JOURNAL OF Evolutionary Biology

doi: 10.1111/j.1420-9101.2010.02176.x

Paleogenomic data suggest mammal-like genome size in the ancestral amniote and derived large genome size in amphibians

C. L. ORGAN*, A. CANOVILLE†, R. R. REISZ‡ & M. LAURIN†

*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA
†UMR 7207, CNRS/MNHN/UPMC, Centre de Recherches sur la Paléobiodiversité et les Paléoenvironnements, Muséum National d'Histoire Naturelle, Département Histoire de la Terre, Paris, France
‡Department of Biology, University of Toronto Mississauga, Mississauga, ON, Canada

Keywords:

amniotes;
diapsids;
genome;
paleogenomics;
synapsids.

Abstract

An unsolved question in evolutionary genomics is whether amniote genomes have been expanding or contracting since the common ancestor of this diverse group. Here, we report on the polarity of amniote genome size evolution using genome size estimates for 14 extinct tetrapod genera from the Paleozoic and early Mesozoic Eras using osteocyte lacunae size as a correlate. We find substantial support for a phylogenetically controlled regression model relating genome size to osteocyte lacunae size (P of slopes < 0.01, $r^2 = 0.65$, phylogenetic signal (λ) = 0.83). Genome size appears to have been homogeneous across Paleozoic crown-tetrapod lineages (average haploid genome size 2.9–3.7 pg) with values similar to those of extant mammals. The differentiation in genome size and underlying architecture among extant tetrapod lineages likely evolved in the Mesozoic and Cenozoic Eras, with expansion in amphibians, contractions along the diapsid lineage, and no directional change within the synapsid lineage leading to mammals.

Introduction

Amniote genomes are both constrained and diverse across organizational tiers – a characteristic that has likely been a key to their evolutionary success (Organ *et al.*, 2008). For example, part of the adaptive success of early amniotes was an increased genetic repertoire, which had its origin in earlier vertebrates that underwent several rounds of whole genome duplication (Dehal & Boore, 2005). And, although whole genome sequencing projects consistently find that amniote species generally have just over 20 000 genes (Hillier *et al.*, 2004; Clamp *et al.*, 2007; Warren *et al.*, 2008, 2010; Hellsten *et al.*, 2010), the density of repetitive elements (Shedlock, 2006; Shedlock *et al.*, 2007) and number, shape, and size of chromosomes (Burt, 2002; Murphy *et al.*, 2005; Olmo, 2005) differ greatly across amniotes. Other aspects of genome biology, such as the diversity and function of conserved noncoding elements, remain a mystery in the larger context of amniotes (Gardiner *et al.*, 2006).

How these differences arose is a fundamental question in evolutionary genomics that can only fully understood by knowing the direction that evolution has taken from a common ancestor, a question that typically requires fossil data (Laurin, 2010). Increased comparative sampling of amniote species will help resolve these questions (G10KCOS, 2009), yet there are currently many more mammalian genomes sequenced than reptilian genomes. In addition to the chicken genome (Hillier *et al.*, 2004), the zebra finch (*Taeniopygia guttata*) and turkey (*Meleagris gallopavo*) genomes have been sequenced (Dalloul *et al.*, 2010; Warren *et al.*, 2010), and the green anole (*Anolis carolinensis*) genome is finished and under analysis. Because of its phylogenetic position, the *Anolis* genome will be critical for understanding broader evolutionary processes that have shaped the amniote genome.

Some questions will remain difficult to answer without fossil data, however. For example, mammals have genomes roughly twice the size of birds. How this difference evolved is related to the evolutionary dynamics of repetitive content, the size and number of introns

Correspondence: Chris Organ, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. Tel.: +617 496 9389; fax: +617 495 5667;
e-mail: corgan@oeb.harvard.edu

(Gregory, 2005), and the presence or absence of selection in relation to population size (Lynch, 2007). Yet, insights from chicken–mammal or chicken–mammal-*Anolis* comparisons are limited taxonomically and temporally (as only living species are sampled). The polarity of genome size evolution from the common amniote ancestor has remained a mystery as a consequence. Two alternative hypotheses can be proposed concerning the genome of the last common ancestor of amniotes (Shedlock *et al.*, 2007): (i) The ancestor had a small genome, within the range of extant birds, with a low density of transposable elements and (ii) the ancestor had a medium (mammal-sized) genome characterized by a high density of repetitive content. The extant sister group to amniotes is Amphibia, which have very large genomes (Gregory, 2010). When they are also considered, two more hypotheses may be proposed: (iii) The ancestor had a large (frog-sized) genome characterized by a high density of repetitive content and possibly higher orders of ploidy and (iv) The ancestor had an enormous (salamander-sized) genome characterized by a high density of repetitive content and possibly higher orders of ploidy.

In each scenario, genome size expands or contracts because of retroelement proliferation or extinction. Comparative genomic analysis of BAC-end sequences suggests that the common ancestor of amniotes had a high density of interspersed and simple sequence repeats (SSRs) (Shedlock *et al.*, 2007). This result suggests a polarity in the evolution of the amniote genome and implies that a gradual loss of retroelements explains the small genome of birds and some reptiles, whereas the genome size in mammals represents the ancestral condition. Contradicting these findings, recent research on retroelement decay suggests that genome size in ancestral mammals may have been much larger than those observed for extant species (Rho *et al.*, 2009).

It is often difficult to understand how traits evolve by looking at extant species alone, because we have $n = 1$ in terms of sampling through time (Laurin, 2010). Paleontological evidence greatly improves our understanding of how organismal traits evolved and the condition of those traits in the common ancestors of extant groups (e.g. Ji *et al.*, 2009). The same principle holds for genome biology. Moreover, genomic insights derived from fossils offer an independent test of hypotheses generated from sequencing-based studies. Here, we leverage paleohistological data to untangle the macroevolutionary patterns of genome size in Amniota. We use an approach that capitalizes on the well-known relationship between cell size and genome size (Gregory & Hebert, 1999; Gregory, 2000, 2001a,b, 2002), that we reassess here using an expanded dataset, and use Bayesian phylogenetic comparative methods to estimate the genome sizes of 14 extinct tetrapod taxa, which can be used to test the hypotheses described earlier by providing an independent estimate for the ancestral size of the amniote genome.

