

Origin of Avian Genome Size and Structure in Nonavian Dinosaurs

Chris L. Organ, Andrew M. Shedlock, Andrew Meade, Mark Pagel, and Scott V. Edwards

Table of Contents

Notes on methods and results	3	
Histological and genome size data		3
Sources for phylogenetic trees	5	
Regression model	6	
Model/method checking		7
Phylogenetically corrected test for differences among groups		8
Inference of genome size		8
Inference of interspersed repetitive element density		10
Trees and character matrices		13
Reference information for sequence data used in repeat analysis		
18		
References	21	

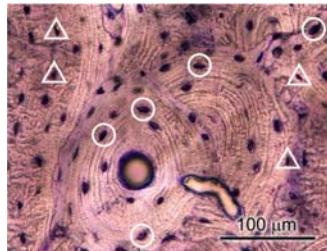
Supplementary Figures

Figure 1: Overview of Methods	2
-------------------------------	---

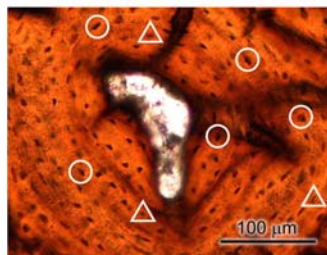
Supplementary Tables

Table 1: Summary of Bone Histology Data	4
Table 2: Evaluation of Bayesian GLS Regression Inferences	
8	
Table 3: Summary of GLS regressions for Cell Size and Genome Size	9
Table 4: Summary of Inferences for Dinosaurs	9
Table 5: Summary of Repetitive Data	11
Table 6: Summary of Regressions for Genome Size and Repetitive Element Density	
12	
Table 7: Data and Taxa for Trees 1 and 2	13
Table 8: Data and Taxa for Trees 3 and 4	15

Supplementary Figure 1: Overview of Methods



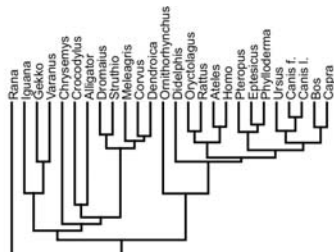
Step 1. We measured the size of lacunae in thin sections of bone from 26 extant vertebrate species from which genome size is known. Circles indicate types of cells that were measured and triangles indicate cells that were not measured.



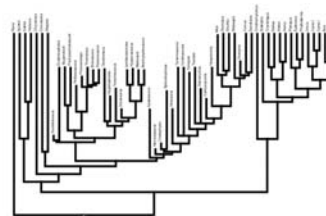
Step 4. We then measured the size of lacunae in thin sections of bone from 31 extinct dinosaur species.

```
TGTGGTGGATTCTCACATTTATAGGTGCCATTATAT
AACCTGTAGGATTCAAGATGTTTCAGAAACAACCCAG
AAATGGACCTTTTTGCATATATTTGATGTTAGCTCA
CTTGTCTGCAGGATGAATATATTTTCATTGAGGTAT
GTTGATAGCAATGTTTAATAGTGACATGTAGTCTCT
CTTGGCATATTTGCATAATTTGCACCAATATGAAA
GCAATACATAAATCAGGGAGCTGTTTGCAGATGAGA
CAGAAAGTTTACTAGTATGAAACAATTTAAAATCTTG
TCAATCTATTTAAGGTTCAAATTTGGTCTATTCATAT
CTCTGACAAAACAATACCAATATGTTTCACCCGTGA
```

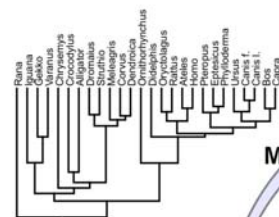
Step 7. RepeatMasker was used to summarize repeat density and GC content in BAC clone sequences covering over 119 Mb downloaded from Genbank and random regions along scaffolds from published genomes on the UCSC genome browser.



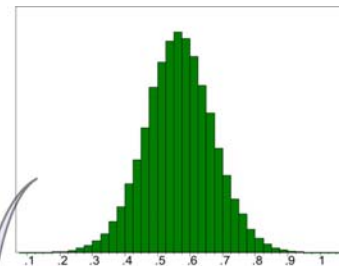
Step 2. A phylogenetic tree of the 26 species was assembled from published sources with branch lengths estimated from the fossil record.



