

RESEARCH ARTICLE

Open Access

Intra-genomic GC heterogeneity in sauropsids: evolutionary insights from cDNA mapping and GC₃ profiling in snake

Kazumi Matsubara^{1,7*}, Shigehiro Kuraku^{2,3}, Hiroshi Tarui^{3,8}, Osamu Nishimura^{3,4}, Chizuko Nishida⁵, Kiyokazu Agata⁴, Yoshinori Kumazawa¹ and Yoichi Matsuda⁶

Abstract

Background: Extant sauropsids (reptiles and birds) are divided into two major lineages, the lineage of Testudines (turtles) and Archosauria (crocodilians and birds) and the lineage of Lepidosauria (tuatara, lizards, worm lizards and snakes). Karyotypes of these sauropsidan groups generally consist of macrochromosomes and microchromosomes. In chicken, microchromosomes exhibit a higher GC-content than macrochromosomes. To examine the pattern of intra-genomic GC heterogeneity in lepidosaurian genomes, we constructed a cytogenetic map of the Japanese four-striped rat snake (*Elaphe quadrivirgata*) with 183 cDNA clones by fluorescence *in situ* hybridization, and examined the correlation between the GC-content of exonic third codon positions (GC₃) of the genes and the size of chromosomes on which the genes were localized.

Results: Although GC₃ distribution of snake genes was relatively homogeneous compared with those of the other amniotes, microchromosomal genes showed significantly higher GC₃ than macrochromosomal genes as in chicken. Our snake cytogenetic map also identified several conserved segments between the snake macrochromosomes and the chicken microchromosomes. Cross-species comparisons revealed that GC₃ of most snake orthologs in such macrochromosomal segments were GC-poor (GC₃ < 50%) whereas those of chicken orthologs in microchromosomes were relatively GC-rich (GC₃ ≥ 50%).

Conclusion: Our results suggest that the chromosome size-dependent GC heterogeneity had already occurred before the lepidosaur-archosaur split, 275 million years ago. This character was probably present in the common ancestor of lepidosaurs and but lost in the lineage leading to *Anolis* during the diversification of lepidosaurs. We also identified several genes whose GC-content might have been influenced by the size of the chromosomes on which they were harbored over the course of sauropsid evolution.

Keywords: GC-content, Lepidosauria, Snake, Macrochromosome, Microchromosome

Background

Molecular phylogenetic analyses have suggested that extant sauropsids (reptiles and birds) are divided into two major groups, the lineage of Testudines (turtles) and Archosauria (crocodilians and birds) and the lineage of Lepidosauria (tuatara, lizards, worm lizards and snakes)

although phylogenetic position of Testudines is still debatable [1-7]. The divergence time between the two lineages has been estimated at around 275 million years [3,7-9]. Most sauropsidan species have karyotypes consisting of macrochromosomes and microchromosomes, as for birds [10-17], except for crocodilian species, whose karyotypes contain no microchromosomes [18,19].

Whole genome sequencing of chicken revealed that the overall GC-content of chromosomes increases as chromosomal size decreases, that is, microchromosomes exhibit a higher GC-content than macrochromosomes

* Correspondence: kazumi.matsubara@canberra.edu.au

¹Department of Information and Biological Sciences, Graduate School of Natural Sciences, Nagoya City University, 1 Yamanohata, Mizuho-cho, Mizuho-ku, Nagoya, Aichi 467-8501, Japan

⁷Present address: Institute for Applied Ecology, University of Canberra, Canberra, ACT 2601, Australia

Full list of author information is available at the end of the article

[20,21]. In a compositional map of GC-content constructed by 100-kb window analysis for the chicken whole genome sequence, most microchromosomes were occupied by GC-rich DNA segments, whereas GC-poor segments were more common in macrochromosomes [22]. The differences of other features such as gene density, distribution in interphase nuclei and rate of nucleotide divergence were also identified between the two chromosomal groups of birds [23-27].

Reptiles are crucial taxon for tracking genome evolution in amniotes [21,28,29]. Intra-genomic GC heterogeneity has been found in reptiles by calculating GC-content in exonic third positions (GC_3) [21,30-33]. Although the use of GC_3 as a proxy for genomic GC-content has been controversial [34], it is known that GC_3 generally reflects the local GC-content of the introns and flanking regions of a gene [21,35-37]. Chojnowski et al. [32] analyzed the GC_3 of more than 6,000 ESTs in the American alligator (*Alligator mississippiensis*) and suggested that the alligator genome has a certain level of GC heterogeneity. They also examined the isochore structure of the red-eared slider turtle (*Trachemys scripta*) and suggested that the isochore structure of the turtle is intermediate between that of a frog and the GC-rich isochore structures of archosaurs and mammals [33]. However, the chromosomal distribution of the GC heterogeneity has not been fully investigated in reptiles.

We previously constructed a cytogenetic map with 90 cDNA clones for the Chinese soft-shelled turtle (*Pelodiscus sinensis*), which revealed that the chromosomes have been highly conserved between the turtle and chicken, with the six largest chromosomes being almost equivalent to each other [38]. GC_3 of the mapped genes showed a heterogeneous distribution, and orthologs exhibited similar GC_3 levels between the turtle, chicken and human, suggesting that the intra-genome GC heterogeneity had already occurred in the last common ancestor of extant amniotes [21]. Furthermore, our results suggested that the turtle microchromosomes tend to contain more GC-rich genes than GC-poor genes, as in chicken [21].

The green anole lizard (*Anolis carolinensis*) is the first reptilian species for which whole genomic sequence has been released [39]. *Anolis* has a homogeneous genome composition compared with other amniotes [37,39] and, unlike chicken, the GC-content is similar between macro- and microchromosomes [39]. However, it remains unknown whether these genomic characteristics are common to other lepidosaurs or not. Snake karyotypes have been highly conserved within the group, and the usual diploid number is $2n = 36$, consisting of eight pairs of macrochromosomes and 10 pairs of microchromosomes [10,40,41]. The chromosome number is largely different from the chicken karyotype ($2n=78$) because of

the remarkable difference in the number of microchromosomes. The snake therefore provides an ideal system for exploring changes in GC-content between macro- and microchromosomes over the course of sauropsid evolution.

Previously we constructed a cytogenetic map with 109 cDNA clones for the Japanese four-striped rat snake, *Elaphe quadrivirgata* (Serpentes, Colubridae) [38,42]. In this study, we have extended cDNA-based chromosome mapping of the snake genes and consequently constructed a cytogenetic map with a total of 183 genes. We compared GC_3 of the mapped snake genes with GC_3 of their orthologs of chicken, green anole lizard, Chinese soft-shelled turtle, human and *Xenopus tropicalis*. This highlighted the chromosome size-dependent GC heterogeneity in the snake genome and the shift of GC-content possibly caused by chromosome rearrangements during sauropsid evolution.

Methods

Selection of EST clones

A cytogenetic map with 109 cDNA clones was constructed in our previous study [33,36]. In the present study, we searched the snake EST clones isolated from the cDNA library constructed from brain [38], selected clones with significant similarity (E-value $< 2e^{-35}$) to human and/or chicken genes in BLASTX [43], and used them for chromosome mapping (Additional file 1 and Additional file 2).

Orthology assessment

We rigorously confirmed orthologies of the snake sequences to their homologs of other vertebrates by constructing molecular phylogenetic trees with the neighbor-joining method [44] using XCED in which the alignment algorithm MAFFT is implemented [45] and with the maximum-likelihood method using PhyML [46]. Sequence IDs of orthologs in six species (*Anolis carolinensis*, *Gallus gallus*, *Pelodiscus sinensis*, *Homo sapiens*, *Mus musculus* and *Xenopus tropicalis*) are included in Additional file 2. When multiple more than one potential ortholog was detected for a snake gene, we used the sequence with the greatest similarity to snake for cross-species comparison of GC-content.

Chromosome preparation and fluorescence in situ hybridization

Cell culture, preparation of R-banded chromosomes and fluorescence *in situ* hybridization (FISH) were performed as described previously [38,47]. Fibroblast cells derived from lung tissues of the Japanese four-striped rat snake were cultured and used for chromosome preparations. DNA probes were labeled by nick translation with biotin-16-dUTP (Roche Diagnostics). The

hybridized cDNA probes were reacted with goat anti-biotin antibodies (Vector Laboratories), and then stained with Alexa488-labeled donkey anti-goat IgG (Molecular Probes).

Calculation of GC-content

GC₃ and GC-content at fourfold degenerate sites (GC₄) were calculated using an original Perl script with the Bioperl module [48]. The calculation was automatically processed on the basis of the open reading frame identified by a pairwise alignment between a translated nucleotide sequence and amino acid sequences of orthologs using BLASTX [43]. When multiple alternative splicing variants were found for one gene, we used only the one that had the longest stretch of sequence homology with its orthologs of other species. We arbitrarily classified genes into GC-rich (GC₃ ≥ 50%) and GC-poor (GC₃ < 50%) genes based on the GC₃.