Materials and methods

Our general approach is to regress genome size on osteocyte lacunae size while accounting for phylogenetic relatedness and use this relationship to make phylogenetically informed predictions (retrodictions) of genome size in early extinct tetrapod taxa. Osteohistological thin sections were obtained for the following extant species: mudpuppy (*Necturus maculosus*), fire salamander (*Salamandra salamandra*), marbled salamander (*Ambystoma opacum*), seal salamander (*Desmognathus monticola*), cane toad (*Rhinella marina*), toad (*Bufo bufo*), mallard duck (*Anas platyrhynchos*), snapping turtle (*Chelydra serpentina*), Hermann's tortoise (*Testudo hermanni*), tuatara (*Sphenodon punctatus*), green lacerta lizard (*Lacerta viridis*), black spiny-tailed lizard (*Uromastyx acanthinura*), African lion (*Panthera leo*), dromedary camel (*Camelus dromedarius*), red fox (*Vulpes vulpes*), and dwarf armadillo (*Zaedyus pichiy*). We also generated thin sections for several early crown-tetrapod taxa. These include three Permian amphibians ('lepospondyls'), *Diplocaulus*, *Cardiocephalus*, and *Brachydectes*. There is a debate about the affinities of extant amphibians (Ruta & Coates, 2007; Anderson *et al.*, 2008; Marjanović & Laurin, 2009), but all studies indicate that these 'lepospondyls' are closely related to amniotes; thus, they are highly relevant to this study. A review of the evidence (Marjanović & Laurin, 2009) suggests that *Brachydectes* is the Paleozoic sister group of lissamphibians, but this pattern of relationships has yet to gain wide acceptance. We also sampled several early members of the main clades of Paleozoic amniotes. *Mesosaurus* is one of the earliest reptiles (Laurin & Reisz, 1995), and it is one of the earliest secondarily aquatic amniotes (Canoville & Laurin, 2010). It may be closely related to the procolophonid parareptile *Phaanthosaurus* and thus to turtles (Modesto, 1999; Lee, 2001), although most molecular phylogenies place turtles among diapsids (Rest *et al.*, 2003). Our sample also includes some of the earliest eureptiles, such as the captorhinids *Captorhinus* and *Labidosaurus*, an undetermined Romeriidan (presumably a 'protorothyridid'), all of which have anapsid skulls and fit outside Diapsida (Reisz, 1997). Finally, we sampled several Early Permian synapsids, including the caseid *Angelosaurus*, the varanopid *Mycterosaurus*, *Edaphosaurus boanerges* (one of the earliest herbivorous tetrapods), the sphenacodontids *Dimetrodon* and *Sphenacodon* (Reisz *et al.*, 1992), and the Early Jurassic cynodont *Bienotherium*. Thus, our taxonomic sample includes taxa of various habitats (terrestrial vs. aquatic) and trophic levels (herbivores and carnivores). See Tables S1 and S2 for more information about these data and taxa.

The volumes of osteocyte lacunae were estimated assuming ellipsoid ($4/3 \times \pi \times$ width axis radius$^2 \times$ length axis radius), with measurements taken in the program ImageJ (Rasband, 1997–2009) and Adobe Photoshop. To reduce the variance in the sampling protocol, and to help ensure that osteocytes were measured about their

mid-axis, only the largest cells in each section were measured. These data were combined with an osteocyte lacuna volume dataset from the literature (Organ & Shedlock, 2009), yielding a total of 54 extant tetrapod species. Haploid genome size data were obtained from http://www.genomesize.com (Gregory, 2010). Both forms of data were natural log transformed before analysis.

Our phylogeny was created in Mesquite v2.7 (Maddison & Maddison, 2009) and the Stratigraphic Tools package (Josse *et al.*, 2006). We used topologies and divergence times from the literature for the following groups: Amphibia (Evans *et al.*, 2004; Marjanović & Laurin, 2007), Mammalia (Springer *et al.*, 2004), Reptilia (Rest *et al.*, 2003), Squamata (Rest *et al.*, 2003; Vidal & Hedges, 2005; Wiens *et al.*, 2006; Kumazawa, 2007), Serpentes (Lawson *et al.*, 2005; Lee *et al.*, 2007; Vidal *et al.*, 2007), Testudines (Fujita *et al.*, 2004; Near *et al.*, 2005; Fritz & Bininda-Emonds, 2007), and Aves (Ericson *et al.*, 2006; Brown *et al.*, 2007). Branch lengths are in units of time and follow the standard geologic timescale (Walker & Geissman, 2009).

The program JMP v8 (SAS Institute, 2008) was used to explore data without reference to phylogeny. Phylogenetically controlled analyses were performed using the program BayesTraits (http://www.evolution.rdg.ac.uk), which accounts for the evolutionary nonindependence of trait data using phylogenetic generalized least squares (PGLS) (Pagel, 1997, 1999). BayesTraits uses a Markov chain Monte Carlo (MCMC) algorithm to produce posterior distributions of regression models while estimating the phylogenetic signal (λ) of the data, given the tree. The variance–covariance matrix derived from the tree (but scaled by λ) is used to phylogenetically normalize the data (and hence the residuals) during the estimation of the regression models. To integrate the extremely large genome sizes of salamanders in our analysis, we develop a multiple regression model derived from the general case $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + e$, where α and β are the parameters of the model and e is the error term. Consider $y = \alpha + \beta_1 x_1 + \beta_2 d + \beta_3 x_1 d + e$, which is the interaction of the continuous variable (osteocyte lacunae size) with a dummy variable (d) that separates one group (here the salamanders) from the other data points (other tetrapods) by assigning salamanders $d = 0$, and $d = 1$ to all other taxa. For example, for salamanders ($d = 0$), the model reduces to $y = \alpha + \beta_1 x_1 + e$, whereas the model for other taxa is $y = \alpha + \beta_1 x_1 + \beta_2 + \beta_3 x_1 + e$, or $y = (\alpha + \beta_2) + (\beta_1 + \beta_3)\ x_1 + e$. This is easily implemented using multiple regression in the phylogenetic least squares approach while estimating the degree to which the patterns in the data are predicted by the phylogeny (λ).

