primates		1.6984		2.2	**8**
Pholidota		1.6997		2.9	<1
Lagomorpha		1.6998		4.1	**7**
Rodentia		1.6982		2.4	**9**
		1.6983		4.3	
		1.7008		1.6	<1
		1.7006		0.3	2
		1.7000		2.1	<4[e]
		1.7000		1.9	2
		1.6991		1.3	
		1.6989		2.9	**6**[f]
Carnivora		1.6986		3.3	**7**
		1.6990		2.7	**6**
Perissodactyla		1.6990		4.0	**8**
Artiodactyla		1.7000		3.9	**7**[g]

[a] This column lists rounded-out figures of the corresponding column from Table 3. Values equal to, or higher than 6 (typical of the "general distribution") are shown in bold type
[b] Data from Thiery et al. (1976)
[c] No assessment available from fractionation results
[d] The light satellite DNA was not taken into account in calculating $\langle\rho\rangle$
[e] Based on the results of Rynditch et al. (1991)
[f] From the recentrifugation experiment of Fig. 6
[g] Data from Cortadas et al. (1977)

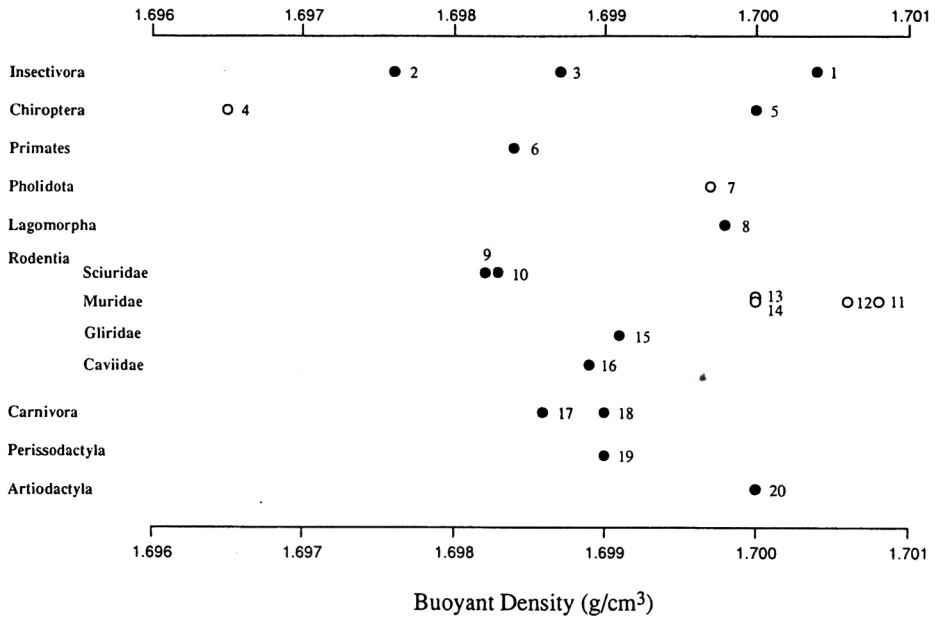
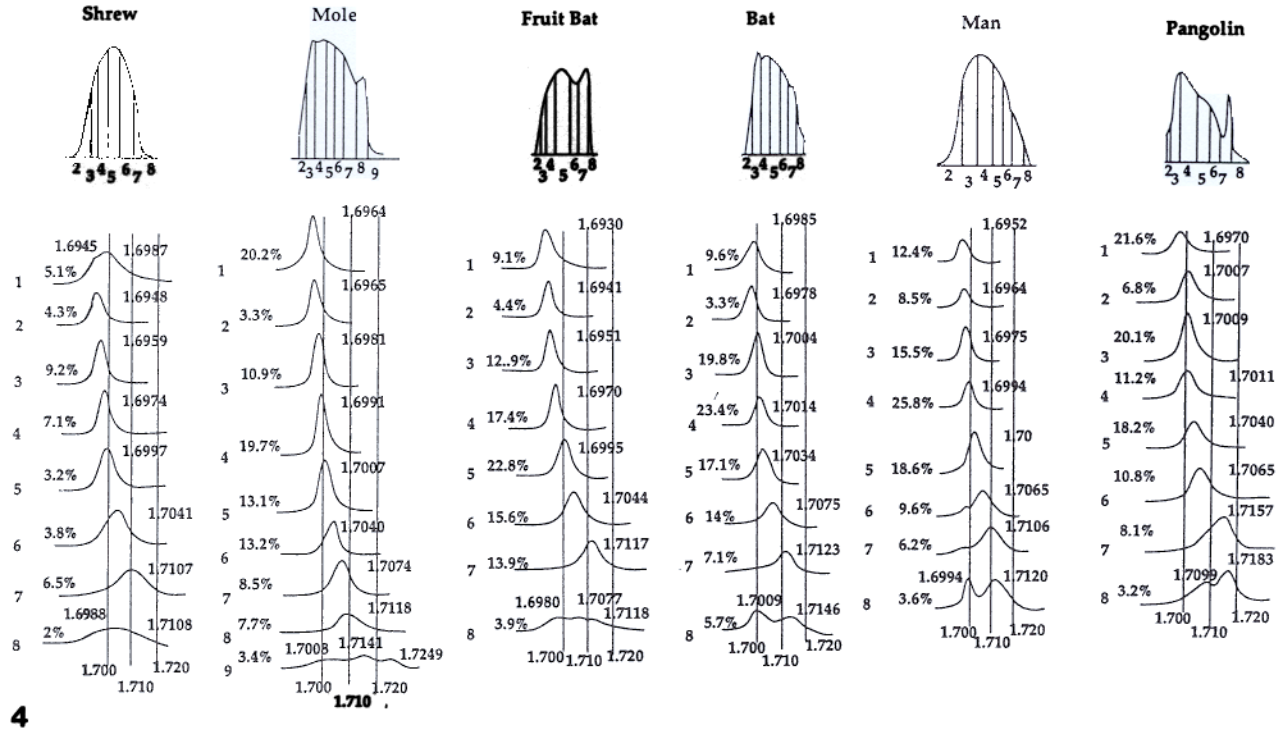


Fig. 3. Modal buoyant densities of the mammalian DNAs investigated in the present work. *Closed* and *open symbols* refer to DNAs comprising or not comprising GC-rich components higher than 1.710 g/cm^3 in buoyant density (Table 3). The presence of such components was not established by preparative fractionation in the case of hedgehog, woodchuck, and dormouse (Table 2).

g/cm^3). Except for carnivores, different modal buoyant density values were found within individual orders. Indeed, the three insectivores span almost a 3 mg/cm^3 range and the two chiropters almost a 4 mg/cm^3 range. Among rodents, woodchuck and squirrel were close to 1.698 g/cm^3; rat, mouse, hamster, and mole rat exhibited values around 1.700 g/cm^3; dormouse and guinea pig were close to 1.699 g/cm^3.

