Available online at www.sciencedirect.com
ScienceDirect
Gene 400 (2007) 9–15
www.elsevier.com/locate/gene

Isochore pattern and gene distribution in the chicken genome

Maria Costantini [1], Miriam Di Filippo [1], Fabio Auletta, Giorgio Bernardi *

Laboratory of Molecular Evolution, Stazione Zoologica Anton Dohrn, 80121 Naples, Italy

Received 4 April 2007; received in revised form 18 May 2007; accepted 24 May 2007
Available online 13 June 2007
Received by I. King Jordan

Abstract

We report here investigations on the isochore pattern and the distribution of genes in the chromosomes of chicken. In spite of large differences in genome size and karyotype, the compositional properties and the gene distribution of the chicken genome are very similar to those recently published for the human genome, which is a good representative of most mammalian genomes. In fact, this similarity, which extends to the relative amounts and, also, to a large extent at least, to the average base composition of isochore families, is most interesting in view of the very large distance of mammals and birds for a common ancestor, which goes back to 310–340 million years ago. This raises important questions about genome evolution in vertebrates.

Keywords: Birds; Chicken; Genes; Genomes; Isochores

1. Introduction

The first compositional analysis of avian genomes was obtained by ultracentrifugation in CsCl density gradients (Thiery et al., 1976). Both avian DNAs investigated, from chicken and sea-gull, showed compositional patterns that were remarkably similar to those of mammalian DNAs. A much more detailed analysis (Cortadas et al., 1979) of chicken DNA using preparative Cs_2SO^4/Ag^+ and Cs_2SO_4/BAMD (BAMD is 3,6-bis(acetato–mercury–methyl)dioxane) revealed that 88% of the chicken genome was made up of four "major DNA components", DNA fractions that were similar in relative amounts and base composition to those of mammalian genomes. The demonstration of the compositional homogeneity over at least 200 kb around the ovalbumin gene (Cortadas et al., 1979) indicated that the DNA molecules (about 100 kb in size) forming the major components derived from much longer, fairly homogeneous chromosomal regions previously identified in mammalian genomes, the isochores (Macaya et al., 1976). In other words, the chicken genome, like the mammalian genome, was a mosaic of isochores that belonged to four major families, which were called L1, L2, H1, and H2 because of their similarity with the corresponding isochore families of mammals. The remaining 12% of the genome was formed by seven minor and/or satellite components, two of which were later identified as derived from two additional isochore families, H3 (Bernardi et al., 1985) and H4 (Bernardi, 1989). While the H3 family was also present in the mammalian genome, the H4 family was only present in chicken (and other avian genomes; see below).

Further work showed (i) that the relative amounts of interspersed repeats were different in different chicken isochores and different from those of the corresponding mammalian isochores (Olofsson and Bernardi, 1983a,b); and (ii) that the compositional patterns of the genomes of birds belonging to eight different orders (both paleognathous and neognathous) were very similar, as were orthologous coding sequences (Kadi et al., 1993). Moreover, a remarkable compositional similarity was found for orthologous genes from human and chicken, especially in the case of GC-poor genes (see Bernardi et al., 1997). Another aspect of the similarity of avian and mammalian genomes concerned their lower CpG and mC levels compared to those of fishes/amphibians (Jabbari et al., 1997; Cacciò et al., 1994; Varriale and Bernardi, 2006).

An analysis of GC_3 levels and codon frequencies of genes from human, chicken and *Xenopus* (Cruvellier et al., 2000) showed that GC-poor genes were characterized by only minor differences in orthologous sets from *Xenopus*, human and

* Corresponding author. Tel.: +39 081 5833402; fax: +39 081 2455807.
E-mail addresses: mcosta@szn.it (M. Costantini), miriam@szn.it (M.D. Filippo), auletta@szn.it (F. Auletta), bernardi@szn.it (G. Bernardi).
[1] These authors contributed equally to this work.