Predictions are made by adding the extinct species to the dataset and estimating the dependent variable (genome size) while accounting for uncertainty in the regression model by sampling the posterior distributions of the regression models derived from extant species. Predictions are transformed by reference to the variance–covariance matrix derived from the tree (but scaled by λ) and are therefore informed by the extinct species' phylogenetic position (topology and branch lengths). The MCMC for deriving the regression models for extant species and the MCMC for deriving the phylogenetically informed predictions lasted 5 001 000 iterations with a burnin of 50 000 and a sample period of 1000. The rate deviation determines the boldness of the proposal procedure of the MCMC, and we chose values consistent with acceptance rates that range between 0.2 and 0.4 (proportion of proposals accepted).

Before MCMC analysis, we used maximum likelihood (ML) to choose among models. Under ML, we used the Akaike information criterion (AIC) to choose between the single PGLS regression model and the PGLS multiple regression model detailed earlier. The AIC = −2 (log likelihood) + 2 K, where K is the number of free parameters (Burnham & Anderson, 2002). The difference between the AIC of the best model (smallest AIC) and each model's AIC is defined as ΔAIC (Δ_i). Akaike weights (w_i) are then calculated as $w_i = \exp(-\Delta_i/2)/\Sigma(\exp(-\Delta_i/2))$, where the model with the highest weight is preferred. The selected model was then analysed in MCMC, where the significance of regression models was assessed by comparing the proportion of the posterior distribution of slope parameters (β) that crossed 0 (the null model). Ancestral values of genome size were estimated using a random walk model (preliminary analysis showed no evidence for a directional model of trait evolution) in BayesTraits by producing posterior predictions of dummy taxa with zero-length branches placed at nodes of interest.

Results

Osteocyte lacunae measured in the extant taxa (Table S1) show variability among species and higher-ranking taxa. For example, we find that the average osteocyte size in the seal salamander (*Desmognathus monticola*, average lacunae volume = 343.58 μm^3, σ = 144.94) is much higher than the green lacerta lizard (*Lacerta viridis*, average lacunae volume = 114.34 μm^3, σ = 40.14) or the dwarf armadillo (*Zaedyus pichiy*, average lacunae volume = 249 μm^3, σ = 113.19). Variability in osteocyte volume within species is relatively high as well, as indicated by the large standard deviation values. Volumes for the extinct amphibians, reptiles and synapsids show the same levels of variability (Fig. 1).

We find substantial support for the multiple regression model over the single regression using AIC weights (Table 1). Bayes factor tests also support the multiple regression model (BF = 33). This phylogenetically controlled multiple regression model relates genome size to osteocyte lacunae volume, while separating tetrapods into two groups: salamanders and other tetrapods (Fig. 2). When analysed using MCMC, the posterior

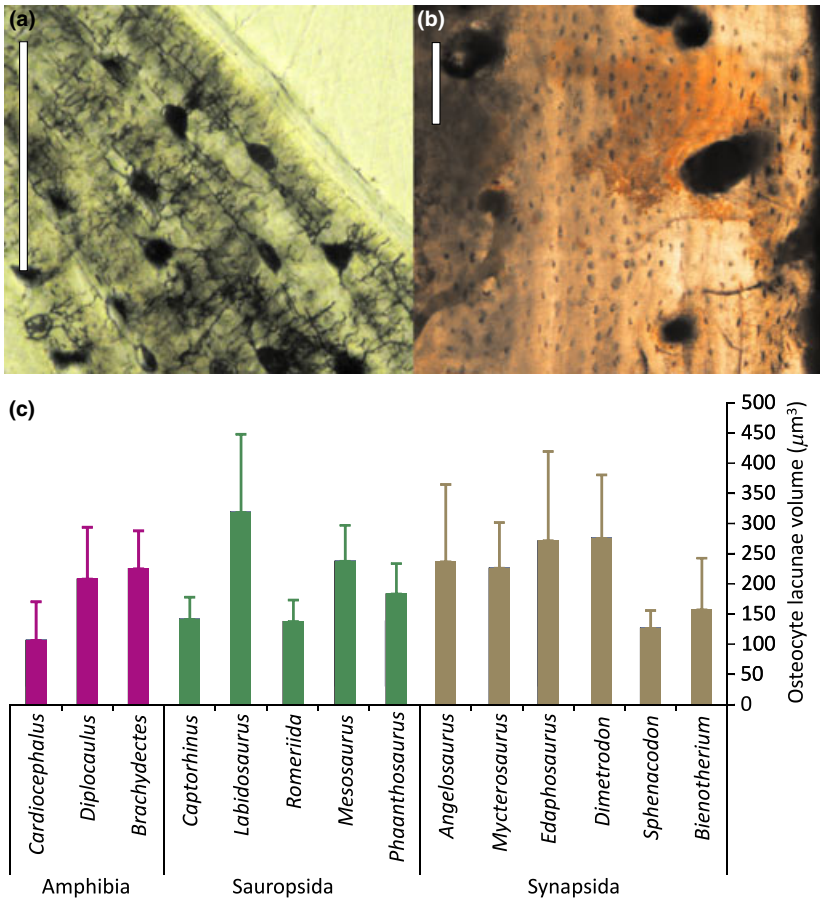

Fig. 1 Example micrographs and histological data used in this study. (a) Micrograph of compact bone tissue from a fire salamander (*Salamandra salamandra*) humerus. (b) Micrograph of compact bone tissue from the femur of an Early Jurassic cynodont (*Bienotherium*). The scale bars for panels a and b are equal to 100 μm. (c) Distributions of osteocyte lacunae volumes for three extinct taxa of 'lepospondyl' amphibians, five sauropsid taxa and six synapsid taxa, representing 348 measurements.