Step 5. The 31 extinct dinosaur species were then included in the phylogenetic tree.

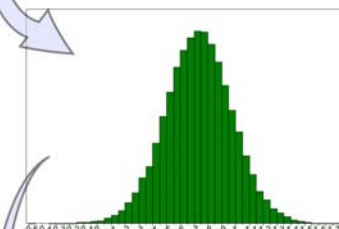


Step 8a and b. We repeated steps 2 and 3, but for repeat density and genome size.



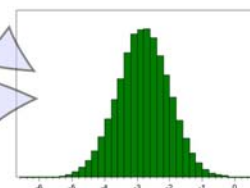
Step 3. Bayesian posterior distribution of regression models for genome size & osteocyte size using a modified version of Continuous (Pagel 1997), which accounts for phylogenetic non-independence among species & for uncertainty about the true regression model.

MCMC



Step 6. We then generated a Bayesian posterior distribution of genome size predictions in the 31 dinosaurs. During this step the Markov chain sampled the distribution of the posterior regression model from the extant species to incorporate uncertainty in the predictions.

MCMC



Step 9. We used the modified version of Continuous to generate a posterior distribution of predictions for interspersed repetitive DNA density in the 31 dinosaurs. Accounting for uncertainty: The posterior predictive distributions of repeat density were derived from a Markov chain that sampled the posterior distributions of the genome size - repetitive DNA regression models estimated from the extant species as well as the posterior predictive distributions of dinosaur genome size from step 6.

MCMC

Notes on Methods and Results

Histology and Genome Size Data

Histological slides from extant animals were obtained from the collections of the Museum of Comparative Zoology at Harvard University. Histological slides from additional extant animals and extinct dinosaurs were collected from the Department of Paleontology at the Museum of the Rockies in Bozeman, Montana. Slides at the Museum of Comparative Zoology were examined at the Imaging Center in the Department of Cellular and Molecular Biology, Harvard University. A Leica DMRB microscope equipped with a digital camera was used to examine the slides. The digital camera was connected to a computer running Openlab 4.0, which was used to calibrate scale and capture digital micrographs. Slides at the Museum of the Rockies were examined in the Gabriel Laboratory for Cellular and Molecular Paleontology (Department of Paleontology). Thin sections were examined using a Nikon Eclipse e600 polarized microscope and a Nikon Sight Ds-L1 digital camera captured digital photos. See Horner and Padian et al, (2001); de Ricqlès and Padian et al. (2003) and other references therein for examples on the methods used to prepare osteological and paleohistological thin sections. Microanatomical terminology follows established nomenclature for bone (de Ricqlès 1980; Francillon-Vieillot, de Buffrénil et al. 1990; Reid 1996; Currey 2002).

Digital images of histological slides taken at 200x magnification were randomly assigned labels and osteocyte lacunae size data were collected blind. A total of 825 measurements were taken from 26 extant vertebrate species and 598 measurements were taken from extinct dinosaur species (Supplementary Table 1). Age classes ranged from juvenile to adult. Osteocyte volume (lacunae) was determined in Adobe Photoshop CS using the measure tool to calculate the length and width of osteocyte lacunae. Volumes were then calculated assuming an ellipsoid shape ($\frac{4}{3} \times \pi \times \text{width axis radius}^2 \times \text{length axis radius}$).

To minimize variation and help make measurements more consistent osteocyte lacunae were measured from compact bone tissue, both primary and secondary (if present), from each slide. Canaliculi were excluded from these measurements. Osteocyte lacunae size has been shown to be relatively uniform in compact bone in the diaphyseal regions of long bones across ontogeny (Canè, Marotti et al. 1982), though osteocyte size is known to vary among bones within an individual (de Ricqlès, Meunier et al. 1991), increasing the variation in our dataset. Only larger cells were measured because osteocyte lacunae are sliced at different planes during sectioning. This procedure helped ensure that cells were measured near their mid-axis (see Fig. 1a in the main text for examples of the types of cells that were measured and excluded). Also of concern is a process called osteocytic osteolysis, known to be induced by parathyroid hormone (Tazawa, Hoshi et al. 2004; Hall 2005), in which standing