Identification of orthologous sequences in the Burmese python

We conducted nucleotide BLAST for whole genome shotgun sequence of Burmese python, *Python molurus bivittatus* [49] using rat snake ESTs as queries. We selected python sequences (consisting of exons, introns and flanking regions) that exhibited high similarities for rat snake ESTs (Additional file 3). We deduced the non-coding regions and the protein coding regions within each python genomic region using Wise2 program [50]. We then calculated GC-content of non-coding regions and GC₃ of coding regions.

Gene location in sequenced genomes

Chromosome locations of chicken, human and mouse orthologs were retrieved from Ensembl (version 56) [51] and NCBI Entrez Gene [52]. In this study, we categorized chicken chromosomes 1–8, Z and W as macrochromosomes, and the remaining chromosomes as microchromosomes. Recently the second version of the genome assembly and annotation of green anole lizard (*Anolis carolinensis*) was released in Ensembl (version 61) [51]. However, locations of about half of all *A. carolinensis* genes are at present annotated only at the scaffold level. We therefore did not compare the chromosome locations of orthologs between the snake and green anole lizard.

Comparison of gene characteristics between chicken macrochromosomes and microchromosomes

We classified the chicken genes into two groups, macrochromosomal genes and microchromosomal genes, and examined the over- and under-representation of gene functions between the two groups by FatiGO [53]. FatiGO detects over-represented functional categories of

Gene Ontology (GO), KEGG pathway, InterPro motif and Swissprot in either group between two gene lists using Fisher's exact test. The Ensembl IDs of all chicken genes whose chromosome locations are known were downloaded and used for comparison by FatiGO.

Results

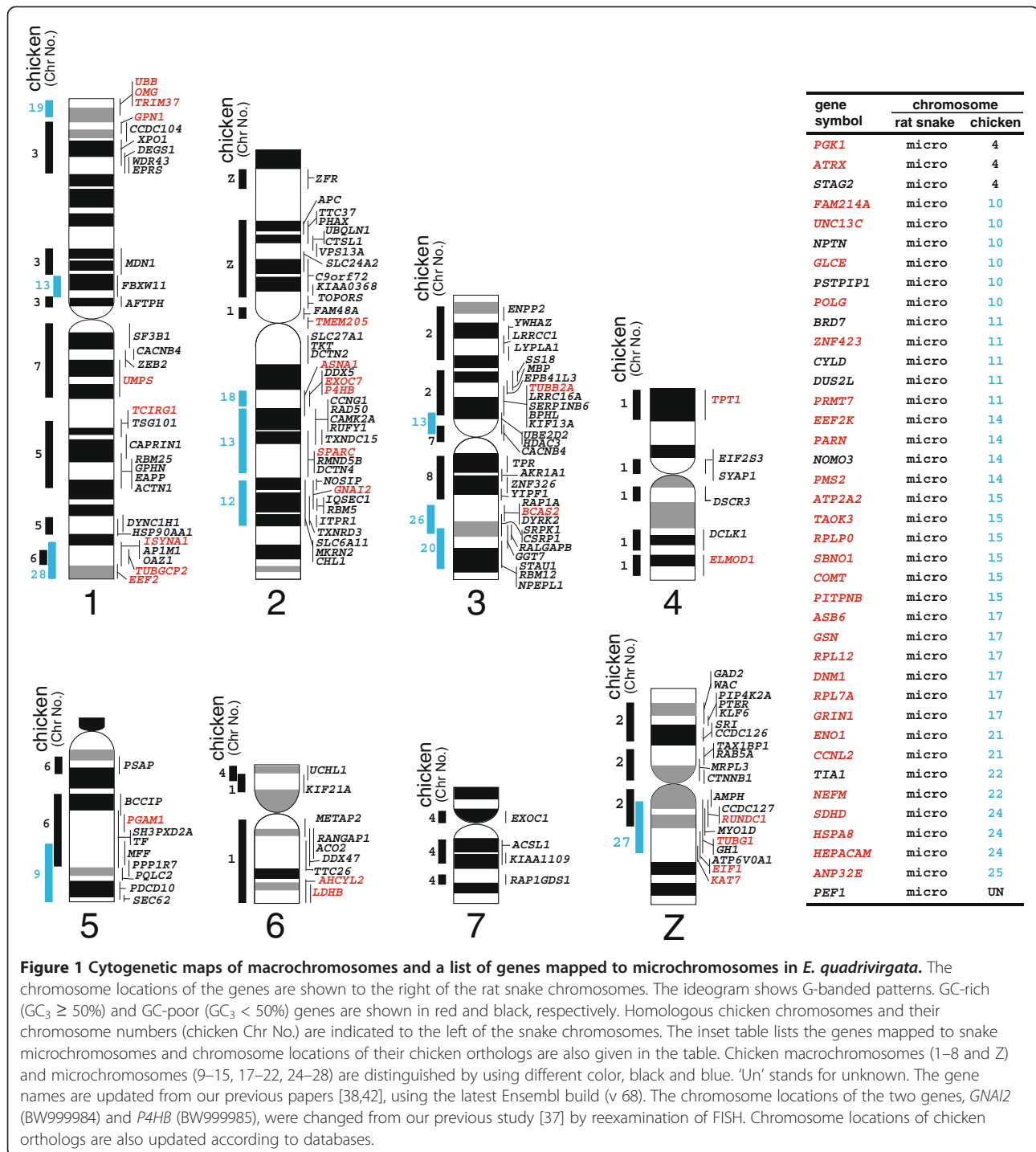
Cytogenetic map of the Japanese four-striped rat snake

Eighty-three cDNA clones were newly mapped to the snake chromosomes, and finally a cytogenetic map with a total of 183 genes was constructed in this study (Figure 1, Additional file 1 and Additional file 2). The nucleotide sequences of the newly mapped EST clones were deposited in GenBank under the accession numbers FS942043-FS942125.

Of the 183 genes, 144 genes were mapped to macrochromosomes (chromosomes 1–7, Z and W chromosomes), and the others were mapped to microchromosomes (Figure 1). Twenty-nine segments in the snake chromosomes 1–7 and the Z chromosome were conserved between the snake and chicken. Most of them had a one to one correspondence to a particular region of chicken chromosomes. However, chromosomal homology for each of chicken chromosomes 1, 2, 4, 6 and 7 was found on more than two snake macrochromosomes, indicating that some inter-chromosomal rearrangements occurred between the snake and chicken macrochromosomes (Figure 1).

Chicken orthologs of 36 snake microchromosomal genes were located on chicken microchromosomes (chromosomes 10, 11, 14, 15, 17, 21, 22, 24 and 25). *PGK1*, *ATRX* and *STAG2* genes on the snake microchromosomes are localized to the short arm of chicken chromosome 4, which was derived from a microchromosome fused with acrocentric chromosome 4 of the avian ancestor [16,17,54]. Since all the snake microchromosomes corresponded to avian microchromosomes, they have likely been retained from the ancestral karyotype of extant sauropsids without dynamic chromosome rearrangements.

Linkage homologies with chicken microchromosomes 9, 10, 12, 13, 18–20 and 26–28 were found on the snake macrochromosomes (Figure 1). For example, chicken chromosomes 19 and 28 were homologous to the distal segments of the short and long arms of snake chromosome 1, respectively. These results confirmed our previous assumption [42] that the large differences of chromosome numbers between the snakes ($2n = 36$) and chicken ($2n = 78$) resulted from frequent chromosome rearrangements containing fusions between macro- and microchromosomes and also between microchromosomes in the lepidosaurian lineage. An alternative explanation is that fissions of macrochromosomes, which

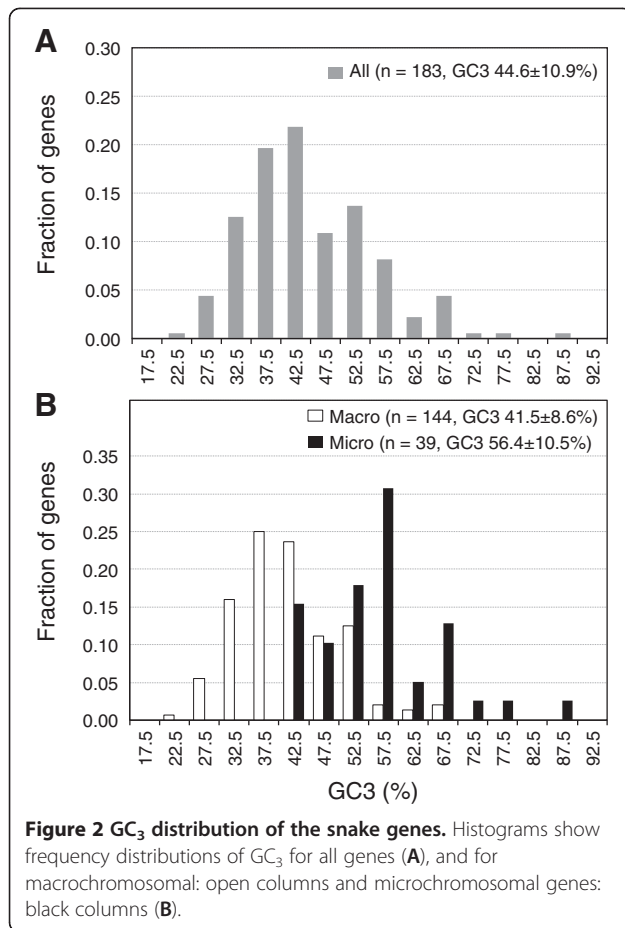


increase microchromosomes, may also have occurred in the lineage leading to birds.