Two points should be stressed here concerning such differences in ρ_o. First, they do not seem to be due to trivial reasons, like the presence of satellite DNAs. Even in the case of hedgehog, in which a satellite shoulder is clearly visible at about 1.705



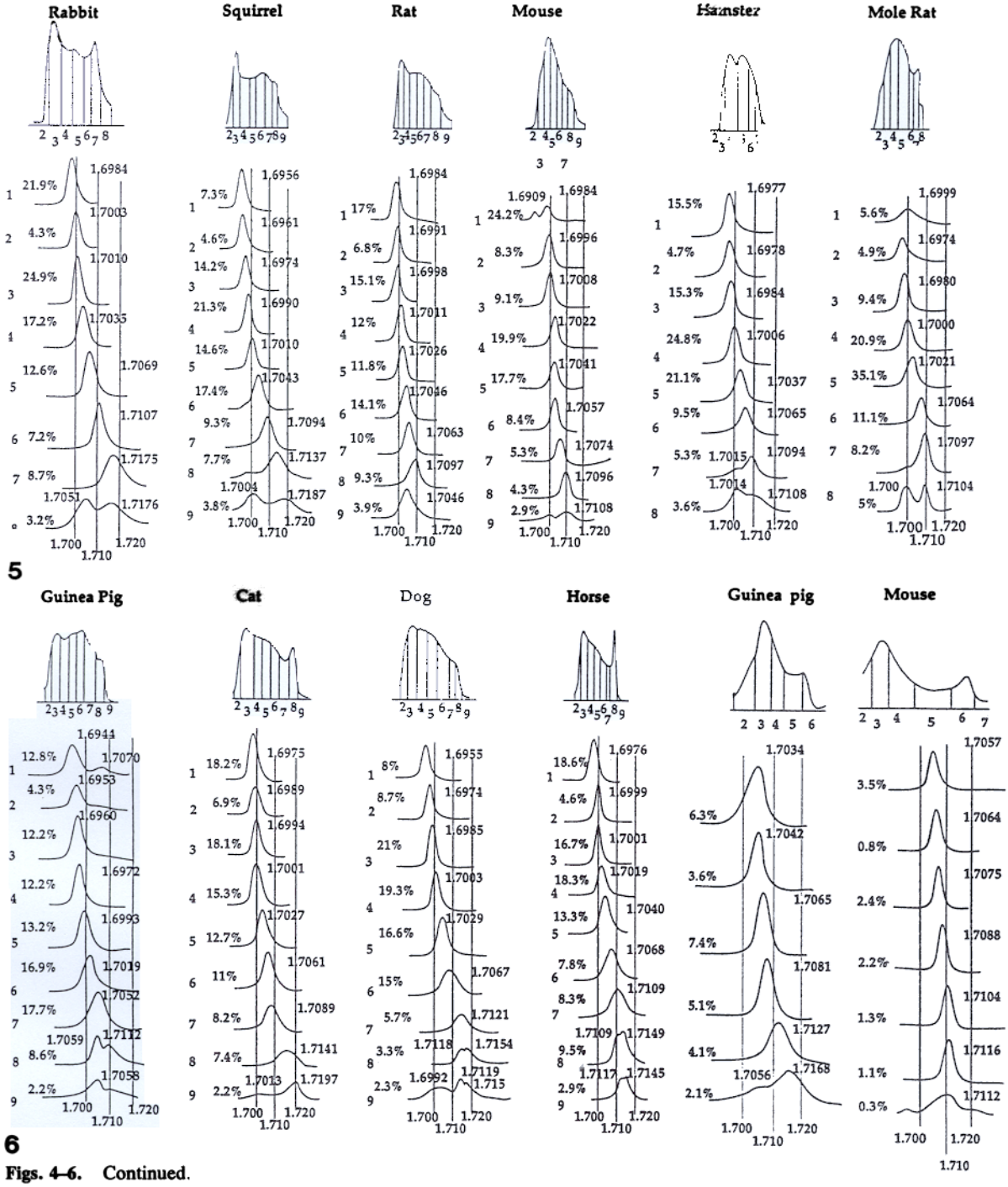
4

Figs. 4–6. Fractionation in Cs_2SO_4/BAMD density gradient. The transmission profiles recorded at 254.7 mμ during fraction collection are shown on the top. The bottom panels show the analytical CsCl profiles of fractions, the relative amounts of DNA present in them, and the modal buoyant densities of the peaks. Fraction 1 corresponds to pelleted DNA. Loads were 12 A_{260} units per tube. Mouse data are from Fig. 1a of Salinas et al. (1986). The last two series of data on Fig. 6 concern recentrifugation of GC-rich fractions in Cs_2SO_4/BAMD density gradients. Indications as above, except for the absence of pellet and for the lower DNA loads. Mouse data are from Fig. 5 of Salinas et al. (1986). Continued on next page.

g/cm^3, this cannot be responsible for the higher ρ_o value (by 1.7 mg/cm^3) compared to mole, nor for the 2.8 mg/cm^3 difference with shrew. Second, at least some differences are accompanied by other distinct compositional features. The infraorder Myomorpha not only shares a higher modal buoyant density compared to the other rodents investigated but also exhibits the lowest CsCl profile asymmetries, $\langle\rho\rangle - \rho_o$, as well as a scarcity of DNA having a modal buoyant density higher than 1.710 g/cm^3. Dormouse DNA showed a lower ρ_o value, 1.6991 g/cm^3, but a similar low asymmetry compared to other myomorph families.

The CsCl profiles of Fig. 2 also differ, as expected, in the number and the buoyant density of satellites showing up as shoulders or separate peaks (see the profiles of hedgehog, fruit bat, pangolin, mouse, mole rat, guinea pig, cat, horse, calf). Cryptic satellites, which cannot be detected in CsCl by definition, were revealed by centrifugation in preparative density gradients in the presence of a sequence-specific ligand, BAMD (Figs. 4–6). A second centrifugation in Cs_2SO_4/BAMD at a different ligand/nucleotide ratio was used in five cases (mouse and guinea pig, see Fig. 6; rabbit, dog and horse) in order to help assess the presence and amount of GC-rich components. The compositional distributions of Figs. 4–6 are presented as histograms in Fig. 7.