osteocytes resorb (and may subsequently deposit) bone around the osteocyte lacunae, changing their size. Although known to occur in a wide range of animals, this process is thought to be limited to situations such as hibernation or pregnancy that require increased mobilization of minerals from the skeleton (Franz-Odenaal, Hall et al. 2005). While such a process may add noise to our osteocyte size dataset, it is clearly not large enough to obscure patterns in osteocyte size across taxonomic groups. Concerning paleohistology, despite color changes caused by infiltration of sediment solute, it is commonly recognized that most fossil bone is unaltered by bacterial invasion or diagenetic factors, leaving histological features relatively unaltered (Horner, de Ricqlès et al. 2000).

Data on genome size were obtained from the Animal Genome Size Database (Gregory 2005). Genome sizes were averaged for species with multiple entries. Genome size and cell size data were natural log transformed; an established practice for normalizing comparisons of genome and cell size due to large ranges in size (Olmo 1983; Gregory 2000; Gregory 2001; Olmo 2003; Vinogradov and Anatskaya 2006).

Supplementary Table 1: Summary of Bone Histology Data

Extant Taxon	Bone Element	N lacunae	Avg Vol (μm^3)	Vol 95% CI	C-Val
<i>Alligator mississippiensis</i>	Femur	33	197.98	25.56	2.49
<i>Ateles paniscus</i>	Turbinates	20	290.3	50.55	3.47
<i>Bos taurus</i>	Tibia	37	207.04	20.45	3.59
<i>Canis familiaris</i>	Turbinates	20	142.97	26.64	3.12
<i>Canis latrans</i>	Femur	40	149.18	19.51	2.82
<i>Capra hircus</i>	Tibia	40	160.86	15.9	3.24
<i>Chrysemys picta</i>	Humerus	40	166.97	24.93	2.8
<i>Corvus corax</i>	Turbinates	20	95.12	9.58	1.21
<i>Crocodylus niloticus</i>	Humerus	37	245.81	31.86	3.21
<i>Dendroica dominica</i>	Turbinates	20	53.27	5.46	1.4
<i>Didelphis virginiana</i>	Femur	40	110.35	9.49	4.12
<i>Dromaius novaehollandiae</i>	Tibia	40	95.22	12.76	1.63
<i>Eptesicus fuscus</i>	Turbinates	20	115.38	17.96	2.37
<i>Gekko sp.</i>	Skull	38	337.2	53.93	2.56
<i>Homo sapiens</i>	Femur	20	274.04	39.05	3.5
<i>Iguana iguana</i>	Skull	27	270.8	47.49	2.89
<i>Meleagris gallopavo</i>	Femur	40	58.82	6.36	1.54
<i>Ornithorhynchus anatinus</i>	Tibia	35	160.04	15.85	3.06
<i>Oryctolagus cuniculus</i>	Femur	40	196.76	20.6	3.18
<i>Phylloderma stenops</i>	Turbinates	20	145.26	27.02	2.48
<i>Pteropus giganteus</i>	Turbinates	18	155.28	25.31	2.22
<i>Rana pipiens</i>	Ilium	40	444.36	56.93	6.76
<i>Rattus norvegicus</i>	Turbinates	20	240.68	36.84	3.36