Table 1 Model selection.

Model	LH	*n*	*K*	AIC	Δ_i	exp(−Δ_i/2)	w_i
PGLS regression	−14.99	54	3	36.48	30.21	0.00	0.00
Mult. PGLS regression	2.49	54	5	6.26	0.00	1.00	1.00

The PGLS regression model was selected using the Akaike information criterion (AIC). The multiple regression model used dummy variables to group salamanders differently from other tetrapods. Δ_i, difference between the AIC of the best fitting model and that of the model i; *K*, number of parameters; LH, log-likelyhood; w_i, Akaike weights; PGLS, phylogenetic generalized least squares.

distributions of slope coefficients deviate strongly from 0, the null hypothesis, indicating substantial support (per cent of β_1 crossing 0 = 0.0001; per cent of β_2 crossing 0 = 0.002; per cent of β_3 crossing 0 = 0.0001). The median for the posterior distribution of r^2 (0.65) indicates that, using this model, osteocyte volume can account for about two-thirds of the variance in genome size in tetrapods. We also find large values for phylogenetic signal (λ, median = 0.83) with our data and the given tree. λ transforms the traits before regression analysis, which is the same as transforming the residuals and indicates that genome size and bone cell size covary according to the tree.

Genome size in extinct taxa can be inferred using the phylogenetically controlled multiple regression model (Table 2). Using this approach, we find that estimated genome sizes for early amniotes and amphibians ('lepospondyls') all range around 3 pg (average genome size 2.94–3.69 pg). When the 'lepospondyls' are estimated with the salamander model, their genome sizes are moderately large (between 3.96 and 9.83 pg). However, there is no *a priori* justification for treating the lepospondyls as having a salamander-like genome given their phylogenetic position, anatomy, and the outlier genome

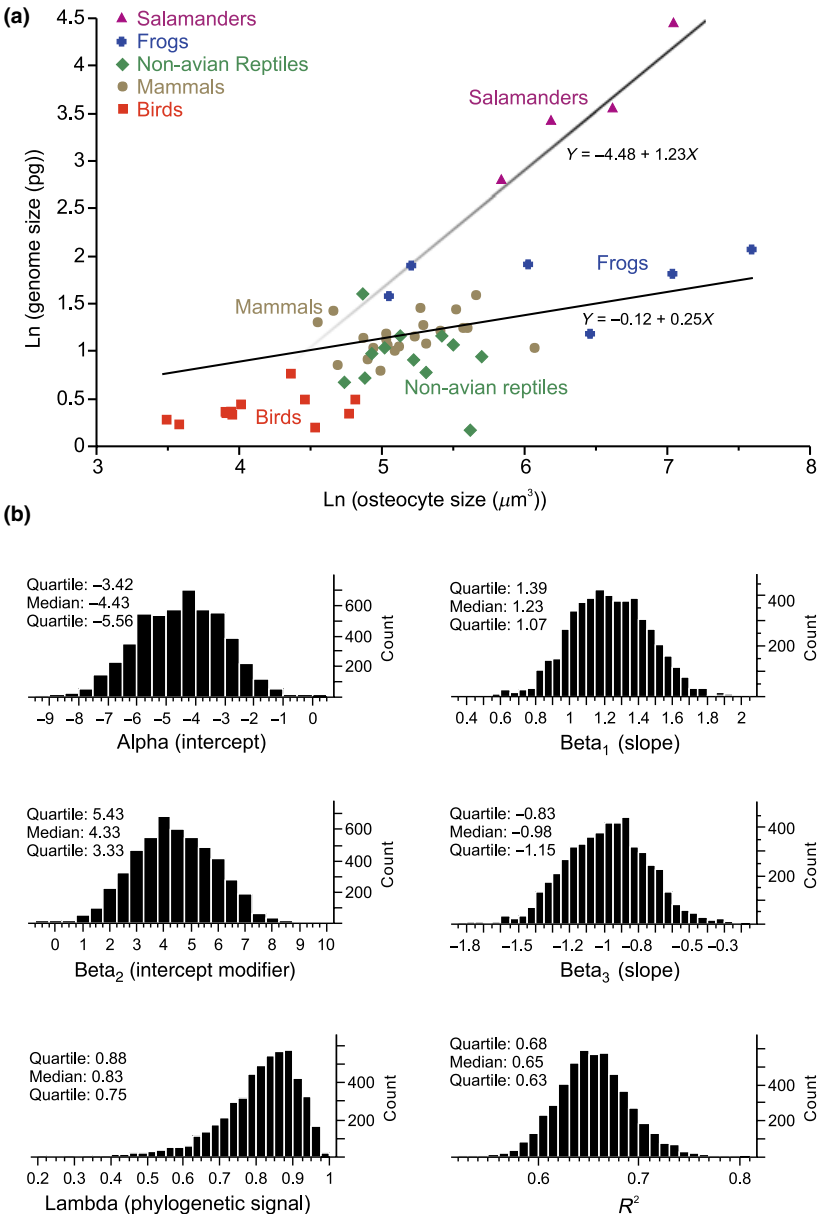

Fig. 2 Phylogenetic generalized multiple regression model relating genome size to osteocyte lacunae volume. (a) The two lines of the multiple regression model, which is Ln (genome size) = −4.48 + 1.23 * Ln (cell size) + 4.36 * (0.1)−0.98 * Ln(cell size) * (0.1); $R^2 = 0.65$. All coefficients are substantially different from 0 (β_1 $P < 0.0001$, β_2 $P = 0.002$, β_3 $P < 0.0001$). (b) The estimated distribution of the estimates of the various parameters of the regression model.

size status of Urodeles. The predicted distributions of genome size for early sauropsids and early synapsids are similar (Table 2), and although the extinct sauropsids are slightly smaller on average than the genome sizes of early synapsids, the difference is not significant (*t*-test $P = 0.23$).