The CsCl density profiles of fractions from preparative Cs_2SO_4/BAMD density gradients (Figs. 4–7) show in most cases symmetrical peaks with increasing densities, except for the last, usually heterogeneous and multimodal, fraction(s). Satellite DNAs were detected in all genomes. In general, they appeared as DNA components deviating in buoyant density from the light-to-heavy progression of main-band fractions. In addition, or alternatively, they were distinguished by the sharpness of the corresponding peaks. Two typical examples of these phenomena are (1) the 1.6994 g/cm^3 human satellite DNA, which shows up as a narrow band in fraction 8, whereas a main band DNA component having exactly the same ρ_o value shows a broad peak and forms fraction 4; and (2) the 1.7046 g/cm^3 rat satellite DNA present in fraction 9, whereas a main band DNA component having exactly the same ρ_o value forms fraction 6. In other cases, the satellite peak could be identified because it appeared as such in the CsCl profile of Fig 2 (see, for example, the 1.7117 peak of fraction 7 of the fruit bat). Needless to say, the satellite DNAs of major interest here were those which could interfere with the assessment of GC-rich main-band DNA.



Figs. 4–6. Continued.

In order to estimate the amount of DNA higher than 1.710 g/cm^3 in ρ_o, the GC-richest preparative fractions of each DNA were assessed for the presence of satellite DNAs using the criteria just mentioned (see Table 3 and its footnotes). Then, the relative amounts of all "nonsatellite" DNA having a ρ_o value above 1.710 g/cm^3 were added up, as shown in Table 3. We assumed that the presence of GC-rich fractions from main-band DNA was confirmed when two subsequent fractions showed peaks corresponding to the same nonsatellite component and banding at a ρ_o value equal to, or higher than, 1.710 g/cm^3. The value of 1.710 g/cm^3 was chosen because it discriminates the DNAs of Myomorpha from those of most other mammals investigated.

Obviously, a sharp cut at 1.710 g/cm^3 is not an ideal procedure. This is clearly shown in an extreme case, in which a peak with a ρ_o value of 1.7097 g/cm^3, representing 9.3% of the rat genome (fraction 8), was not counted, whereas another one with a ρ_o value of 1.7108 representing 2% of the mouse genome (fraction 9) was. In this case, the identity of the compositional distributions of the two murid ge-

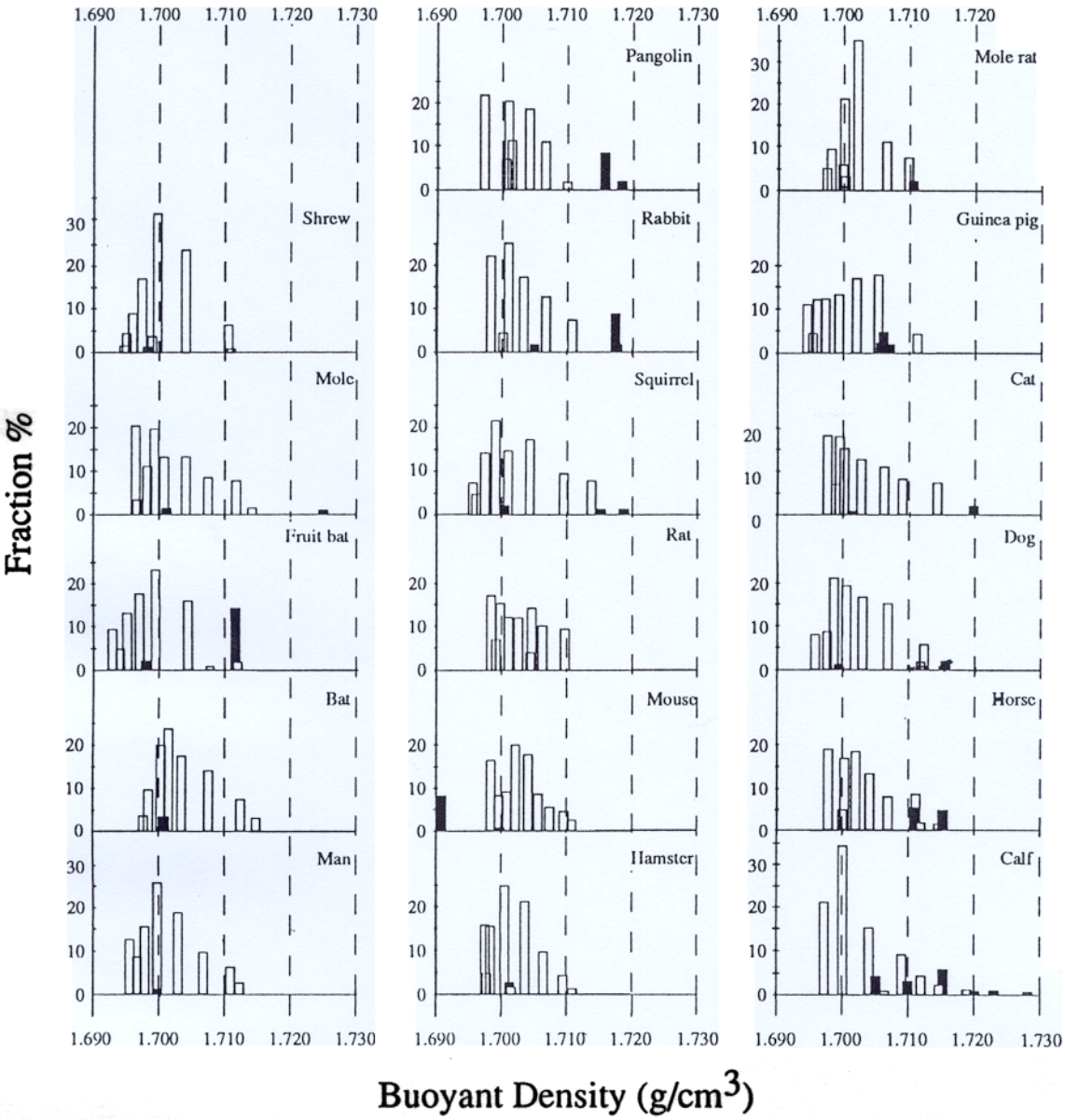


Fig. 7. Compositional distributions of mammalian DNAs. Diagrams are deduced from Figs. 4–6 and from previous data in the case of calf (Cortadas et al. 1977). *Black bars* concern satellite DNAs.

nomes (neglecting the mouse satellite DNA) was demonstrated by a detailed analysis of the CsCl profiles, as well as by the compositional distributions of third codon positions (Bernardi et al. 1988; Mouchiroud and Bernardi 1993). The results from the investigations carried out on the mouse genome (Salinas et al. 1986) can, therefore, be taken as a paradigm for both rat and mouse genomes. In the case of hamster, a more sizable amount (3–4%) of DNA above 1.710 g/cm^3 than that (2%) reported in Table 3 was found in a number of experiments reported elsewhere (Rynditch et al. 1991). A small difference between the genomes of hamster on the one hand and of rat and mouse on the other was confirmed by compositional comparisons of third codon positions (Mouchiroud and Bernardi 1993). Higher values for GC-rich DNA compared to those of Table 3 were found for guinea pig by a recentrifugation experiment which could be compared with a similar one carried on mouse DNA (Fig. 6). Again, the differences found between mouse and guinea pig were confirmed by a comparison of homologous sequences (Mouchiroud and Bernardi 1993).

