The *Anolis* Lizard Genome: An Amniote Genome without Isochores?

Maria Costantini[1,*], Gonzalo Greif[2], Fernando Alvarez-Valin[3] and Giorgio Bernardi[1,4,*]

[1]Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Naples, Italy

[2]Unidad de Biología Molecular, Instituto Pasteur de Montevideo, Montevideo, Uruguay

[3]Sección Biomatemática, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

[4]Science Department, Roma Tre University, Rome, Italy

*Corresponding author: E-mail: maria.costantini@szn.it; gbernardi@uniroma3.it.

Abstract

Two articles published 5 years ago concluded that the genome of the lizard *Anolis carolinensis* is an amniote genome without isochores. This claim was apparently contradicting previous results on the general presence of an isochore organization in all vertebrate genomes tested (including *Anolis*). In this investigation, we demonstrate that the *Anolis* genome is indeed heterogeneous in base composition, since its macrochromosomes comprise isochores mainly from the L2 and H1 families (a moderately GC-poor and a moderately GC-rich family, respectively), and since the majority of the sequenced microchromosomes consists of H1 isochores. These families are associated with different features of genome structure, including gene density and compositional correlations (e.g., GC3 vs flanking sequence GC and intron GC), as in the case of mammalian and avian genomes. Moreover, the assembled *Anolis* chromosomes have an enormous number of gaps, which could be due to sequencing problems in GC-rich regions of the genome. In conclusion, the *Anolis* genome is no exception to the general rule of an isochore organization in the genomes of vertebrates (and other eukaryotes).

Key words: isochores, reptiles, genome structure and evolution.

Introduction

The discovery of compartmentalization in mammalian genomes goes back to more than 40 years ago when, using Cs_2SO_4/Ag^+ ultracentrifugation (Corneo et al. 1968), it was shown that the bovine genome mainly consisted of a small number of families of "main band" (non-satellite, non-ribosomal) DNA molecules 10–20 kb in size (Filipski et al. 1973). This observation was then extended to other eukaryotic genomes (Thiery et al. 1976). The 10–20 kb DNA molecules just mentioned derived, in fact, from much larger DNA stretches, fairly homogeneous in base composition (Macaya et al. 1976), that were called "isochores" for (compositionally) equal landscapes (Cuny et al. 1981; see Bernardi 2004, for a review including later investigations).

The very basic features of isochores are that (1) they belong to a small number of families (five in the human genome: L1, L2, H1, H2, and H3, characterized by increasing GC levels); (2) they are correlated with all structural and functional properties of the genome (such as gene density, replication timing, etc.) that could be tested; (3) they are correlated with the architecture of chromosomes from interphase to metaphase (Bernardi 2015).

Some misunderstandings about isochores (Häring and Kypr 2001; Lander et al. 2001; Belle e al. 2002; Ream et al. 2003; Cohen et al. 2005; Elhaik et al. 2009) were promptly corrected (Bernardi 2001; Clay and Bernardi 2001, 2002, 2005, 2011; Clay et al. 2003; Jabbari et al. 2003). This was not done so far for two articles (Alföldi et al. 2011; Fujita et al. 2011) that claimed that the *Anolis* lizard genome was an amniote genome without isochores. The reason why we did not react quickly to this new misunderstanding was the lack of credibility of this conclusion (which, incidentally, was based on a genome sequence with an enormous number of gaps). Indeed (1) we had shown that an isochores organization was general for vertebrates (Costantini et al. 2009) and that the isochore families present in the genomes studied were very close in terms of base composition (maxima and minima), the only possible difference being the relative amounts; for instance, while all primate genomes show five isochore families (L1, L2, H1, H2, and H3), fish genomes show

fewer families (the Zebrafish genome, for instance, comprises only L1 and L2 families); moreover, we also knew (Thiery et al. 1976) that the genome of *Iguana iguana*, another squamate reptile, only comprised two families of DNA segments with buoyant densities of 1.7015 g/m^3 (85% of the genome) and 1.706 g/M^2 (15% of the genome); these values are those of H1 and H2 isochore families; (2) in the specific case of the *Anolis* genome, our previous results showed the presence of a major family, L2, of a less abundant family, H1, and of very small amounts of L1 and H2 families; these families were characterized by the dinucleotide frequencies typical of those families as present in all other vertebrates tested (Costantini et al. 2009); (3) Alföldi et al. (2011) in a paper including Fujita, Edwards and Ponting among the co-authors had presented a beautiful DAPI banding of *Anolis* chromosomes, a strong evidence of a compositional compartmentalization of the genome (see Medrano et al. 1988; Bernardi 2004, 2015); (4) Fujita et al. (2011) showed at least a weak isochore pattern in which isochores represent 20% of the *Anolis* genome as opposed to 40% of the human genome; in addition, they presented evidence of compositional heterogeneity; (5) the segmentation approach of Fearnhead and Vasileiou (2009) used by Fujita et al. (2011) rests on comparisons of local compositional heterogeneities with chromosomal heterogeneities, following Cohen et al. (2005), a procedure shown to be misleading by Clay and Bernardi (2005).