Intra-genomic heterogeneity of GC₃ in snake

We calculated the GC₃ for deduced protein-coding regions of the 183 snake genes (Additional file 2). The average and standard deviation of GC₃ of the snake genes were 44.6% and 10.9% (Figure 2A) and a similar

result was obtained when GC₄ was analyzed (data not shown). The averages of chicken, the soft-shelled turtle and human orthologs were 51.4 ± 13.0% (mean ± standard deviation), 46.5 ± 12.6% and 53.9 ± 16.8%, respectively (Additional file 4). The average and standard deviation of GC₃ of snake genes were thus somewhat smaller than those of the other amniotes. The distribution of GC₃ of the snake genes was also relatively narrow (Figure 2A). In



order to examine bimodality of the distribution, we compared the fit of a “Gaussian model” and “sum of two Gaussians model” by Extra sum-of-squares F test and Akaike’s Information Criterion implemented in GraphPad Prism (GraphPad Software). As a result, “sum of two Gaussians model” showed a better fit in both test ($p < 0.005$ in Extra sum-of-squares F test and 73.28% probability in Akaike’s Information Criterion). This suggests that GC₃ of the snake genes exhibits a bimodal distribution.

GC₃ was compared between macro- (n = 144) and microchromosomal genes (n = 39) to test for the presence of chromosome size-dependent GC₃ heterogeneity in the snake genome (Figure 2B). The average GC₃ was $41.5 \pm 8.6\%$ (mean \pm standard deviation) and $56.4 \pm 10.5\%$ for macrochromosomal and microchromosomal genes, respectively. The average GC₃ of microchromosomal genes was thus significantly higher than for macrochromosomal genes (Mann-Whitney’s *U*-test, $P < 0.01$). 74.4% (29 out of 39 genes) of the microchromosomal genes were GC-rich, whereas 81.9% (118 out of 144 genes) of the macrochromosomal genes were GC-poor (Table 1).

GC-rich isochores are known to have a clear association with R-bands (and particularly T bands) in mammals and

birds [55-58]. In R-banded metaphases of rat snake, most microchromosomes showed R-positive bands (Additional file 5). However, R-positive bands also observed on most macrochromosomal regions. Thus, there was no clear correlation between R-band and GC₃ of the mapped genes in rat snake.

Correlation between GC₃ and GC-content of non-coding regions in snakes

Snake karyotypes are highly conserved among the species [10,40,41], and the Japanese four-striped rat snake (*Elaphe quadrivirgata*) and the Burmese python (*Python molurus bivittatus*) have the same chromosome composition [42]. We compared the GC₃ with the GC-content of non-coding regions in the python to examine whether snake GC₃ reflects the local genomic GC-content (Figure 3A and Additional file 3). High positive correlation was found between GC₃ and GC-content of non-coding regions (n = 176) (Spearman’s rank correlation, $r = 0.73$, $P < 0.01$). In order to consider the differences of base composition among species [36,59,60], we also compared GC₃ of orthologs between the python and the rat snake (n = 182) (Figure 3B and Additional file 3). Strong correlation was found ($r = 0.90$; $P < 0.01$). These results suggest that the two species have similar genomic compositions and snake GC₃ reflects the local genomic GC-content.

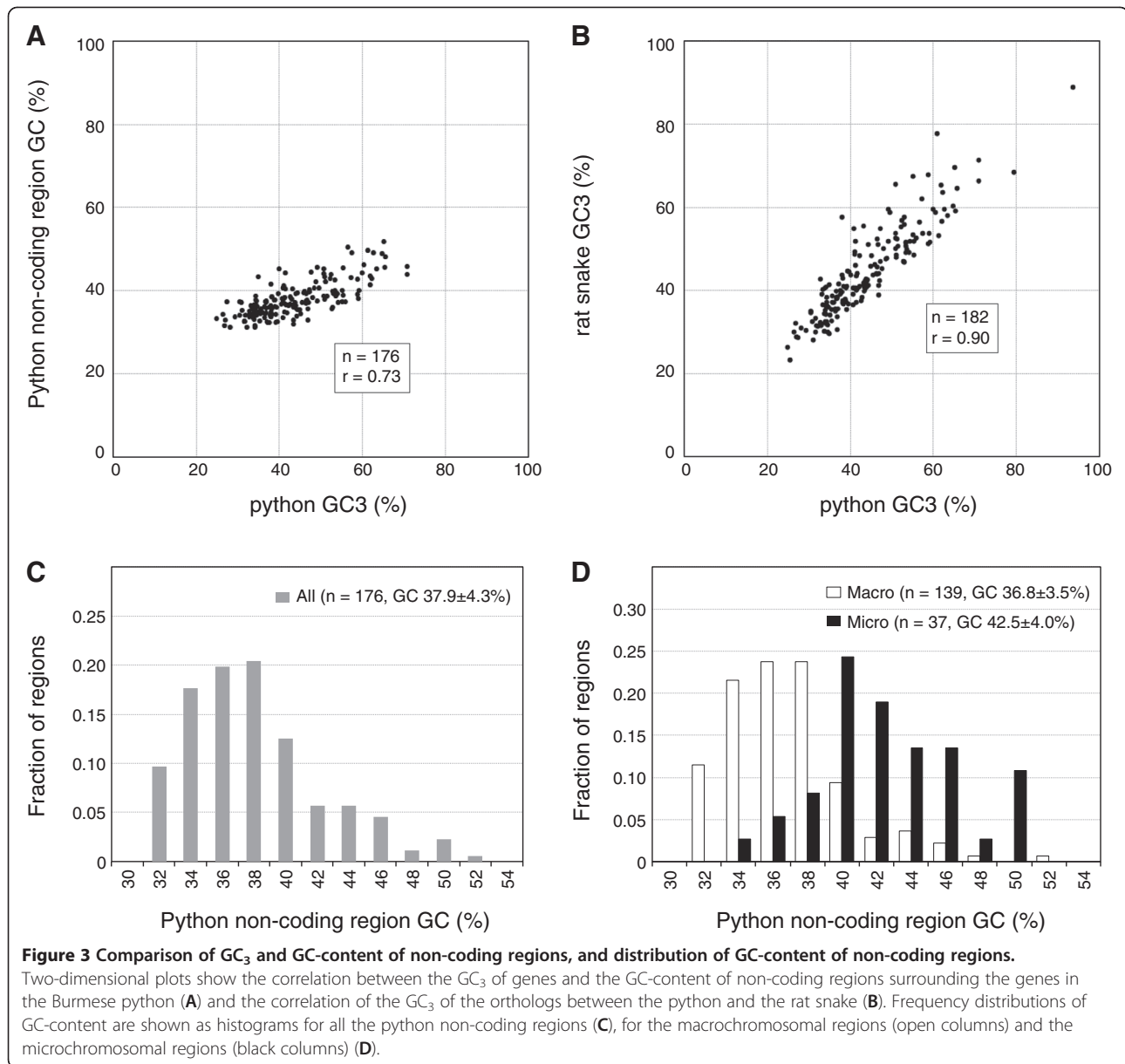
The GC-content of the python non-coding regions showed a narrow distribution with a low average ($37.9 \pm 4.3\%$, mean \pm standard deviation) in contrast to GC₃ (Figure 3A and C). We divided the python non-coding regions into macrochromosomal (n = 139) and microchromosomal regions (n = 37) on the postulate that the chromosome locations of all orthologs were conserved between the rat snake and the python, and compared the GC-content between the two chromosomal groups (Figure 3D). The average GC-content of microchromosomal regions ($42.5 \pm 4.0\%$, mean \pm standard deviation) were significantly higher than those of macrochromosomal regions ($36.8 \pm 3.5\%$) (Mann-Whitney’s *U*-test, $P < 0.01$).