The uncertainty in the quantitative estimates of nonsatellite DNA banding above 1.710 g/cm^3 is, however, mainly due to the widely different amounts and buoyant densities of heavy satellites in different mammals, as well as to the difficulty of quantitative estimations of relative amounts of DNA in the small GC-richest fractions. Another, less important, reason for uncertainty was the dif-

ferent amounts of satellite DNAs in different genomes and the fact that relative amounts are referred to total and not to main-band DNA. For these reasons, the significance of differences in the actual values ranging from 6 to 10% is uncertain and can only be estimated by compositional comparisons of coding sequences (see Mouchiroud and Bernardi 1993). Under these circumstances, no attempt was made to assess the very GC-rich ribosomal DNA, which only represents 0.25–0.5% of the genome of mammals. (See Meunier-Rotival et al. 1979.)

In spite of these problems, the approach used was satisfactory insofar as it could discriminate two classes of genomes: Those which comprise more than 6% and those which comprise less than 4% main-band DNA components higher in ρ_o than 1.710 g/cm^3. Results on the compositional distributions of coding sequences (third codon positions) and comparison of data from homologous sequences basically confirmed the conclusions drawn here in the cases in which this approach could be applied (Mouchiroud et al. 1988; Mouchiroud and Bernardi 1993).

Discussion

The Compositional Patterns of Mammalian Genomes: the "General" and the "Special" Patterns

In this work, our goal was to assess differences in compositional patterns at the DNA level among and within mammalian orders and to consider their relevance for mammalian phylogeny. The major features of the compositional distributions of large DNA fragments are (1) the modal buoyant density, ρ_o; (2) the mean buoyant density, $\langle\rho\rangle$, and the CsCl profile asymmetry, $\langle\rho\rangle - \rho_o$; and (3) the relative amounts of DNA fragments having a ρ_o value above a certain level, which was taken here as 1.710 g/cm^3. Among these features, the first one is significantly different for the three families belonging to the insectivores, an order known to group together diverse families, and for the two suborders of chiropters. Smaller differences were found among the three infraorders of rodents and even within the infraorder of Myomorpha. These differences might, however, be due to differences in DNA methylation, which causes a decrease of buoyant density (Kirk 1967). This explanation certainly does not hold, however, for Myomorpha, because compositional differences matching those reported here were found at the coding sequence level (see Mouchiroud and Bernardi 1993), and this is unlikely to be the only reason for the difference between chiropters, because the difference in buoyant density is too large. It should be noted that modal buoyant densities were already studied by Arrighi et al. (1970) in 91 species from 10 eutherian orders (those studied here, except for Pholidota, plus Edentata and Dermoptera). In the six species identical to those studied here, the reported ρ_o values were systematically lower by 1–2 mg/cm^3, but the range of all values was the same as found by us. Unfortunately, these authors did not report the CsCl profiles of their DNAs. As for the second feature, the low asymmetry of CsCl profiles, $\langle\rho\rangle - \rho_o$, clearly distinguishes the Myomorpha and pangolin (if the CsCl profile of the latter is corrected for satellite DNAs) from all other mammals investigated. The third feature, the amount of DNA having a $\rho_o >$ 1.710 g/cm^3, appears to be the most indicative one. (See below.)

Using these criteria, we have identified one "general" and three "special" compositional patterns. The "general" compositional pattern of mammalian genomes is basically characterized by the presence of a relatively large amount (6–10%) of main-band DNA higher than 1.710 g/cm^3 in ρ_o. This pattern is shared by species belonging to eight out of the nine orders investigated—namely, Insectivora, Chiroptera, Primates, Lagomorpha, Rodentia, Carnivora, Perissodactyla, and Artiodactyla.

A second compositional pattern, the "myomorph pattern," is that shown by a rodent infraorder, Myomorpha (Colbert and Morales 1991), comprising four families investigated here, murids, cricetids, spalacids, and glirids. This pattern exhibits a narrower compositional distribution of DNA fragments compared to the general distribution. As a consequence, DNA fragments are scarce above a buoyant density of 1.710 g/cm^3 (although less so in hamster; see Rynditch et al. 1991), whereas ρ_o values of unfractionated DNAs are high (1.700 g/cm^3; although less so, 1.6991 g/cm^3, for dormouse, which has a rather uncertain taxonomic position; Catzeflis et al. 1992).

Compositional patterns resembling those of Myomorpha, in that they lack GC-rich components, are exhibited by pangolin, a species belonging to the single genus (*Manis*) of the order Pholidota, and by the fruit bat, which belongs to the Pteropodidae family of Chiroptera, and whose DNA exhibits the lowest modal buoyant density of all mammalian DNAs investigated here (1.6965 g/cm^3). These two patterns are, however, different from each other. Indeed, while in the first case, the distribution is narrower than the general distribution (disregarding the very GC-rich satellite; see Table 3), as already described for Myomorpha, in the case of the fruit bat the distribution is broad (even after correction

Table 3. An analysis of DNA fractions higher than 1.710 g/cm^3 in modal buoyant density

