CrossMark

REVIEW PAPER

The Isochores as a Fundamental Level of Genome Structure and Organization: A General Overview

Maria Costantini[1] · Héctor Musto[2]

Abstract The recent availability of a number of fully sequenced genomes (including marine organisms) allowed to map very precisely the isochores, based on DNA sequences, confirming the results obtained before genome sequencing by the ultracentrifugation in CsCl. In fact, the analytical profile of human DNA showed that the vertebrate genome is a mosaic of isochores, typically megabase-size DNA segments that belong to a small number of families characterized by different GC levels. In this review, we will concentrate on some general genome features regarding the compositional organization from different organisms and their evolution, ranging from vertebrates to invertebrates until unicellular organisms. Since isochores are tightly linked to biological properties such as gene density, replication timing, and recombination, the new level of detail provided by the isochore map helped the understanding of genome structure, function, and evolution. All the findings reported here confirm the idea that the isochores can be considered as a "fundamental level of genome structure and organization." We stress that we do not discuss in this review the origin of isochores, which is still a matter of controversy, but we focus on well established structural and physiological aspects.

Keywords Base composition · Invertebrates · Isochores · Vertebrates · Unicellulars

✉ Maria Costantini
maria.costantini@szn.it

[1] Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy

[2] Laboratorio de Organización y Evolución del Genoma, Unidad de Genómica Evolutiva, Facultad de Ciencias, 11400 Montevideo, Uruguay

Introduction to Isochores

The ultracentrifugation in Cs_2SO_4 density gradients in the presence of sequence-specific ligands (e.g., Ag^+) was shown to lead to a high resolution of mammalian DNAs according to base composition much before genome sequencing (Corneo et al. 1968). These findings offered new perspectives in the study of the organization of eukaryotic genomes, taking the place of DNA reassociation kinetics based on the separation of single- and double-stranded DNA on hydroxyapatite (Bernardi 1965; Britten and Kohne 1968). More than 50 years ago, calf thymus DNA, the standard eukaryotic DNA, was shown to be remarkably more heterogeneous in base composition than bacterial DNAs (Meselson et al. 1957). Interestingly, high-resolution ultracentrifugation of this DNA showed a discontinuous compositional heterogeneity of the main band, consisting of three families of DNA molecules, also separating the GC-rich satellites (Filipski et al. 1973). The families of DNA molecules were then shown in the other mammalian genomes (including the human genome) explored and were defined as fairly homogeneous DNA stretches (Macaya et al. 1976; Thiery et al. 1976) called isochores (Cuny et al. 1981) for compositionally equal landscapes. The first family was then resolved into two families, L1 and L2; the second and the third families were called H1 and H2, respectively; and another quantitatively small family, H3 was identified (Zerial et al. 1986), neglecting the satellite DNAs (~2% of the genome) and ribosomal DNAs (~0.5% of the genome). The isochore families were characterized by their increasing GC levels from L1 to H3.

Twenty-five years after these findings, also thanks to the availability of the complete sequence of the human genome, different computational approaches have been used to disprove or redefine isochores (Eyre-Walker and

Hurst 2001; Häring and Kypr 2001; Lander et al. 2001; Nekrutenko and Li 2001; Cohen et al. 2005). In particular, Eyre-Walker and Hurst (2001) accepted the existence of the isochores even if they claimed that the question of why there is large-scale variation in base composition along mammalian and avian chromosomes is far from resolved because none of the available hypotheses adequately explained all the data. Lander et al. (2001) studied the draft of the human genome sequence to determine strict isochores. Their results ruled out a strict notion of isochores as compositionally homogeneous, whereas being a substantial variation at many different scales. They concluded that although isochores did not appear to merit the prefix `iso,' the genome clearly contained large regions of distinctive GC content, suggesting to redefine isochore concept so as to partition rigorously the genome into regions. Häring and Kypr (2001) demonstrated that the isochores should be defined in unambiguous molecular terms if they were used for an up-to-date genome structure characterization. They calculated the GC content variations along the DNA molecules of the human chromosomes 21 and 22, and found the variations to be higher everywhere compared to the randomized sequences. According to their findings, the GC content was certainly not homogeneous on the isochore scale in the two human chromosomes. In addition, no significant difference between the two human molecules and the genome of *Escherichia coli* have been found, regarding the GC content variations. Hence, no isochores were both present in the DNA molecules of the human chromosomes 21 and 22, or the isochores were also present in the genome of *E. coli*. Nekrutenko and Li 2001 studied the compositional patterns of eukaryotic genomes, using large amounts of long genomic sequences and developing a simple measure (called the compositional heterogeneity or variability index) to compare the differences in compositional heterogeneity between long genomic sequences. They reported the following: (i) The extent of the compositional heterogeneity in a genomic sequence strongly correlated with its GC content in all multicellular eukaryotes studied independently from the genome size. (ii) The human genome appeared to be highly compositionally heterogeneous both within and between individual chromosomes, going much beyond the predictions of the isochore model. (iii) All genomes of multicellular eukaryotes examined were compositionally heterogeneous, although they also contained compositionally uniform segments, or isochores. (iv) The human (or mammalian) genome was characterized by the presence of very high GC regions, exhibiting unusually high compositional heterogeneity and containing few long homogeneous segments (isochores). According to their findings, they concluded that GC-poor isochores were longer than GC-rich ones, indicating that the genomes of multicellular organisms were much more heterogeneous

in nucleotide composition than depicted by the isochore model, thereby leading to a looser definition of isochores. Cohen et al. (2005) addressed the question of the validity of the isochore theory through a rigorous sequence-based analysis of the human genome. By the selection criteria used in this study (distinctiveness, homogeneity, and minimal length of 300 kb), 1857 genomic segments have been identified, warranting the label of isochores. These putative isochores were not uniformly scattered throughout the genome and cover about 41% of the human genome. A four-family model of putative isochores has been found and these families were GC poor, with mean GC contents of 35, 38, 41, and 48% and not resembling the classical five isochore families. Moreover, due to large overlaps among the families, it was impossible to classify genomic segments into isochore families reliably, according to compositional properties alone. These findings undermined the utility of the isochore theory and seem to indicate that the theory may have reached the limits of its usefulness as a description of genomic compositional structures.