Because the conclusion of Alföldi et al. (2011) and Fujita et al. (2011) was in apparent contradiction with our previous results (and even with data presented by the authors themselves), we decided to analyze further the *Anolis* genome. Indeed, formally disproving that conclusion could reinforce the notion that an isochore structure is absolutely general in vertebrate genomes (as well as in other multicellular eukaryotes; see Bernardi 2004, for a review). Moreover, it was an interesting challenge for us to more precisely confirm an isochore organization in a genome full of gaps.

Materials and Methods

Genome and Isochore Mapping

The sequences of the chromosomes from *Anolis carolinensis* genome were downloaded from Ensembl Genome Browser (Genome assembly: AnoCar2.0 (GCA_000090745.1)). The chromosomal sequences of the available genome assembly were partitioned into non-overlapping 100-kb windows, and their GC levels were calculated using the program draw_chromosome_gc.pl (Pavlíček et al. 2002; Pačes et al. 2004; http://genomat.img.cas.cz), an approach which provides a visual overview of GC-rich and GC-poor regions along chromosomes.

To study in detail other typical features that characterize the compositional patterns of isochores, we selected contigs that were at least 20 kb in length (21,596 contigs). For the protein-coding genes located in these contigs, the GC level was calculated for exons, introns, as well as for the flanking regions. The correlations among these measures were estimated. We also calculated the number of coding sequences, exons, and protein coding densities for each one of the isochore families.

Results and Discussion

A compositional map of *Anolis* chromosomes using a non-overlapping windows of 100 kb is presented in figure 1. As far as macrochromosomes are concerned, the map shows a large number of isochores belonging to the L2 family that are interspersed with H1 isochores, the latter being not only less abundant in the map but also mostly present in very short stretches with, however, a number of larger stretches preferentially located at telomeres and centromeres. It should be stressed, however, that the compositional map of figure 1 presents an enormous number of gaps. Indeed, the number of gaps is equal to that of the sequenced regions, over 40,000, of which only ~21,000 larger than 20 kb. It should be stressed that gaps are usually associated with sequencing problems in GC-rich regions (Lander et al. 2001) and that they may represent 20% of the *Anolis* genome. Indeed, if we use the most recent estimate of the genome size of *Anolis*, 2.20pg, (Peterson et al. 1994; incidentally, in agreement with the oldest estimate, 2.30 pg, by Atkin et al. 1965), this indicates that the sequenced part of the genome, 1.78 Gb (Alföldi et al. 2011), namely 1.74 pg, only corresponds to 80% of the genome, the remaining 20% being present in gaps.

If we now consider microchromosomes, the seven out of 12 for which sequences are available show a very different compositional organization, in that (1) one only comprises L2 isochores; (2) one is a mixture of L2 and H1 isochores with a predominance of the former; and (3) five practically only consist of H1 isochores. The presence of GC-richer (and gene-richer; see below) microchromosomes was already found in chicken (Costantini et al. 2009) and other birds but not in *Accipitridae*, that practically lack microchromosomes, the corresponding sequences being present at telomeres of macrochromosomes (Federico et al. 2005).