Cross-species comparison of GC₃ between orthologs of amniotes

We examined the frequency distribution of GC₃ of orthologs in green anole lizard, chicken, Chinese soft-

Table 1 Relationships between GC-content and chromosome location of the snake genes

	GC-rich gene (GC ₃ ≥50)	GC-poor gene (GC ₃ ≤50)	Total
Macrochromosomal gene	26	118	144
Microchromosomal gene	29	10	39
Total	55	128	183

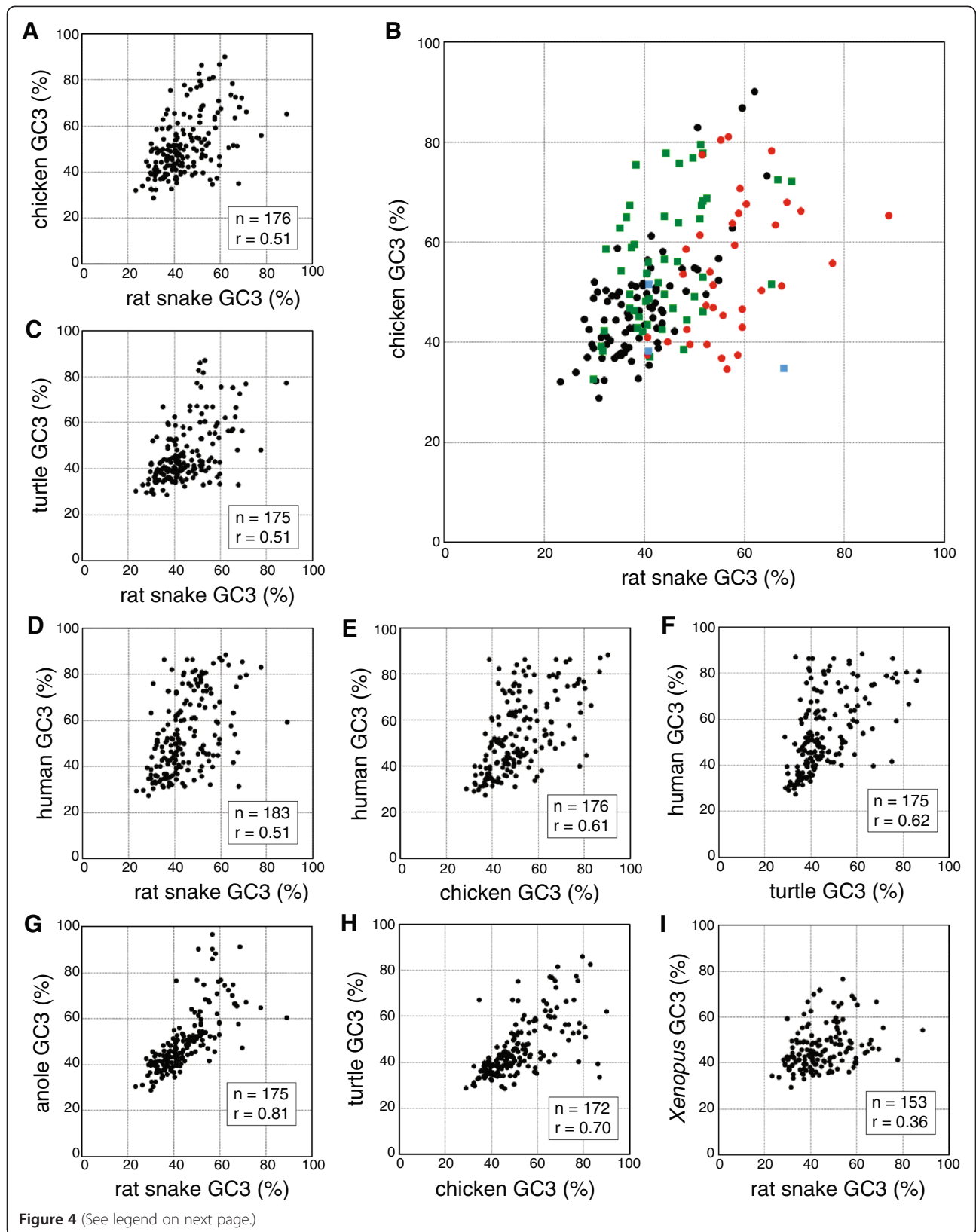


shelled turtle, human, mouse and an amphibian species *X. tropicalis* (Additional file 4). In previous study, we analyzed the frequency distribution of GC₃ with massive genes in these species except for anole lizard. The distribution patterns of GC₃ of this study were similar to those in Figure 2 of our previous study [21] although the gene set of this study contains a somewhat higher proportion of GC-poor genes. Thus we thought that GC₃ of the gene set of this study could be used as a representative of GC₃ of the whole genes in each species.

Then, to examine whether orthologs exhibit similar levels of GC₃ between species, we compared the GC₃ of the snake genes with orthologs in other species (Figure 4 and Additional file 2). Moderate positive correlations

were found for snake-chicken (n = 176), snake-turtle (n = 175) and snake-human (n = 183) comparisons with correlation coefficients $r = 0.51$ (Spearman's rank correlation, $P < 0.01$). Similarly, moderate positive correlations were found for chicken-human (n = 176) ($r = 0.61$; $P < 0.01$) and turtle-human (n = 175) ($r = 0.62$; $P < 0.01$), respectively. Much higher correlation was found for snake-anole lizard (n = 175) ($r = 0.81$; $P < 0.01$) and chicken-turtle (n = 172) ($r = 0.70$; $P < 0.01$), respectively. In contrast, only weak correlation was found for snake-*X. tropicalis* (n = 153) ($r = 0.36$; $P < 0.01$).

For the snake-chicken pair, we also compared the GC₃ of orthologs with four categories: 1) genes located on



(See figure on previous page.)

Figure 4 Cross-species comparison of GC₃ between orthologs. Two-dimensional plots of GC₃ for orthologous gene pairs are shown for rat snake-chicken (A, B), rat snake-Chinese soft-shelled turtle (C), rat snake-human (D), chicken-human (E), Chinese soft-shelled turtle-human (F), rat snake-green anole lizard (G), chicken-Chinese soft-shelled turtle (H) and rat snake-*Xenopus tropicalis* (I). At (B), orthologs are divided into the following four categories as described in text: 1) genes located on macrochromosomes in both species (black dots), 2) genes on microchromosomes in both species (red dots), 3) genes on macrochromosomes in the snake and on microchromosomes in the chicken (green dots), and 4) genes on microchromosomes in the snake and on macrochromosomes in the chicken (blue dots, i.e., *PGK1*, *ATRX* and *STAG2*; see Figure 1).

macrochromosomes in both species (n = 85), 2) genes located on microchromosomes in both species (n = 35), 3) genes located on snake macrochromosomes and on chicken microchromosomes (n = 49), and 4) genes located on snake microchromosomes and on chicken macrochromosomes (n = 3) (Figure 4B). Four orthologs whose chromosome locations are unknown were excluded from this comparison. The average GC₃ of the snake and chicken orthologs were 39.1 ± 8.0% (mean ± standard deviation) and 47.1 ± 10.9% in the 1st group, 57.4 ± 9.8% and 54.9 ± 13.7% in the 2nd group, and 43.9 ± 8.7% and 55.7 ± 12.8% in the 3rd group, respectively (Figure 5). We were not able to compare the GC₃ statistically in the 4th group because only three genes were classified into this group.

Comparison of gene functions between macro- and microchromosomes

To examine the relationships between gene functions and chromosome size-dependent GC bias in chicken, we assessed over-representation of functional categories of Gene Ontology (GO), KEGG pathway, InterPro motif and Swissprot in either of two gene groups: the 10,053 genes on chicken macrochromosomes (namely, chromosomes 1–8, Z and W) and 6,297 genes on chicken microchromosomes (namely, chromosomes 9–28). Chromosomes 29–31 and 33–38 were excluded from analysis because of their absence from the assembled genome. No genes have been assigned to chromosome 32, although the genome sequence was anchored to this chromosome. Fourteen GO terms or InterPro domains were over-represented in either macrochromosomal genes or microchromosomal genes (Figure 6 and Additional file 6: Table S1). Over-representation of "chromatin" (GO:0000785), "nucleosome" (GO:0000786) and "chromosomal part" (GO:0044427) on macrochromosomes largely depended on the abundance of members of the histone gene family on chromosome 1 (Additional file 6: Table S9, S11 and S12). Over-representation of "immunoglobulin C1-set" (IPR003597) and "MHC protein complex" (GO:0042611) in microchromosomes came from clustering of many immune genes on chromosome 16 (Additional file 6: Table S8 and S14). Similarly, over-representation of "keratin" (IPR003461) was due to the abundance of keratin-like gene family members on

chromosomes 25 and 27 (Additional file 6: Table S13). Over-representation of the other eight categories in either macrochromosomes or microchromosomes was not attributable to an excess of genes on particular chromosomes (Additional file 6: Table S1-S7, S10 and S15).

Discussion

Chromosome size-dependent GC heterogeneity is a pan-Sauropsida characteristic

We constructed a cytogenetic map of the Japanese four-striped rat snake, which contained 183 genes. Our

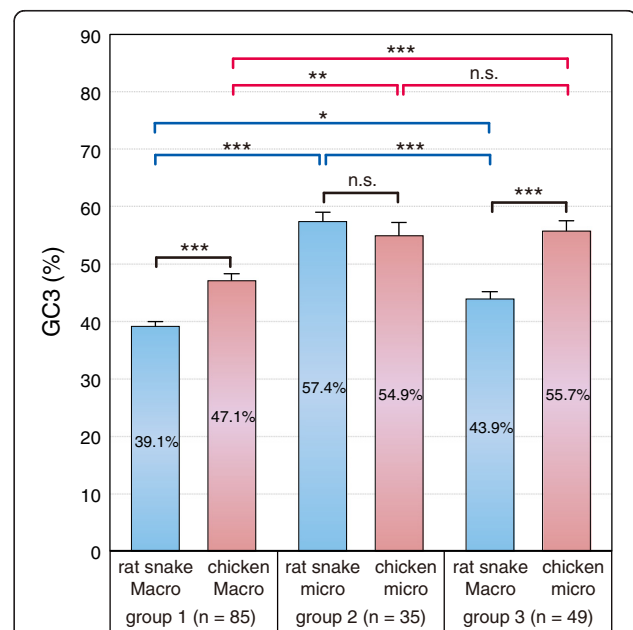


Figure 5 Comparison of the GC₃ between the species and among the gene groups. Histograms show mean ± SEM of GC₃.

