cDNA-based gene mapping and GC_3 profiling in the soft-shelled turtle suggests a chromosomal size-dependent GC bias shared by sauropsids.

Shigehiro Kuraku,[1*] Junko Ishijima,[2] Chizuko Nishida-Umehara,[2,3] Kiyokazu Agata,[4,5] Shigeru Kuratani,[1] and Yoichi Matsuda[2,3]

[1]Laboratory for Evolutionary Morphology, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan; Tel: +81-78-3063064; Fax: +81-78-3063370; E-mail: venezia@cdb.riken.jp; [2]Laboratory of Animal Cytogenetics, Division of Genome Dynamics, Creative Research Initiative "Sousei", Hokkaido University, North 10 West 8, Kita-ku, Sapporo 060-0810, Japan; [3]Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo, 060-0810, Japan; [4]Laboratory for Evolutionary Regeneration Biology, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, 650-0047, Japan

[*]*Correspondence.*

[5]Present address: *Laboratory for Molecular Developmental Biology, Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto, 606-8502, Japan*

Running title: GC heterogeneity in the turtle genome

Key words: GC-content, sauropsida, turtle, microchromosome.

Abstract

Mammalian and avian genomes comprise several classes of chromosomal segments that vary dramatically in GC-content. Especially in chicken, microchromosomes exhibit a higher GC-content and a higher gene density than macrochromosomes. To understand the evolutionary history of the intra-genome GC heterogeneity in amniotes, it is necessary to examine the equivalence of this GC heterogeneity at the nucleotide level between these animals including reptiles, from which birds diverged. We isolated cDNAs for 39 protein-coding genes from the Chinese soft-shelled turtle, *Pelodiscus sinensis*, and performed chromosome mapping of 31 genes. The GC-content of exonic third positions (GC_3) of *P. sinensis* genes showed a heterogeneous distribution, and exhibited a significant positive correlation with that of chicken and human orthologs, indicating that the last common ancestor of extant amniotes had already established a GC-compartmentalized genomic structure. Furthermore, chromosome mapping in *P. sinensis* revealed that microchromosomes tend to contain more GC-rich genes than GC-poor genes, as in chicken. These results illustrate two modes of genome evolution in amniotes: mammals sophisticated the genomic configuration in which GC-rich and GC-poor regions coexist in individual chromosomes, whereas sauropsids (reptiles and birds) refined the chromosomal size-dependent GC compartmentalization in which GC-rich genomic fractions tend to be confined to microchromosomes.

Introduction

Mammalian and avian genomes have been revealed, by means of chromosome banding and density gradient centrifugations, to be composed of several classes of chromosomal segments that differ in GC-content, which are called 'isochores' (Bernardi *et al.* 1985). Although the evolutionary origin and intrinsic nature of this GC heterogeneity is not fully understood (Eyre-Walker & Hurst 2001), the existence of intra-genome GC heterogeneity was recently confirmed by analyses of whole-genome sequences in the chicken as well as human, mouse and rat (International Human Genome Sequence Consortium [IHGSC] 2001; Mouse Genome Sequencing Consortium [MGSC] 2002; International Chicken Genome Sequencing Consortium [ICGSC] 2004; Rat Genome Sequencing Project Consortium [RGSPC] 2004; also see Figure 1a).

Karyotypes of extant sauropsids (reptiles and birds) generally consist of two major components: macrochromosomes and microchromosomes (Burt 2002; Norris *et al.* 2004). In chicken, cytogenetic observations indicate that microchromosomes exhibit a higher gene density (McQueen *et al.* 1998; Smith *et al.* 2000), a higher density of CpG islands (McQueen *et al.* 1996) and a higher GC-content than macrochromosomes (Auer *et al.* 1987; Andreozzi *et al.* 2001). The whole genomic sequence of chicken has yielded trends consistent with the above, and, especially, suggested that the global GC-content of chromosomes increases exponentially with the reduction in chromosomal size (ICGSC 2004; also see Figure 1b), whereas this tendency is not seen in mammals and teleosts (Figures 1c-e). These features of the chicken genome suggest that avian microchromosomes might be the counterparts of mammalian GC-rich chromosomal segments (Andreozzi *et al.* 2001). However, it is not clear whether the intra-genome GC

heterogeneity observed in mammals and birds was derived from a common ancestor, or was the result of a convergence that occurred independently in the two lineages. Reptiles could provide valuable information for addressing this question.

The existence of intra-genomic GC heterogeneity in reptiles has not been fully confirmed by chromosome banding studies (Holmquist 1989) and density gradient centrifugation (Thiery *et al.* 1976; Hughes *et al.* 2002). However, some recent studies at the nucleotide level suggest that GC heterogeneity exists in reptilian genomes, based on variations in GC-contents in exonic third positions (GC_3) and introns of a limited number of genes (Hughes *et al.* 1999; Belle *et al.* 2002; Hamada *et al.* 2003). Here, the GC_3 of a gene is expected to positively correlate with the GC_3 of the genomic region where the gene is located, as confirmed in mammalian and avian genomes (Clay *et al.* 1996; Musto *et al.* 1999). However, the paucity of sequence information on reptilian species has inhibited understanding of the physical configuration of reptilian genomes and the evolutionary origin of heterogeneity in base composition.

In this study, we cloned and sequenced cDNAs of protein-coding genes from the Chinese soft-shelled turtle, *Pelodiscus sinensis*, and localized them to chromosomes by fluorescent *in situ* hybridization (FISH). The physical evidence of chromosomal configurations in the turtle, as suggested by cDNA-based approaches and comparison with chromosomal configurations in other amniotes, has highlighted two modes of genome evolution in amniotes.

Materials and methods

Isolation and sequencing of cDNAs with degenerate primers

Total RNA isolated from whole embryos of stage 14 *P. sinensis* was reverse transcribed into cDNA using an oligo(dT) primer and SuperScript III (Invitrogen). These cDNAs were used as templates for PCR amplification with the FastStart High Fidelity PCR System (Roche). The sense and antisense degenerate primers were designed based on the conserved amino acid residues in the multiple alignments constructed as described below, and are as shown in Table 1. PCR was conducted as follows: 2 min denaturation step at 94 °C; then 10 cycles of 94 °C for 15 s, 48 °C for 30 s, and 72 °C for 2 min; followed by 30 cycles of 94 °C for 15 s, 56 °C for 30 s, and 72 °C for 2 min. Modifications were made when required, depending on the presumed length of amplicons and the T_m value of the primers used. The PCR products were purified using MinElute (Qiagen) and cloned into a pT7Blue vector (Novagen). More than three independent clones per gene were sequenced using a 3100 Genetic Analyzer or 3730XL DNA Analyzer (Applied Biosystems). Upstream and downstream regions of isolated cDNAs were cloned and sequenced by 5′ and 3′ rapid amplification of cDNA ends (Frohman *et al.* 1988).

Estimation of numbers of synonymous and non-synonymous substitutions

Nucleotide sequences of orthologous gene pairs of chicken-turtle were manually aligned on the XCED program (Katoh *et al.* 2002) based on alignments of the amino acid sequences of the proteins they encode. K_s and K_a were calculated with the codon-based maximum-likelihood method (Goldman & Yang 1994) and with the method of Nei and Gojobori (1986). Computations were processed using a PAML 3.1 package (Yang

1997).

Calculation of GC-content

cDNA sequences for human, mouse, rat, chicken, tiger pufferfish (*F. rubripes*), *C. intestinalis*, *Drosophila melanogaster* and *Caenorhabditis elegans* were downloaded from Ensembl (version 34 - Oct, 2005; URL: http://www.ensembl.org/; Hubbard *et al.* 2005). Those for sheep (*O. aries*), axolotl (*Ambystoma mexicanum*), *X. tropicalis*, zebrafish (*D. rerio*), and amphioxus (*Branchiostoma belcheri*) were downloaded from GenBank (version 148.0). After redundant sequences, which are thought to be derived from a single gene, has been assembled with Phrap (URL: http://www.phrap.com/), GC_3 and GC_4 were calculated using a Perl script with the bioperl module (Stajich *et al.* 2002). The calculation was automatically processed based on the open reading frame identified with a pairwise alignment between translated nucleotide sequences and corresponding amino acid sequences with BLASTX (Altschul *et al.* 1997). Analysis of variance (ANOVA) for statistical analysis of similarities in distribution of GC-contents was conducted with non-parametric rank tests on the assumption that the overall distribution of GC_3 for all the genes in one species does not have a normal distribution.

Chromosome preparation and FISH

Fibroblast cells derived from embryos of *P. sinensis* were cultured and used for chromosome preparations. Preparation of R-banded chromosomes and FISH were performed as described previously (Matsuda & Chapman 1995; Suzuki *et al.* 1999).