This debate prompted us to map the isochores, as originally defined (Macaya et al. 1976), in the finished sequence of the human genome (International Human Genome Sequencing Consortium 2004). The results obtained were in agreement with those obtained by equilibrium sedimentation and confirm the existence of five isochore families, going much farther in mapping isochores on chromosomes and leading to a resolution of chromosomal bands (Costantini et al. 2006). In more detail, a very simple approach has been used to segment the human chromosomes *de novo*, based on assessments of GC and its variation within and between adjacent regions. Scanning the GC profiles of human chromosomes from any starting point using a non-overlapping window of 100 kb, a mosaic of sequences has been found (ranging from 200 kb to several megabases), characterized by different GC levels and by a remarkable compositional homogeneity. A complete coverage of the human genome, neglecting the remaining gaps, revealed the presence of ~3200 isochores. The isochore pattern is, expectedly, different from chromosome to chromosome. However, when isochores are pooled in bins of 1% GC (Fig. 1), isochore families stand out. This is evident for isochore families L1, L2, and H1, but also visible for the H2 and H3 families, which are present in small amounts in the genome. The relative amounts of DNA in isochore families were 19, 37, 31, 11, and 3% for L1, L2, H1, H2, and H3 isochores, respectively, again in fair agreement with previous results (Macaya et al. 1976; Cuny et al. 1981).

This new level of detail provided by the isochore map helps to understand genome and chromosome structure, function, and evolution. The isochores were characterized by a different gene density. In fact, the genes are not distributed randomly in the human genome: indeed, gene density

Fig. 1 Distribution of iso-
chores according to GC levels
in vertebrate genomes. The
histograms show the distribu-
tion (by weight) of isochores as
pooled in bins of 0.5% GC for
chicken, zebrafish, pufferfish,
and human. Genome sizes are
calculated from the sums of
isochores. *Colors* represent the
five isochore families. Modified
from Costantini et al. 2009.
(Color figure online)
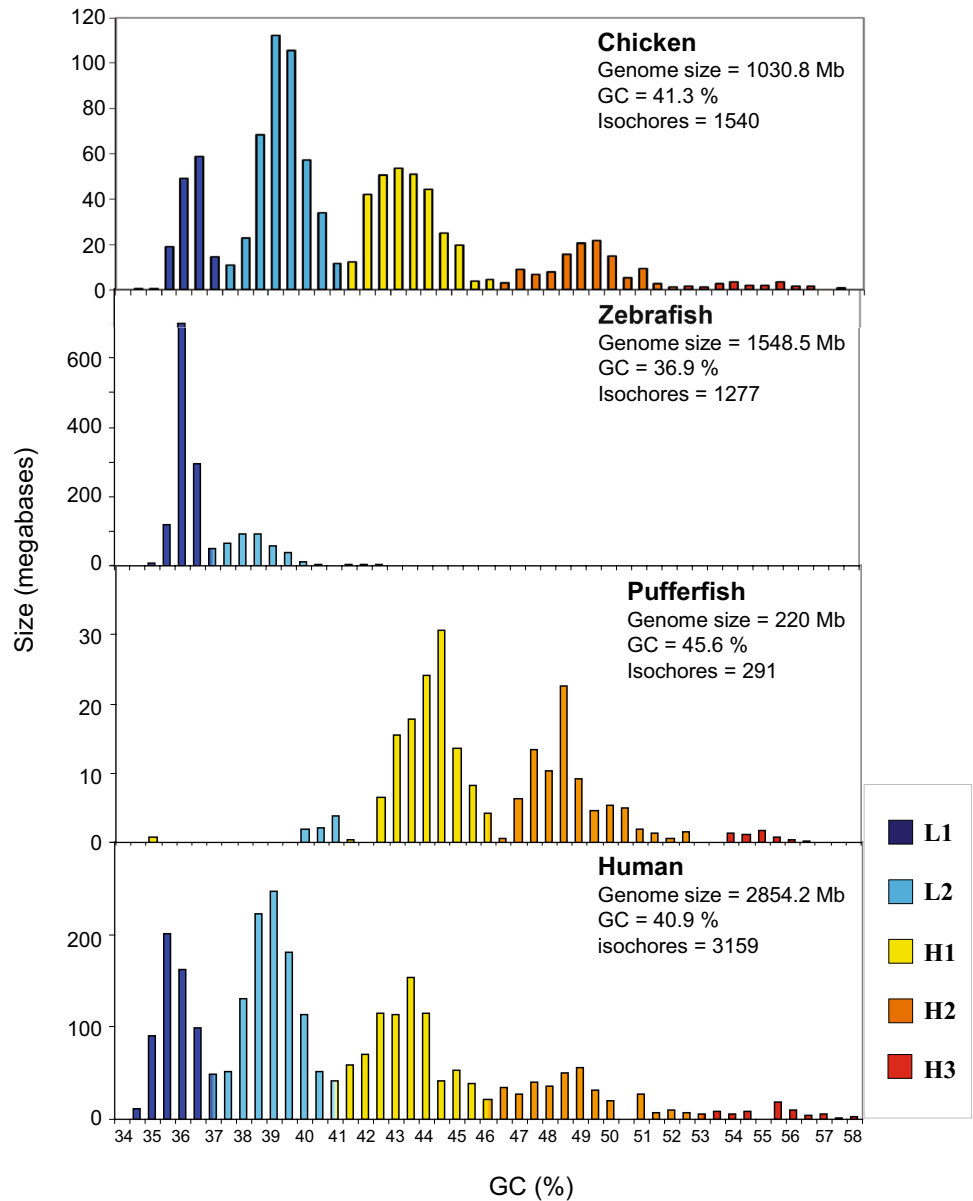
is low in GC-poor isochores, increases with increasing of GC in H1 and H2 and reaches a maximum in H3, even if this isochore family represents a small amount in the human genome. Because of these properties, the GC-poor isochores are called "genome desert" while the GC-rich ones are called "genome core." These two gene spaces are characterized by several different properties (for review, see Bernardi 2004), the most remarkable ones being the correlations of isochore families not only with gene density but also with replication timing, recombination, location, and chromatin structure in interphase nuclei, chromatin being "open" in the genome core and "closed" in the genome desert (Saccone et al. 2002). Isochore borders were identified on the basis of marked compositional differences and H3 isochores were always flanked by GC-poorer isochores, and L1 isochores were always flanked by GC-richer isochores, as expected (p value < 0.001). This is also the predominant situation found in the cases of H2, H1, and L2 isochores, respectively. However, these families also exhibited "transition isochores" in several cases, where one flanking isochore was higher, and the other lower. Very large GC differences at borders (such as L1/H3 borders) were rare, thus leading to the formation of blocks of isochores from closer families (e.g., L1/L2). Further investigations on the molecular basis of the classical Giemsa and Reverse bands in human chromosomes revealed that the ~3200 isochores of the human genome were assembled in high (850-band)- resolution bands, and the latter in low

(400-band)- resolution bands, so forming the nested mosaic structure of chromosomes (Costantini et al. 2007c). In more detail, the borders of both sets of chromosomal bands have been defined at the DNA sequence level on the basis of the map of isochores, representing the highest-resolution ultimate bands. The isochore-based level of definition (100 kb) of chromosomal bands was much higher than the cytogenetic definition level (2–3 Mb), thereby giving the solution of the long-standing problem of the molecular basis of chromosomal bands defined on the basis of compositional DNA properties alone. Previously, this problem has been also dealt by Niimura and Gojobori (2002), demonstrating that Giemsa-dark bands were the regions in which the GC content is relatively lower than that of the surrounding regions. These results implied the relationship between isochores and chromatin structures, inferring a different mechanism of isochore formation, and so propose that the functional constraint for retaining compact chromatin would be one contributor to forming isochores.