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.05.025

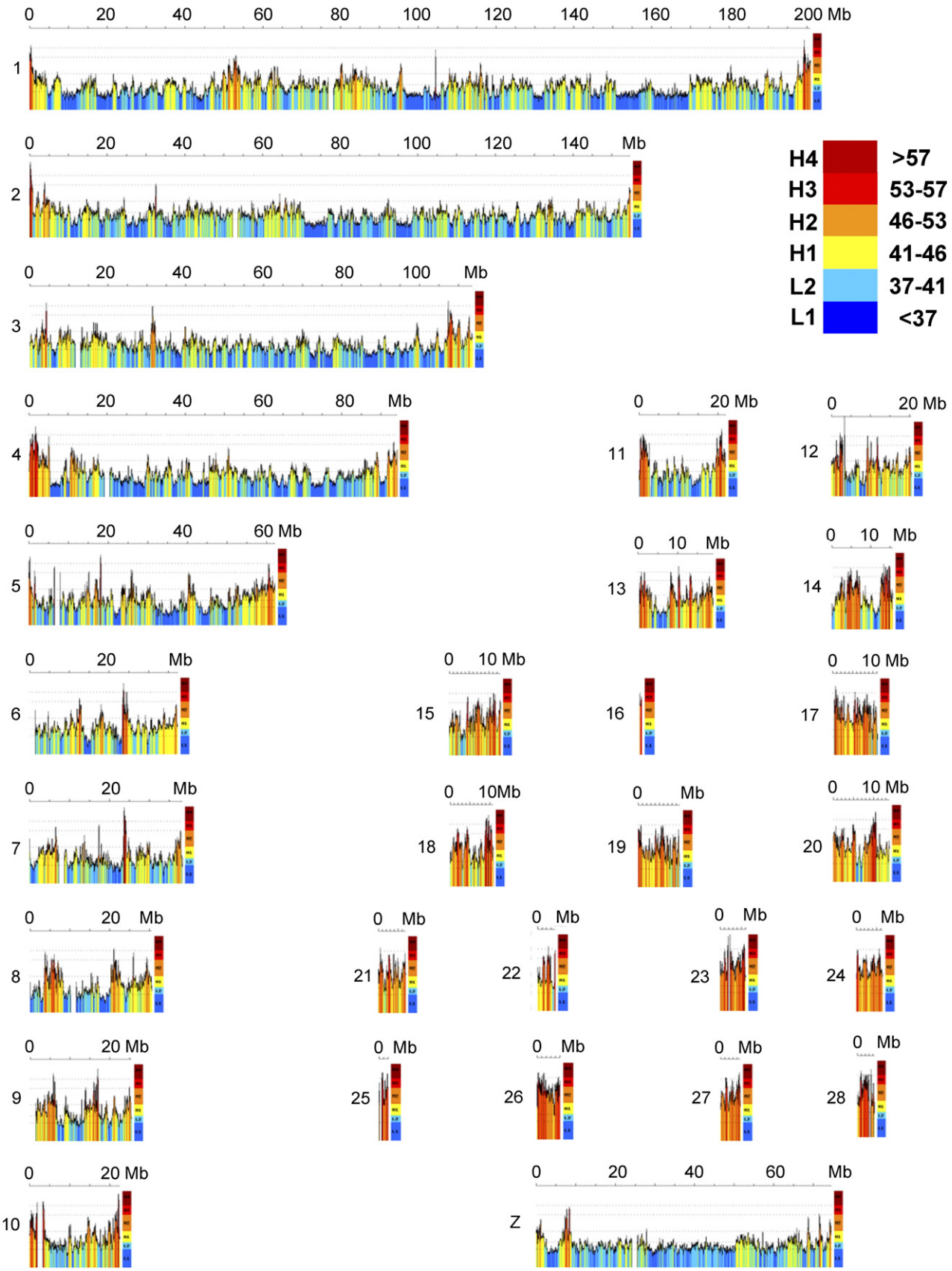

Fig. 1. Compositional overview of chicken chromosomes. The color-coded map shows 100 kb moving window plots using the program draw_chromosome_gc.pl (Pačes et al., 2004) (http://genomat.img.cas.cz). The color code spans the spectrum of GC levels in six steps, indicated by broken horizontal lines, from ultramarine blue (GC-poorest L1 isochores) to dark red (GC-richest H4 isochores). Chromosomes 1 to 10 are the macrochromosomes, chromosomes 11 to 32 the microchromosomes. Chromosomes not shown comprise chromosomes W and 32, because they are very short, and chromosomes 29, 30 and 31, because they were not yet sequenced.

chicken, a remarkable result in view of the very many nucleotide substitutions that occurred over the long evolutionary times separating these species from their common ancestor. In contrast, GC-rich genes showed large differences between *Xenopus* and warm-blooded vertebrates, but only relatively small differences between chicken and human, the independent changes that occurred in avian and mammalian genes being similar in average composition.