5-bromodeoxyuridine (BrdU) was incorporated into chromosomes during the late replication stage for differential staining, and R-banded chromosomes were obtained by exposing chromosome slides to UV light after staining with Hoechst 33258. DNA probes were labeled by nick translation with biotin-16-dUTP (Roche) using a standard protocol. Plasmids with insert cDNA longer than 0.7 kb were used as templates for labeling. The hybridized cDNA probes were reacted with goat anti-biotin antibodies (Vector Laboratories), and then stained with fluorescein-labeled donkey anti-goat IgG (Nordic Immunology). The slides were stained with 0.50 μg/ml propidium iodide for observation.

Gene mapping information for human, mouse and chicken chromosomes

Chromosomal locations of human and mouse genes were retrieved from NCBI Entrez Gene (URL: http://www.ncbi.nih.gov/entrez/query.fcgi?db=gene) and Ensembl. Mapping information of chicken genes was based on the previous studies (Suzuki *et al.* 1999; Guttenbach *et al.* 2000; Schmid *et al.* 2000) and Ensembl.

Results

Identification of novel cDNAs in the Chinese soft-shelled turtle, Pelodiscus sinensis

With the reverse transcription-polymerase chain reaction (RT-PCR) using degenerate primers, we isolated and sequenced cDNA derived from 39 protein-coding genes

located in the nuclear genome of *P. sinensis* (Table 2). The total length of the sequenced cDNA fragments was 38,324 bp (9,527 amino acids). These sequences were deposited in GenBank under accession numbers AB188346-AB188384. Orthology to homologous genes reported in other vertebrates was rigorously confirmed for each gene by molecular phylogenetic trees constructed with the neighbor-joining method (Saitou & Nei 1987) and the maximum-likelihood method (Felsenstein 1981; Yang 1997). In these phylogenetic studies, we did not detect any gene duplications unique to the turtle lineage, indicating that the *P. sinensis* genome possesses a highly similar gene repertoire to that of other amniotes for these genes (data not shown).

Estimated number of synonymous substitutions between turtle and chicken

Including the sequences available in the public nucleotide sequence database GenBank (version 148.0), we selected 56 genes that satisfied the criteria that only a single ortholog should be found in *P. sinensis*, chicken, human and mouse, and that the orthologous sequences aligned between *P. sinensis* and chicken should be longer than 300 bp. For each pair, we estimated the number of synonymous (K_s) and non-synonymous substitution (K_a) between turtle and chicken (Table 3). The total length used for calculations was 26,268 bp (8,756 codons). The average K_s was 0.96 (standard deviation, 0.58; n = 56) under the maximum-likelihood method (Goldman & Yang 1994), and was 0.68 (standard deviation, 0.26; n = 56) under the method of Nei and Gojobori (1986). Positive selection ($K_a/K_s > 1$) was not detected in any of the 56 gene pairs (Table 3).

Distribution of GC_3 in turtle and other chordates

We calculated the GC_3 for 125 *P. sinensis* genes. In addition to the genes found in GenBank, we used cDNA sequences already deposited in the NCBI dbEST category (accession nos. AU312239-AU312301; Matsuda *et al.* 2005) with deduced protein-coding regions longer than 200 bp. The GC_3 of *P. sinensis* genes exhibited a bimodal distribution with an average of 55.7% and a standard deviation of 16.0% (Figure 2a), which was similar to that in chicken, human and sheep (*Ovis aries*) (standard deviation 16.3%-16.8%; Mann-Whitney U test, $P < 0.01$; Figures 2b, c). In contrast, non-amniotic vertebrates and invertebrates showed a unimodal distribution of GC_3 with a narrow standard deviation (7%-11%), although the averages varied extremely between species (40%-70%; Figures 2e-h). Mouse and rat also exhibited a unimodal distribution with averages of approximately 60%, and a much smaller standard deviation compared with the other amniotes (11%-12%; Figure 2d).

Cross-species GC_3 comparison between orthologs

To examine whether each turtle gene possessed a similar GC_3, we compared the GC_3 of the turtle genes with that of 56 genes from chicken (Figure 3a) and human (Figure 3b), for which we identified 1:1 ortholog pairs (Table 3). The turtle-chicken GC_3 comparison showed a significant positive correlation, with a correlation coefficient $r = 0.84$ (Spearman's rank correlation, $P < 0.001$), as did the turtle-human comparison with $r = 0.61$ ($P < 0.001$). Similarly, the human-chicken comparison exhibited a significant positive correlation with our gene set ($r = 0.63$, $P < 0.001$). In contrast, turtle-*Xenopus*

tropicalis and human-*X. tropicalis* comparisons of GC_3 for these 56 genes did not yield a significant correlation ($r = 0.09$ and $r = 0.11$, respectively) (Figures 3c, d).

Gene mapping on P. sinensis chromosomes

In this study, we treated chicken chromosomes 1-8, Z, and W, and the *P. sinensis* chromosomes 1-6 as macrochromosomes, and the rest of chromosomes as microchromosomes, following previous studies (McQueen *et al.* 1996; Matsuda *et al.* 2005; for karyotypic configuration of *P. sinensis* and chicken, see Discussion). We performed FISH mapping of 31 *P. sinensis* genes using cDNA clones isolated in this study as probes. Seventeen of the 31 genes were localized to the macrochromosomes, and the remaining 14 genes were localized to microchromosomes (Table 4). The five largest turtle chromosomes each corresponded to one chicken chromosome; chromosome numbers were equivalent between the turtle and chicken, with one exceptional case of the *PRRX1* gene (Table 4). The *FGF10* gene was localized to the turtle chromosome 6 (Table 4), which corresponded to the chicken sex Z chromosome (Matsuda *et al.* 2005). All the genes on the turtle chromosomes 7 and 8 were localized to the chicken chromosomes 7 and 6, respectively (Table 4). These results indicated that the eight largest turtle chromosomes each correspond to one chicken chromosome (chicken chromosomes 1–7 and chromosome Z), and confirmed that there is high level of conserved synteny along chromosomes between the turtle and chicken.

Comparison of GC_3 between macrochromosomes and microchromosomes of turtle and chicken

To corroborate the high GC-content in microchromosomes, previously shown by chromosome banding (Auer *et al.* 1987; Andreozzi *et al.* 2001) and whole genome sequencing (ICGSC 2004; also see Figure 4a), we compared the GC_3 between genes on macrochromosomes and microchromosomes, in chicken (Figure 4b) and *P. sinensis* (Figure 4c). We used 59 genes previously reported (Matsuda *et al.* 2005) and 31 genes mapped in this study (Table 4). The average GC_3 was 51.6% and 60.1% in chicken (Figure 4b), and 50.1% and 57.7% in *P. sinensis*, for macrochromosomes and microchromosomes, respectively. In *P. sinensis*, about 51.1% (23 out of 45 genes) of GC-rich genes ($GC_3 \geq 50\%$) resided on microchromosomes (47.9% in chicken) (Figure 5b), whereas 31.1% (14 out of 45 genes) of GC-poor genes ($GC_3 < 50\%$) resided on microchromosomes (24.7% in chicken) (Figure 5c).

Discussion

Molecular phylogeny and evolutionary distance in sauropsids

Phylogenetic relationships between reptilian orders and birds have been controversial for decades (see Zardoya & Meyer 2001, for review). However, recent molecular phylogenetic analyses, using nuclear DNA-coded and mitochondrial DNA-coded genes, support a tree topology that places turtles closer to the Archosaurians (birds and crocodilians) than to the Lepidosaurians (tuataras, snakes, and lizards), in contrast to the archaic tree topology that positions turtles at a basal branch within the sauropsids

(Zardoya & Meyer 1998; Hedges & Poling 1999; Kumazawa & Nishida 1999; Cao *et al.* 2000; Rest *et al.* 2003; Iwabe *et al.* 2005).

In this study, we estimated the number of synonymous substitutions (K_s) in genes of the Chinese soft-shelled turtle, *P. sinensis*, and chicken, to measure the evolutionary distance between the two species at the molecular level, using 56 carefully chosen gene pairs that are conserved as a single ortholog in turtle, chicken, human and mouse (Table 3). The number of synonymous substitutions represents the amount of neutral substitutions accumulated in both lineages (Miyata & Yasunaga 1980) and, accordingly, serves as a standard index of the evolutionary distance between the two species. Our K_s estimation for the turtle-chicken gene pairs was 0.96 with the maximum-likelihood method (Goldman & Yang 1994), whereas previous studies using the same method have estimated the K_s for human-mouse gene pairs as 0.56 (RGSPC 2004) and for human-chicken gene pairs as 1.66 (ICGSC 2004). The difference between these figures is consistent with the above evolutionary hierarchy of amniote phylogeny, given that neutral substitution rates along these lineages have not dramatically changed. Additionally, it serves as a standard of evolutionary distance among these taxa and an indicator of orthology between genes in species belonging to these taxa.