Chromosome replication timing is biphasic (early–late) in the cell cycle of vertebrates and of most (possibly all) eukaryotes. The extended, detailed replication timing maps that are available (namely those of human chromosomes 6, 11q, and 21q) have been compared with chromosomal bands as visualized at low (400 bands), high (850 bands), and highest (3200 isochores) resolution (Costantini et al. 2007c). The replicons located in a given isochore showed either all early or all late replication timing and that early-replicating isochores are short and GC-rich and late-replicating isochores are long and GC-poor. In the vast majority of cases, replicons are clustered in isochores, which are themselves most often clustered in early- or late-replication timing zones and may often reach the size of high-resolution bands and, very rarely, even that of low-resolution bands.

A recent investigation compared maps of the Topologically Associating Domains (TADs) and of the Lamina Associated Domains (LADs) with the corresponding isochore maps of mouse and human chromosomes (Jabbari and Bernardi 2017). This approach revealed that (1) TADs and LADs correspond to isochores, being the genomic units that underlie chromatin domains; (2) TADs and LADs are well conserved in mammalian genomes because of the evolutionary conservation of isochores; (3) chromatin domains, which correspond to GC-poor isochores, interact with other domains also corresponding to GC-poor isochores even if located far away on the chromosomes, whereas chromatin domains corresponding to GC-rich isochores show more localized chromosomal interactions, many of which are inter-chromosomal. In conclusion, this investigation establishes a link between DNA sequences and chromatin architecture, explains the evolutionary conservation of TADs and LADs, and provides new information on the spatial distribution of GC-poor/gene-poor and GC-rich/gene-rich chromosomal regions in the interphase nucleus.

Furthermore, an analysis of di- and tri-nucleotide densities in the isochores from the five families showed large differences. Densities of di- and tri-nucleotides were assessed on human DNA sequences 100 kb in size as derived from different isochore families (Costantini and Bernardi 2008b). Indeed, among dinucleotides, the "AT set," ApA, TpT, ApT, and TpA, showed a remarkable decrease from the L1 to H3 families. In contrast, the "GC set," CpC, GpG, CpG, and GpC, showed an increase, the CpG density reaching a fivefold higher level in H3 compared with L1 isochores. The same happened with trinucleotides. These different "short-sequence designs" (i) account for the fractionation of human DNA (and vertebrate DNA in general) when using sequence-specific ligands in density gradients, (ii) are very similar in whole isochores and in the corresponding intergenic sequences and introns, (iii) are reflected in different codon usages, (iv) lead to amino acid differences that increase the thermal stability of the proteins encoded by genes located in increasingly GC-rich isochore families, and (v) correspond to different chromatin structures. The significance of dinucleotide properties for local DNA structure (mainly in stacking energies) has been known for a long time (Dickerson 1992), but the periodicities of ApA/TpT/TpA and GpC in connection with position and stability of nucleosomes has been stressed only more later (Segal et al. 2006). The different densities of di- and tri-nucleotides suggested that chromatin structure may be different at the level of isochores belonging to different families. This has been demonstrated by mapping DNase-I hypersensitive sites and showing that the density of these sites on the human genome increases with increasing GC of isochores (Di Filippo and Bernardi 2008).

In the next sections, we will approach the general problem of the organization and evolution of genomes at the sequence level, ranging from vertebrates to invertebrates until unicellular organisms.

Vertebrates

The availability of a number of fully sequenced genomes ranging from fishes to mammals allowed to analyze the structure of their genomes, approaching the general problem of the organization and evolution of vertebrate genomes at the sequence level. Costantini et al. (2009) analyzed the genomes of the chicken *Gallus gallus*; four fishes, zebrafish (*Brachydanio rerio*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*) and pufferfish (*Tetraodon nigroviridis*); Eutherians not yet explored, namely chimpanzee (*Pan troglodytes*), mouse (*Mus musculus*), and dog (*Canis familiaris*); a Marsupial, the opossum

Monodelphis domestica; a Monotreme, (the platypus *Ornithorhynchus anatinus*); and a Reptile (the lizard *Anolis carolinensis*). Some comparative data from an Amphibian (*Xenopus tropicalis*) were also obtained, even if these genome sequences are only available as scaffolds.

Concerning the fishes, the two major compositional features of these genomes obtained by the ultracentrifugation approach previously used, namely the wide intergenomic spread of base composition and the narrow intragenomic distribution, were confirmed on the basis of sequences. Zebrafish was practically made up of only L1 and L2 isochores, with a predominance of the former family, whereas the isochores of pufferfish consist of H1 and H2 families with a minor presence of L2 and H3 isochores (Fig. 1; Costantini et al. 2007b). The compositionally "intermediate" genomes of medaka and stickleback consist essentially of L2/H1 and H1/H2 isochores, respectively, the first family being predominant in each case. Both the narrow intragenomic distribution and the wide compositional spread of fish genomes can be understood in two ways. First, as an adaptation to environmental and physiological factors (water temperature, metabolic rate, and oxygen consumption), as is the case in some prokaryotes (Musto et al. 2006; Romero et al. 2009; Naya et al. 2002), or second as a result of the mutational bias characteristic of each genome. Given the lack of experimental data, for the moment, we favor the second explanation.

The narrow intragenomic distribution may be visualized as an adaptation to a particular ecological niche; on the other hand, the wide compositional spread suggests the existence of compositional transitions, involving whole genomes and responding to changes in environmental conditions and new adaptations. These transitions may occur among genomes of fishes belonging to different orders, or even to different families independent of geological time (Bernardi and Bernardi 1990), leading to different patterns of isochore families.