Along a different line, the experimental assessment of the GC level of the chromosomal bands, led to the identification in human chromosomes of the GC-richest and of the GC-poorest bands, which were predominantly localized in telomeres and in internal regions, respectively (see Saccone et al. 1992, 1993). The same approach (Andreozzi et al., 2001) showed that the GC-richest isochores of chicken are localized not only on a large number of microchromosomes in agreement with previous findings on the concentration of CpG islands and genes on microchromosomes (McQueen et al., 1996, 1998), but also on almost all telomeric bands of macrochromosomes. On the other hand, the GC-poorest isochores are generally localized in the internal regions of macrochromosomes and are almost absent in microchromosomes. Interestingly, in the *Accipitridae* (diurnal raptors), an avian family that shows no very large chromosomes and only a very small number of microchromosomes, the gene-rich regions are prevalently located in the few microchromosomes and in the telomeric regions of the middle-sized chromosomes (Federico et al., 2005). The distinct localization of the GC-richest and the GC-poorest bands initially observed in human chromosomes appears, therefore, to be a general feature of chromosomes from warm-blooded vertebrates. Again as in the case of mammals, the gene-richest/GC-richest chromosomal regions of chicken (and of *Accipitridae*) are predominantly distributed in internal locations of interphase nuclei, whereas the gene-poorest/GC-poorest DNA regions are close to the nuclear envelope (Saccone et al., 2002).

In this paper we present the first isochore map and a gene density analysis of the chicken genome. A similar analysis was recently published for fish genomes (Costantini et al., 2007, in press). Together with the previously reported isochore map of the human genome (Costantini et al., 2006) the isochore map of chicken provides a new approach to the study of vertebrate evolution.

2. Methods

2.1. Isochore mapping

The methodology used for isochore mapping was described by Costantini et al. (2006). The entire chromosomal sequences of the finished genome assembly for *G. gallus* (UCSC Release galGal3 http://genome.ucsc.edu), were partitioned into non-overlapping 100 kb windows, and their GC levels calculated using the program draw_chromosome_gc.pl (Pavliček et al., 2002; Pačes et al., 2004; http://genomat.img.cas.cz).

As far as the name of each isochore band is concerned we decided to use a convention in which the first number in the name represents the chromosome number the following two letters are the initials of the scientific name of the chicken, and the last number identifies the band (see Supplementary Tables T1).

2.2. Gene distribution

The chicken genes were retrieved from Ensembl (http://www.ensembl.org/Gallus_gallus/index.html ReleaseWASHUC2, May 2006). This set of genes (12,663 known genes) was used to download gene coordinates in order to analyze the gene density in the chicken isochore families.

The partial, putative, synthetic construct, predicted, not experimental, hypothetical protein, r-RNA, t-RNA, ribosomal and mitochondrial genes were eliminated and then the cleanup software was applied (Grillo et al., 1996), a fast computer program for cleaning nucleotide sequence databases of redundancies. For the remaining genes a computer program implemented by us (available upon request) was used in order to identify the coding sequences with start and stop codon and to eliminate internal stop codons so as to calculate the GC, GC_1, GC_2 and GC_3 values. Applying our protocol, we obtained a very small number of genes (1522 genes); so we decided to retrieve chicken genes from GenBank (Release 157, 15 December 2006). Applying our protocol to clean this gene set we obtained 4435 coding sequences for the compositional analysis.