GC_3 as a reflection of genomic GC level

In contrast to the homogeneous distribution of GC-content with sharp peaks in non-amniotic species, such as the tiger pufferfish (*Fugu rubripes*) and the tunicate (*Ciona intestinalis*) (Aparicio *et al.* 2002; Dehal *et al.* 2002), amniotes, such as human and chicken, show intra-genome GC heterogeneity; however, GC-content in rodent

genomes is somewhat homogeneous (Mouchiroud *et al.* 1988; MGSC 2002; RGSPC 2004) (Figure 1a). In this study, we focused on protein-coding regions rather than unavailable genomic sequences, because the GC_3 of a specific gene is expected to correlate with the global GC-content of the genomic region where the gene is found (Clay *et al.* 1996; Musto *et al.* 1999). Another advantage of focusing on protein-coding regions is that orthologies between corresponding chromosomal segments in different species are easily detectable with molecular phylogenetic analyses of genes harbored within these segments. In contrast, by focusing on protein-coding regions, we cannot incorporate the GC-content of non-coding DNA sequences, which make up a considerable proportion of a genome, into our present analysis. However, cross-species analysis of coding regions is an effective tool for focusing on orthologous genomic fractions derived from common ancestors by excluding the influence of the lineage-specific expansion of some specific genomic regions, such as repetitive elements. Another concern for imaginable pitfalls is that, in examining intra-genome GC heterogeneity, it is preferable to focus on a single species as a representative of the taxonomic group under investigation, because inter-species GC variation may mask the intra-genome GC landscape. For example, variable peaks of GC_3 distribution occur between the tiger pufferfish (*F. rubripes*) and the zebrafish (*Danio rerio*), which both belong to a single group of teleostei (Figure 2f).

Equivalence of the GC heterogeneity among amniotes

We isolated and sequenced 39 novel cDNAs from *P. sinensis*. Our subsequent GC_3 calculation for various chordates revealed that the GC_3 in *P. sinensis* exhibited a broad

and bimodal distribution, which had a strong resemblance to that in chicken and non-rodent mammalians, such as human and sheep (Figures 2a-c), but not to that in non-amniotic species (Figures 2e-h) or rodents (Figure 2d). Similar results were obtained when GC-contents at four-fold degenerate sites (GC_4) were analyzed (data not shown).

The next question we addressed was whether the GC_3 distribution in turtle was derived from the common ancestor of mammalians, reptiles and birds, or acquired secondarily in independent lineages. Thus, we performed a cross-species comparison of GC_3 between orthologs found in turtle, chicken and human. The results clearly showed that each turtle gene possesses a similar level of GC_3 to its ortholog in human and chicken, with statistically significant positive correlations (Figures 3a, b). The higher levels of GC_3 correlation between the turtle-chicken pair compared with the turtle-human pair can be explained by the lower levels of neutral substitutions and translocations that accumulated in the turtle-chicken pair, which may have caused a secondary decay of ancestral intra-genome GC-bias. Our ortholog set also exhibited a significant positive correlation between GC_3 and GC-content of the surrounding genomic region (10 kb on each side) in chicken (Spearman's $r = 0.65$, $P < 0.005$) (Figure 6). To rule out the possibility that the limited number of genes has yielded misleading results because of a biased choice of genes, we then confirmed that the GC_3 distribution in the set of orthologous genes examined significantly resembled that obtained with an original large set of cDNAs in human and chicken, respectively (Mann-Whitney U test, $P < 0.01$ for both; data not shown). Despite that our gene set is limited in number, it seems sufficient to speculate the overall features of the turtle genome. In conclusion, taking into account the results of the present analysis with 56

orthologous genes, along with previous observations that GC_3 levels have been highly conserved in a large set of orthologous genes between human and chicken (Kadi *et al.* 1993; Bernardi 2000), the most parsimonious interpretation is that orthologous genomic regions in these three species have maintained similar GC-contents which were derived from the last common ancestor of mammals, birds, and reptiles.

In contrast, we detected no significant correlation in a comparison of GC_3 between turtle and *X. tropicalis* for the above ortholog pairs (Figures 3c, d); however, Bernardi (2000) have reported a weak positive correlation in GC_3 between human and *X. laevis*. At present, whether the origin of intra-genome GC heterogeneity antedated the common ancestors of amniotes and amphibians remains unanswered.

Differences in gene density between macrochromosomes and microchromosomes

The diploid chromosome number of *P. sinensis* is 2n = 66, which consists of six pairs of macrochromosomes and 27 pairs of microchromosomes (Matsuda *et al.* 2005), whereas the chicken karyotype (2n = 78) consists of nine pairs of macrochromosomes, including the ZW sex chromosomes, and 29 pairs of microchromosomes. In chicken, it was suggested that microchromosomes contain up to 50% of the genes in the genome and represent about 23% of total genomic DNA, indicating that the gene density on microchromosomes is two to three times higher than that on macrochromosomes (Smith *et al.* 2000). In this study, we found that 41.1% (37 out of 90 genes) of turtle genes localize to microchromosomes (Figure 5a). This figure resembles that observed in chicken (37.4%; 3132 out of 8380 genes; Figure 5a). Although this trend is now roughly confirmed with genomic sequences only in chicken (ICGSC 2004), further efforts will

be required to unveil a gene distribution on turtle chromosomes. Additionally, the genes we used represent all the macrochromosomes and some microchromosomes in chicken and *P. sinensis* (Table 4), which suggests these genes serve as random markers for whole genomic regions.

Difference in GC-content between macrochromosomes and microchromosomes

Unlike in mammals, the average intra-chromosomal GC-content in chicken increases with the reduction in chromosomal length (Figure 1b). In this context, our gene mapping detected a difference in GC_3 distribution between macrochromosomes and microchromosomes (Figure 4). In *P. sinensis* and chicken, GC-poor genes are two to three times more likely to reside on macrochromosomes than on microchromosomes (Figure 5c), whereas GC-rich genes tend to reside equally on macrochromosomes and microchromosomes (Figure 5b). Thus, macrochromosomes tend to contain more GC-poor genes (Figure 5d), whereas microchromosomes tend to contain more GC-rich genes (Figure 5e). In contrast, there is no significant correlation between GC_3 and the size of chromosomes harboring them in human and mouse (data not shown), which is consistent with the analysis at the genomic level (Figures 1b-e). Accordingly, chromosomal size-dependent GC compartmentalization seems to be unique to sauropsids whose karyotypes consist of macrochromosomes and microchromosomes.

Insight into the evolutionary history of intra-genome GC heterogeneity

We conclude that the base composition in the turtle genome has a strong resemblance to

its chicken counterpart; that is, the turtle genome exhibits a high level of intra-genome GC heterogeneity, and a higher proportion of GC-rich genes on microchromosomes (Figure 5e). This conclusion is incompatible with previous observations using chromosome banding studies (Holmquist 1989) and density gradient centrifugation (Thiery *et al.* 1976; Hughes *et al.* 2002), which reported the lower level of GC heterogeneity in the turtle genome. Apart from the problems in sensitivity of the above indirect methods, the effect of lineage-specific events such as expansion of repetitive elements with extreme GC-contents might reconcile this difference. After all, to understand the evolutionary history of intra-genome GC bias, studies focusing on protein-coding regions might be more likely to detect features derived from the common ancestors.

The intra-genome distribution of GC-rich and GC-poor regions has not been clarified in other sauropsids. However, the karyotype of the common ancestor of extant sauropsids is thought to have contained both macrochromosomes and microchromosomes (Burt 2002; Norris *et al.* 2004), although some lineages underwent frequent secondary fusion of microchromosomes resulting in no or few microchromosomes as seen in the reptilian family Crocodylidae and the avian family Falconiformes (Cohen & Gans 1970; De Boer & Sinoo 1984). In contrast, chromosome sizes are relatively uniform and there is no striking bias in inter-chromosomal GC-content in most mammals. These facts indicate that sauropsids adopted a chromosomal size-dependent GC compartmentalization strata, whereas mammals maintained the system in which GC-rich and GC-poor regions coexist on individual chromosomes in a highly juxtaposed manner (Figure 7). This hypothesis has yet to be verified by further large-scale studies, not only in turtle, but also in other sauropsids.

Furthermore, it is important to clarify whether monotremes, marsupials and amphibians have a similar pattern of intra-genome GC distribution to eutherians, in order to speculate on the ancestral configuration for the amniote genome by adding outgroup polarity to the present scheme.

The genomic landscape of base composition cannot be comprehensively realized without elaborate genomic sequencing. However, our approaches of cDNA sequencing followed by cDNA-based gene mapping and *in silico* GC_3 calculation have revealed a shared GC heterogeneity of the Chinese soft-shelled turtle *P. sinensis* and chicken. In this study, as a source of material in GC profiling in the turtle, we cloned and sequenced cDNAs of limited number of protein-coding genes, especially because we observed in our preliminary studies that a high-throughput sequencing of cDNAs, such as expression sequence tags (ESTs), did not always reproduce GC distribution of whole gene repertoires in chicken and human, possibly due to some experimental biases or correlation between GC level of genes and their expression levels. Therefore, we carefully examined in chicken and human whether our gene set reproduces GC distribution of whole genes and even of whole genomes. As long as such attention is paid, our approach serves as an informative tool for surveying genomic features in non-model organisms for which there is a limited amount of genomic sequence information.