The chicken genome has a size about 1/3 of the human genome and all isochore families were very slightly shifted toward GC-rich values compared to the human distribution (Costantini et al. 2007a). The differences concerned the existence of a minor isochore family, H4, which is absent in mammals. These data still lack some microchromosomes, all of which are known to be very GC-rich (Andreozzi et al. 2001). Taking into account that the last common ancestor of mammals and birds is estimated to go back to 310–340 Mya (van Rheede et al. 2006), that mammals and birds emerged at different times from different reptilian lines (from Therapsids, about 220 Mya, and from Dinosaurs, about 150 Mya, respectively), that the genome size of birds is about one-third that of mammals, and that the karyotype is profoundly different in most avian species, the similarities between the human

and the chicken genome are particularly striking. These similarities can be explained, as proposed by Bernardi and Bernardi (1986), to the necessity of stabilizing the genome mainly in the gene-rich part of it, which is characterized by an open chromatin structure (Saccone et al. 2002). The existence of H4 isochores might be related to the higher body temperature (about 41 °C) of birds compared to that of mammals (about 37 °C), as suggested by Kadi et al. (1993), although some other explanations cannot be completely excluded [see, for example, Duret et al. 2006]. Obviously, the similarities of the compositional patterns and of other genome properties, such as the GC level of isochore families, in the very distant genome of birds and mammals raise an important evolutionary question concerning the fate of the extremely large number of neutral or nearly neutral changes that occurred, as explained by the neo-selectionist theory of genome evolution (Bernardi 2007).

As expected, two Primates (human and chimpanzee) and a Carnivore (dog) showed a large similarity in the relative amounts of the isochore families, whereas in mouse L1 isochores were poorly represented and H3 isochores were essentially absent (Costantini et al. 2009). In opossum, L1 isochores were much more represented than in Eutherians, and GC-rich isochores H2 and H3 were very scarce. This pattern might be due to interspersed repeats that represent about 50% of this genome. In contrast to the genomes of Eutherians and chicken, which showed an average GC level of about 41%, and to the GC-poorer genome of opossum (~38% GC), the platypus genome (genome size of about 2.4 Gb, only 18% of which are assembled, the remaining sequences being available as supercontigs) showed a high GC level of 43.4%. This genome essentially consisted of L2 and H1 isochores with a small amount of H2 isochores, a result due in part to the missing assembly of GC-rich microchromosomes. The GC profile of the unassembled sequences was superimposed on the isochore profile of the platypus genome, showing that the unassembled parts essentially corresponded to GC-rich chromosomal regions, as well as in the case of chicken.

The reptile *Anolis carolinensis*, is indeed heterogeneous in base composition, since its macrochromosomes comprise isochores mainly from the L2 and H1 families, with the majority of the sequenced microchromosomes consisting of H1 isochores (Costantini et al. 2016). Unfortunately, only scaffolds were available for the genomes of the amphibian *Xenopus tropicalis*. When 100-kb segments of these scaffolds were binned and compared with similar histograms for the human and medaka genomes, the compositional heterogeneities of the *Xenopus* genome were found to be much lower than that of the human genome and rather close to that of the genome of medaka, the compositionally closest fish genome.

The average GC levels of isochores families were remarkably conserved, in spite of the different relative amounts of isochore families found within and among vertebrate classes. Because of their possible functional relevance in connection with chromatin structure (Costantini and Bernardi 2008b), dinucleotide frequencies were also assessed and the observed/expected ratios were extremely close between human, mouse, opossum, and platypus in each of the isochore families. The average size of isochores in the different families also showed a remarkable conservation in all vertebrates, from fish to human (see Table 1). The conservation of isochore size may be linked to the role played by isochores in chromosome structure and replication.

Invertebrates

The compositional organization of the invertebrate genomes explored comprised Nematodes, Arthropods, Echinoderma, and Chordata (Cammarano et al. 2009), having very different sizes of their genomes (Fig. 2). *Ciona intestinalis*, a Urochordate, which is the closest ancestral species to vertebrates (Delsuc et al. 2006), presented the existence of a major L1 and a minor L2 isochore family (de Luca di Roseto et al. 2002). The genome of the nematode *Caenorhabditis elegans* also displayed a very GC-poor genome, consisting essentially of L1 isochores with only a very small amount of L2 isochores. Interestingly, in the Platyhelminth *Schistosoma mansoni*, an isochore-like structure was found (although biased toward low GC values), associated with several features found in the human genome (Lamolle et al. 2016). Considering the insects, a sequence analysis

of the genomes of *Anopheles gambiae* and *Drosophila melanogaster* revealed that *Anopheles* DNA is more heterogeneous and GC-richer than *Drosophila* DNA (Jabbari and Bernardi 2004). The gene concentration across the *Anopheles* genome was characterized by low levels in the GC-poor part of the genome and a threefold increase in the GC-richest part; this gene density gradient was approximately half that of Drosophila (Jabbari and Bernardi 2000, 2004). Then with the availability of the complete sequences of the chromosomes, three *Drosophila* species (*D. melanogaster, D. simulans, D. yakuba*) and *A. gambiae* showed three isochore families: a minor L2 family, a predominant H1 family, and a H2 family which is barely represented in *Drosophila*, but is rather abundant in *Anopheles*. L1 family appeared as a very minor component (about 1%). The genome organization in terms of isochore families of insects seems to be very similar to those of a fish, stickleback.

The compositional patterns of scaffolds or contigs (for which the complete sequences of assembled chromosomes were not available) from *Branchiostoma floridae, Strongylocentrotus purpuratus, Aedes aegypti, Tribolium castaneum*, and *Daphnia pulex* revealed their compositional distributions generally narrow, covering a range of about 5% GC, with the exception of *T. castaneum*, in which case the range was about 10% GC, the center of distribution being lower (33%) than in the other cases. The average GC levels of the isochore families from the invertebrates investigated were very close to each other and to the corresponding values of vertebrates (Table 2). The isochore families in invertebrate genomes were generally characterized by GC levels that were identical or very close to those of vertebrates. Differences in dinucleotide patterns were found among invertebrates, as well as between invertebrates and vertebrates; in the latter case, the most salient feature was the CpG shortage which is due to the methylation of C in CpG followed by its deamination to T. No correlation was found between isochore size and genome size in spite of the very large genome size range explored so far, stressing their possible correlation with the structure and replication of chromosomes, as suggested by Costantini and Bernardi (2008a). The relative amounts of isochore families are different in different genomes, because of the different environmental factors playing a role in determining compositional patterns of genome. The gene concentration followed the general trends previously found for vertebrates, increasing with increasing GC of isochore families.