The rat snake and chicken orthologs are grouped as described in text: group 1) genes located on macrochromosomes in both species, group 2) genes located on microchromosomes in both species, and group 3) genes located on macrochromosomes in the snake and on microchromosomes in the chicken. Because only three genes were classified into the 4th group, indicated by blue dots in Figure 4B, the GC₃ of this group was not compared statistically and thus omitted from this figure. Horizontal lines above bars show differences. Significances for comparison between species (black lines) and for comparison between groups (blue and red lines) are examined by Mann-Whitney's U-test and the Kruskal-Wallis test with Dunn's post test, respectively. *P<0.05, **P<0.01, ***P<0.001 and n.s., not significant.

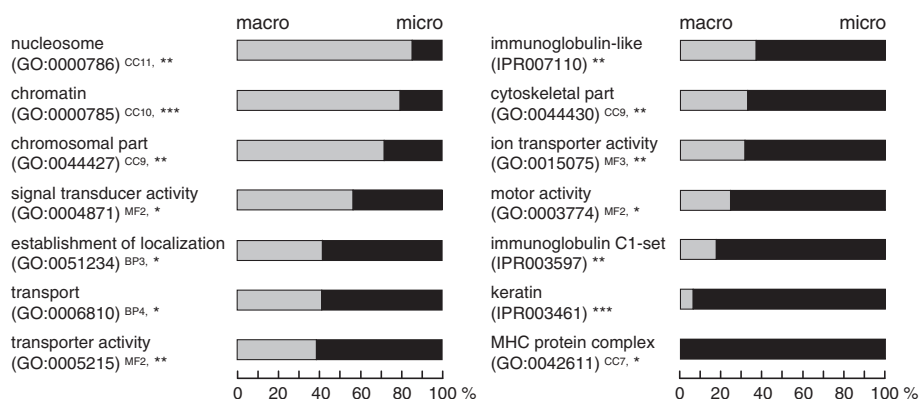


Figure 6 Statistically over-represented Gene Ontology or InterPro categories between chicken macrochromosomal and microchromosomal genes.

Over-representation of gene categories was examined between the genes on chicken macrochromosomes (chromosomes 1–8, Z and W) and microchromosomes (chromosomes 9–28). The proportion of the macro- and microchromosomal genes is shown in each over-represented category of Gene Ontology (GO) or InterPro (IPR). The classifications of GO categories are indicated by superscripts: CC, cellular component; MF, molecular function; BP, biological process; Numbers indicate the levels of GO hierarchy. Significance for each category is assessed with Fisher's exact test; * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$.

cytogenetic map covered most macrochromosomal regions and at least one gene was mapped to each microchromosome on the basis of the homologies with chicken chromosomes. Our map also showed linkage homologies with most chicken chromosomes, namely 1–15, 17–22, 24–28 and Z. These results make it possible to infer the global GC heterogeneity of the snake genome and the shift of GC-content caused by chromosome rearrangements during sauropsid evolution.

We calculated GC_3 to investigate the intra-genomic GC heterogeneity in the rat snake. The GC_3 of the snake genes exhibited a bimodal distribution (Figure 2A). Such a bimodal distribution of GC_3 was also observed in the genomes of the Chinese soft-shelled turtle (*P. sinensis*), chicken and non-rodent mammals [21]. This result suggests that the GC_3 heterogeneity is a common feature of amniote genomes. However, the standard deviation of GC_3 in snake genes was somewhat lower than that of other amniotes [21], and GC-content of python non-coding regions also showed a narrow distribution, as observed in the green anole lizard [37]. Thus the heterogeneities of base composition have probably decreased in lepidosaurian lineages over evolutionary time.

Although the standard deviation of GC_3 of snake genes was relatively small, our results suggest that snake microchromosomes contain a higher proportion of GC-rich genes than macrochromosomes, as observed in both the Chinese soft-shelled turtle and chicken [21]. Recently, chromosome size-dependent GC heterogeneity was also identified in the red-eared slider turtle (*Trachemys scripta elegans*) and the Nile crocodile (*Crocodylus niloticus*) using chromosome flow sorting technique [61]. Chromosome size-dependent GC heterogeneity therefore seems to be a widespread characteristic in sauropsids

whose karyotypes consist of macrochromosomes and microchromosomes, and possibly originated in the common ancestor of sauropsids. Interestingly, the green anole lizard, whose karyotype consists of 6 pairs of macrochromosomes and 12 pairs of microchromosomes [39,62], does not show such marked biases in GC-content between macro- and microchromosomes. This suggests that the chromosome size-dependent GC heterogeneity has disappeared in the lineage leading to the anole lizard. Lepidosauria is a species-rich group consisting of about 8,000 species, and the karyotypes are also diversified within the group. Further investigation in various lepidosaurian species may help clarify the relationship between GC-content and the karyotype.

Disparity between lepidosaurs and the turtle-archosaurs

The correlation coefficient of GC_3 between the rat snake and chicken orthologs are lower than that between the Chinese soft-shelled turtle and chicken (Figure 4). One explanation for the lower correlation is that the phylogenetic distance between the snake and chicken is larger than between the turtle and chicken. However, the divergence time between turtles and birds is estimated to be more than 231 million years, which is not largely different from the time of the lepidosauria-archosauria split, 275 million years ago [3,7-9]. Therefore we consider the effect of the large differences of karyotypes, especially the number of microchromosomes, between the snake and the other two species.

The chromosome numbers are largely different between the rat snake ($2n = 36$) and chicken ($2n = 78$). In contrast, the karyotype of the Chinese soft-shelled turtle, which consists of nine pairs of macrochromosomes and 24 pairs

of microchromosomes ($2n = 66$), is very similar to the chicken karyotype [38]. Our previous study demonstrated by comparative gene mapping that the chromosomes have been highly conserved between chicken and the turtle [21,38]. Chicken microchromosomes were considered to extensively retain the ancestral linkage groups of genes [63], and Nakatani *et al.* [64] also suggested that many chicken microchromosomes (i.e., chromosomes 11, 15, 19, 20, 21, 22, 23, 24, 27 and 28) have one-to-one correspondence to ancestral proto-chromosomes of the gnathostome ancestor [64]. These results lead us to infer that chromosome rearrangements have occurred more frequently in the snake lineage than in the chicken lineage, that is, chromosome number has been reduced by frequent chromosome fusions between macro- and microchromosomes and also between microchromosomes in the snake lineage.

Eleven chromosome segments homologous to chicken microchromosomes were localized to the snake macrochromosomes in this study (Figure 1). The GC_3 of the snake orthologs on these macrochromosomal segments were lower than those of their chicken orthologs on microchromosomes (green dots of Figure 4B and 5). For instance, 16 of 21 snake orthologs mapped on chromosome 2q, which is homologous to chicken chromosomes 18, 13 and 12, were GC_3 -poor ($GC_3 < 50\%$), whereas 9 of 17 chicken orthologs on chicken chromosomes 18, 13 and 12 were GC_3 -rich ($GC_3 \geq 50\%$) (Additional file 2). These results suggest that changes in chromosome sizes caused the differences of GC_3 levels between the chicken microchromosomal genes and their snake orthologs on macrochromosomal segments derived from the ancestral microchromosomes (Figure 7).

Impact of chromosome fissions/fusions on GC-content

What mechanisms were involved in the changes of the GC-content of the genes after the fusion of microchromosomes into macrochromosomal complement? It has been suggested that the GC-content is primarily influenced by local recombination rates via GC-biased gene conversion [65,66]. Under this model, A or T is displaced by G or C through mismatch repair when an AT/GC heteroduplex is formed at recombining regions. Accordingly, AT/GC heterozygotes produce more GC than AT gametes, thus conferring predominance of GC alleles in frequently recombining regions.