Acknowledgements

We are very grateful to Ryo Usuda for technical assistance in cDNA cloning and

sequencing, and to Kanako O. Koyanagi and Kazutaka Katoh for valuable comments and discussion. This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology in Japan and the Grant for Biodiversity Research of the 21st Century COE (A14).

References

Altschul SF, Madden TL, Schaffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389-3402.

Andreozzi L, Federico C, Motta S *et al.* (2001) Compositional mapping of chicken chromosomes and identification of the gene-richest regions. *Chromosome Res* **9:** 521-532.

Aparicio S, Chapman J, Stupka E *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297:** 1301-1310.

Auer H, Mayr B, Lambrou M, Schleger W (1987) An extended chicken karyotype, including the NOR chromosome. *Cytogenet Cell Genet* **45:** 218-221.

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 (2000) Isochores and the evolutionary genomics of vertebrates. *Gene* **241:** 3-17.

Bernardi G, Olofsson B, Filipski J *et al.* (1985) The mosaic genome of warm-blooded vertebrates. *Science* **228:** 953-958.

Burt DW (2002) Origin and evolution of avian microchromosomes. *Cytogenet Genome Res* **96:** 97-112.

Cao Y, Sorenson MD, Kumazawa Y, Mindell DP, Hasegawa M (2000) Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. *Gene* **259:** 139-148.

Clay O, Caccio S, Zoubak S, Mouchiroud D, Bernardi G (1996) Human coding and noncoding DNA: compositional correlations. *Mol Phylogenet Evol* **5:** 2-12.

Cohen MM, Gans C (1970) The chromosomes of the order Crocodilia. *Cytogenetics* **9:** 81-105.

De Boer LEM, Sinoo RP (1984) A karyological study of Accipitridae (Aves: Falconiformes), with karyotypic description of 16 species new to cytology. *Genetica* **65:** 89-107.

Dehal P, Satou Y, Campbell RK *et al.* (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298:** 2157-2167.

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

Felsenstein, J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17:** 368-376.

Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* **85:** 8998-9002.

Goldman, N, Yang, Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11:** 725-736.

Graves, JA, Westerman, M (2002) Marsupial genetics and genomics. *Trends Genet* **18:** 517-521.

Grützner F, Deakin J, Rens W, El-Mogharbel N, Marshall Graves JA (2003) The monotreme genome: a patchwork of reptile, mammal and unique features? *Comp Biochem Physiol A Mol Integr Physiol* **136:** 867-881.

Guttenbach M, Nanda I, Brickell PM *et al.* (2000) Chromosomal localization of the genes encoding ALDH, BMP-2, R-FABP, IFN-gamma, RXR-gamma, and VIM

in chicken by fluorescence in situ hybridization. *Cytogenet Cell Genet* **88:** 266-271.

Hamada K, Horiike T, Ota H, Mizuno K, Shinozawa T (2003) Presence of isochore structures in reptile genomes suggested by the relationship between GC contents of intron regions and those of coding regions. *Genes Genet Syst* **78:** 195-198.

Hedges SB, Poling LL (1999) A molecular phylogeny of reptiles. *Science* **283:** 998-1001.

Holmquist GP (1989) Evolution of chromosome bands: molecular ecology of noncoding DNA. *J Mol Evol* **28:** 469-486.

Hubbard T, Andrews D, Caccamo M *et al.* (2005) Ensembl 2005. *Nucleic Acids Res* **33:** D447-453.

Hughes S, Clay O, Bernardi G (2002) Compositional patterns in reptilian genomes. *Gene* **295:** 323-329.

Hughes S, Zelus D, Mouchiroud D (1999) Warm-blooded isochore structure in Nile crocodile and turtle. *Mol Biol Evol* **16:** 1521-1527.

International Chicken Genome Sequencing Consortium (ICGSC) (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432:** 695-716.

International Human Genome Sequence Consortium (IHGSC) (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921.

Iwabe N, Hara Y, Kumazawa Y *et al.* (2005) Sister group relationship of turtles to the bird-crocodilian clade revealed by nuclear DNA-coded proteins. *Mol Biol Evol* **22:** 810-813.

Jaillon O, Aury JM, Brunet F *et al.* (2004) Genome duplication in the teleost fish

Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431:** 946-957.

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.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30:** 3059-3066.

Kumazawa Y, Nishida M (1999) Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. *Mol Biol Evol* **16:** 784-792.

Matsuda Y, Chapman VM (1995) Application of fluorescence in situ hybridization in genome analysis of the mouse. *Electrophoresis* **16:** 261-272.

Matsuda Y, Nishida-Umehara C, Tarui H *et al.* (2005) Highly conserved linkage homology between birds and turtles: Bird and turtle chromosomes are precise counterparts of each other. *Chromosome Res* **13:** 601-615.

McQueen HA, Fantes J, Cross SH, Clark VH, Archibald AL, Bird AP (1996) CpG islands of chicken are concentrated on microchromosomes. *Nat Genet* **12:** 321-324.

McQueen HA, Siriaco G, Bird AP (1998) Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Res* **8:** 621-630.

Mouse Genome Sequencing Consortium (MGSC) (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520-562.

Miyata T, Yasunaga T (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from

homologous nucleotide sequences and its application. *J Mol Evol* **16:** 23-36.

Mouchiroud D, Gautier C, Bernardi G (1988) The compositional distribution of coding sequences and DNA molecules in humans and murids. *J Mol Evol* **27:** 311-320.

Musto H, Romero H, Zavala A, Bernardi G (1999) Compositional correlations in the chicken genome. *J Mol Evol* **49:** 325-329.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418-426.

Norris TB, Rickards GK, Daugherty CH (2004) Chromosomes of tuatara, *Sphenodon*, a chromosome heteromorphism and an archaic reptilian karyotype. *Cytogenet Genome Res* **105:** 93-99.

Phillips MJ, Penny D (2003) The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol Phylogenet Evol* **28:** 171-185.

Rest JS, Ast JC, Austin CC *et al.* (2003) Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol Phylogenet Evol* **29:** 289-297.

Rat Genome Sequencing Project Consortium (RGSPC) (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428:** 493-521.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4:** 406-425.

Schmid M, Nanda I, Guttenbach M *et al.* (2000) First report on chicken genes and chromosomes 2000. *Cytogenet Cell Genet* **90:** 169-218.

Smith J, Bruley CK, Paton IR *et al.* (2000) Differences in gene density on chicken macrochromosomes and microchromosomes. *Anim Genet* **31:** 96-103.

Stajich JE, Block D, Boulez K *et al.* (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* **12:** 1611-1618.

Suzuki T, Kurosaki T, Shimada K *et al.* (1999) Cytogenetic mapping of 31 functional genes on chicken chromosomes by direct R-banding FISH. *Cytogenet Cell Genet* **87:** 32-40.

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

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13:** 555-556.

Zardoya R, Meyer A (1998) Complete mitochondrial genome suggests diapsid affinities of turtles. *Proc Natl Acad Sci USA* **95:** 14226-14231.

Zardoya R, Meyer A (2001) The evolutionary position of turtles revised. *Naturwissenschaften* **88:** 193-200.

Figure legends

Figure 1. Overview of intra-genome GC-content in chordates revealed by whole genome sequencing. (**a**) Distribution of GC-content in non-overlapping 20 kb windows for human (blue), mouse (yellow), chicken (red), tiger pufferfish (*Fugu rubripes*) (green), and tunicate (*Ciona intestinalis*) (grey). For simplicity, we excluded data for rat in which intra-genome GC heterogeneity highly resembles that of mouse (RGSPC 2004). (**b**-**e**) Chromosomal length and global GC-content. Overall GC-content (%) and chromosome length are plotted for chicken (**b**), human (**c**), mouse (**d**), and green pufferfish (*Tetraodon nigroviridis*; Jaillon *et al.* 2004) (**e**). Chromosomal GC-content was calculated as the proportion of guanine or cytosine within the length of nucleotide sequence that had already been determined. Chromosomes with less than 70% sequencing coverage were excluded from the analysis. Note that the horizontal axes are not to equal scale. Genomic sequences sorted by chromosomes and estimated chromosome lengths were retrieved from Ensembl.