According to the compositional data obtained from the invertebrate genomes, an isochore structure appears to be general for all metazoans explored, raising the question if all eukaryotic genomes are characterized by an isochore structure. A subsequent work on unicellular eukaryotes clarified this point (Costantini et al. 2016).

Table 1 Average sizes in megabases (Mb) of isochore families from vertebrates and invertebrates

Average size (Mb)	L1	L2	H1	H2	H3
Vertebrates					
Human	0.9	0.9	0.8	0.7	0.7
Zebrafish	0.8	0.6			
Medaka		2.8	0.9		
Stickleback			2.1	0.7	
Pufferfish			0.9	0.7	
Invertebrates					
C. intestinalis	0.5	0.1	0.1		
C. elegans	3.4	0.5			
D. melanogaster	1.3	0.5	1.6	0.5	
D. simulans	0.9	0.4	2.0	0.5	
D. yakuba	1.3	0.4	1.7	0.5	
A. gambiae	0.6	0.7	1.3	1.0	

Fig. 2 Distribution of isochores according to GC levels in invertebrate genomes. The histograms show the distribution (by weight) of isochores as pooled in bins of 0.5% GC for *C. intestinalis, C. elegans, D. melanogaster*, and *A. gambiae*. Genome sizes are calculated from the sums of isochores. *Colors* represent the five isochore families. (Modified from Cammarano et al. 2009). (Color figure online)
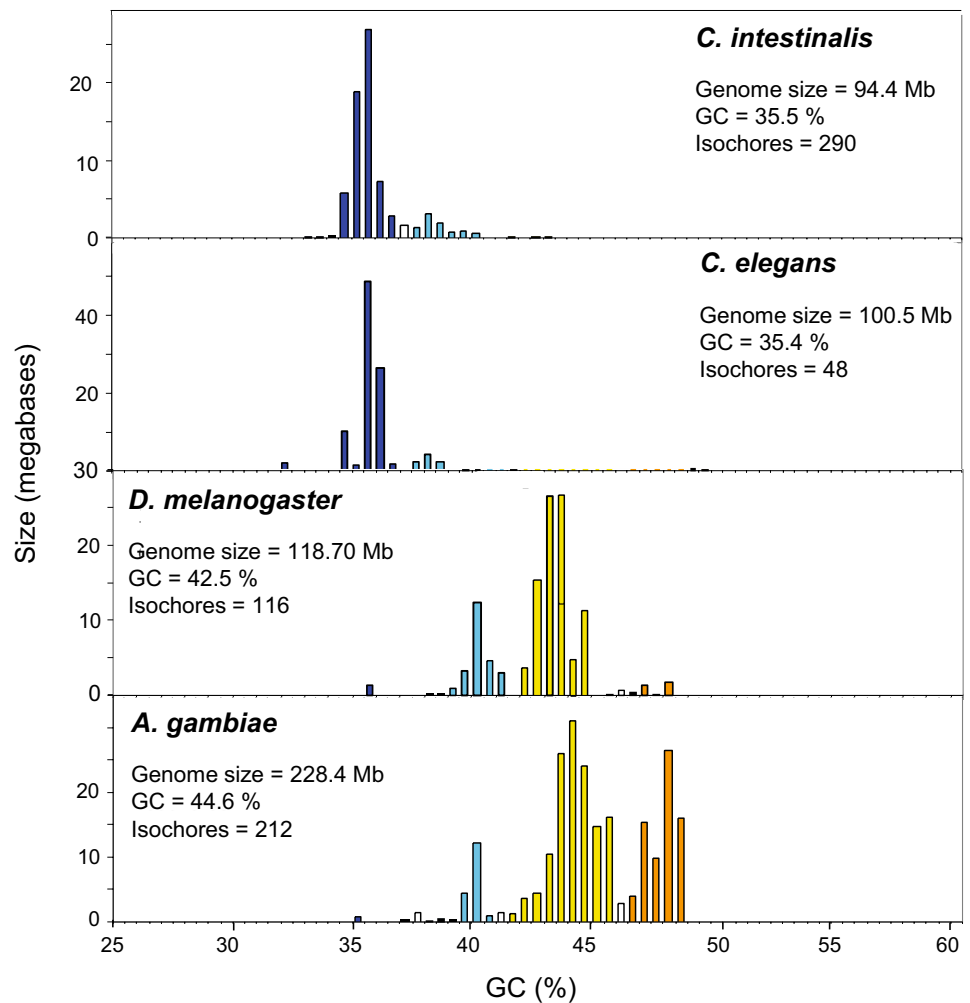
Unicellulars

The data reported until now revealed that the genomes of multicellular eukaryotes are compartmentalized in mosaics of isochores, belonging to a small number of families characterized by different average GC levels, by different gene concentrations, different chromatin structures, and different replication timings in the cell cycle.

A question raised by these basic results concerns how far back in evolution the compartmentalized organization of the eukaryotic genomes arose. Several findings encouraged to extend the investigations on compositional genome organization of unicellular eukaryotes. Previous results on the genomes from a small number of unicellular eukaryotes provided the first indication that a compositional compartmentalization was not only present in the genomes of multicellular eukaryotes, but also in those of some protozoa (Thiery et al. 1976; Pollak et al. 1982; McCutchan et al. 1984; Isacchi et al. 1993; Karlin et al. 1993; Sharp and Lloyd 1993; Musto et al. 1994; Rodríguez-Maseda and Musto 1994; Dujon 1996; Dekker 2007).