Figure 2 displays a histogram of the results just described in terms of DNA amounts as distributed in bins of 0.5% GC. This histogram is only slightly narrower in GC level, from 39.5% to 43% GC, compared with that already published using the *Anolis* scaffolds (from 39% to 46% GC; Costantini et al. 2009). The *Anolis* genome presents a predominant amount of DNA characterized by GC levels of 39–41% (corresponding to isochores from the L2 family) as well as a smaller amount of DNA ranging from 41% to 43% GC (corresponding to isochores from the H1 family). To get a clear view of the distribution the contigs' composition and their relationship to isochore families, contigs were grouped into bins whose
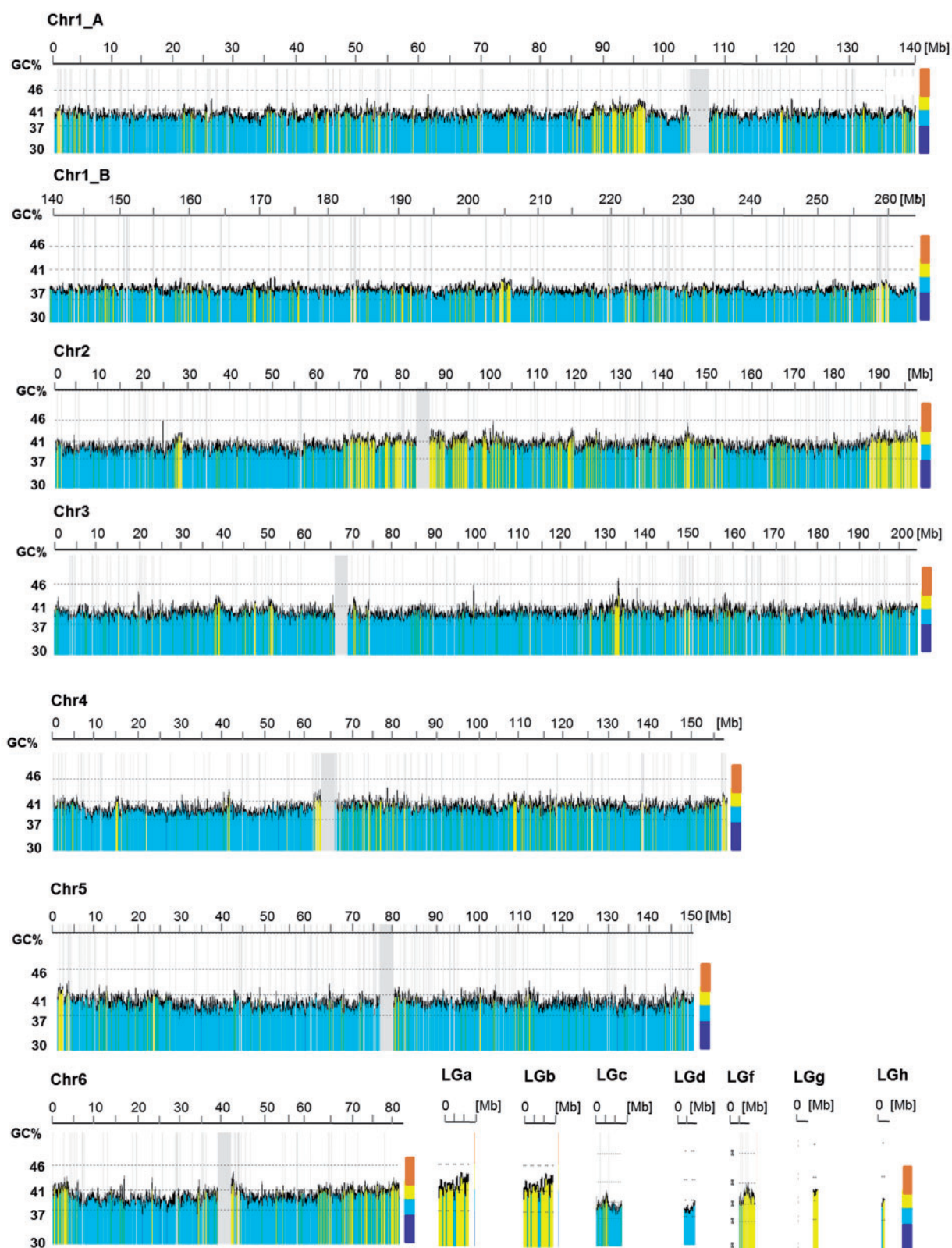
FIG. 1.—Compositional overview of *Anolis* chromosomes. The color-coded map shows 100-kb moving window plots using the program draw_chromosome_gc.pl (Pačes et al. 2004) (http://genomat.img.cas.cz). The color code spans the spectrum of GC levels. The ordinate values correspond to the minima GC values of isochore families. Gray vertical lines correspond to the gaps still present in the sequences, gray vertical regions to the centromeres.
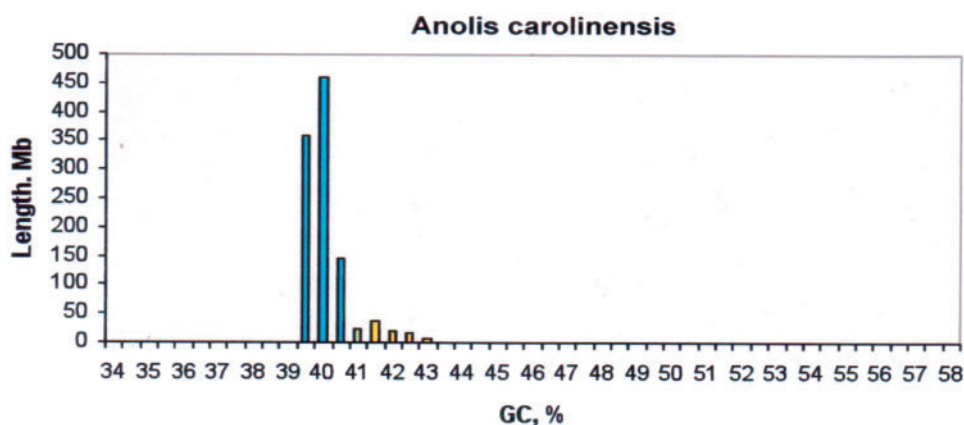
FIG. 2.—Distribution of isochores according to GC levels. The histograms show the distribution (by weight) of isochores as pooled in bins of 0.5% GC. Total amounts of sequences are calculated from the sums of isochores; colors represent the isochore families, according to Costantini et al. (2006).
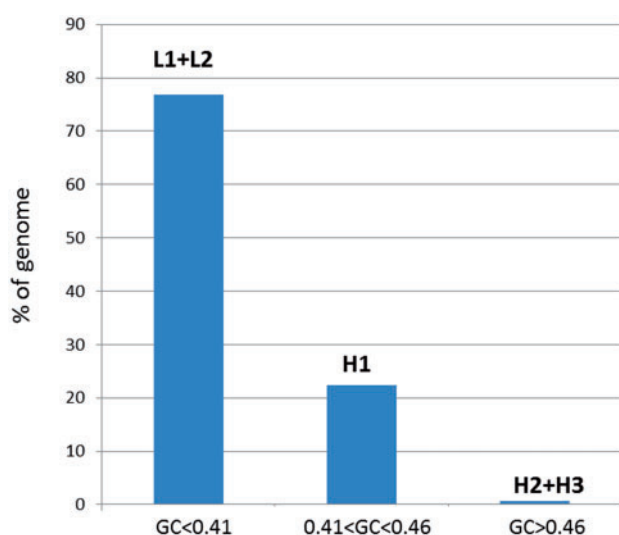