Figure 2. Distribution of GC-content in exonic third positions (GC_3) in various chordates. Distribution of GC_3 are shown as histograms for genes of the Chinese soft-shelled turtle (*P. sinensis*) (n = 125) (**a**); chicken (n = 8380) (**b**); human (n = 23636) and sheep (*Ovis aries*) (n = 1436) (**c**); mouse (n = 22376) and rat (n = 21009) (**d**); *Xenopus tropicalis* (n = 1815) and axolotl (*Ambystoma mexicanum*) (n = 4257) (**e**); zebrafish (*Danio rerio*) (n = 3653) and tiger pufferfish (*Fugu rubripes*) (n = 23285) (**f**); amphioxus (*Branchiostoma belcheri*) (n = 199) and tunicate (*Ciona intestinalis*) (n = 12588) (**g**); and *Drosophila melanogaster* (n = 12862) and *Caenorhabditis elegans* (n =

16341) (**h**).

Figure 3. Cross-species comparison of GC_3 between orthologs. Two-dimensional plots of GC_3 for 56 ortholog pairs are shown for turtle-chicken (**a**) and turtle-human (**b**). Two-dimensional plots of GC_3 comparison for *P. sinensis* and *Xenopus tropicalis* (**c**) and human and *X. tropicalis* (**d**) are also shown for 41 genes whose *X. tropicalis* orthologs were found in the Ensembl (URL: http://www.ensembl.org/).

Figure 4. Distribution of GC-content on macrochromosomes and microchromosomes. Distribution of global GC-content of genomic sequences on macrochromosomes and microchromosomes of chicken is shown as a histogram in 20 kb non-overlapping windows (**a**); distribution of GC-content in exonic third positions (GC_3) for 90 *P. sinensis* genes (**b**); and distribution of GC_3 for 8380 chicken cDNAs (**c**).

Figure 5. GC-content of turtle and chicken genes in relation to their location on macrochromosomes and microchromosomes. Circle graphs representing the relative proportion of genes located on macrochromosomes and microchromosomes are shown for all genes analyzed (**a**); GC-rich genes ($GC_3 \geq 50\%$) (**b**); and GC-poor genes ($GC_3 < 50\%$) (**c**). Relative proportions of GC-rich and GC-poor genes out of those located on macrochromosomes (**d**) and microchromosomes (**e**) are also shown.

Figure 6. Correlation of GC-content in exonic third positions (GC_3) and surrounding genomic regions in chicken. Two-dimensional plots of GC_3 for chicken cDNAs and GC-contents of the genomic regions (10 kb on each side) in which the gene is located

are shown for 53 genes whose flanking genomic sequences were found in the Ensembl Chicken Genome Server (URL: http://www.ensembl.org/Gallus_gallus/), out of 56 genes used in this study.

Figure 7. Schematic representation of chromosomal evolution and transition of base composition in amniote phylogeny. Hypothesized evolutionary model for macrochromosome and microchromosome karyotypic configuration and intra-genome GC heterogeneity is illustrated in accordance with phylogenetic relationships revealed by recent molecular phylogenetic analyses (Phillips & Penny 2003; Rest *et al.* 2003; Iwabe *et al.* 2005). The intensity of monochrome tone in the chromosome image represents the relative extent of intra-chromosomal GC-content. Lepidosaurians tend to have microchromosomes (Norris *et al.* 2004) as do most other sauropsids; however, their GC profiles have not yet been obtained. Marsupials possess only large chromosomes (Graves & Westerman 2002), and thus are expected to show a eutherian mode of intra-genome base composition. Karyotypes of monotremes are denoted as a 'patchwork' of other mammals and reptiles because of the co-existence of large and small chromosomes (Grützner *et al.* 2003). However, there are no reports on GC profiles in these non-eutherian mammals. See text for details about rodents and crocodiles.

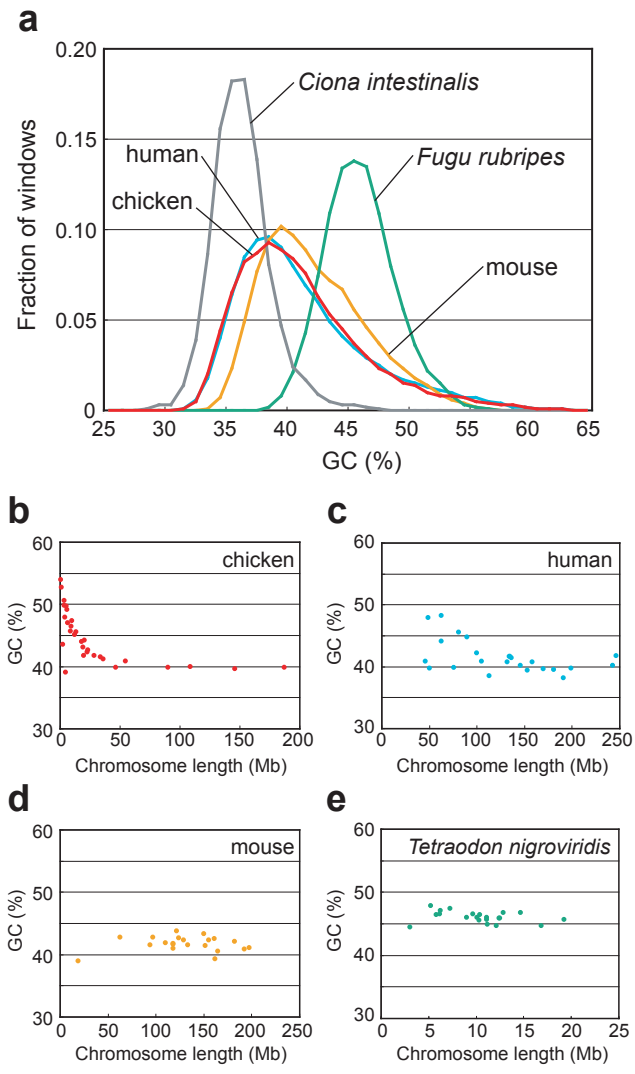
a
0.20
0.15
0.10
0.05
0
Fraction of windows
Ciona intestinalis
human
chicken
Fugu rubripes
mouse
25 30 35 40 45 50 55 60 65
GC (%)
b
chicken
GC (%)
60 50 40 30
0 50 100 150 200
Chromosome length (Mb)
c
human
GC (%)
60 50 40 30
0 50 100 150 200 250
Chromosome length (Mb)
d
mouse
GC (%)
60 50 40 30
0 50 100 150 200 250
Chromosome length (Mb)
e
Tetraodon nigroviridis
GC (%)
60 50 40 30
0 5 10 15 20 25
Chromosome length (Mb)

Figure 1

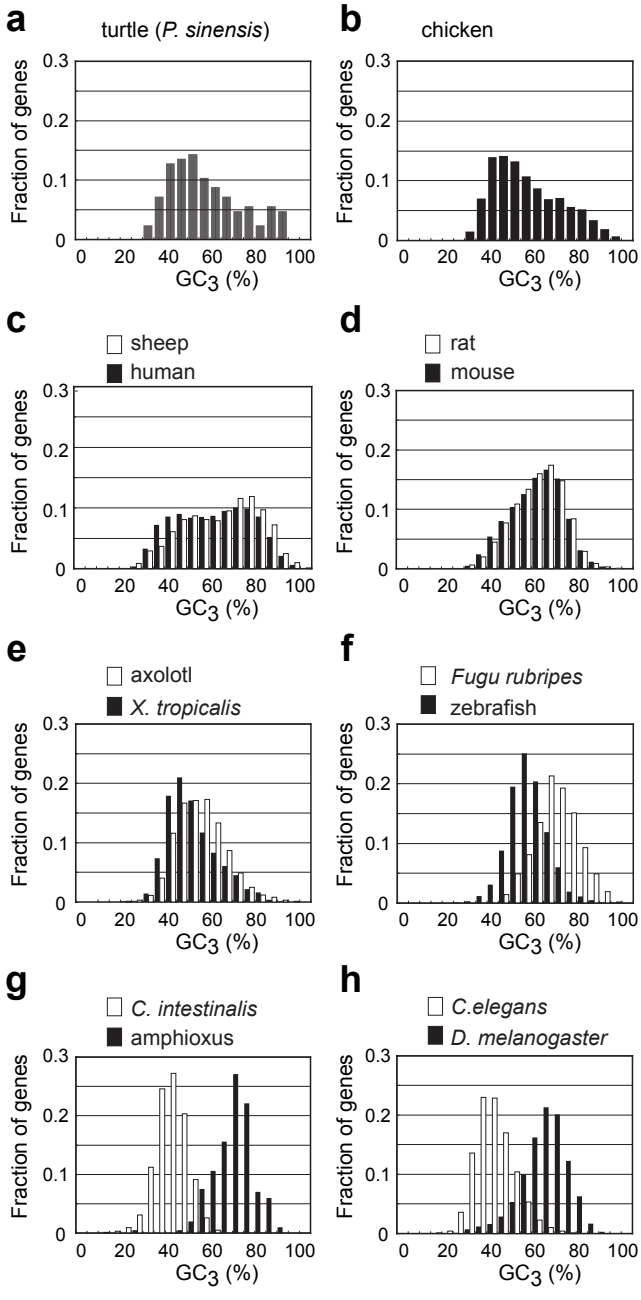
a
turtle (P. sinensis)
b
chicken
c
sheep
human
d
rat
mouse
e
axolotl
X. tropicalis
f
Fugu rubripes
zebrafish
g
C. intestinalis
amphioxus
h
C.elegans
D. melanogaster
Fraction of genes
0
0.1
0.2
0.3
0 20 40 60 80 100
GC3 (%)