3. Results

3.1. Isochore map

Fig. 1 shows the GC profiles of the chicken chromosomes using a fixed window of 100 kb. This window was chosen because plots of average standard deviations of GC against window size indicate the existence of a plateau that begins around 100 kb and extends to over 500 kb (Fig. 2; see also Costantini et al., 2006, for a similar result in the case of the human genome). Isochore borders were identified on the basis

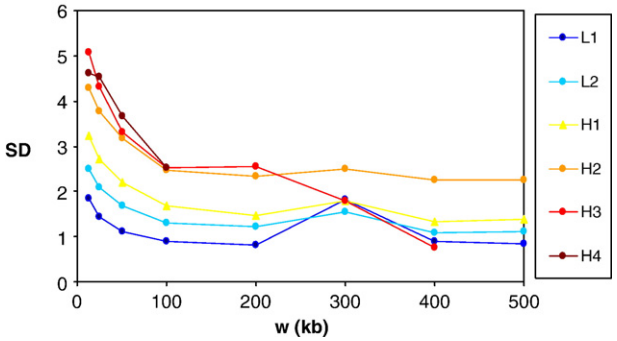

Fig. 2. Plots of average standard deviations (SD) of GC within isochore families versus window size (w). Plots are shown for all isochores comprising at least four windows, after partitioning isochores into families according to GC level. The plots summarize the compositional variations within isochore families. The variations are consistently increasing for window sizes below 100 kb, and settle down to a plateau for larger window sizes. This fact justifies graining at 100 kb. Marginally higher standard deviations around 300 kb windows may be due to a larger number of isochore borders seen through this window (see also Costantini et al., 2006). In the case of the H4 family the plot shows points only up to 400 kb, because isochores are much shorter than in the case of the other families.

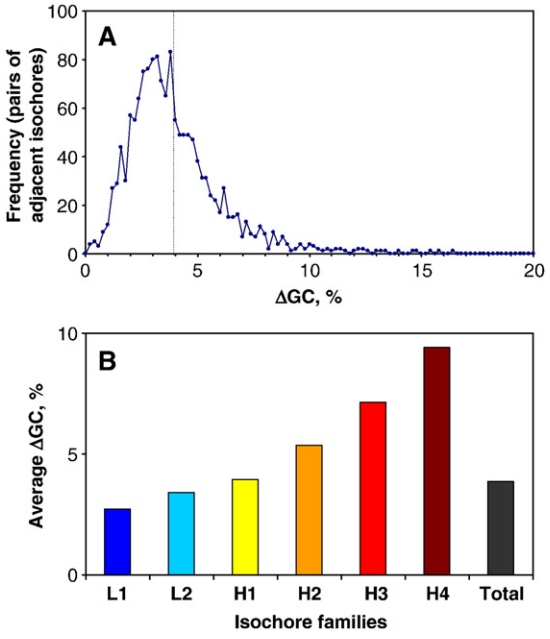

Fig. 3. ΔGC distribution of adjacent isochores. The frequency plot (A) shows the jumps in GC between adjacent isochores, in intervals of 0.2% GC. The mean difference is 3.9% GC (dashed vertical line). The bar plot (B) shows the average ΔGC concerning isochores from each of the six families.

of marked compositional differences that ranged from 2.7% to 9.4% GC for isochores belonging to different families, the average value being 3.9% GC (Fig. 3), a value identical to that found in the human genome (Costantini et al., 2006).

Chromosomes 1 to 10, the macrochromosomes, are mosaics of GC-poor and GC-rich isochores, the latter prevailing at telomeres. The other chromosomes, the microchromosomes, are characterized by the prevalence of GC-rich regions. Chromosomes W and 32 are very short (260 bp and 1028 bp, respectively) and are not represented in Fig. 1. Chromosomes 29, 30 and 31 are also not represented because they have not been sequenced yet.

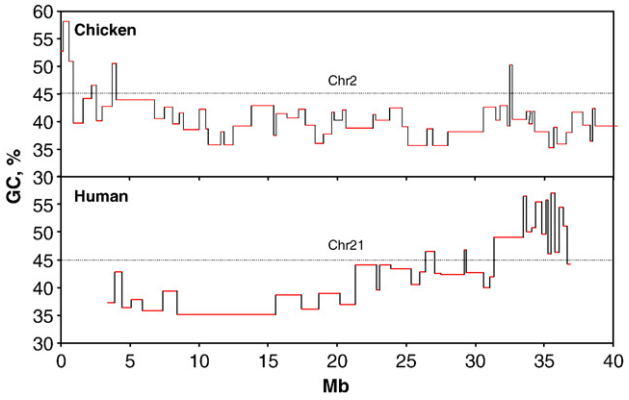

Fig. 4. An example of isochores in chicken chromosomes. The isochores (horizontal red lines) identified on a stretch from chicken chromosome 2 totaling about 40 Mb is shown. Human chromosome 21 (which was approximately the same size) is also shown for sake of comparison. A broken line at 45% GC is shown in both plots as a reference.