We mapped the means of the predictive distributions for the extinct taxa onto the phylogeny (Fig. 3). For this mapping, all inferences were obtained using the dummy variable that groups early taxa with non-salamanders. This mapping highlights the large genome sizes of extant amphibians relative to all early (and other extant) tetrapods. We used the predicted genome sizes to infer the ancestral amniote genome size using a random walk model. We estimate that the ancestral genome size was 3.34 pg ($\sigma = 0.34$), or 7.45 pg ($\sigma = 1.17$) if the predictions used salamander predictive model (although this would be difficult to justify).

Table 2 Inferences of haploid genome size in extinct tetrapods.

	Estimated genome size (pg)		Estimated genome size (pg), salamander biased	
Taxa	Mean	SD	Mean	SD
Extinct amphibians				
Cardiocephalus	3.03	0.53	3.96	1.47
Diplocaulus	3.54	0.62	8.77	2.59
Brachydectes	3.69	0.52	9.83	2.63
Extinct sauropsids				
Captorhinus	2.99	0.51	5.18	1.81
Labidosaurus hamatus	3.65	0.63	13.78	3.75
Romeriida Incertae sedis	2.94	0.42	4.94	1.67
Mesosaurus	3.4	0.54	9.64	2.83
Phaanthosaurus	3.17	0.55	7.01	2.26
Extinct synapsids				
Angelosaurus	3.51	0.63	9.96	2.98
Mycterosaurus	3.46	0.52	9.36	2.67
Edaphosaurus boanerges	3.66	0.62	11.82	3.33
Dimetrodon	3.67	0.57	11.99	3.31
Sphenacodon	3.03	0.43	4.71	1.64
Bienotherium	3.33	0.69	6.34	2.28

The means and standard deviations from the posterior predictive distributions. The first two columns of statistics were obtained by grouping the extinct taxa with non-salamanders in the multiple regression model. The last two columns were obtained by grouping the extinct taxa with salamanders in the multiple regression model.

Discussion

We present a multiple regression model relating genome size to bone cell size in extant tetrapods (Fig. 2) that shows a different slope for salamanders compared with other tetrapods. A similar pattern has also been reported for blood cells (Olmo & Morescalchi, 1978; Gregory, 2001b), suggesting that the extremely large genomes of salamanders relate to cell morphology differently than do the genomes of other tetrapods. We use the multiple regression model to estimate the genome sizes for 14 extinct tetrapod taxa from measurements of osteocyte lacunae in their fossil remains. Our results substantially improve estimates of ancestral amniote genome architecture that are based only on extant taxa, because we employ taxa sampled near the early divergence in Amniota (between synapsids and saurposids), which dates to about 310–345 Ma ago (Marjanović & Laurin, 2007). Few crown-tetrapods (Laurin, 1998) older than those sampled here could have been used in our analysis. The oldest undisputed reptiliomorphs date from the Late Carboniferous (about 315 Ma ago) and although several Carboniferous amphibians are known, many are represented by flattened, poorly preserved bone that has been prepared by acid-etching and removal of the bone to produce casts. The few well-preserved Carboniferous crown-tetrapod remains are sufficiently rare to make curators reluctant to any destructive study, such as required for making histological thin sections. The tetrapod crown is only slightly older than Amniota, probably dating to about 332–360 Ma ago. Thus, estimates for the tetrapod crown should be well-constrained by taxa that date from 35–80 Ma after the origin of the clade. It is unlikely that trends could seriously bias our estimates of crown-tetrapod and amniote genome size given the geological age of the sampled taxa (most date from the Early Permian, 280–300 Ma ago).

Our results suggest that the massive genome sizes observed in extant salamanders (average = 35.90 pg, $\sigma = 1.05$, $n = 217$) are likely derived, because the extinct 'lepospondyl' amphibians in our dataset have much smaller cell sizes and consequently estimated genome sizes (means for the predictive distributions range from 3.03–3.69 pg). We find that 'lepospondyls' had genome sizes smaller than those of frogs (average = 4.68 pg, $\sigma = 0.13$, $n = 248$) and more comparable with early amniotes. This finding is consistent with previous non-phylogenetically controlled analyses of lungfish and amphibians (Thomson & Muraszko, 1978) and suggests that the modern amphibian genome has expanded during its evolutionary history, by enlarging introns as well as increasing repetitive content (Smith *et al.*, 2009), especially in early salamanders. Most of this increase in genome size must have occurred sometime between the late Carboniferous (time of divergence between *Brachydectes* and lissamphibians) and the Cenozoic (the time at which much of the lissamphibian diversification had already occurred (Marjanović & Laurin, 2007). Better constraining the timing of this event would require sections of Mesozoic lissamphibians.

Our estimation of the ancestral amniote genome size at 3.34 pg is close to the average observed for extant mammals (average = 3.45 pg, $\sigma = 0.81$, $n = 613$), thereby matching predictions based on the study of repetitive elements (Shedlock *et al.*, 2007; Janes *et al.*, 2010). These results are incongruent with the hypothesis that genome size contracted in parallel mammalian lineages during the Cenozoic (Rho *et al.*, 2009). Furthermore, we show evidence that early reptiles (sauropsids) had genome sizes (average = 3.23 pg, $\sigma = 0.3$, $n = 5$) larger than the average for extant non-avian reptiles (average = 2.24 pg, $\sigma = 0.04$, $n = 320$), suggesting genomic contraction on the line to extant reptiles, a trend most accentuated in birds. On the synapsid side of the amniote tree, we find evidence that the genome size of early members did not differ from the genome size seen in mammals (the only extant synapsids).