FIG. 3.—Relative amounts of the different compositional regions. All genomic contigs (at least 20 kb in length) were grouped according to their GC level. Owing to its very low representation, the GC poorest fraction (GC <37%), that corresponds to isochore family L1, was pooled with the fraction ranging from 37% to 41% (isochore family L2). Similarly, all contigs whose GC levels were >46% (corresponding to families H2 and H3) were considered together.

boundaries correspond to those previously defined for isochores in vertebrates (Costantini et al. 2006).

The relative abundance of each contig group is presented in figure 3. Note that the GC poorest GC fraction (GC <37%), that represents isochore family L1, was pooled with the fraction ranging from 37% to 41% (isochore family L2). Similarly, all contigs the GC contents of which were >46% (corresponding to H2 and H3 families) were considered together. Figure 3 shows that GC-poor regions represent the

predominant fraction in the sequenced *Anolis* genome. Nevertheless GC-richer isochore represent a substantial proportion of genome, 23% in the case of the H1 family and a smaller one, 2%, in the case of the H2 + H3 families.

The abundance estimates of each isochore family presented in figure 3 were obtained on the basis of the analysis of contigs >20 kb. If, instead, a more restrictive criterion is used, and contigs >50 kb are analyzed, the results remain basically unchanged (L1 = 3.7%, L2 = 78%, H1 = 18.2%, and H2 + H3 = 1.4%), but using contigs >100 kb the percentages are: L1 = 2.64%, L2 = 83%, and H1 + H2 + H3 = 13.5%. The reason of the decrease in the estimated proportion of GC-richer and GC-poorest isochores in larger contigs (i.e., L1, H1, and H2) is due to the fact that the larger the contig, the more probable a mixture of two different isochore families leading to average GC levels.

We have also checked what happens in the six macrochromosomes when one changes the window size (in the sliding window analysis) from 100 to 300 kb. The results are very clear. In Chr1 the proportion of windows that fall in H isochores decreases from 15% to 6%, in Chr2 from 29% to 26%, in Chr3 from 11% to 5%, in Chr4 from 17% to 10%, Chr5 from 8% to 3%, and in Cr6 from 19% to 15%.