Species	Fraction	Relative amount (%)		ρ_o[a] (g/cm^3)	>1.710[b] (%)
		Fraction	Peaks		
	7	6.5			
	8	1.9			
			1.5	1.6988 s[c]	
			0.4	1.7108	6.9
Mole	8	7.7		1.7118	
	9	3.4			
			1.2	1.7008 s[c]	
			1.2	1.7142	
			1.0	1.7249 s[d]	8.9
Fruit bat	7	13.9			
			0.4	1.6980 s[c]	
			0.5	1.7060	
			13.0	1.7117 s[e]	
	8	3.9			
			1.6	1.6980 s[c]	
			0.5	1.7077	
			1.8	1.7118 s[e]	<1
Bat	7	7.1			
			0.2	1.7000 s[c]	
			6.9	1.7123	
	8	5.7			
			2.9	1.7009 s[c]	
			2.8	1.7146	9.7
Man	7	6.2			
			0.3	1.6990 s[c]	
			5.9	1.7106	
	8	3.6			
			1.0	1.6994 s[c]	
			2.6	1.7120	8.5
Pangolin	7	8.1			
			2.1	1.7090	
			6.0	1.7157 s[d]	
	8	3.2			
			1.5	1.7099	
			1.7	1.7183 s[d]	<1
Rabbit	6	7.2		1.7107	
	7	8.7		1.7175 s[d]	
	8	3.2			
			1.6	1.7051 s[c]	
			1.6	1.7176 s[d]	7.2
Squirrel	8	7.7		1.7137	
	9	3.8			
			2.0	1.7004 s[c]	
			1.0	1.7150	
			0.8	1.7187 s[d]	8.7
Rat					<1
Mouse	9	2.9			
			0.5	1.7000 s[c]	
			2.4	1.7108	2.4
Hamster	8	3.6			
			2.0	1.7014 s[c]	
			1.6	1.7108	1.6[f]
Mole rat	8	5.0			
			3.0	1.7000 s[c]	
			2.0	1.7104	2
Guinea pig	8	8.6			
			4.0	1.7059 s[c]	
			4.6	1.7112	
	9	2.2			
			1.7	1.7058 s[c]	
			0.5	1.7130	5.1[g]

Table 3. Continued

Species	Fraction	Relative amount (%) Fraction	Peaks	ρ_o[a] (g/cm^3)	>1.710[b] (%)
	8	7.4	6.5		
			0.9		
		2.2			
			0.4	1.7013 s[c]	
			0.2	1.7120	
			1.6	1.7197 s[d]	6.7
Dog	7	5.7		1.7121	
	8	3.3			
			1.6	1.7118 s[g]	
			1.7	1.7154 s[g]	
		2.3			
			1.0	1.6992 s[c]	
			0.7	1.7119 s[g]	
			0.6	1.7150 s[g]	5.7
Horse	7	8.3		1.7109	
		9.5			
			5.0	1.7109 s[g]	
			4.5	1.7149 s[g]	
	9	2.9			
			1.5	1.7117	
			1.4	1.7145 s[g]	8.3

[a] s indicates satellite DNAs, as identified using the criteria discussed in the text and in the following footnotes.
[b] Relative amount of nonsatellite DNA having a modal buoyant density higher than 1.710 g/cm^3
[c] Peak too low in ρ_o for the fraction in which it is present
[d] Peak too high in ρ_o to belong to main-band DNA. Also, in some cases, large separation in ρ_o from preceding fraction
[e] Peak corresponding to a satellite in CsCl profile (see Fig. 2)
[f] A higher value, close to 4%, can be estimated from the experiments of Rynditch et al. (1991)
[g] Sharp peak

for the GC-rich satellite), but shifted to lower values as a whole. In this case, the heaviest main-band peak is at 1.7077 g/cm^3, but small, additional amounts of GC-richer DNA might be hidden under the very GC-rich satellite DNA. The possible contribution of methylation to the density shift of fruit bat DNA remains to be determined.

Phylogenetic Implications of Compositional Patterns: Some General Considerations

Two considerations suggest that the compositional patterns of warm-blooded vertebrates may have a phylogenetic relevance. (1) The two compositional patterns which were studied in a number of species, the general and the myomorph patterns, are characterized by a remarkable stability over geological time (see the following two sections). Such a stability was first detected in early work (Thiéry et al. 1976) but received its strongest support from results obtained on compositional patterns at the coding sequence level (Mouchiroud et al. 1988; Bernardi et al. 1988; Mouchiroud and Bernardi 1993), as well as by analyses of the buoyant density properties of unfractionated DNA (Bernardi and Bernardi 1990b). Interestingly, a conserved mode of chromosomal change is also observed in these two groups of mammals (O'Brien and Seuanez 1988). (2) The invariance of the compositional patterns of eight birds belonging to eight different orders (Kadi et al. 1993) matches the widely accepted monophyly of birds (Ansari et al. 1988). In this case, even more than in that of mammals, there is a remarkable conservation in the karyotypes of the entire class (Ansari et al. 1988).

It should be noted that the stability of compositional patterns in warm-blooded vertebrates is in remarkable contrast with their relative instability in cold-blooded vertebrates (Bernardi and Bernardi 1990a,b, 1991). In fact, this situation and the compositional transition of the genome between cold- and warm-blooded led to the development of the concepts of the conservative and transitional (or shifting) mode in genome evolution (Bernardi et al. 1988).

Under these circumstances, differences in compositional patterns of warm-blooded vertebrates may have a phylogenetic relevance, which may, in turn, be expected to concern general rather than detailed features such as those studied by the analysis of homologous sequences. This contribution may, however, be very valuable if one considers the

number of unsolved problems in mammalian phylogeny.

Phylogenetic Implications: the General Compositional Pattern of Mammalian Genomes

The wide spread of the general compositional pattern of mammals suggests by itself its presence in a common ancestor for at least the eight orders in which it was found. Otherwise, one should postulate a much less likely convergence from several distinct ancestral compositional patterns into the common pattern found in the present-day mammals from the orders under consideration. It has been argued elsewhere (Bernardi 1993b) that this ancestral pattern was probably reached gradually, at the end of the evolution of early mammals, and that its striking compositional heterogeneity might be responsible for genome instability leading to karyotypic changes that, in turn, might have contributed to the mammalian radiation.

While the general compositional pattern of mammalian genomes has been shown to concern species from eight of the 17 orders of Eutheria, preliminary results indicate that species from at least one more order, Cetacea, also share the general pattern (F. Kadi, U. Arnason, G. Sabeur, and G. Bernardi, paper in preparation). In other words, the conclusion should be drawn that the majority of eutherian orders exhibit the general pattern, thus justifying this term. The general pattern suggests a common origin not only for lagomorphs, rodents, primates, chiropters, insectivores, and carnivores (see, however, the next section for the special cases of Myomorpha and of the megachiropteran), as proposed on the basis of paleontology (Fig. 8), but also of artiodactyls, cetaceans, and perissodactyls.

Phylogenetic Implications: the Special Compositional Patterns of Mammalian Genomes

As far as the compositional patterns other than the general pattern are concerned, two basic possibilities exist (Mouchiroud et al. 1988). Either these "special" patterns have an independent ancestral origin or they are derived from the general pattern.