Figure 2

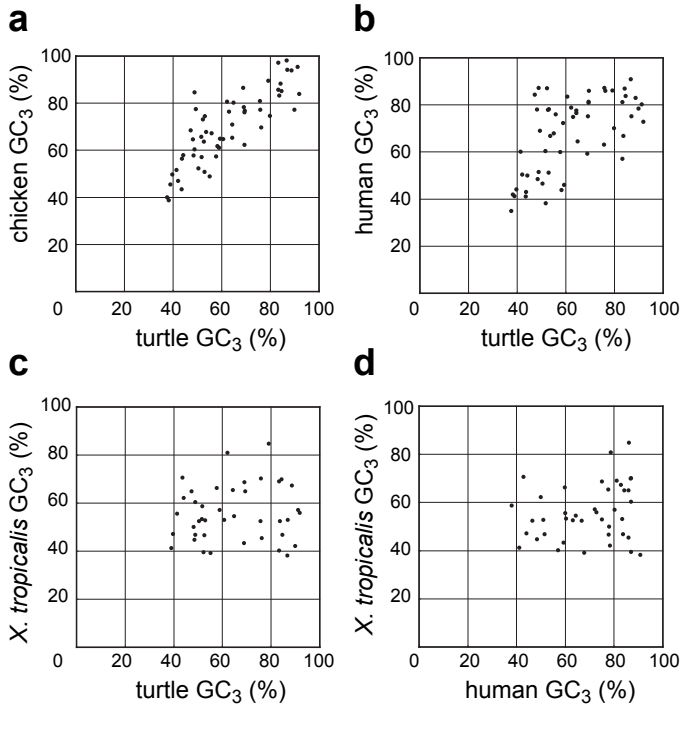
a
b
c
d
chicken GC3 (%)
human GC3 (%)
X. tropicalis GC3 (%)
X. tropicalis GC3 (%)
turtle GC3 (%)
turtle GC3 (%)
turtle GC3 (%)
human GC3 (%)
0
20
40
60
80
100

Figure 3

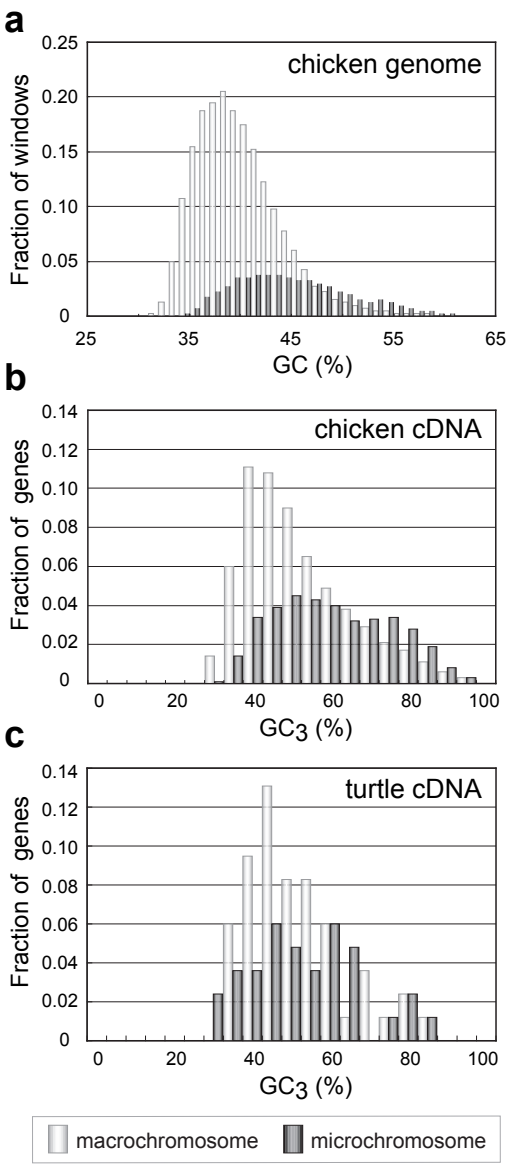
a
chicken genome
Fraction of windows
0.25
0.20
0.15
0.10
0.05
0
25
35
45
55
65
GC (%)
b
chicken cDNA
Fraction of genes
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
0
20
40
60
80
100
GC3 (%)
c
turtle cDNA
Fraction of genes
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
0
20
40
60
80
100
GC3 (%)
macrochromosome
microchromosome

Figure 4

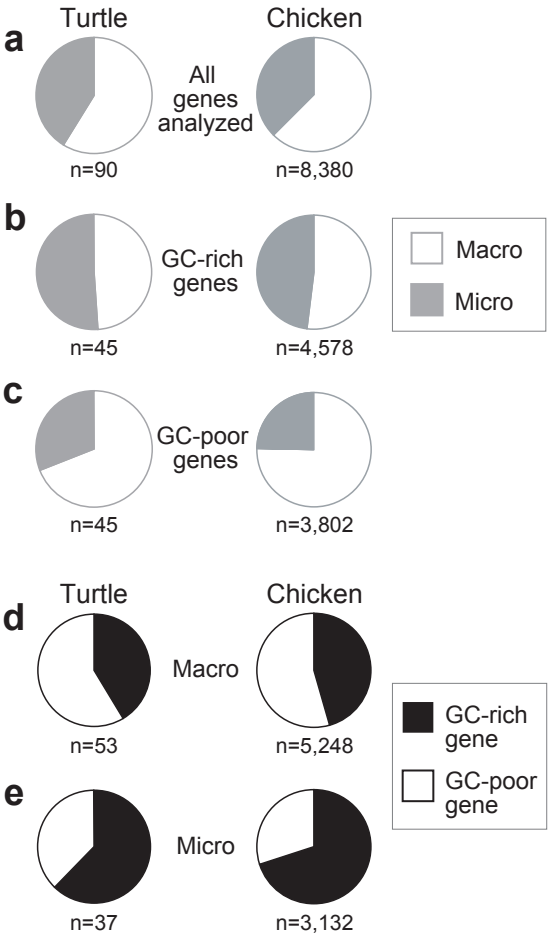
a
Turtle
Chicken
All genes analyzed
n=90
n=8,380
b
GC-rich genes
n=45
n=4,578
Macro
Micro
c
GC-poor genes
n=45
n=3,802
d
Turtle
Chicken
Macro
n=53
n=5,248
GC-rich gene
GC-poor gene
e
Micro
n=37
n=3,132

Figure 5

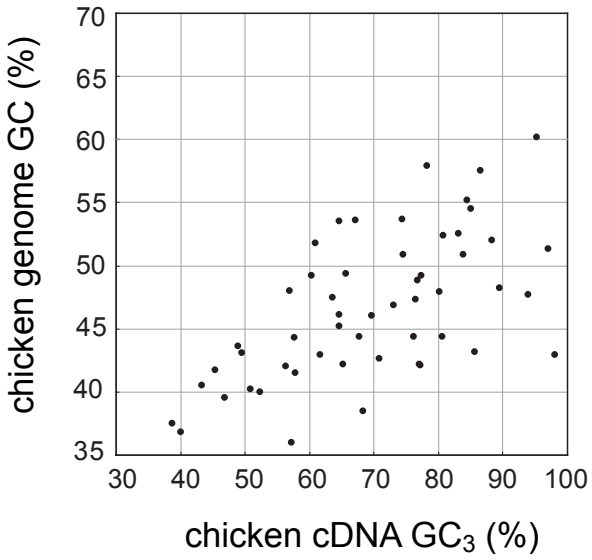
chicken genome GC (%)
70
65
60
55
50
45
40
35
30
40
50
60
70
80
90
100
chicken cDNA GC3 (%)