Figure 4 shows the number of coding sequences as plotted against their GC, GC_1, GC_2, GC_3. These histograms differ from one another in the distribution of the coding sequences. The GC plot covers a range of ~35 to ~73% GC, with a peak at ~45% GC and shows a strong asymmetry toward high GC values. The GC_1 and GC_2 plots are essentially symmetrical ranging from ~35% to ~75%, and from ~25% to ~65%, respectively, the maxima being at ~55% for GC_1 and ~37% for GC_2. Expectedly, GC_3 covered a very wide range 30–100%, the maxima being at ~40–45% GC_3. While the GC_1 and GC_2 peaks are in the range found for all vertebrates (including warm-blooded vertebrates), the
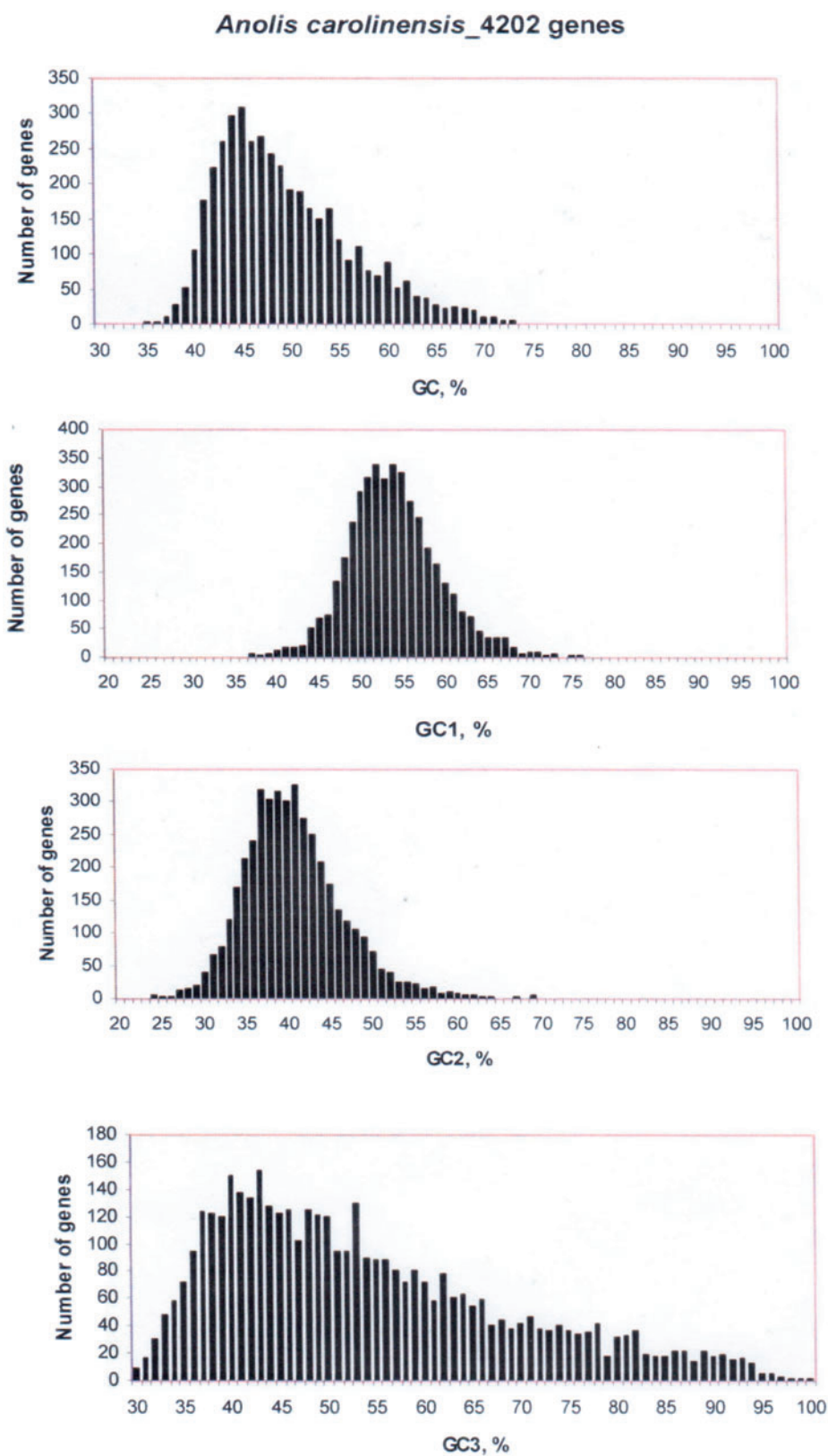
Fig. 4.—Distribution of GC, GC1, GC2, and GC3 of 4,202 genes from *Anolis*.