1. The case for separate ancestry is definitely strong for the compositional pattern of pangolin. Indeed, the order Pholidota may have been among the first ones (together with the order Edentata) to differentiate among all eutherian groups (Novacek 1992; see Fig. 8), and might, therefore, have arisen from a different ancestor compared to all or most other mammalian orders.

2. In the case of different isochore patterns of the fruit bat and of the bat, it should be recalled that Megachiroptera, like the fruit bat, have been hypothesized (Pettigrew 1986, 1991) to be more closely aligned with primates than with Microchiroptera, such as the vespertilionid bat investigated here. This view is, however, contradicted by a body of paleontological and molecular evidence in favor of bat monophyly (see Stanhope et al. 1992, and references quoted therein). As far as isochore patterns are concerned, a striking difference definitely exists between fruit bat and bat. The pattern of the former clearly is, however, very different from that of primates, who share the general pattern with Microchiroptera. One should then draw the conclusion that the shift in the compositional pattern of the fruit bat took place after the appearance of a common ancestor for Mega- and Microchiroptera—namely, that it derived from the general pattern. In view of the stability of compositional patterns in warm-blooded vertebrates, of the very large difference in buoyant density exhibited by the DNA of fruit bat, and of the proposal that the two suborders of chiropters are only distantly related (Smith 1976), the explanation of a separate origin should be, however, kept open.

3. In the case of Myomorpha, Mouchiroud et al. (1988) already considered the two possibilities of a derived and of an ancestral origin for Myomorpha. The predominant opinion from paleontology (Novacek 1992) is that Myomorpha have a common ancestry with Sciuromorpha and Histricomorpha—namely, with infraorders that exhibit the general pattern, in which case the myomorph pattern would be derived from the general pattern. The view of a derived origin of the myomorph patterns would be in agreement (Fig. 9A) with the stratigraphic record indicating a derivation of the infraorder Myomorpha from the suborder Sciurognathi (Carroll 1988) and would not contradict the proposal that rodents and lagomorphs (which exhibit the "general pattern") belong to the same superorder Glires (Novacek 1990).

Molecular data have been interpreted, however, to suggest (Easteal 1990; Li et al. 1990; Bulmer et al. 1991) that Myomorpha branched off before the divergence among carnivores, lagomorphs, artiodactyls, and primates—a divergence of orders which all share the general pattern. In such a case, Myomorpha would have had a separate ancestor (Fig. 9B), a possibility already considered by Mouchiroud et al. (1988).

If, however, guinea pig (which exhibits the general pattern) diverged before the separation of primates and artiodactyls from Myomorpha, as suggested by Graur et al. (1991) (see also Allard et al.

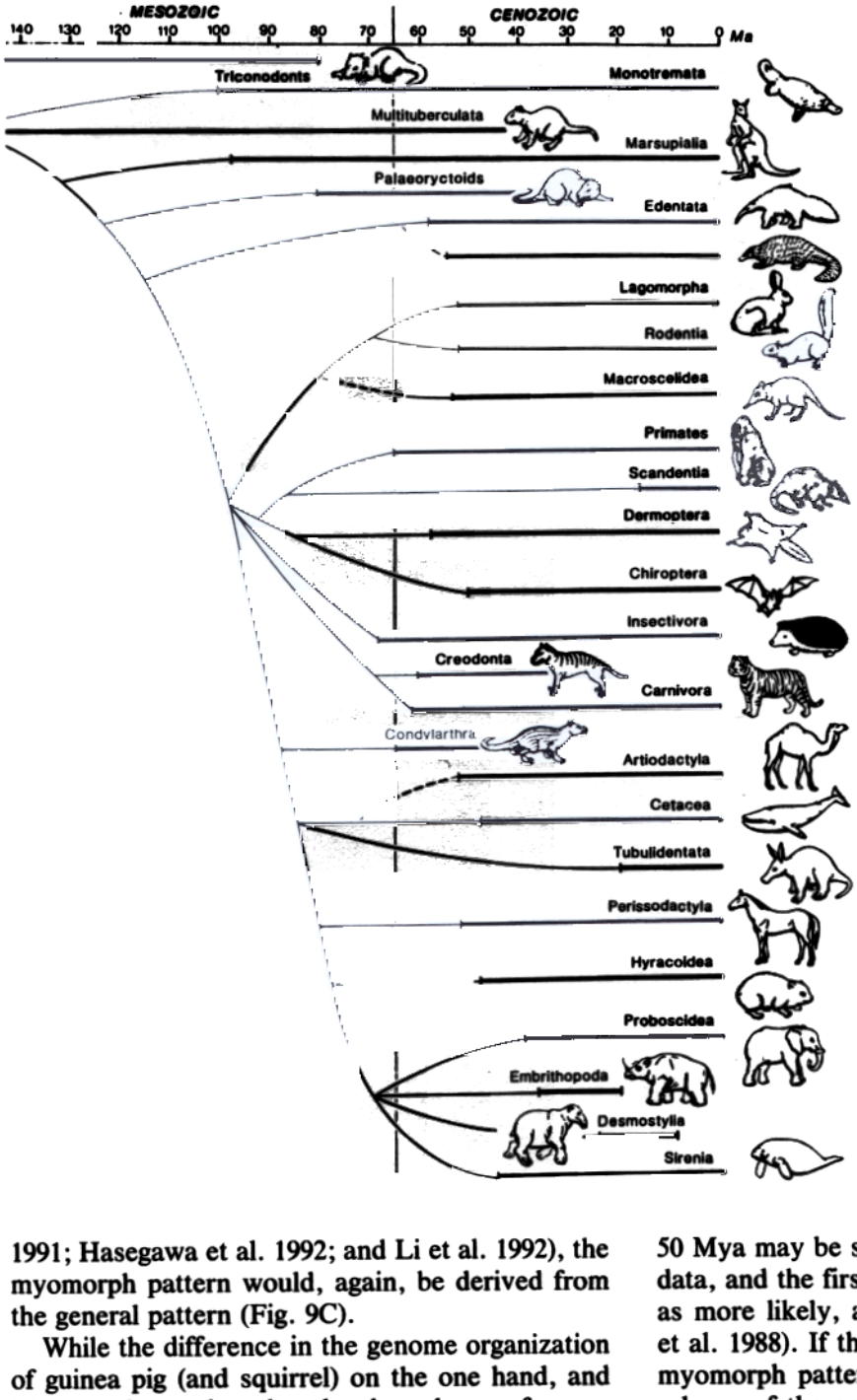


Fig. 8. A phylogenetic tree showing relationships among the major mammalian clades. The *solid horizontal bars* indicate the age range of the clade on the basis of dated first appearance in the fossil record; *solid lines* indicate the branching sequence, although the date of the actual splitting event can only be inferred from the relationships of the clades and their known ages. *Dashed lines* indicate relatively more ambiguous relationships. (From Novacek 1992.)