Recombination rate is negatively correlated with the size of chromosome arms in human and chicken [20,67]. In chicken in particular, recombination rate per unit physical length is much higher in microchromosomes than in macrochromosomes [20]. Recombination rates per physical length are thus expected to be lower in the snake macrochromosomal segments derived from the ancestral microchromosomes than those in their homologous chicken

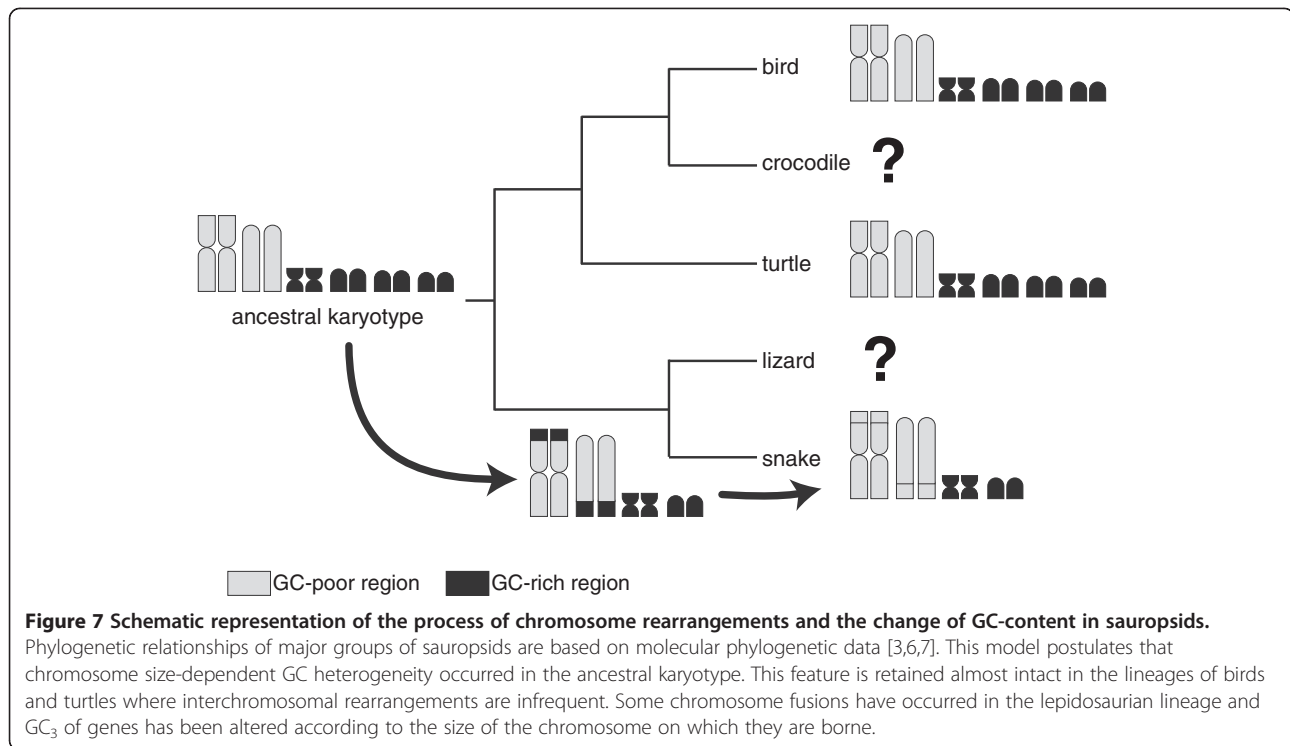
microchromosomes. The chromosome size-dependent difference in the recombination rate thus seems to have caused the decrease of GC-content in the snake macrochromosomal genes derived from the microchromosomes of the common ancestor of sauropsids.

Empirical evidence of chromosome length-driven evolution of GC-content has been shown in marsupial and monotreme species [36]. In contrast, chromosome size-dependent GC heterogeneity was not clearly demonstrated in eutherian species, likely because frequent chromosome rearrangements have obscured the history of GC-content changes [36]. In many eutherian lineages, however, intra-chromosomal GC heterogeneity has been reported [21,36,68,69]. In chicken, intra-chromosomal GC heterogeneity is not as prominent as in eutherians [22], and there is no intra-chromosomal GC heterogeneity known in the anole genome [37,39]. Our approach could not demonstrate whether there was intra-chromosomal GC heterogeneity in snake genome because of the insufficient sequence data provided by cDNA sequencing. It is necessary to conduct whole genome sequencing and assemble the sequences into a chromosome scale in order to clarify the relationship between intra- and inter-chromosomal GC heterogeneity in the snake genomes.

When overall GC-content and recombination rates were compared among various vertebrate species, there was no clear correlation between GC-content and recombination rates [70]. For example, the recombination rate of chicken is about two times higher than that of zebra finch, although GC-content is almost equal between the two species [70,71]. Whereas rat snake microchromosomal genes show similar level of GC_3 to chicken microchromosomal genes, GC_3 of rat snake macrochromosomal genes is significantly lower than chicken macrochromosomal genes (Figure 5). Thus other factors might exert influences on overall genomic GC-content. In bacteria, genomic GC-content have been subject to natural selection but not to biased gene conversion [72]. Modes other than biased gene conversion were also proposed for evolution of genome composition in vertebrates [68]. Further consideration is therefore necessary for the evolution of overall genome composition in snakes.

Any biological partitioning between macrochromosomes and microchromosomes?

Many literatures reported existing correlations between gene function and base compositions of the genes, the genomes and the promoter regions [73-75]. The difference of global GC-content between the macro- and microchromosomes may potentially cause the biased distribution of gene functions between the chromosomes: some proteins containing more amino acids for GC-rich codons due to functional constraints may be more advantageous in being encoded in microchromosomes



than in macrochromosomes. To test this possibility, we investigated functional difference between chicken macro- and microchromosomal genes on the basis of the frequencies of appearance of functional categories (Figure 6). Over-representations found in six categories could be caused by clustering of particular gene families in short genomic stretches. These gene families may be less informative to test the hypothesis that genes are selected based on the size of chromosomes on which they are encoded. The over-representations of the other eight categories were independent from an excess of genes on particular chromosomes, implying that there could be some functional differences between macro- and microchromosomal genes.

In order to examine the relationship between the apparent localization of functional gene categories and the chromosome size-dependent GC-content, we investigated the GC-content of the gene sets assorted in the eight over-represented categories: GO:0004871, GO:0051234, GO:0006810, GO:0005215, IPR007110, GO:0044430, GO:0015075 and GO:0003774 (Additional file 6: Table S16). As shown in Figure 6 and Additional file 6: Table S1, seven of the eight categories were over-represented in microchromosomes. If most genes assorted in these GO or InterPro categories have high GC-content, it is likely that functional compartmentalization of genes is influenced by the difference of global GC-content between macro- and microchromosomes, i.e., some functionally categorized gene groups with high GC-content have been

selectively preserved on microchromosomes. The average GC-content of the genes of the seven categories ranged from 43.9 to 48.1% and most of them were lower than the average of all chicken genes (47.4%) (Additional file 6: Table S16). In the other six categories (GO:0042611, GO:0044427, GO:0000785, GO:0000786, IPR003461 and IPR003597), there was no clear correlation between abundance in macro- or microchromosomes and GC-content. For an example, GO:0044427, GO:0000785 and GO:0000786 were over-represented in macrochromosomes, but their GC-contents were relatively high (Additional file 6: Table S1 and S16). These results imply that functional differences of genes did not correlate with the global difference of GC-content between macro- and microchromosomes. Further characterization of functional difference between macro- and microchromosomal genes, as well as its correlation with the general GC trend in chromosomal environments should await more extensive analyses using multiple species in future.

Conclusion

In this study, we constructed a cytogenetic map with 183 genes in the Japanese four-striped rat snake, and calculated GC₃ across all chromosomes. Our results revealed cytogenetic evidence that snake microchromosomal genes tend to have higher GC₃ than macrochromosomal genes, as found in the chicken and the Chinese soft-shelled turtle, a feature apparently lost in the genome of anole lizard. By comparing GC₃ of orthologs

between snake and chicken, we show that the GC-content of genes is correlated with the size of chromosomes on which the genes are harbored. This chromosome size-dependent GC heterogeneity is particularly apparent in snake genes that have been translocated from microchromosomes to macrochromosomes since snakes and birds shared a common ancestor, some 275 million years. The addition of whole genome sequencing and karyotypes from wide variety of sauropsidan species will provide the fine-scale picture of timing and mode of GC shift accompanying karyotypic evolution in this important group of vertebrates.

Additional files

Additional file 1: FISH mapping of cDNA clones in *Elaphe quadrivirgata*. FISH mapping of six cDNA clones in *Elaphe quadrivirgata*.

Additional file 2: IDs and GC₃ of 184 orthologous gene sets. IDs and GC₃ of 184 orthologous gene sets.

Additional file 3: IDs and GC-content of the python contigs that include the orthologs of the rat snake ESTs. IDs and GC-content of the python contigs that include the orthologs of the rat snake ESTs.

Additional file 4: GC₃ distribution of the orthologs in other vertebrates. GC₃ distribution of the orthologs in other vertebrates.

Additional file 5: R-banded karyotype of *Elaphe quadrivirgata*. R-banded karyotype of *Elaphe quadrivirgata*.