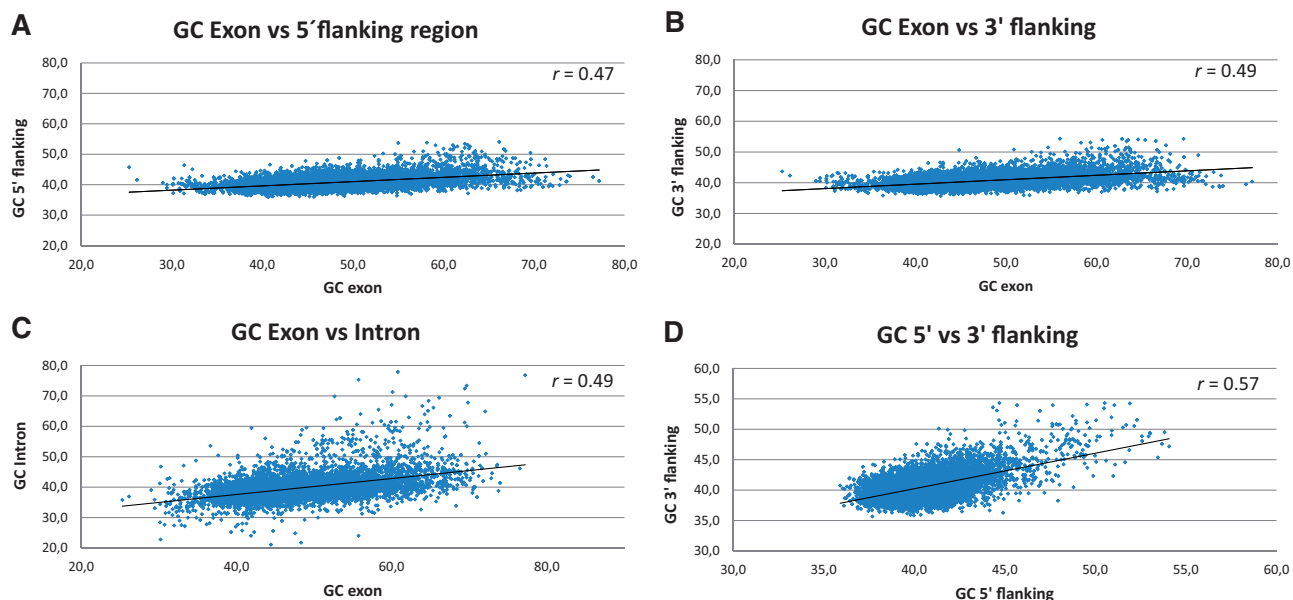
Fig. 5.—Linear correlation between GC levels in different genome regions. The total GC level of exons is plotted against the GC levels of the corresponding genomic region where the gene is embedded. 5′ and 3′ parts were calculated separately and are presented in *A* and *B*, respectively. (*C*) depicts the relationship in GC levels between exons and introns and (*D*) the relationship between genomic regions located up and downstream the coding region. To diminish the effect of large statistical errors (a problem raised by those genes that are contained in small contigs) only the genes surrounded by at least 20 kb were included in these plots. The correlation coefficients are indicated in the figures.
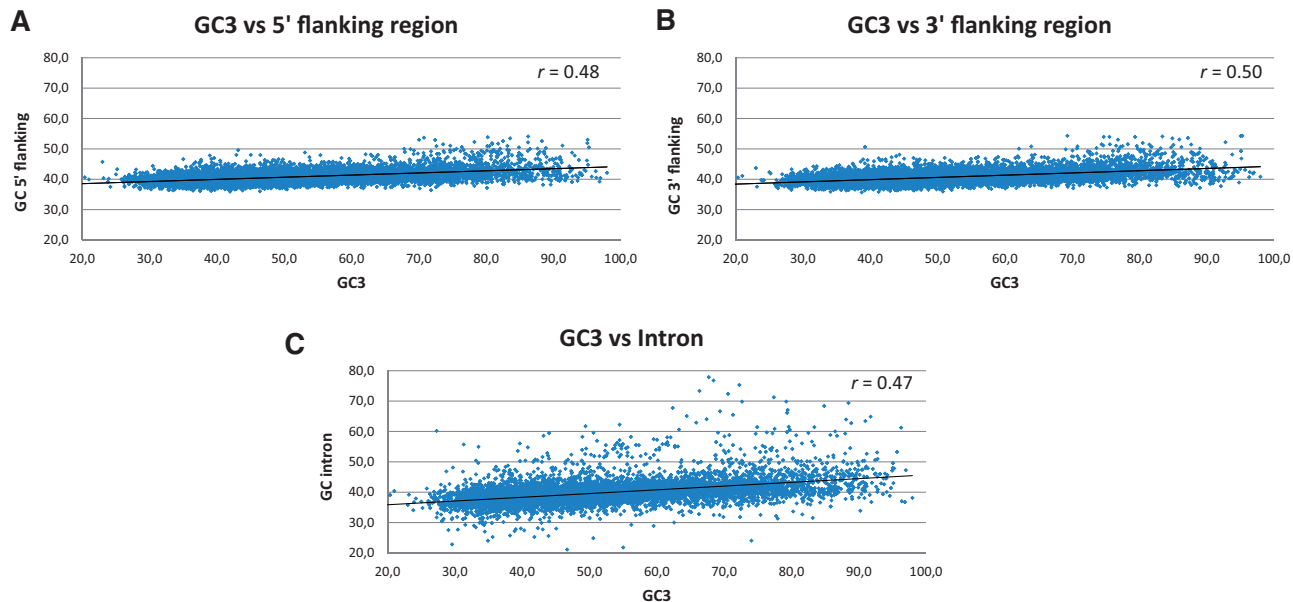