1991; Hasegawa et al. 1992; and Li et al. 1992), the myomorph pattern would, again, be derived from the general pattern (Fig. 9C).

While the difference in the genome organization of guinea pig (and squirrel) on the one hand, and Myomorpha on the other, has been known for several years now (Bernardi et al. 1988), the evidence for the early divergence of guinea pig (Graur et al. 1991) relative to Myomorpha is far from compelling (Luckett and Hartenberger 1993). The choice remains, therefore (Mouchiroud et al. 1988), between the first two possibilities, which cannot be reconciled because the stratigraphic record for Myomorpha does not go beyond 50 Mya (Carroll 1988), whereas Li et al. (1990) place the rodent (in fact, the murid) divergence at 100 Mya.

At present, the absence of a fossil record for over 50 Mya may be seen as outweighing the molecular data, and the first possibility should be considered as more likely, as already proposed (Mouchiroud et al. 1988). If the general pattern is primitive, the myomorph pattern could be derived from it by a release of the compositional constraints operating on the "tails" of the compositional distribution of coding sequences and third codon positions. This would, indeed, push such "tails" toward 50% GC—it would lower the high values and raise the low values of the compositional distribution. (See Mouchiroud and Bernardi 1993.)

The second, unorthodox scheme, with Myomorpha having a separate ancestor, deserves, however, further studies because of its very interesting implications. Indeed, if correct, not only Pholidota, but also Myomorpha, might have a (perhaps common)

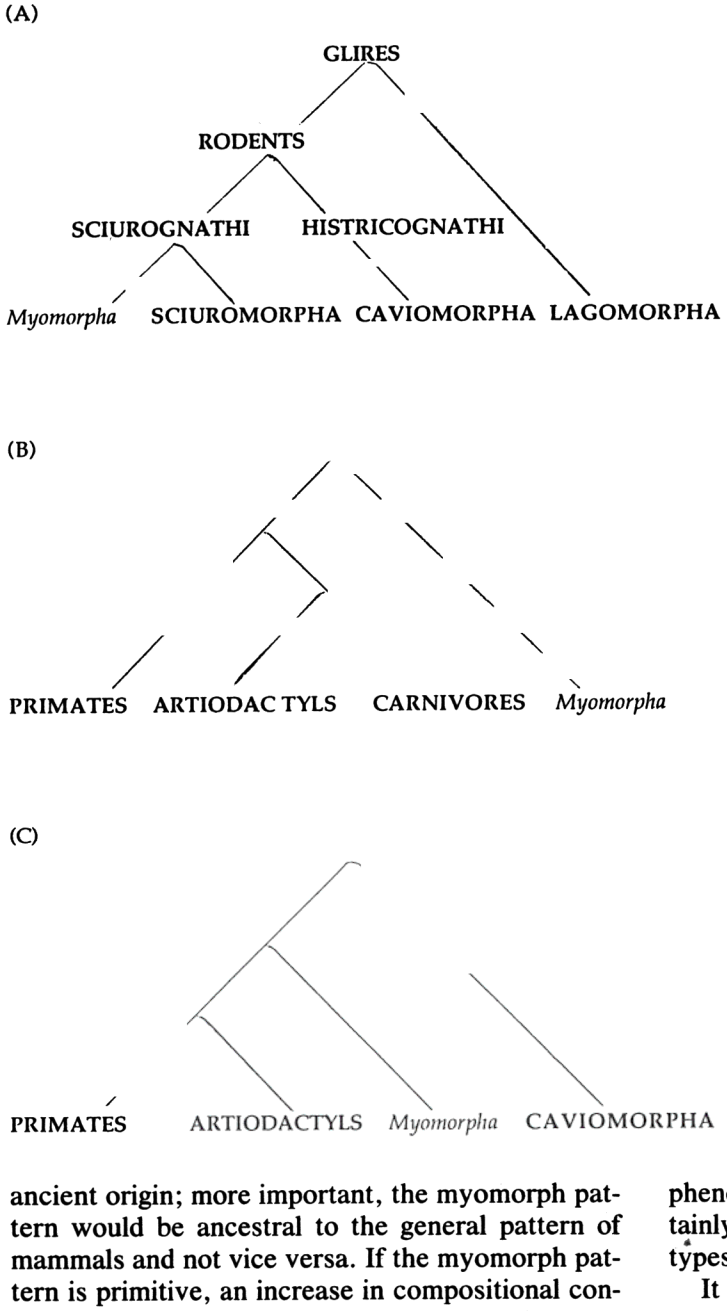


Fig. 9. Schemes of origins of Myomorpha according (**A**) to paleontological views (Carroll 1988; Novacek 1990), (**B**) to Li et al. (1990), and to Bulmer et al. (1991) and (**C**) to Li et al. (1992).

ancient origin; more important, the myomorph pattern would be ancestral to the general pattern of mammals and not vice versa. If the myomorph pattern is primitive, an increase in compositional constraints should have taken place between the myomorph and the general patterns, leading to an increased heterogeneity in the latter case. Interestingly, in such a case, the myomorph pattern would be intermediate, as expected, between the low heterogeneity pattern of cold-blooded vertebrates (Bernardi and Bernardi 1990a,b, 1991) and the high heterogeneity pattern of most mammals, although, obviously, much closer to the latter.

While the DNA compositional patterns of the megachiropteran and of myomorphs do not allow us to decide at present whether they are primitive or derived patterns, they are so different from those of the suborder Microchiroptera and of other rodent infraorders, respectively, that their current taxonomical level should be raised. Indeed, genome phenotypes (Bernardi and Bernardi 1986) are certainly no less important than morphological phenotypes.

It should be noted that the results presented here were obtained before the controversies on rodents and chiropters arose, and that more detailed analyses from the species investigated, as well as analyses of DNAs from other species chosen in view of the current controversies, might shed further light on the problems discussed above. Moreover, the present work should be extended to the mammalian orders which were not explored here. Indeed, there is a possibility for additional differences to be found in other orders, like Monotremata and Marsupialia, which have a very different past history compared to Eutheria, or like Dermoptera, Tubulidentata, and Hyracoidea, which have a controversial phylogenetic status. The analysis of compositional patterns at the DNA level, used here, and the combination of compositional heterogeneities and CsCl profile

asymmetries (G. Bernardi and G. Bernardi, paper in preparation) lend themselves especially well to a rapid screening of a large number of species.