How was the ancestral amniote genome structured? Base-pair composition provides a basic metric for describing genomes. The base-pair fabric of the ancestral amniote genome was likely low in global GC content (42%), similar to the condition seen in mammals (Shedlock *et al.*, 2007). Emerging evidence from the genomes of reptiles and monotremes also suggests an ancestral amniote genome that was structurally diverse

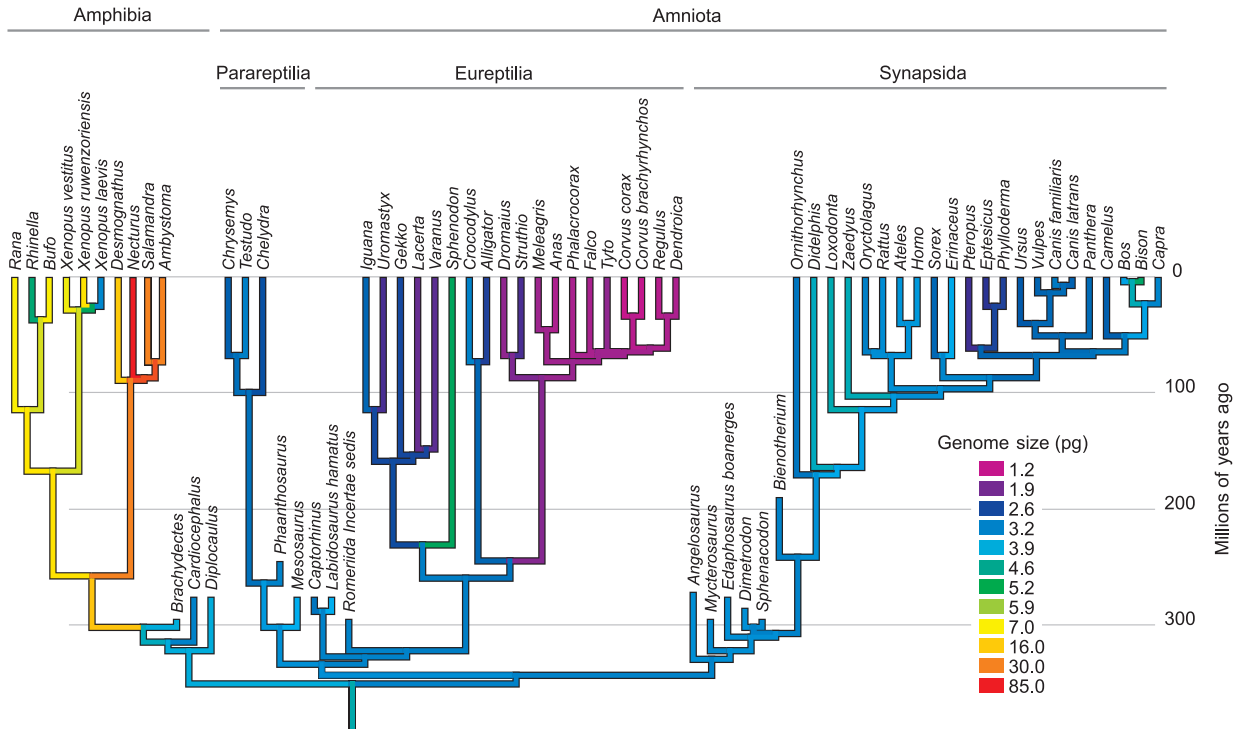

Fig. 3 Inferred genome sizes using the multiple regression model for the extinct amniotes and amphibians. The means of the predictive distributions mapped onto a phylogeny. For this mapping, inferences on all early taxa are derived from the multiple regression model with extinct amphibians and amniotes coded as not salamander-like.

with mammal-like abundances of retroelements and simple sequence repeats (Warren *et al.*, 2008; Janes *et al.*, 2010). Specifically, the repetitive landscape of the ancestral amniote genome was likely abundant in AT-rich SSRs, chicken repeat 1 (CR1) and mammalian interspersed repeat (MIR) retroelements (Shedlock *et al.*, 2007). The comparative findings of Shedlock *et al.* (2007) suggest that long interspersed elements (LINEs) and related mobile elements found in mammals replaced CR1 elements during the early evolution of amniotes. Multiple sequential reductions in repeat diversity probably also occurred in reptiles. Our findings are consistent with these conclusions, because the size and abundance of repetitive elements and introns are the most influential determinants of genome size in animals (Waltari & Edwards, 2002; Shedlock, 2006).

Animal genomes are compartmentalized into chromosomes, and although our results cannot directly speak to the ancestral amniote karyotype, comparisons among extant species do provide some insights. For example, many small-scale chromosomal rearrangements and deletions likely occurred sometime in neognath birds after the neognath/paleognath divergence (Chapus & Edward, 2009). In the same study, small deletions were found to occur more widely in reptiles than small insertions, which may help explain the contraction of the avian genome by 50% compared with other reptiles. Paleogenomic work similar to that presented here has dated a large fraction of this contraction to sometime within early saurischian dinosaur evolution between 230 and 250 million years ago (Organ *et al.*, 2007). Further genomic contractions also occurred within Aves (Gregory *et al.*, 2009), suggesting several episodes of genome size reduction on the line to birds as opposed to a single event. The expanded sampling of extant and extinct taxa of this study strengthens the conclusion that the avian genome underwent strong size reduction by better constraining the ancestral amniote and tetrapod genome size.

In this report, we estimated genome sizes for 14 extinct tetrapod taxa and used these estimates to constrain the genome size of the common amniote ancestor. Because genome size in animals is determined to a great extent by the repetitive landscape and the size of introns (Waltari & Edwards, 2002; Shedlock, 2006), these results speak to the structure of the ancestral genome. Furthermore, our results match the predictions of repetitive element analysis in reptiles – that the ancestral amniote genome was mammalian-sized (Waltari & Edwards, 2002; Shedlock, 2006; Shedlock *et al.*, 2007), strengthening the conclusions of both studies. Our findings are all the more relevant for interpreting the evolutionary genomic

dynamics of extant reptiles, because with the *Anolis* genome project nearing completion, we will finally be able to understand the non-avian reptile genome from a comparative standpoint. Plans are being laid for sequencing the genomes of other birds, lepidosaurs and turtles by various research groups, and our results should help place them in an evolutionary context by polarizing the evolutionary trends of genome size and repetitive element abundance in amniotes.