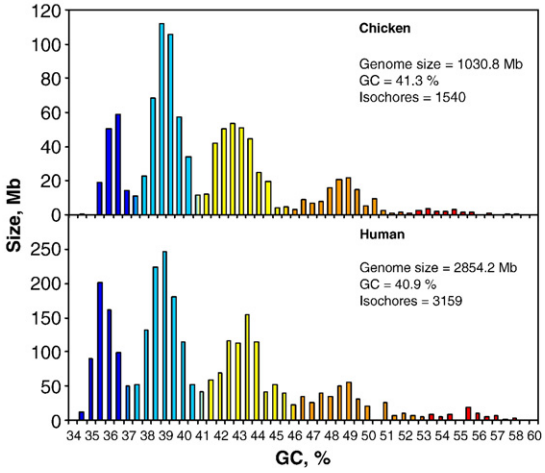

Fig. 5. Distribution of isochores according to GC levels. The histogram shows the distribution (by weight) of isochores as pooled in bins of 0.5% GC. Colors represent the five isochore families. The H4 family does not appear in the figure because of its very small relative amount. Values at minima (histogram bars with mixed colors) were split between the two neighboring families. A comparable plot for human isochores is also displayed.

The data presented in Fig. 1 and the criteria given by Costantini et al. (2006) allowed us to construct an isochore map of the chicken genome (Fig. 4). We estimate the total number of isochores (neglecting the still existing gaps) as 1540, about half the isochore number (~3200) of the human genome. The sum of these isochores is 1030.8 Mb, a value lower than that estimated by other approaches (Smith and Burt, 1998), 1.2 Gb the difference being, however, accounted for by the microchromosomes not yet sequenced by centromeric DNA.

3.2. Isochore families

When isochores are pooled in bins of 0.5% GC (Fig. 5), isochore families stand out. This is evident for isochore families L1, L2 and H1, but it also visible for the H2 and H3 families, which are present in small amounts in the genome. Supplementary Table T1 provides the coordinates on the UCSC map

Table 1
Percentage and average GC levels of isochore families from chicken and human (A, B). Average size of isochores belonging to these families C)

	L1	L2	H1	H2	H3	H4
A) Percentage						
Chicken	17.4	38.5	30.3	12.0	(1.6)[a]	(0.2)[a]
Human	19.0	37.0	31.0	11.0	3.0	
B) Average GC						
Chicken	36.6	39.3	43.4	48.8	54.7	57.9
Human	36.0	38.9	43.1	48.7	54.5	
C) Average size (Mb)						
Chicken	0.70	0.81	0.60	0.55	(0.34)[b]	(0.26)[b]
Human	0.9[c]	0.9	0.8	0.7	0.7	

[a] These values are underestimated (see text).
[b] These values concern only a small part of H3 and H4 isochores (see text).
[c] This value concerns the isochores lower than 3 Mb in size.

(galGal3). Table 1 presents the relative amount and the GC level of each isochore family (see also Supplementary Fig. S1), as well as the average size of isochores and compares these results with those from the human genome (Costantini et al., 2006). These findings show that the relative amounts and the average GC level of isochores belonging to the same family are remarkably similar between chicken and human. It should be noticed, however, that the values for chicken H3 and H4 isochores were severely underestimated because some microchromosomes, which are especially rich in these isochores families, have not been sequenced so far. A similarity was also found for the average size of isochores belonging to corresponding families, with systematically slightly lower values for chicken isochores. Moreover, the "long isochores" of the human L1 family, which represent the GC-poorest tail of this family, were practically absent in the small genome of chicken (Fig. 6), as it also was the case for the pufferfish genome (Costantini et al., 2007, in press; see Discussion).

In chromosomes, H3 isochores were always flanked by GC-poorer isochores, and L1 isochores were always flanked by GC-richer isochores, as expected. These were also the predominant situations found in the cases of isochores L2 and of H2 or H1 isochores, respectively, where the most abundant flanking isochores always belonged to the next families so leading to the formation of blocks of isochores from close families (*e.g.* L1/L2, etc.; see Supplementary Fig. S2). Moreover, in several cases these families also exhibited "transition isochores", where one flanking isochore was higher, the other lower (Supplementary Fig. S3). These situations were already met when studying human isochores (Costantini et al., 2006, 2007).