Acknowledgments. We wish to thank our colleagues Giacomo Bernardi, Dominique Mouchiroud, and Emile Zuckerkandl for discussions, Dr. R. Fons (Laboratoire Arago, Banyuls-sur-Mer, France) for shrews, Dr. E. Nevo (Tel Aviv University, Tel Aviv, Israel) for mole rats, Dr. Geneviève Fourel (Institut Pasteur, Paris, France) for providing woodchuck DNA, Dr. Ariel Lustig (Biozentrum, Basel, Switzerland) for securing a multiplexer-pickup for our analytical ultracentrifuge, and Dr. J. Filipski for his help with analytical ultracentrifugation.

References

Aïssani B, D'Onofrio G, Mouchiroud D, Gardiner K, Gautier C, Bernardi G (1991) The compositional properties of human genes. J Mol Evol 32:497–503

Allard MW, Myamoto MM (1991) Tests for rodent polyphyly. Nature 353:610–611

Ansari HA, Takagi N, Sasaki M (1988) Morphological differentiation of sex chromosomes in three species of ratite birds. Cytogenet. Cell Genet 47:185–188

Arrighi FE, Mandel M, Bergendahl J, Hsu TC (1970) Buoyant densities of DNA of mammals. Biochem Genet 4:367–376

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

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

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

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

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

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

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

Bernardi G (1993b) Genome organization and species formation in vertebrates. J Mol Evol (in press)

Bettecken T, Aïssani B, Müller CR, Bernardi G (1992) Compositional mapping of the human dystrophin gene. Gene 122: 329–35

Bulmer M, Wolfe KH, Sharp PM (1991) Synonymous nucleotide substitution rates in mammalian genes: implications for the molecular clock and the relationship of mammalian orders. Proc Natl Acad Sci USA 88:5974–5978

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

Catzeflis FM, Aguilar J-P, Jaeger J-J (1992) Muroid rodents: phylogeny and evolution. Trends Ecol Evol 7:122–126

Colbert EH, Morales M (1991) Evolution of the vertebrates, 4th ed. Wiley-Liss, New York

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

Cortadas J, Macaya G, Bernardi G (1977) An analysis of the bovine genome by density gradient centrifugation: fractionation in Cs_2SO_4/3,6 bis (acetato-mercurimethyl) dioxane density gradient. Eur J Biochem 76:13–19

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

D'Onofrio G, Mouchiroud D, Aïssani B, Gautier C, Bernardi G (1991) Correlations between the compositional properties of human genes, codon usage and aminoacid composition of proteins. J Mol Evol 32:504–510

Easteal S (1990) The pattern of mammalian evolution and the relative rate of molecular evolution. Genetics 124:165–173

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

Graur D, Hide WA, Li W-H (1991) Is the guinea-pig a rodent? Nature 351:646–652

Hasegawa M, Cao Y, Adachi J, Yano T-A (1992) Rodent polyphyly? Nature 355:595

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

Kadi F, Mouchiroud D, Sabeur G, Bernardi G (1993) The compositional patterns of the avian genomes and their evolutionary implications. J Mol Evol (submitted)

Kay ERM, Simmons NS, Dounce AL (1952) An improved preparation of sodium desoxyribonucleate. J Am Chem Soc 74: 1724–1726

Kirk JTO (1967) Effect of methylation of cytosine residues on the buoyant density of DNA in caesium chloride solutions. J Mol Biol 28:171–712

Kopecka H, Macaya G, Cortadas J, Thiery JP, Bernardi G (1978) Restriction enzyme analysis of satellite DNA components from the bovine genome. Eur J Biochem 84:189–195

Li W-H, Gouy M, Sharp PM, D'Huigin C, Yang Y-W (1990) Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. Proc Natl Acad Sci USA 87:6703–6707

Li W-H, Hide WA, Graur D (1992) Origin of rodents and guinea-pigs. Nature 359:277–278

Luckett WP, Hartenberger J-L (1993) Monophyly or polyphyly of the order Rodentia: possible conflict between morphological and molecular interpretations. J Mammal Evol (in press)

Macaya G, Cortadas J, Bernardi G (1978) An analysis of the bovine genome by density gradient centrifugation. Eur J Biochem 84:179–188

Meunier-Rotival M, Cortadas J, Macaya G, Bernardi G (1979) Isolation and organization of calf ribosomal DNA. Nucleic Acids Res 6:2109–2123

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, Bernardi G (1988) The compositional distribution of coding sequences and DNA molecules in humans and murids. J Mol Evol 27:3311–320

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

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

Mouchiroud D, Bernardi G (1993) Compositional properties of coding sequences and mammalian phylogeny. J Mol Evol (in press)

Novacek MJ (1990) Morphology, paleontology and the higher clades of mammals. In: Genoways HH (ed) Current mammalogy vol 2. Plenum, New York, pp 507–543

Novacek MJ (1992) Mammalian phylogeny: shaking the tree. Nature 356:121–125

Nowak RM, Paradiso J (1983) Walker's mammals of the world, 4th ed. Johns Hopkins University Press, Baltimore and London.

O'Brien SJ, Seuanez HN (1988) Mammalian genome organization: an evolutionary view. Ann Rev Genet 22:323–351

Pettigrew JD (1986) Flying primates? Megabats have the advanced pathway from eye to midbrain. Science 231:1304–1306

Pettigrew JD (1991) Wings or brain? Convergent evolution in the origin of bats. Syst Zool 40:199–216

Rynditch A, Kadi F, Geryk J, Zoubak S, Svoboda J, Bernardi G (1991) The isopycnic, compartmentalized integration of Rous sarcoma virus sequences. Gene 106:165–172

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 JD (1976) Comments on flight and evolution of bats. In: Hecht MR, Goody PC, Hecht BM (eds) Major patterns of vertebrate evolution. Plenum Press, New York, pp 427–437

Stanhope MJ, Czelusniak J, Si J-S, Nickerson J, Goodman M (1992) A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly. Mol Phylogenet Evol 1:148–160

Sueoka N (1959) A statistical analysis of deoxyribonucleic acid distribution in density gradient centrifugation. Proc Natl Acad Sci USA 45:1480–1490

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

Zerial M, Salinas J, Filipski J, Bernardi G (1986) Gene distribution and nucleotide sequence organization in the human genome. Eur J Biochem 160:479–48