Figure 6

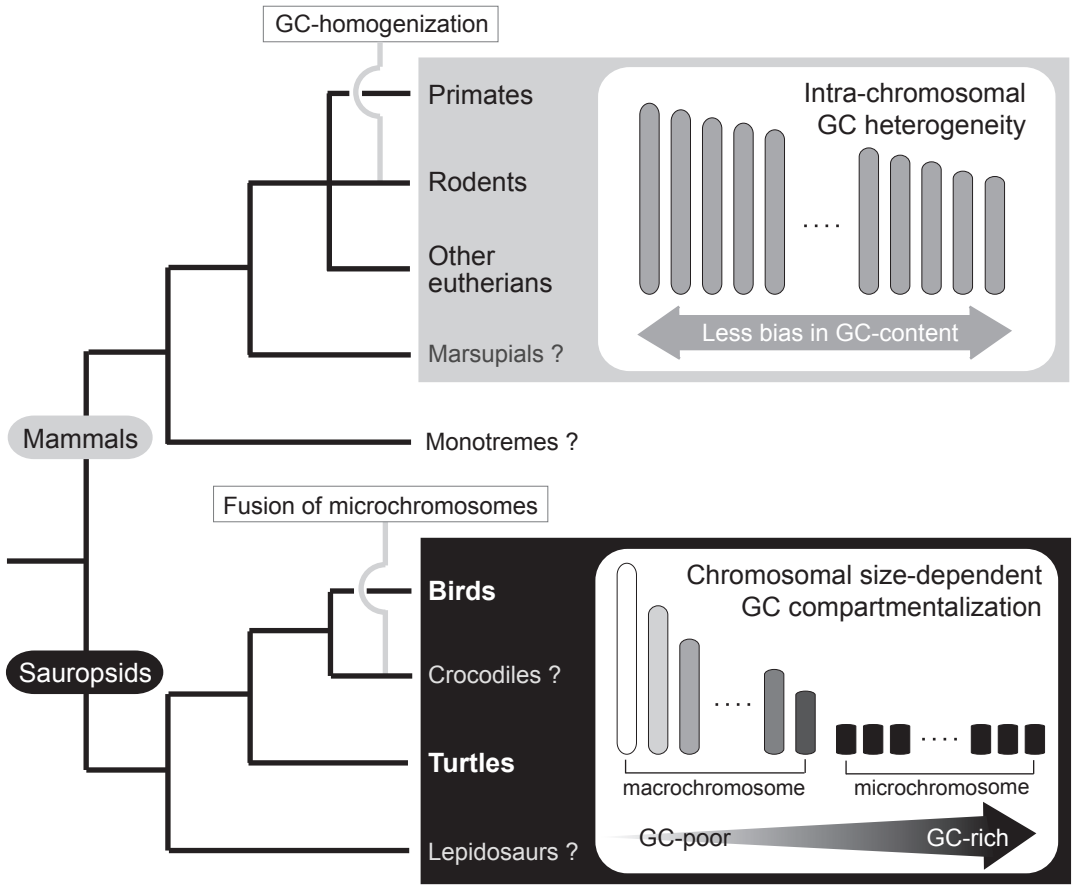
GC-homogenization
Primates
Rodents
Other eutherians
Marsupials ?
Intra-chromosomal GC heterogeneity
Less bias in GC-content
Mammals
Monotremes ?
Fusion of microchromosomes
Birds
Crocodiles ?
Turtles
Lepidosaurs ?
Sauropsids
Chromosomal size-dependent GC compartmentalization
macrochromosome
microchromosome
GC-poor
GC-rich

Figure 7

Table 1. Degenerate oligonucleotide primers used for isolation of unidentified cDNAs

Gene[a]	Forward primer (5´-3´)	Reverse primer (5´-3´)
HOXD13	CACTTCGGSAAYGGNTAYTAYWSNTG	CGCTTCATNCKNCKRTTYTGRAACCA
PRRX1 (Prx1)	GCGCAAGCTMGNAARAAYTTYWSNGT	CGGTTCGCDATNSWRTTNGCCATRTT
EN1	TGGCCAGCNTGGGTNTAYTGYACNMG	GTGCTGTGGTTRTANARNCCYTGNGC
EMX2	TACACGAAYCCNGAYYTNGTNTTYGC	GGRAACCANACYTTNACYTGNGTYTC
PAX3	GAARATHGTNGARATGGCNCAYCAYG	CCGGGGDATNARRTGRTTRAANGCCAT
PAX7	GAARATHGTNGARATGGCNCAYCAYG	CCTGGTAANARRTGRTTRAANGCNGC
TCF7L2 (TCF4)	GGAAYGCNTTYATGYTNTAYATGAARGA	TTACCATAGTTRTCNCKNGCNSWCCA
RARA/RARG[b]	GAACAAGGTAACNMGNAAYMGNTGYCA	CCWGCRTTRTGCATYTGNGTNCKRTT
RARB	CGCGTNTAYAARCCNTGYTTYGTNTG	GTCAARTCRTCNARYTCNGCNGTCA
MYOD1 (MyoD)	GGGCTTGYAARGCNTGYAARMGNAAR	GCGTTCCTAARDATNTCNACYTTNGG
GLI2	GAGAARAARGARTTYGTNTGYMGNTGG	TCGAGATAGGRTCRTANSWRTCNGC
GLI3	GAGAARAARGARTTYGTNTGYMGNTGG	GACCRTGNACNGTYTTNACRTGYTT
SHH/IHH/DHH[b]	TNACNGARGGNTGGGAYGARGAYG	TAGTRNACCCARTCRAANCCNGCYTC
WNT2B/4/11[b]	TAAATGTCACGGAGTAWSNGGNWSNTG	GACNNARCARCACCARTGRAAYTTRCA
WNT3A	GGNATHCARGARTGYCAYCARCAYTT	CCACAACATAGNARNTYRCANCCRTC
WNT5A	GGTGCNAARACNGGNATHAARGARTG	GACNNARCARCACCARTGRAAYTTRCA
WNT7A	TAGGAGAGMGNACNGTNTTYGGNA	GCGCWRCANGTRTTRCAYTTNACRTA
WNT8A	TTCCTGATWACNGGNCCNAARGCNTA	GACNNARCARCACCARTGRAAYTTRCA
BMP2/4/7[b]	GTGCCGCCNTAYATGYTNGAYYTNTA	TCDATCCARTCRTTCCANCCNACRTC
LFNG/RFNG/MFNG[b]	GACGTGTTYATHGCNGTNAARACNAC	TGCAGGTTCTCNARRTGNSWRTGRAA
FGFR1/2/4[b]	CCTTTWGGNGARGGNTGYTTYGGNCA	GAGCCNCCNARNGTRAADATYTCCCA
CDH1 (E-cadherin)	TACGNGAYTGGGTNATHCCNCCNAT	GATGAARTTNCCDATYTCRTCNGGRTT
TWIST2 (dermo-1)	CAGGTGYTNCARWSNGAYGARATG	CACCTTCCATNCKCCANACNSWRAA
FOXA1 (HNF3α)	GAGCNGTNAARATGGARGGNCAY	GGTARCANCCRTTYTCRAACATRTTNC

[a]Gene names were indicated as gene symbols designated for human orthologs. Famous aliases for gene names are also added in parenthesis. [b]A single set of degenerate primers were used for amplification of multiple phylogenetically close paralogs. Degenerative nucleotides are shown following IUB code.

Table 2. List for cDNAs newly isolated from *P. sinensis* in this study.

Gene symbol[a]	Gene name	Accession Number
HOXD13	homeobox D13	AB188346
PRRX1	paired related homeobox 1	AB188347
EN1	engrailed homolog 1	AB188348
EMX2	empty spiracles homolog 2	AB188349
PAX3	paired box gene 3	AB188350
PAX7	paired box gene 7	AB188351
TCF7L2	transcription factor 7-like2 (T-cell factor 4)	AB188352
RARA	retinoic acid receptor, alpha	AB188353
RARB	retinoic acid receptor, beta	AB188354
RARG	retinoic acid receptor, gamma	AB188355
MYOD1	myogenic factor 3 (myoD)	AB188356
GLI2	GLI-Kruppel family member GLI2	AB188357
GLI3	GLI-Kruppel family member GLI3	AB188358
IHH	Indian hedgehog	AB188359
WNT2B	wingless-type MMTV integration site family, member 2B	AB188360
WNT3A	wingless-type MMTV integration site family, member 3A	AB188361
WNT4	wingless-type MMTV integration site family, member 4	AB188362
WNT5A	wingless-type MMTV integration site family, member 5A	AB188363
WNT7A	wingless-type MMTV integration site family, member 7A	AB188364
WNT8A	wingless-type MMTV integration site family, member 8A	AB188365
WNT11	wingless-type MMTV integration site family, member 11	AB188366
BMP7	bone morphogenetic protein 7	AB188367
LFNG	lunatic fringe homolog	AB188368
RFNG	radical fringe homolog	AB188369
MFNG	manic fringe homolog	AB188370
FGFR1	fibroblast growth factor receptor 1	AB188371
FGFR2	fibroblast growth factor receptor 2	AB188372
FGFR4	fibroblast growth factor receptor 4	AB188373
CDH1	cadherin 1, type 1 (E-cadherin)	AB188374
THOC2	THO complex 2	AB188375
RPL23	ribosomal protein L23	AB188376
RPS20	ribosomal protein S20	AB188377
RPL30	ribosomal protein L30	AB188378
TWIST2	twist-related dermis-expressed protein 1 (Dermo-1)	AB188379
SKI	v-ski sarcoma viral oncogene homolog (c-ski)	AB188380
TCF7L1	transcription factor 7-like 1 (TCF3)	AB188381
FOXA1	forkhead box A1 (HNF3alpha)	AB188382
DHH	desert hedgehog	AB188383
DDX46	DEAD box polypeptide 46 (Prp5-like)	AB188384

[a]Gene names were indicated as gene symbols designated for human orthologs.