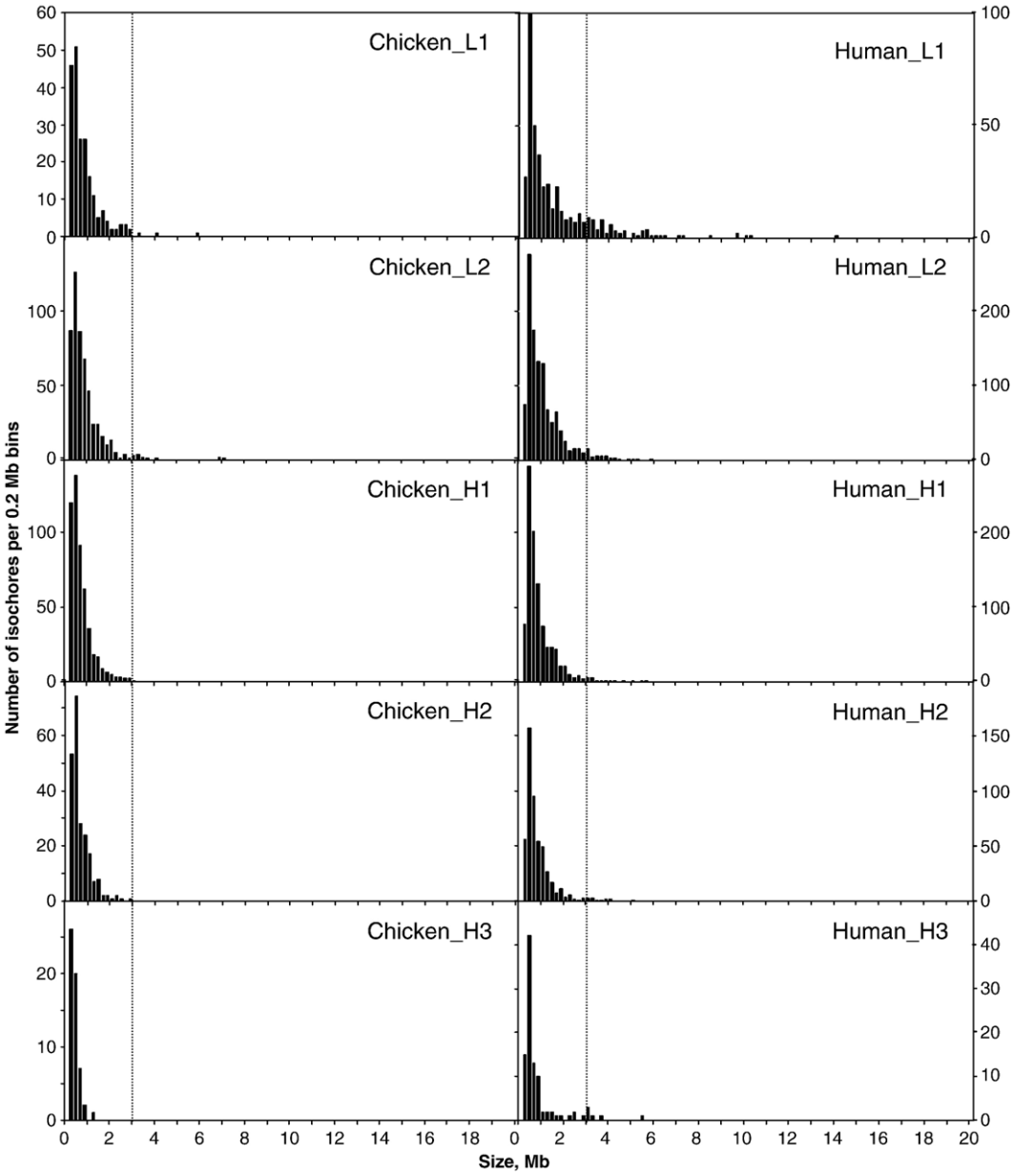

Fig. 6. Size distributions of the chicken isochores are compared with human isochores from the corresponding families. A vertical line at 3 Mb is drawn as a reference. In all isochore families, but especially in family L1, long isochores are present in the human genome, but they are absent or very scarce in the chicken genome.