The different groups of unicellular organisms studied by Costantini et al. (2013) exhibited a diversity of genome compositional patterns, ranging from very weak to very strong compartmentalization (Fig. 3). These findings indicated that unicellular eukaryotes encompassed a wide range of genomic composition and heterogeneity. In fact, the average GC range of the genomes of these species was very broad (as broad as that of prokaryotes) and individual compositional patterns cover a very broad range from very narrow to very complex. Both features were not surprising for organisms very far from each other both in terms of phylogenetic distances and of environmental life conditions. More in detail, in the group of marine Algae, the green alga *Ostreococcus tauri* and the red alga *Cyanidioschyzon merolae* showed very GC-rich genomes, being the DNA centered at 59–60% GC and at 55–56% GC, respectively. More in detail, *O. tauri* genome had still previously surprised for its heterogeneity, a feature which is not only unusual but also perplexing from an evolutionary perspective (Derelle et al. 2006; Palenik et al. 2007). In fact, two chromosomes, 2 and 19, were different from the others, in terms

Table 2 Average GC (%) of isochore families from vertebrates (Costantini et al. 2009) and invertebrates (Cammarano et al. 2009), and of subcomponents from unicellular genomes (Costantini et al. 2013)

Average GC (%)	L1	L2	H1	H2	H3
Vertebrates					
Human	36.0	38.9	43.1	48.7	54.5
Chimp	36.0	38.9	43.2	48.6	55.0
Dog	35.9	38.9	43.2	48.7	55.8
Mouse	36.5	39.4	43.6	48.1	54.4
Platypus		40.0	42.9	47.9	54.9
Opossum	36.0	38.5	42.9	48.6	55.6
Chicken	36.6	39.3	43.4	48.8	54.7
Zebrafish	36.0	38.2			
Medaka		39.9	42.3		
Stickleback			44.2	47.3	
Pufferfish			44.4	48.2	54.7
Invertebrates					
C. intestinalis	35.5	38.3	41.7		
C. elegans	35.4	37.9			
D. melanogaster	35.3	40.0	42.9	47.2	
D. simulans	35.6	39.9	43.1	47.0	
D. yakuba	35.7	40.1	43.0	47.0	
A. gambiae	35.8	39.6	44.2	47.2	
Unicellulars					
O. tauri					59.4[a]
C. merolae				51.8	55.4
T. pseudonana				47.0	
P. tricornutum				48.9	
S. cerevisiae		38.5			
C. glabrata	35.7	38.9	43.5		
A. gossypii				50.7	53.8
C. neoformans				48.9	
T. brucei			43.2	47.5	
T. cruzi				49.6	54.7
T. gondii				52.7	55.0
D. discoideum	28.4[a]				
P. falciparum	19.4[a]				
P. chabaudi	22.5[a]				
P. berghei	22[a]				
P. knowlesi	35.0	38.5	42.8		
P. vivax	32.6	39.1	43.7	46.2	53.7
Overall average	35.6	39.1	43.2	48.5	54.8
Standard deviation	0.90	0.68	0.65	1.60	0.62

[a] These genomes were excluded from the calculation of the overall average because of their extreme values

of organization for chromosome 2 and function for chromosome 19. Both of these chromosomes have lower GC content than the 59% GC of the other 18 chromosomes. Chromosome 2 is composed primarily of two blocks, one with a GC content similar to that of the other chromosomes and the other with a markedly lower GC content (52%). The average GC content of the entire chromosome 2 amounts to 55%. Likewise, the GC content of chromosome 19 (54%) is similar to the atypical region of chromosome 2.

The marine diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (also studied in Bowler et al. 2008) showed medium GC genomes, consisting of components centered at 47 and 49% GC, respectively.

The genomes of fungi exhibited very different GC ranges: *Saccharomyces cerevisiae* and *Candida glabrata* showed GC-poor genomes, consisting of DNA components centered at 38–39%, accompanied in the case of *C. glabrata* by a minor component ranging from 34 to 38% GC and also by a very minor GC-richer component in the 42–46% GC range. In contrast, the other two fungi *Ashbya gossypii* and *Cryptococcus neoformans* showed GC-richer genomes: the first one centered at about 52–53% GC, the second one centered at 55% GC, whereas exhibited one component centered at 48–49% GC.

From a phylogenetic viewpoint, the protists were an exceptionally diverse group. Indeed, the genome-wide distances and times of divergence between two protozoan groups are many times larger than those of the most divergent metazoans. In particular, *Trypanosoma brucei* and *Trypanosoma cruzi* exhibited GC-rich genomes: the first one was essentially formed by a component centered at 48% GC, and by minor GC-poorer component; while on the other hand, the second species showed two main components, the first one centered at 48% GC, and a second, smaller one at 54% GC. The situation was more striking in the case of *Plasmodium* species. In fact, the malaria parasite *Plasmodium vivax* exhibited a genome covering a broad compositional spectrum (28–55% GC) with two major components centered at about 44 and 49% GC, whereas in an exceedingly sharp contrast, *Plasmodium chabaudi, Plasmodium berghei*, and *Plasmodium falciparum*, having genome sizes very close to that of *P. vivax*, showed very GC-poor genomes with single major components centered at 24, 22, and 19.4% GC, respectively. Only the *P. falciparum* genome showed some minor components ranging from 20 to 32% GC. *Plasmodium knowlesi* exhibited a genome pattern intermediate between *P. falciparum* and *P. vivax*, showing two major components centered at about 39% and 43% GC as well as a smaller component at 35% GC. The parasitic protist *Toxoplasma gondii* consisted of one major component centered at 52% GC and a smaller component at 55% GC, whereas the Amoeba *Dictyostelium discoideum* showed one major component centered at 28% GC. The average GC levels of the genome subcomponents from the unicellulars investigated were very close to each other and to the corresponding values of vertebrates and invertebrates (Table 2).

The analysis of the unicellular genomes has been prompted by several considerations: (i) the range of