Additional file 6: Table S1. Functional categories over-represented in either gene group, chicken macrochromosomal genes or microchromosomal genes. **Table S2.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0051234 and their chromosome locations and GC-contents. **Table S3.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0006810 and their chromosome locations and GC-contents. **Table S4.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0005215 and their chromosome locations and GC-contents. **Table S5.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0003774 and their chromosome locations and GC-contents. **Table S6.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0004871 and their chromosome locations and GC-contents. **Table S7.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0015075 and their chromosome locations and GC-contents. **Table S8.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0042611 and their chromosome locations and GC-contents. **Table S9.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0044427 and their chromosome locations and GC-contents. **Table S10.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0044430 and their chromosome locations and GC-contents. **Table S11.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0000785 and their chromosome locations and GC-contents. **Table S12.** Chicken macrochromosomal and microchromosomal genes assorted into GO:0000786 and their chromosome locations and GC-contents. **Table S13.** Chicken macrochromosomal and microchromosomal genes assorted into IPR003461 and their chromosome locations and GC-contents. **Table S14.** Chicken macrochromosomal and microchromosomal genes assorted into IPR003597 and their chromosome locations and GC-contents. **Table S15.** Chicken macrochromosomal and microchromosomal genes assorted into IPR007110 and their chromosome locations and GC-contents. **Table S16.** The number of genes and mean of GC-content of macro- and microchromosomal genes assorted into each over-represented functional category. The detailed tables for statistically over-represented categories of Gene Ontology or InterPro between the chicken macrochromosomal and microchromosomal genes.

Abbreviations

cDNA: Complementary DNA; GC₃: GC-content of exonic third codon positions; EST: Expressed sequence tag; FISH: Fluorescence *in situ* hybridization; GO: Gene Ontology.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KM, SK, YK and YM designed the study. HT, ON and KA constructed the EST library. KM and CN conducted FISH mapping. KM and SK analyzed the data. KM, SK, YK and YM wrote the paper. All authors read and approved the manuscript.

Acknowledgements

We thank Yasuyuki Hashiguchi for advice on identification of coding and non-coding regions in the python contigs, Yukuto Sato for advice on comparison of the gene functions, Yosuke Kawai for discussion and Denis O'meally for English correction. This work was supported by Grants-in-Aid for Scientific Research (No. 15370001, No. 16086201 and No. 19770001) from the Ministry of Education, Culture, Sports, Science and Technology in Japan, and by JSPS Research Fellowships for Young Scientists to K.M. (No. 215132).

Author details

¹Department of Information and Biological Sciences, Graduate School of Natural Sciences, Nagoya City University, 1 Yamahata, Mizuho-cho, Mizuho-ku, Nagoya, Aichi 467-8501, Japan. ²Department of Biology, University of Konstanz, Universitaetsstrasse 10, Konstanz 78457, Germany. ³Genome Resource and Analysis Unit, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. ⁴Laboratory for Molecular Developmental Biology, Department of Biophysics, Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake, Sakyo-ku, Kyoto, Kyoto 606-8502, Japan. ⁵Department of Natural History Sciences, Faculty of Science, Hokkaido University, Kita 10 Nishi 8, Kita-ku, Sapporo, Hokkaido 060-0810, Japan. ⁶Laboratory of Animal Genetics, Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan. ⁷Present address: Institute for Applied Ecology, University of Canberra, Canberra, ACT 2601, Australia. ⁸Present address: Omics Science Center, Yokohama Institute RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.

Received: 28 January 2012 Accepted: 24 October 2012

Published: 9 November 2012

References

1. Zardoya R, Meyer A: Complete mitochondrial genome suggests diapsid affinities of turtles. *Proc Natl Acad Sci USA* 1998, **95**:14226-14231.
2. Hedges SB, Poling LL: A molecular phylogeny of reptiles. *Science* 1999, **283**:998-1001.
3. Kumazawa Y, Nishida M: Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. *Mol Biol Evol* 1999, **16**:784-792.
4. Cao Y, Sorenson MD, Kumazawa Y, Mindell DP, Hasegawa M: Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. *Gene* 2000, **259**:139-148.
5. Rest JS, Ast JC, Austin CC, Waddell PJ, Tibbetts EA, Hay JM, Mindell DP: Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol Phylogenet Evol* 2003, **29**:289-297.
6. Iwabe N, Hara Y, Kumazawa Y, Shibamoto K, Saito Y, Miyata T, Katoh K: Sister group relationship of turtles to the bird-crocodylian clade revealed by nuclear DNA-coded proteins. *Mol Biol Evol* 2005, **22**:810-813.
7. Kumazawa Y: Mitochondrial genomes from major lizard families suggest their phylogenetic relationships and ancient radiations. *Gene* 2007, **388**:19-26.
8. Benton MJ, Donoghue PCJ: Paleontological evidence to date the tree of life. *Mol Biol Evol* 2007, **24**:26-53.
9. Shedlock AM, Edwards SV: Amniotes (Amniota). In *The timetree of life*. Edited by Hedges SB, Kumar S. New York: Oxford University Press; 2009:375-379.

10. Beçak W, Beçak ML, Nazareth HRS, Ohno S: Close karyological kinship between the reptilian suborder Serpentes and the class Aves. *Chromosoma* 1964, **15**:606–617.
11. Bickham JW, Baker RJ: Chromosome homology and evolution of emydid turtles. *Chromosoma* 1976, **54**:201–219.
12. Bickham JW, Bull JJ, Legler JM: Karyotypes and evolutionary relationships of trionychid turtles. *Cytologia* 1983, **48**:177–183.
13. Norris TB, Rickards GK, Daugherty CH: Chromosomes of tuatara, *Sphenodon*, a chromosome heteromorphism and an archaic reptilian karyotype. *Cytogenet Genome Res* 2004, **105**:93–99.
14. Takagi N, Sasaki M: A phylogenetic study of bird karyotypes. *Chromosoma* 1974, **46**:91–120.
15. Belterman RHR, de Boer LEM: A karyological study of 55 species of birds, including karyotypes of 39 species new to cytology. *Genetica* 1984, **65**:39–82.
16. Shetty S, Griffin DK, Graves JAM: Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Res* 1999, **7**:289–295.
17. Nishida-Umehara C, Tsuda Y, Ishijima J, Ando J, Fujiwara A, Matsuda Y, Griffin DK: The molecular basis of chromosome orthologies and sex chromosomal differentiation in palaeognathous birds. *Chromosome Res* 2007, **15**:721–734.
18. Cohen MM, Gans C: The chromosomes of the order Crocodylia. *Cytogenetics* 1970, **9**:81–105.
19. Kawagoshi T, Nishida C, Ota H, Kumazawa Y, Endo H, Matsuda Y: Molecular structures of centromeric heterochromatin and karyotypic evolution in the Siamese crocodile (*Crocodylus siamensis*) (Crocodylidae, Crocodylia). *Chromosome Res* 2008, **16**:1119–1132.
20. International Chicken Genome Sequencing Consortium (ICGSC): Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 2004, **432**:695–716.
21. Kuraku S, Ishijima J, Nishida-Umehara C, Agata K, Kuratani S, Yoichi M: cDNA-based gene mapping and GC3 profiling in the soft-shelled turtle suggest a chromosomal size-dependent GC bias shared by sauropsids. *Chromosome Res* 2006, **14**:187–202.
22. Costantini M, Filippo MD, Auletta F, Bernardi G: Isochore pattern and gene distribution in the chicken genome. *Gene* 2007, **400**:9–15.
23. McQueen HA, Siriaco G, Bird AP: Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Res* 1998, **8**:621–630.
24. Smith J, Bruley CK, Paton IR, Dunn I, Jones CT, Windsor D, Morrice DR, Law AS, Masabanda J, Sazanov A, Waddington D, Fries R, Burt DW: Differences in gene density on chicken macrochromosomes and microchromosomes. *Anim Genet* 2000, **31**:96–103.
25. Habermann FA, Cremer M, Walter J, Kreth G, von Hase J, Bauer K, Wienberg J, Cremer C, Cremer T, Solovei I: Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Res* 2001, **9**:569–584.
26. Axelsson E, Webster MT, Smith NG, Burt DW, Ellegren H: Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Res* 2005, **15**:120–125.
27. Federico C, Cantarella CD, Scavo C, Saccone S, Bed'Hom B, Bernardi G: Avian genomes: different karyotypes but a similar distribution of the GC-richest chromosome regions at interphase. *Chromosome Res* 2005, **13**:785–793.
28. Shedlock AM, Botka CW, Zhao S, Shetty J, Zhang T, Liu JS, Deschavanne PJ, Edwards SV: Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc Natl Acad Sci USA* 2007, **104**:2767–2772.
29. Janes DE, Organ CL, Fujita MK, Shedlock AM, Edwards SV: Genome evolution in Reptilia, the sister groups of mammals. *Annu Rev Genomics Hum Genet* 2010, **11**:239–264.
30. Belle EM, Smith N, Eyre-Walker A: Analysis of the phylogenetic distribution of isochores in vertebrates and a test of the thermal stability hypothesis. *J Mol Evol* 2002, **55**:356–363.
31. Hamada K, Horiike T, Ota H, Mizuno K, Shinozawa T: Presence of isochore structures in reptile genomes suggested by the relationship between GC content of intron regions and those of coding regions. *Genes Genet Syst* 2003, **78**:195–198.
32. Chojnowski JL, Franklin J, Katsu Y, Iguchi T, Guillelte LJ Jr, Kimball RT, Braun EL: Patterns of vertebrate isochore evolution revealed by comparison of expressed mammalian, avian, and crocodylian genes. *J Mol Evol* 2007, **65**:259–266.
33. Chojnowski JL, Braun EL: Turtle isochore structure is intermediate between amphibians and other amniotes. *Integr Comp Biol* 2008, **48**:454–462.
34. Elhaik E, Landan G, Graur D: Can GC content at third-codon positions be used as a proxy for isochore composition? *Mol Biol Evol* 2009, **26**:1829–1833.
35. Clay OK, Bernardi G: GC3 of genes can be used as a proxy for isochore base composition: A reply to Elhaik et al. *Mol Biol Evol* 2011, **28**:21–23.
36. Romiguier J, Ranwez V, Douzery EJP, Galtier N: Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes. *Genome Res* 2010, **20**:1001–1009.
37. Fujita MK, Edwards SV, Ponting CP: The Anolis lizard genome: an amniote genome without isochores. *Genome Biol Evol*, in press.
38. Matsuda Y, Nishida-Umehara C, Tarui H, Kuroiwa A, Yamada K, Isobe T, Ando J, Fujiwara A, Hirao Y, Nishimura O, Ishijima J, Hayashi A, Saito T, Murakami T, Murakami Y, Kuratani S, Agata K: Highly conserved linkage homology between birds and turtles: Bird and turtle chromosomes are precise counterparts of each other. *Chromosome Res* 2005, **13**:601–615.
39. Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, Russell P, Lowe CB, Glor RE, Jaffe JD, Ray DA, Boissinot S, Shedlock AM, Botka C, Castoe TA, Colbourne JK, Fujita MK, Moreno RG, Los Hallers BF, Haussler D, Heger A, Heiman D, Janes DE, Johnson J, de Jong PJ, Koriabine MY, Lara M, Novick PA, Organ CL, Peach SE, Poe S, Pollock DD, de Queiroz K, Sanger T, Searle S, Smith JD, Smith Z, Swofford R, Turner-Maier J, Wade J, Young S, Zadissa A, Edwards SV, Glenn TC, Schneider K, Moser JB, Lander ES, Breen M, Ponting CP, Lindblad-Toh K: The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 2011, **477**:587–591.
40. Beçak W, Beçak ML: Cytotaxonomy and chromosomal evolution in Serpentes. *Cytogenetics* 1969, **8**:247–262.
41. Singh L: Evolution of karyotypes in snakes. *Chromosoma* 1972, **38**:185–236.
42. Matsubara K, Tarui H, Toriba M, Yamada K, Nishida-Umehara C, Agata K, Matsuda Y: Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proc Natl Acad Sci USA* 2006, **103**:18190–18195.
43. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, **25**:3389–3402.
44. Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987, **4**:406–425.
45. Katoh K, Kuma K, Toh H, Miyata T: MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 2005, **33**:511–518.
46. Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003, **52**:696–704.
47. Matsuda Y, Chapman VM: Application of fluorescence *in situ* hybridization in genome analysis of the mouse. *Electrophoresis* 1995, **16**:261–272.
48. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigan C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehväslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E: The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* 2002, **12**:1611–1618.
49. Castoe TA, de Koning AJ, Hall KT, Yokoyama KD, Gu W, Smith EN, Feschotte C, Uetz P, Ray DA, Dobry J, Bogden R, Mackessy SP, Bronikowski AM, Warren WC, Secor SM, Pollock DD: Sequencing the genome of the Burmese python (*Python molurus bivittatus*) as a model for studying extreme adaptations in snakes. *Genome Biol* 2011, **12**:406. URL: <http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AEQU>.
50. Wise2 - Intelligent algorithms for DNA searches. URL: <http://www.ebi.ac.uk/Tools/Wise2/index.html>.
51. Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Gilbert J, Hammond M, Herrero J, Hotz H, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Kokocinski F, London D, Longden I, McVicker G, Melsopp C, Meidl P, Potter S, Proctor G, Rae M, Rios D, Schuster M, Searle S, Severin J, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodward C, Birney E: Ensembl 2005. *Nucleic Acids Res* 2005, **33**:D447–453. URL: <http://www.ensembl.org/>.
52. NCBI Entrez Gene. URL: <http://www.ncbi.nlm.nih.gov/gene/>.
53. Al-Shahrour F, Diaz-Uriarte R, Dopazo J: FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes.