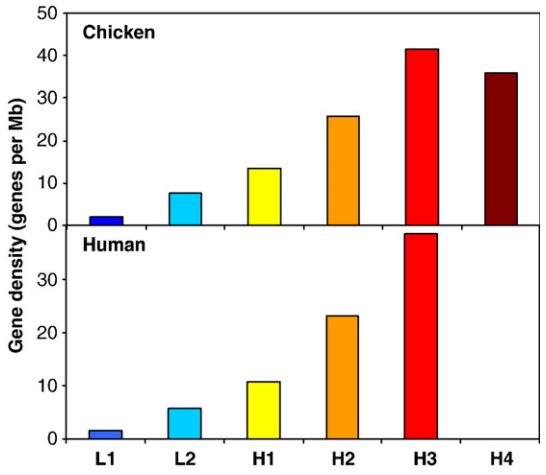

Fig. 7. Gene density. The gene density in the isochore families of the human and chicken genome are compared.

3.3. Gene density

As in the human genome, gene density of the chicken genome increases with the increasing GC of isochore families from L1 to H3 (see Fig. 7). The lower value of isochore family H4 is likely to be due to an assessment problem linked to the very small relative amount of this family sequenced so far. Indeed, at present some microchromosomes, which are especially GC-rich, have not been sequenced not yet (see page 9). As a consequence the genes localized in these very GC-rich regions are not available. In fact if one considers the Supplementary Fig. S4 A–B, there is a very striking difference especially in the GC and GC_3 profiles of the chicken genes in comparison with the human genes. The GC-rich genes are missing, due to the microchromosomes sequences not yet available.

We analyzed the GC, GC_1, GC_2 and GC_3 distribution of chicken genes and compared then with those of human genes (see Supplementary Fig. S4 A, B). The striking differences in the profiles, especially the GC and GC_3 profiles, are obviously due to the microchromosomes sequences not yet available (see above).

4. Discussion

The compositional similarity of avian and mammalian genomes, first observed thirty years ago (Thiery et al., 1976) was not only confirmed but also considerably extended by subsequent work (Kadi et al., 1993). The present investigations, based as they were on sequences, provide, as expected, more precise results compared to those previously obtained from ultracentrifugation. As a result, the following properties were found to be very similar: (i) the relative amounts of isochores belonging to different families; (ii) the average GC level of the isochore families; and (iii) to some extent, the average size of isochores belonging to different families. The differences concerned the existence of a minor isochore family, H4, which is absent in mammals, and the fact that the average size of isochores belonging to different families was systematically slightly lower compared to human isochores.

It is of interest that the size reduction especially affected the GC-poorest isochores, this being an indication that intergenic sequences may vary more in this isochore family compared to the other families. The drastic reduction in the size of the genome as well as of intergenic and intronic undergone by birds could explain the smaller size of isochores. Another difference concerned the practical absence of "long isochores" in the GC-poorest isochore family of chicken, in contrast to the situation found in the human genome. A similar result was also found in the genome of pufferfish. In both cases, the reduction in genome size seems to be responsible for the absence of "long isochores" which should be visualized as expanded genome regions.

The similarities between the human and the chicken genome are particularly striking if one considers 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 can be explained, however, if the parallel, yet independent, compositional evolution of mammals and birds are due, as proposed by Bernardi and Bernardi (1986), to the necessity of stabilizing the genome (more precisely the gene-rich part of it, which is characterized by an open chromatin structure; Saccone et al., 2002), a subject discussed in detail elsewhere. If such is the case, the existence of H4 isochores might be related (as suggested by Kadi et al., 1993) to the higher body temperature (about 41 °C) of birds compared to that of mammals (about 37 °C). 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. An explanation is provided by the neo-selectionist theory of genome evolution (see Bernardi, 2007 for a detailed discussion on the evolution of the vertebrate genome).

Acknowledgements

We thank Giacomo Bernardi for comments and Oliver Clay for helpful discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2007.05.025.

References

Andreozzi, L., et al., 2001. Compositional mapping of chicken chromosomes and identification of gene-richest regions. Chromosome Res. 9, 521–532.

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

Bernardi, G., 2007. The neo-selectionist theory of genome evolution. PNAS 104 (20), 8385–8390.