Fig. 3 Distribution by weight
of DNA segments according
to GC levels in the *red* alga
C. merolae, in diatoms *P. tricor-
nutum*, in fungi *S. cerevisiae*
and *A. gossypii*, and in protists
T. brucei and *P. falciparum*.
(Modified from Costantini et al.
2013). (Color figure online)
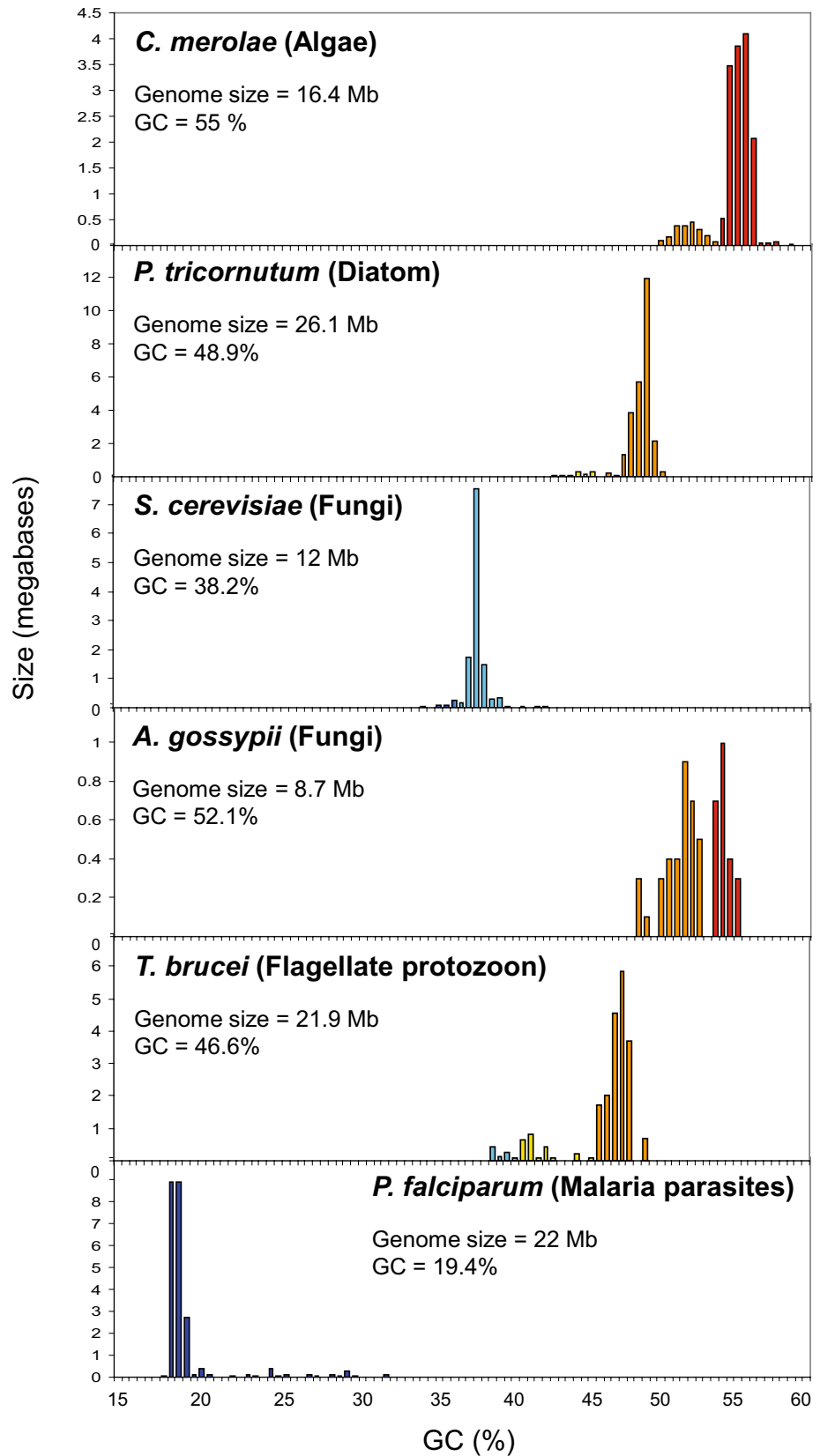
genome sizes of unicellular eukaryotes is even broader than that of metazoans (Costantini et al. 2009; Cammarano et al. 2009); (ii) the range of average GC levels of the genomes of unicellular eukaryotes is as broad as that of prokaryotes (Bernardi and Bernardi 1990; Katz et al. 2012); (iii) the chromatin structure of unicellular eukaryotes may be organized in a different way compared to that of multicellular eukaryotes (*Saccharomyces cerevisiae* lacks histone H1; *Trypanosomes* have H1 histone but the chromatin does not reach high levels of compaction during mitosis); (iv) the environmental conditions under which unicellular eukaryotes live are much more diverse than those of vertebrates and also of invertebrates; and (v) unicellular eukaryotes lack the very complex regulatory system involved in the developmental process of multicellular eukaryotes.

General Conclusions

All the data reviewed here demonstrated that genome compartmentalization can be considered a very general feature of all eukaryotes. In fact, preparative centrifugation in density also demonstrated the existence of isochores in plant genomes (Montero et al. 1990), later confirmed analyzing the complete genome sequences of Arabidopsis thaliana (Zhang and Zhang 2004).

Different levels of compartmentalization are probably linked with increasing regulatory complexity and/or other functional requirements to which organisms are bound. These findings suggest, in line with previous observations reported in Costantini and Bernardi (2008a), the following conclusions: (i) the high similarity of GC levels of isochore families may be due to their composition, linked to chromatin structure; (ii) the increasing variability in isochore patterns from warm- to cold-blooded vertebrates and to invertebrates may be correlated with the environmental factors, which were able to affect genome organization and functions; (iii) the distribution of genes seems to be dictated by the need of a certain genomic context, whose composition influences the transcriptional activity, and also the structure and function of the encoded proteins.

Two additional conclusions are of very great interest. The first one concerns the differences found between free-living and parasitic unicellular eukaryotes. The second one is the fact that the GC levels found in unicellular eukaryotes are very close (with two exceptions) to those of isochore families from multicellular eukaryotes. Indeed, the first point suggests some compositional adaptation of the genomes of parasitic unicellular organisms, and the second a correlation with chromatin structure.

References

Andreozzi L, Federico C, Motta S, Saccone S, Sazanova AL, Sazanov AA, Smirnov AF, Galkina SA, Lukina NA, Rodionov AV, Carels N, Bernardi G (2001) Compositional mapping of chicken chromosomes and identification of gene-richest regions. Chromosome Res 9:521–532

Bernardi G (1965) Chromatography of nucleic acids on hydroxyapatite. Nature 206:779–783

Bernardi G (2004) Structural and evolutionary genomics. Natural selection in genome evolution. Elsevier, Amsterdam

Bernardi G (2007) The neo-selectionist theory of genome evolution. Proc Natl Acad Sci USA 104(20):8385–8390

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

Bowler C, Allen AE, Badger JH et al (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244

Britten RJ, Kohne DE (1968) Repeated sequences in DNA. Science 161:529–540

Cammarano R, Costantini M, Bernardi G (2009) The isochore patterns of invertebrate genomes. BMC Genomics 10:538

Cohen NT, Dagan L, Graur D (2005) GC composition of the human genome: In search of isochores. Mol Biol Evol 22:1260–1272

Corneo G, Ginelli E, Soave C, Bernardi G (1968) Isolation and characterization of mouse and guinea pig satellite deoxyribonucleic acids. Biochemistry 7(12):4373–4379

Costantini M, Bernardi G (2008a) Replication timing, chromosomal bands and isochores. Proc Natl Acad Sci USA 105(9):3433–3437

Costantini M, Bernardi G (2008b) The short-sequence design of isochores from the human genome. Proc Natl Acad Sci USA 105(37):13971–13976

Costantini M, Clay O, Auletta F, Bernardi G (2006) An isochore map of human chromosomes. Genome Res 16(4):536–541