Acknowledgments

C. Organ thanks S. V. Edwards for postdoctoral support, and M. Pagel and N. Hobbs for discussions that improved this research. The authors are very grateful to M. M. Loth, H. Lamrous and M. Lemoine for some histological preparations and Pr. A. de Ricqlès who has lent some histological slides of extinct taxa for this study.

References

Anderson, J.S., Reisz, R.R., Scott, D., Fröbisch, N.B. & Sumida, S.S. 2008. A stem batrachian from the Early Permian of Texas and the origin of frogs and salamanders. *Nature* **453**: 515–518.

Brown, J.W., Payne, R.B. & Mindell, D.P. 2007. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on ericson *et al*. *Biol. Lett.* **3**: 257–259.

Burnham, K.P. & Anderson, D.R. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York.

Burt, D.W. 2002. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96**: 97–112.

Canoville, A. & Laurin, M. 2010. Evolution of humeral microanatomy and lifestyle in amniotes, and some comments on paleobiological inferences. *Biol. J. Linn. Soc.* **100**: 384–406.

Chapus, C. & Edward, S.V. 2009. Genome evolution in Reptilia: in silico chicken mapping of 12,000 BAC-end sequences from two reptiles and a basal bird. *BMC Genomics* **10**: 1–13.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M.F. *et al.* 2007. Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Nat. Acad. Sci.* **104**: 19428–19433.

Dalloul, R.A., Long, J.A., Zimin, A.V., Aslam, L., Beal, K., Ann Blomberg, L. *et al.* 2010. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* **8**: e1000475.

Dehal, P. & Boore, J.L. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**: e314.

Ericson, P.G., Anderson, C.L., Britton, T., Elzanowski, A., Johansson, U.S., Kallersjo, M., Ohlson, J.I., Parsons, T.J., Zuccon, D. & Mayr, G. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol. Lett.* **2**: 543–547.

Evans, B.J., Kelley, D.B., Tinsley, R.C., Melnick, D.J. & Cannatellae, D.C. 2004. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Mol. Phylogenet. Evol.* **33**: 197–213.

Fritz, U. & Bininda-Emonds, O.R.P. 2007. When genes meet nomenclature: tortoise phylogeny and the shifting generic concepts of *Testudo* and *Geochelone*. *Zoology* **110**: 298–307.

Fujita, M.K., Engstrom, T.N., Starkey, D.E. & Shaffer, H.B. 2004. Turtle phylogeny: insights from a novel nuclear intron. *Mol. Phylogenet. Evol.* **31**: 1031–1040.

G10KCOS 2009. Genome 10K: a proposal to obtain whole-genome sequence for 10 000 vertebrate species. *J. Hered.* **100**: 659–674.

Gardiner, E.J., Hirons, L., Hunter, C.A. & Willett, P. 2006. Genomic data analysis using DNA structure: an analysis of conserved nongenic sequences and ultraconserved elements. *J. Chem. Inf. Model.* **46**: 753–761.

Gregory, T.R. 2000. Nucleotypic effects without nuclei: genome size and erythrocyte size in mammals. *Genome* **43**: 895–901.

Gregory, T.R. 2001a. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev. Camb. Philos. Soc.* **76**: 65–101.

Gregory, T.R. 2001b. The bigger the C-value, the larger the cell: genome size and red blood cell size in vertebrates. *Blood Cells Mol. Dis.* **27**: 830–843.

Gregory, T.R. 2002. A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves. *Evolution* **56**: 121–130.

Gregory, T.R. 2005. Genome size evolution in animals. In: *The Evolution of the Genome* (T.R. Gregory, ed.), pp. 4–71. Elsevier Academic Press, Boston.

Gregory, T.R. 2010. Animal genome size database. Available at: http://www.genomesize.com/.

Gregory, T.R. & Hebert, P.D.N. 1999. The modulation of DNA content: proximate causes and ultimate consequences. *Genome Res.* **9**: 317–324.

Gregory, T.R., Andrews, C.B., McGuire, J.A. & Witt, C.C. 2009. The smallest avian genomes are found in hummingbirds. *Proc. R. Soc. B Biol Sci* **276**: 3753–3757.

Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V. *et al.* 2010. The genome of the Western Clawed frog *Xenopus tropicalis*. *Science* **328**: 633–636.

Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P. *et al.* 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695–716.

Janes, D.E., Organ, C.L., Fujita, M.K., Shedlock, A.M. & Edwards, S.V. 2010. Genome evolution in Reptilia, the sister group of mammals. *Annu. Rev. Genet.* **11**: 239–264.

Ji, Q., Luo, Z.-X., Zhang, X., Yuan, C.-X. & Xu, L. 2009. Evolutionary development of the middle ear in Mesozoic therian mammals. *Science* **326**: 278–281.

Josse, S., Moreau, T. & Laurin, M. 2006. Stratigraphic tools for Mesquite. Available at: http://mesquiteproject.org/packages/stratigraphicTools/.

Kumazawa, Y. 2007. Mitochondrial genomes from major lizard families suggest their phylogenetic relationships and ancient radiations. *Gene* **388**: 19–26.

Laurin, M. 1998. A reevaluation of the origin of pentadactyly. *Evolution* **52**: 1476–1482.

Laurin, M. 2010. Assessment of the relative merits of a few methods to detect evolutionary trends. *Syst. Biol.* **59**: 689–704.

Laurin, M. & Reisz, R.R. 1995. A reevaluation of early amniote phylogeny. *Zool. J. Linn. Soc.* **113**: 165–223.

Lawson, R., Slowinski, J.B., Crother, B.I. & Burbrink, F.T. 2005. Phylogeny of the Colubroidea (Serpentes): new evidence from mitochondrial and nuclear genes. *Mol. Phylogenet. Evol.* **37**: 581–601.