Fig. 6.—The GC_3 levels of exons is plotted against the GC levels of 5′ (*A*) and 3′ (*B*) flanking regions and that of introns (*C*). For further details see the legends of figure 4.

distribution of GC_3, although very wide (30–100% GC_3) has a maximum in a lower range, 40–50%, compared, for instance, with the human genome (70–80%).

Figures 5 and 6 present the correlations that are distinctive features of an isochore organization, namely those between exon GC and intron GC (of the same genes) and with flanking genomic regions located downstream and upstream from the genes. The correlation between 5′ and 3′ flanking regions is also presented. These figures show that in all cases the correlation coefficients, although weaker than in mammals, are

A — Number of exons per Mb

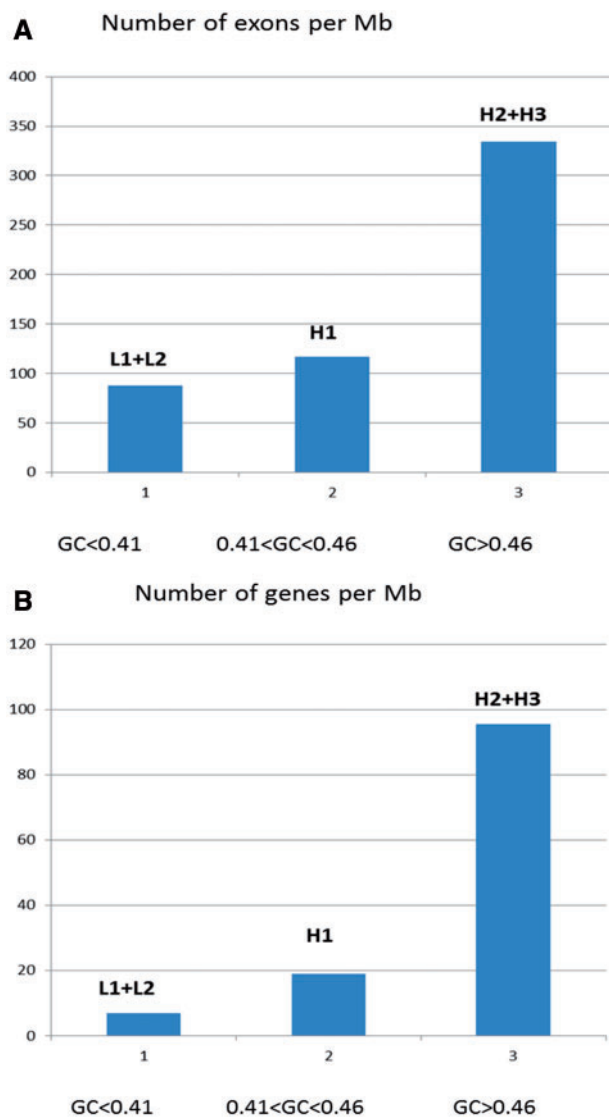## B — Number of genes per Mb

Fig. 7.—Coding sequence density. Number of exons (A) and number of genes per Mb. As in figure 6, contigs from L1 and L2 were pooled as were those from H2 and H3.

statistically very significant, showing that the genome of *Anolis* also displays the typical correlations that characterize vertebrate isochores.

Finally, figure 7 shows the coding densities in different isochore families (number of genes and exons per Mb). The figure shows a ~10-fold increase in the number of genes per Mb, from 8 in L1 + L2, to almost 100 in the GC-richest genomic segments. These results are in striking contradiction with the failure of Fujita et al. (2011) to find any increase in coding density in GC-rich genomic regions.

Needless to say, the results presented here show the distinctive signature of an isochore organization and contradict the claim that the *Anolis* genome is a genome without isochores. The compositional map of the *Anolis*

macrochromosomes consists of isochores mainly belonging to two families, a major L2 family and a less abundant H1 family. Genomic segments belonging to other isochore families (namely L1, H2, and H3) represent a minor fraction. On the other hand, the seven (out of the 12) microchromosomes for which sequences are available are predominantly formed by H1 isochores. In fact, as already mentioned, microchromosomes correspond to telomeric regions in birds (*Accipitridae*) that practically do not have microchromosomes (Federico et al. 2005).

Among the possible reasons why Fujita et al (2011) failed to find isochores in *Anolis*, the most likely one is the segmentation algorithm they have used along with the peculiar definition of isochores upon which the method is based (Fearnhead and Vasileiou 2009). Specifically, this algorithm identifies a genomic fragment as an isochore when two conditions are met: the segment should be at least 300 kb long and its GC variability must be lower than that of the chromosome where the segment is contained.

The first requirement is really difficult to meet in a genome the assembly of which has such a degree of fragmentation. Indeed, as explained both in Alföldi et al. (2011) and in http://www.ensembl.org/Anolis_carolinensis/Info/Annotation, this is indeed a very fragmented genome, consisting of 41,986 contigs and 2,143 scaffolds. The number of contigs $\geq$ 300 kb is only 157, whereas that of contigs 100–300 kb in size is 4,170 and those <100 kb is 38,000. Fujita et al. (2011) have used scaffolds (many of which of course are >300 kb) that comprise many gaps, even if, as a safety criterion, they discarded from their analysis scaffolds containing >20% of gaps. This, in our opinion, may introduce distortions because the real size of gaps (inside scaffolds) is unknown and since they are estimated assuming that the distances between the (consecutive) contigs used to build the scaffold in questions is similar to that in Gallus (in the conserved syntenic blocks) forgetting that the chicken genome size, 1.25 pg (Wright and Gregory 2014) is almost half of the *Anolis* genome, 2.20 pg. Under these conditions it is clear that looking for genomic stretches having fairly homogenous GC levels which are $\geq$ 300 kb will lead to a underestimation of H1 family isochore, because fragmentation is more pronounced in GC-rich regions (Lander et al. 2001).

The second requirement is even more problematic. This definition (Cohen et al. 2005) is inconsistent in many ways, as already pointed out (see Clay and Bernardi 2005). Among the several inconsistencies that this definition implies, we will mention the most obvious one. If one takes a GC-rich isochore from the human genome and places it on chromosome 19, it will be not identified as an isochore since this chromosome is GC-rich and relatively homogeneous. However, if the same genomic fragment is placed in GC-poor human chromosome, it will be readily identified as such. In other words, this misleading definition of isochore leads to flagrant inconsistencies. As Clay and Bernardi (2005) pointed out, the baseline of GC

variability that should be used to test isochore homogeneity is that of the whole genome including both macro- and micro-chromosomes, when the latter are present.