Table 3. Estimated number of synonymous and non-synonymous substitutions in orthologous gene pairs of turtle and chicken.

Gene symbol[a]	Accession number for *P. sinensis*	Aligned length (nt)	ML		Nei-Gojobori[d]	
			Ka	*Ks*	*Ka*	*Ks*
HOXD13	AB188346[b]	615	0.02	0.92	0.02	0.78
MSX1	AB124572	735[c]	0.04	1.01	0.05	0.54
MSX2	AB181139	594	0.02	0.60	0.02	0.53
PRRX1	AB188347[b]	540	0.02	1.17	0.02	0.79
EN1	AB188348[b]	312	0.01	2.65	0.02	0.61
EMX2	AB188349[b]	540	0.01	0.53	0.01	0.54
PAX3	AB188350[b]	1254	0.01	0.55	0.01	0.59
PAX7	AB188351[b]	1308	0.01	0.82	0.01	0.68
PAX9	AB181136	714	0.11	1.29	0.12	0.98
LEF1	AB124566	1107[c]	0.02	0.47	0.02	0.55
TCF7L2	AB188352[b]	303	0.00	0.19	0.00	0.21
RARA	AB188353[b]	804	0.01	1.83	0.02	0.44
RARB	AB188354[b]	1092	0.01	0.35	0.01	0.40
RARG	AB188355[b]	843	0.04	1.94	0.05	0.54
MYOD1	AB188356[b]	600	0.05	1.02	0.06	0.64
GLI2	AB188357[b]	1125	0.04	0.68	0.04	0.72
GLI3	AB188358[b]	324	0.00	0.41	0.00	0.42
SP5	AB124563	1098[c]	0.04	1.44	0.05	0.50
SHH	AB181135	564	0.01	0.49	0.01	0.33
IHH	AB188359[b]	477	0.06	1.15	0.07	0.41
FGF8	AB124574	465	0.02	0.43	0.02	0.23
FGF10	AB124573	579[c]	0.02	0.34	0.02	0.38
WNT2B	AB188360[b]	936	0.02	1.54	0.02	0.71
WNT3A	AB188361[b]	813	0.01	0.99	0.01	0.81
WNT4	AB188362[b]	408	0.03	1.06	0.03	0.52
WNT5A	AB188363[b]	828	0.02	1.00	0.02	0.91
WNT7A	AB188364[b]	822	0.00	0.81	0.00	0.93
WNT8A	AB188365[b]	936	0.07	1.68	0.09	1.08
WNT11	AB188366[b]	408	0.01	0.97	0.02	0.79
BMP2	AB181137	651	0.07	0.34	0.07	0.40
BMP4	AB181138	885	0.08	0.92	0.09	0.57
BMP7	AB188367[b]	663	0.01	1.10	0.01	1.40
LFNG	AB188368[b]	771	0.02	1.42	0.03	0.94
RFNG	AB188369[b]	771	0.04	0.67	0.05	0.80
MFNG	AB188370[b]	657	0.12	0.96	0.13	0.77
FGFR1	AB188371[b]	597	0.05	1.55	0.06	0.48
FGFR2	AB188372[b]	981	0.00	0.53	0.00	0.63
FGFR4	AB188373[b]	750	0.04	1.79	0.06	0.57
APCDD1	AB124565	1545[c]	0.05	0.87	0.05	1.06
CRABP1	AB124564	411[c]	0.02	0.46	0.02	0.46
CTNNB1	AB124575	2346[c]	0.00	0.55	0.00	0.61
CDH1	AB188374[b]	1917	0.19	3.31	0.23	0.75
GAPD	AB124567	1023[c]	0.06	0.65	0.06	0.63
EEF1A1	AB124568	1386[c]	0.00	0.58	0.00	0.61
LDHA	AF363794	996[c]	0.06	0.63	0.06	0.75
LDHB	AF363795	1002[c]	0.05	0.79	0.05	0.84
THOC2	AB188375[b]	891[c]	0.01	0.65	0.01	0.76
PLP1	AF369033	555[c]	0.04	0.90	0.05	0.55
RAG2	AF369089	1476[c]	0.11	0.86	0.05	0.55
GHR	AF211173	1806[c]	0.16	0.60	0.15	0.65
TSHB	AY618874	402[c]	0.11	0.94	0.11	0.98
PRNP	AB088368	765[c]	0.39	1.87	0.39	1.76
TYR	AB024280	1569[c]	0.10	0.84	0.10	0.91
RPL23	AB188376[b]	315[c]	0.00	0.78	0.00	0.74
RPL30	AB188378[b]	345[c]	0.00	0.53	0.00	0.61
RPS20	AB188377[b]	354[c]	0.00	0.57	0.00	0.64

Sequences of chicken orthologs were retrieved from Ensembl. [a]Gene symbols designated for human orthologs are indicated. [b]Sequenced in this study. [c]Contains a whole coding sequence. [d]Numbers of synonymous and non-synonymous substitutions were estimated with the method of Nei and Gojobori (1985). ML, maximum likelihood.

Table 4. Gene mapping on *P. sinensis* chromosomes

Gene symbol[a]	*P. sinensis*		Chromosomal location of orthologs		
	Acc. No.	Location	Chicken	Human[g]	Mouse[g]
MFNG	AB188370[b]	1p	1[f]	22q13.1	15qE2
WNT11	AB188366[b]	1q	1[d]	11q13.5	7qF1
RARB	AB188354[b]	2p	2[f]	3p24.2	14qA2
SHH	AB181135	2p	2p[c]	7q36.3	5qA3
APCDD1	AB124565	2q	2q[c]	18q11.22	18qD3
BMP2	AB181137	3p	3[e]	20p12.3	2qF3
MSX1	AB124572	4q	4[f]	4p16.2	5qB2
LEF1	AB124566	4q	4q[c]	4q25	3qH1
MYOD1	AB188356[b]	5p	5q[d]	11p15.1	7qB3
BMP4	AB181138	5q	5[e]	14q22.2	14qC1
PAX9	AB181136	5q	5q[c]	14q13.3	12qC2
PRRX1	AB188347[b]	5q	8[f]	1q32.3	1qH1
FGF10	AB124573	6q	Z[f]	5p12	13qD23
HOXD13	AB188346[b]	MIC(7q)	7q[c]	2q31.1	2qC3
SP5	AB124563	MIC(7q)	7q[c]	2q31.1	2qC3
EN1	AB188348[b]	MIC(7q)	7[f]	2q14.2	1qE2
GLI2	AB188357[b]	MIC(7q)	7q[c]	2q14.2	1qE2
FGF8	AB124574	MIC(8-10)	6[f]	10q24.32	19qD1
FGFR2	AB188372[b]	MIC(8-10)	6[f]	10q26.13	7 (63.0 cM)
EMX2	AB188349[b]	MIC(8-10)	6[f]	10q26.11	19qD3
PAX3	AB188350[b]	MIC(8-10)	9[f]	2q35-q37	1 (44.0 cM)
CRABP1	AB124564	MIC(11-)	10[f]	15q25.1	9qC
DDX46	AB188384[b]	MIC(11-)	13[f]	5q31.1	13 (B1)
WNT2B	AB188360[b]	MIC(11-)	26[f]	1p13.2	3qF3
SKI	AB188380[b]	MIC(11-)	21[f]	1p36.33	4qE2
PAX7	AB188351[b]	MIC(11-)	21[f]	1p36.13	4qD3
LFNG	AB188368[b]	MIC(11-)	14[f]	7p22.3	5qG1
TWIST2	AB188379[b]	MIC(11-)	7[f]	2q37.3	1 D
BMP7	AB188367[b]	MIC(11-)	20[f]	20q13.31	2qH3
WNT7A	AB188364[b]	MIC(11-)	12[f]	3p25.1	6qD2
WNT5A	AB188363[b]	MIC(11-)	12[f]	3p14.3	14qB

[a]Gene names were indicated as gene symbols designated for human orthologs. [b]Sequenced in this study. [c]Mapped with FISH in this study. [d]Suzuki *et al.* (1999). [e]Guttenbach *et al.* (2000). [f]Mapped *in silico* in this study. [g]Mapping information on mouse and human chromosomes were retrieved from NCBI Entrez Gene. MIC, microchromosome. Number of distinguishable microchromosomes and its arm (7q, 8-10, and 11-) are indicated in parentheses.