Costantini M, Di Filippo M, Auletta F, Bernardi G (2007a) Isochore pattern and gene distribution in the chicken genome. Gene 400:9–15

Costantini M, Auletta F, Bernardi G (2007b) Isochore patterns and gene distributions in fish genomes. Genomics 90(3):364–371

Costantini M, Clay O, Federico C, Saccone S, Auletta F, Bernardi G (2007c) Human chromosomal bands: nested structure, high-definition map and molecular basis. Chromosoma 116(1):29–40

Costantini M, Cammarano R, Bernardi G (2009) The evolution of isochore patterns in vertebrate genomes. BMC Genomics 10:146

Costantini M, Alvarez-Valin F, Costantini S, Cammarano R, Bernardi G (2013) Compositional patterns in the genomes of unicellular eukaryotes. BMC Genomics 14:755

Costantini M, Greif G, Alvarez-Valin F, Bernardi G (2016) The Anolis lizard genome: a genome without isochores? Genome Biol Evol 8:1048–1055

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

de Luca di Roseto G, Bucciarelli G, Bernardi G (2002) An analysis of the genome of *Ciona intestinalis*. Gene 295:311–316

Dekker J (2007) GC- and AT-rich chromatin domains differ in conformation and histone modification status and are differentially modulated by Rpd3p. Genome Biol 8:R11615

Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439:965–968

Derelle E, Ferraz C, Rombauts S et al (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc Natl Acad Sci USA 103:11647–11652

Di Filippo M, Bernardi G (2008) Mapping Dnase I-hypersensitive sites on human isochores. Gene 419:62–65

Dickerson RE (1992) DNA structure from A to Z. Methods Enzymol 211:67–111

Dujon B (1996) The yeast genome project: what did we learn? Trends Genet 12:263–270

Duret L, Eyre-Walker A, Galtier N (2006) A new perspective on isochore evolution. Gene 385:71–74

Eyre-Walker A, Hurst LD (2001) The evolution of isochores. Nat Rev Genet 2:549–555

Filipski J, Thiery JP, Bernardi G (1973) An analysis of the bovine genome by Cs2SO4-Ag + density gradient centrifugation. J Mol Biol 80:177–197

Häring D, Kypr J (2001) No isochores in the human chromosomes 21 and 22? Biochem Biophys Res Commun 280:567–573

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931–945

Isacchi A, Bernardi G, Bernardi G (1993) Compositional compartmentalization of the nuclear genomes of *Trypanosoma brucei* and *Trypanosoma equiperdum*. FEBS Lett 335:181–183

Jabbari K, Bernardi G (2000) The distribution of genes in the *Drosophila* genome. Gene 247:287–292

Jabbari K, Bernardi G (2004) Comparative genomics of *Anopheles gambiae* and *Drosophila melanogaster*. Gene 333:183–186

Jabbari K, Bernardi G (2017) An isochore framework underlies chromatin architecture. PLoS One 12(1):e0168023

Kadi F, Mouchiroud D, Sabeur G, Bernardi G (1993) The compositional patterns of the avian genomes and their evolutionary implications. J Mol Evol 37:544–551

Karlin S, Blaisdell BE, Sapolsky RJ, Cardon L, Burge C (1993) Assessment of DNA in homogeneities in yeast chromosome III. Nucleic Acid Res 21:703–711

Katz LA, Grant JR, Parfey LW, Burleigh JG (2012) Turning the crown upside down: gene tree parsimony roots the eukaryotic tree of life. Syst Biol 61:653–660.

Lamolle G, Protasio A, Iriarte A, Jara E, Simón D, Musto H (2016) An sochore-like structure in the genome of the flatworm *Schistosoma mansoni*. Genome Biol Evol 8:2312–2318

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh WES et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

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

McCutchan TF, Dame JB, Miller LH, Barnwell J (1984) Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. Science 225:808–811

Meselson M, Stahl FW, Vinograd J (1957) Equilibrium sedimentation of macromolecules in density gradients. Proc Natl Acad Sci USA 43:581–588.

Montero LM, Salinas J, Matassi G, Bernardi G (1990) Gene distribution and isochore organization in the nuclear genome of plants. Nucleic Acids Res 18(7):1859–1867

Musto H, Rodríguez-Maseda H, Bernardi G (1994) The nuclear genomes of African and American trypanosomes are strikingly different. Gene 141:63–69

Musto H, Naya H, Zavala A, Romero H, Alvarez-Valín F, Bernardi G (2006) Genomic GC level, optimal growth temperature, and genome size in prokaryotes. Biochem Biophys Res Commun 18:347(1):1–3.

Naya H, Romero H, Zavala A, Alvarez B, Musto H (2002) Aerobiosis increases the genomic guanine plus cytosine content (GC%) in prokaryotes. J Mol Evol 55(3):260–264

Nekrutenko A, Li WH (2001) Assessment of compositional heterogeneity within and between eukaryotic genomes. Genome Res 10:1986–1995

Niimura Y, Gojobori T (2002) In silico chromosome staining: reconstruction of Giemsa bands from the whole human genome sequence. Proc Natl Acad Sci USA 99(2):797–802

Palenik B, Grimwood J, Aerts A et al (2007) The tin eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc Natl Acad Sci USA 104:7705–7710

Pollak Y, Katzen A, Spira D, Golenser J (1982) The genome of Plasmodium falciparum I: DNA composition. Nucleic Acid Res 10:539–546

Rodríguez-Maseda H, Musto H (1994) The compositional compartments of the nuclear genomes of *Trypanosoma brucei* and *T. cruzi*. Gene 151:221–224

Romero H, Pereira E, Naya H, Musto H (2009) Oxygen and guanine-cytosine profiles in marine environments. J Mol Evol 69(2):203–206

Saccone S, Federico C, Andreozzi L, D'Antoni S, Bernardi G (2002) Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. Gene 300:169–178

Segal E, Fondufe-Mittendorf Y, Chen L, Thåström AC, Field Y, Moore YK, Wang JZ, Widom J (2006) A genomic code for nucleosome positioning. Nature 442:772–778

Sharp PM, Lloyd AT (1993) Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure. Nucleic Acid Res 21:179–183

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

van Rheede T, Bastiaans T, Boone DN, Hedges SB, de Jong WW, Madsen O (2006) The platypus is in its place: nuclear genes and indels confirm the sister group relation of monotremes and Therians. Mol Biol Evol 23:587–597

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

Zhang R, Zhang CT (2004) Isochore structures in the genome of the plant *Arabidopsis thaliana*. J Mol Evol 59(2):227–238