Bernardi, G., Bernardi, G., 1986. Compositional constraints and genome evolution. J. Mol. Evol. 24, 1–11.
Bernardi, G., et al., 1985. The mosaic genome of warm-blooded vertebrates. Science 228, 953–958.
Bernardi, G., Hughes, S., Mouchiroud, D., 1997. The major compositional transitions in the vertebrate genome: a review. J. Mol. Evol. 44, S41–S51.
Cacciò, S., Perani, P., Saccone, S., Kadi, F., 1994. Single-copy sequence homology among the GC-richest isochores of the genomes from warm-blooded vertebrates. J. Mol. Evol. 39, 331–339.
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.
Costantini, M., Clay, O., Auletta, F., Bernardi, G., 2006. An isochore map of human chromosomes. Genome Res. 16, 536–541.
Costantini, M., Clay, O., Federico, C., Saccone, S., Auletta, F., Bernardi, G., 2007. Human chromosomal bands: nested structure, high-definition map and molecular basis. Chromosoma 116 (1), 29–40.
Costantini, M., Auletta, F., Bernardi, G., in press. Isochore patterns and gene distributions in fish genomes. Genomics. doi:10.1016/j.ygeno.2007.05.006.
Cruvellier, S., D'Onofrio, G., Bernardi, G., 2000. The compositional transition between the genomes of cold- and warm-blooded vertebrates: codon frequencies in orthologous genes. Gene 261, 71–83.
Federico, C., Cantarella, C.D., Scavo, C., Saccone, S., Bed'Hom, B., Bernardi, G., 2005. Avian genomes: different karyotypes but a similar distribution of the GC-richest chromosome regions at interphase. Chromosome Res. 13 (8), 785–793.
Grillo, G., Attimonelli, M., Liuni, S., Pesole, G., 1996. CLEANUP: a fast computer program for removing redundancies from nucleotide sequence databases. Comput. Appl. Biosci. 12, 1–8.
Jabbari, K., Cacciò, S., Païs de Barros, J.-P., Desgrès, J., Bernardi, G., 1997. Evolutionary changes in CpG and methylation levels in vertebrate genomes. Gene 205, 109–118.
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.
Macaya, G., Thiery, J.P., Bernardi, G., 1976. An approach to the organization of eukaryotic genomes at a macromolecular level. J. Mol. Biol. 108, 237–254.
McQeen, H.A., Fantes, J., Cross, S.H., Clark, V.H., Archibald, A.L., Bird, A.P., 1996. CpG islands of chicken are concentrated on microchromosomes. Nat. Genet. 12, 321–324.
McQueen, H.A., Siriaco, G., Bird, A.P., 1998. Chicken microchromosomes are hyperacetylated, early replicating and gene rich. Genome Res. 8, 621–630.
Olofsson, B., Bernardi, G., 1983a. Organization of nucleotide sequences in the chicken genome. Eur. J. Biochem. 130, 241–245.
Olofsson, B., Bernardi, G., 1983b. The distribution of CR1, an Alu-like family of interspersed repeats in the chicken genome. Biochim. Biophys. Acta 740, 339–341.
Pačes, J., Zika, R., Pavliček, A., Clay, O., Bernardi, G., 2004. Representing GC variation along eukaryotic chromosomes. Gene 333, 135–141.
Pavliček, A., Pačes, J., Clay, O., Bernardi, G., 2002. A compact view of isochores in the draft human genome sequence. FEBS Lett. 511, 165–169.
Saccone, S., De Sario, A., Della Valle, G., Bernardi, G., 1992. The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. Proc. Natl. Acad. Sci. U. S. A. 89, 4913–4917.
Saccone, S., De Sario, A., Wiegant, J., Raap, A.K., Della Valle, G., Bernardi, G., 1993. Correlations between isochores and chromosomal bands in the human genome. Proc. Natl. Acad. Sci. U. S. A. 90, 11929–11933.
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.
Smith, J., Burt, D.W., 1998. Parameters of the chicken genome (*Gallus gallus*). Anim. Genet. 29, 290–294.
Thiery, J.P., 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, D.N., Hedges, S.B., de Jong, W.W., 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.
Varriale, A., Bernardi, G., 2006. DNA methylation and body temperature in fishes. Gene 385, 111–121.