Lee, M.S.Y. 2001. Molecules, morphology, and the monophyly of diapsid reptiles. *Contrib. Zool.* **70**: 1–18.
Lee, M.S.Y., Hugall, A.F., Lawson, R. & Scanlon, J.D. 2007. Phylogeny of snakes (Serpentes): combining morphological and molecular data in likelihood, Bayesian and parsimony analyses. *System. Biodivers.* **5**: 371–389.
Lynch, M. 2007. *The Origins of Genome Architecture*. Sinauer Associates, Sunderland.
Maddision, W.P. & Maddison, D.R. 2009. Mesquite: a modular system for evolutionary analysis. Version 2.7. Available at: http://mesquiteproject.org.
Marjanović, D. & Laurin, M. 2007. Fossils, molecules, divergence times, and the origin of lissamphibians. *Syst. Biol.* **56**: 369–388.
Marjanović, D. & Laurin, M. 2009. The origin(s) of modern amphibians: a commentary. *Evo Biology* **36**: 336–338.
Modesto, S.P. 1999. Observations on the structure of the Early Permian reptile *Stereosternum temidum* Cope. *Palaeontol. Afr.* **35**: 7–19.
Murphy, W.J., Larkin, D.M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L. *et al.* 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**: 613–617.
Near, T., Meylan, P. & Shaffer, H.B. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *Am. Nat.* **165**: 137–146.
Olmo, E. 2005. Rate of chromosome changes and speciation in reptiles. *Genetica* **125**: 185–203.
Olmo, E. & Morescalchi, A. 1978. Genome and cell size in frogs: a comparison with salamanders. *Experientia* **34**: 44–46.
Organ, C.L. & Shedlock, A.M. 2009. Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biol. Lett.* **5**: 47–50.
Organ, C.L., Shedlock, A.M., Meade, A., Pagel, M. & Edwards, S.V. 2007. Origin of avian genome size and structure in nonavian dinosaurs. *Nature* **446**: 180–184.
Organ, C.L., Moreno, R.G. & Edwards, S.V. 2008. Three tiers of genome evolution in reptiles. *Integr. Comp. Biol.* **48**: 494–504.
Pagel, M.D. 1997. Inferring evolutionary processes from phylogenies. *Zool. Scr.* **26**: 331–348.
Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884.
Rasband, W.S. 1997–2009. *ImageJ*, 1.42v edn. U. S. National Institutes of Health, Bethesda. Available at: http://rsb.info.nih.gov/ij/.
Reisz, R.R. 1997. The origin and early evolutionary history of amniotes. *Trends Ecol. Evol.* **12**: 218–222.
Reisz, R.R., Berman, D. & Scott, D. 1992. The cranial anatomy of *Secodontosaurus obtusidens*, an unusual mammal-like reptile (Synapsida: Sphenacodontidae) from the Lower Permian of Texas. *Zool. J. Linn. Soc.* **104**: 127–184.
Rest, J.S., Ast, J.C., Austin, C.C., Waddell, P.J., Tibbetts, E.A., Hay, J.M. & Mindell, D.P. 2003. Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol. Phylogenet. Evol.* **29**: 289–297.
Rho, M., Zhou, M., Gao, X., Kim, S., Tang, H. & Lynch, M. 2009. Independent mammalian genome contractions following the KT boundary. *Genome Biol. Evol.* **2009**: 2–12.
Ruta, M. & Coates, M.I. 2007. Dates, nodes and character conflict: addressing the lissamphibian origin problem. *J. Syst. Paleontol.* **5**: 69–122.
SAS Institute 2008. *JMP, Version 8*. SAS Institute Inc., Cary, NC.
Shedlock, A.M. 2006. Phylogenomic investigation of CR1 LINE diversity in reptiles. *Syst. Biol.* **55**: 902–911.
Shedlock, A.M., Botka, C.W., Zhao, S., Shetty, J., Zhang, T., Liu, J.S., Deschavanne, P.J. & Edwards, S.V. 2007. Phylogenomics of non-avian reptiles and the structure of the ancestral amniote genome. *Proc. Nat. Acad. Sci.* **104**: 2767–2772.
Smith, J., Putta, S., Zhu, W., Pao, G., Verma, I., Hunter, T., Bryant, S., Gardiner, D., Harkins, T. & Voss, S.R. 2009. Genic regions of a large salamander genome contain long introns and novel genes. *BMC Genomics* **10**: 19.
Springer, M.S., Stanhope, M.J., Madsen, O. & de Jong, W.W. 2004. Molecules consolidate the placental mammal tree. *Trends Ecol. Evol.* **19**: 430–438.
Thomson, K.S. & Muraszko, K. 1978. Estimation of cell size and DNA content in fossil fishes and amphibians. *J. Exp. Zool.* **205**: 315–320.
Vidal, N. & Hedges, S.B. 2005. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. *C. R. Biol.* **328**: 1000–1008.
Vidal, N., Delmas, A.-S., David, P., Cruaud, C., Couloux, A. & Hedges, S.B. 2007. The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. *C. R. Biol.* **330**: 182–187.
Walker, J.D. & Geissman, J.W. 2009. 2009 Geologic Time Scale. *GSA Today* **19**: 60–61.
Waltari, E. & Edwards, S.V. 2002. Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *Am. Nat.* **160**: 539–552.
Warren, W.C., Hillier, L.W., Marshall Graves, J.A., Birney, E., Ponting, C.P., Grützner, F. *et al.* 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**: 175–183.
Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Kunstner, A. *et al.* 2010. The genome of a songbird. *Nature* **464**: 757–762.
Wiens, J.J., Brandley, M.C. & Reeder, T.W. 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snake-like body form in squamate reptiles. *Evolution* **60**: 123–141.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Bone histology and genome size data for extant species.

Table S2 Summary of bone histology data for extinct species.

Table S3 Genome size statistics in extant taxa.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received 3 August 2010; revised 10 October 2010; accepted 12 October 2010