Another point worth mentioning is the way Fujita et al (2011) interpret their results in relation to the proportion of the *Anolis* genome that would be organized in isochores. Using Fearnhead and Vasileiou (2009) algorithm they found that only ~15% of the *Anolis* genome would be organized in "classical" isochores, and they consider this as a small fraction. However, using the same isochore definition of Cohen et al. (2005), only 41% of the human genome would be organized in isochores.

We will not comment here on other aspects of the article by Fujita et al. (2011), especially on their evolutionary considerations, except to say that the lack of correlation between GC-level and body temperature claimed by Belle et al. (2002) and Ream et al. (2003) was contradicted by Jabbari et al. (2003) and Clay et al. (2003) in papers that are not quoted.

In conclusion, the genome of *Anolis* does not make an exception to the general rule of an isochore organization, which concerns all vertebrate and also the invertebrate genomes tested so far (Cammarano et al. 2009), isochores remaining "a fundamental level of genome organization" (Eyre-Walker and Hurst 2001).

Acknowledgments

Literature Cited

Alföldi J, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477:587–591.

Atkin NB, Mattinson G, Beçak W, Ohno S. 1965. The comparative DNA content of 19 species of placental mammals, reptiles, and birds. Chromosoma 17:1–10.

Belle EM, Smith N, Eyre-Walker A. 2002. Analysis of the phylogenetic distribution of isochores in vertebrates and a test of the thermal stability hypothesis. J Mol Evol. 55:356–363.

Bernardi G. 2001. Misunderstanding about isochores. Gene 276:3–13.

Bernardi G. 2004. Structural and evolutionary genomics: natural selection in genome evolution. Amsterdam: Elsevier.

Bernardi G. 2015. Chromosome architecture and genome organization. PLoS One 10(11):e0143739.

Cammarano R, Costantini M, Bernardi G. 2009. The isochore patterns of invertebrate genomes. BMC Genomics 10:538–550.

Clay O, Arhondakis S, D'Onofrio G, Bernardi G. 2003. LDH-A and α-actin as tools to assess the effects of temperature on the vertebrate genome: some problems. Gene 317:157–160.

Clay O, Bernardi G. 2001. The isochores in human chromosomes 21 and 22. Biochem Biophys Res Commun. 285:855–856.

Clay O, Bernardi G. 2002. Isochores: dream or reality? Trends Biotechnol. 20:237.

Clay O, Bernardi G. 2005. How not to search for isochores: a reply to Cohen et al. Mol Biol Evol. 22:2315–2317.

Clay O, Bernardi G. 2011. GC3 of genes can be used as a proxi for isochore base composition: a reply to Elhaik et al. Mol Biol Evol. 28:21–23.

Cohen N, Dagan T, Stone 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 DNA's. Biochemistry 7: 4373–4379.

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

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

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.

Elhaik E, Landau G, Graur D. 2009. Can GC content at third-codon positions be used as a proxy for isochore composition? Mol Biol Evol. 26:1829–1833.

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

Fearnhead P, Vasileiou D. 2009. Bayesian analysis of isochores. J Am Stat Assoc. 104:132–141.

Federico C, et al. 2005. Avian genomes: different karyotypes but a similar distribution of the GC-richest chromosome regions at interphase. Chromosome Res. 13:785–793.

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.

Fujita MK, Edwards SV, Ponting CP. 2011. The Anolis lizard genome: an amniote genome without isochores. Genome Biol Evol. 3:974–984.

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

Jabbari K, Clay O, Bernardi G. 2003. GC_3 heterogeneity and body temperature in vertebrates. Gene 317:161–163.

Lander EA, 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.

Medrano L, Bernardi G, Couturier J, Dutrillaux B, Bernardi G. 1988. Chromosome banding and genome compartmentalization in fishes. Chromosoma 96:178–183.

Pačes J, Zika R, Pavlíček A, Clay O, Bernardi G. 2004. Representing GC variation along eukaryotic chromosomes. Gene 333:135–141.

Pavlíček A, Pačes J, Clay O, Bernardi G. 2002. A compact view of isochores in the draft human genome sequence. FEBS Lett. 511:165–169.

Peterson DG, Stack SM, Healy JL, Donohoe BS, Anderson LK. 1994. The relationship between synaptonemal complex length and genome size in four vertebrate classes (Osteichthyes, Reptilia, Aves, Mammalia). Chromosome Res. 2:153–162.

Ream RA, Johns GC, Somero GN. 2003. Base compositions of genes encoding α-actin and lactate dehydrogenase-A from differently adapted vertebrates show no temperature-adaptive variation in G+C content. Mol Biol Evol. 20:105–110.

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

Wright NA, Gregory TR, Witt CC. 2014. Metabolic 'engines' of flight drive genome size reduction in birds. Proc Biol Sci. 281:20132780.

Associate editor: Takashi Gojobori