Bioinformatics 2004, **20**:578–580. URL: <http://babelomics.bioinfo.cipf.es/functional.html>.

54. Shibusawa M, Nishibori M, Nishida-Umehara C, Tsudzuki M, Masabanda J, Griffin DK, Matsuda Y: **Karyotypic evolution in the Galliformes: an examination of the process of karyotypic evolution by comparison of the molecular cytogenetic findings with the molecular phylogeny.** *Cytogenet Genome Res* 2004, **106**:111–119.
55. Holmquist GP: **Chromosome bands, their chromatin flavors, and their functional features.** *Am J Hum Genet* 1992, **51**:17–37.
56. Saccone S, De Sario A, Della Valle G, Bernardi G: **The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes.** *Proc Natl Acad Sci USA* 1992, **89**:4913–4917.
57. Saccone S, De Sario A, Wiegant J, Raap AK, Della Valle G, Bernardi G: **Correlations between isochores and chromosomal bands in the human genome.** *Proc Natl Acad Sci USA* 1993, **90**:11929–11933.
58. Saccone S, Cacciò S, Kusuda J, Andreozzi L, Bernardi G: **Identification of the gene-richest bands in human chromosomes.** *Gene* 1996, **174**:85–94.
59. Nabholz B, Künstner A, Wang R, Jarvis ED, Ellegren H: **Dynamic evolution of base composition: causes and consequences in avian phylogenomics.** *Mol Biol Evol* 2011, **28**:2197–2210.
60. Bucciarelli G, Di Filippo M, Costagliola D, Alvarez-Valin F, Bernardi G, Bernardi G: **Environmental genomics: a tale of two fishes.** *Mol Biol Evol* 2009, **26**:1235–1243.
61. Kasai F, O'Brien PC, Ferguson-Smith MA: **Reassessment of genome size in turtle and crocodile based on chromosome measurement by flow karyotyping: close similarity to chicken.** *Biol Lett* 2012, **8**:631–635.
62. Benirschke K, Hsu TC: *Chromosome atlas: fish, amphibians, reptiles and birds volume 1.* New York: Springer; 1971.
63. Burt DW: **Origin and evolution of avian microchromosomes.** *Cytogenet Genome Res* 2002, **96**:97–112.
64. Nakatani Y, Takeda H, Kohara Y, Morishita S: **Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates.** *Genome Res* 2007, **17**:1254–1265.
65. Duret L, Eyre-Walker A, Galtier N: **A new perspective on isochore evolution.** *Gene* 2006, **385**:71–74.
66. Duret L, Galtier N: **Biased gene conversion and the evolution of mammalian genomic landscapes.** *Annu Rev Genomics Hum Genet* 2009, **10**:285–311.
67. Meunier J, Duret L: **Recombination drives the evolution of GC-content in the human genome.** *Mol Biol Evol* 2004, **21**:984–990.
68. Costantini M, Cammarano R, Bernardi G: **The evolution of isochore patterns in vertebrate genomes.** *BMC Genomics* 2009, **10**:146.
69. Costantini M, Clay O, Auletta F, Bernardi G: **An isochore map of human chromosomes.** *Genome Res* 2006, **16**:536–541.
70. Kai W, Kikuchi K, Tohari S, Chew AK, Tay A, Fujiwara A, Hosoya S, Suetake H, Naruse K, Brenner S, Suzuki Y, Venkatesh B: **Integration of the genetic map and genome assembly of fugu facilitates insights into distinct features of genome evolution in teleosts and mammals.** *Genome Biol Evol* 2011, **3**:424–442.
71. Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, Webster MT, Ost T, Schneider M, Kempnaers B, Ellegren H: **The recombination landscape of the zebra finch *Taeniopygia guttata* genome.** *Genome Res* 2010, **20**:485–495.
72. Hildebrand F, Meyer A, Eyre-Walker A: **Evidence of selection upon genomic GC-content in bacteria.** *PLoS Genet* 2010, **6**:e1001107.
73. Bajic VB, Tan SL, Christoffels A, Schönbach C, Lipovich L, Yang L, Hofmann O, Kruger A, Hide W, Kai C, Kawai J, Hume DA, Carninci P, Hayashizaki Y: **Mice and men: their promoter properties.** *PLoS Genet* 2006, **2**:e54.
74. D'Onofrio G, Ghosh TC, Saccone S: **Different functional classes of genes are characterized by different compositional properties.** *FEBS Lett* 2007, **581**:5819–5824.
75. Ren L, Gao G, Zhao D, Ding M, Luo J, Deng H: **Developmental stage related patterns of codon usage and genomic GC content: searching for evolutionary fingerprints with models of stem cell differentiation.** *Genome Biol* 2007, **8**:R35.

doi:10.1186/1471-2164-13-604

Cite this article as: Matsubara et al.: Intra-genomic GC heterogeneity in sauropsids: evolutionary insights from cDNA mapping and GC₃ profiling in snake. *BMC Genomics* 2012 **13**:604.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