<i>Struthio camelus</i>	Tibia	40	83.83	9.72	2.16
<i>Ursus arctos</i>	Femur	40	173.4	19.5	2.75
<i>Varanus sp.</i>	Femur	40	138.3	14.66	2.05
Extinct Dinosaurs	Bone Element	N lacunae	Avg Vol (μm^3)	Vol 95% CI	C-Val
<i>Allosaurus fragilis</i>	Femur	20	68.56	11.67	-
<i>Apatosaurus sp.</i>	Scapula	20	130.49	22.84	-
<i>Brachylophosaurus canadensis</i>	Femur	20	121.21	19.01	-
<i>Coelophysis bauri</i>	Fragment	20	68.22	11.19	-
<i>Confuciusornis sanctus</i>	Skull	20	111.83	13.01	-
<i>Deinonychus antirrhopus</i>	Radius	20	61.09	9.53	-
<i>Diatryma gigantea</i>	Tibia	16	112.19	24.36	-
<i>Dryosaurus altus</i>	Femur	20	175.18	28.75	-
<i>Einiosaurus procurvicornis</i>	Humerus	20	277.05	48.41	-
<i>Euoplocephalus tutus</i>	Tibia	20	237.63	48.25	-
<i>Herrerasaurus ischigualastensis</i>	Tibia	20	46.84	6.56	-
<i>Hesperornis sp.</i>	Fragment	14	65.72	10.65	-
<i>Hypacrosaurus stebingeri</i>	Metatarsal	40	200.29	36.92	-
<i>Hypsiloophodon foxii</i>	Centrum	10	226.2	44.82	-
Lambeosaur	Humerus	12	194.25	30.21	-
<i>Maiasaura peeblesorum</i>	Tibia	20	179.67	22.35	-
Dinornithid	Metatarsal	20	76.77	10.69	-
Ornithomimid	Femur	20	42.47	5.88	-
<i>Orodromeus makelai</i>	Femur	20	164.84	31.7	-
<i>Oviraptor philoceratops</i>	Femur	20	183.2	43.51	-
Pachycephalosaurid	Tibia	20	122.88	23.37	-
<i>Protoceratops andrewsi</i>	Tibia	20	72.06	9.93	-
<i>Psittacosaurus mongoliensis</i>	Tibia	20	78.58	18.02	-
<i>Scutellosaurus lawleri</i>	Fragment	14	209.12	41.74	-
<i>Spinostropheus gautieri</i>	Vertebra	8	89.05	24.68	-
<i>Stygimoloch spinifer</i>	Skull	20	253.5	32.6	-
<i>Styracosaurus albertensis</i>	Parietal	20	173.16	19.53	-
<i>Tenontosaurus tilletti</i>	Femur	20	96.83	13.32	-
<i>Triceratops horridus</i>	Epoccipital	15	359.33	104.49	-
<i>Troodon formosus</i>	Tibia	20	47.8	7.92	-
<i>Tyrannosaurus rex</i>	Tibia	20	94.59	11.94	-

Sources for Phylogenetic Trees

Mesquite v1.11 (Maddison and Maddison 2006) was used to create a character matrix of osteocyte size and genome size for 26 extant taxa. The phylogenetic topology was drawn from sources in Cracraft and Donoghue (2004), Weishampel, Dodson et al. (2004), and Springer, Madsen et al. (2004). Branch lengths were estimated from fossil occurrences and phylogenetic analyses using fossils to estimate divergence time (Carroll 1964; Hoffstetter 1969; Reisz 1977; Nowak 1979; Carroll 1988; Alifanov 1989; Pascual, Archer et al. 1992; Laurin and Reisz 1995; Rose 1996; Wang and H. 1996; Brochu 1997; deBraga and O Rieppel 1997; Archibald, Averianov et al. 2001; Carroll 2001; Luo, Crompton et al. 2001; van Tuinen and Hedges 2001; Ji, Luo et al. 2002; Archibald 2003; Evans 2003; Rest, Ast et al. 2003; Santini and Tyler 2003; Benton 2004; Moyà-Solà, Köhler et al. 2004; Padian, Horner et al. 2004; Rowe 2004; Smith, Peterson et al. 2004; Stiassny, Wiley et al. 2004; Xijun, Wang et al. 2004; Asher, Meng et al. 2005), with the exception of the paleognath - neognath split in Aves, which was based on sequence data from the mitochondrial genome (Härlid, Janke et al. 1998).

A second tree was drawn based on a Dinosauria supertree (Pisani, Yates et al. 2002) that contained extant and extinct taxa with branch lengths estimated in time (Sereno, Wilson et al. 2004; Weishampel, Dodson et al. 2004 and chapters contained therein). Geologic time follows Gradstein, Ogg et al. (2004). Two more trees and character matrices, one with and the other without the extinct dinosaurs were created for genome size vs. interspersed repetitive elements using the references cited above.

Tree Variations: Various tree topologies (*Chrysemys* as a basal amniote and as the sister group to archosaurs; and *Herrerasaurus* as a basal saurischian and as a basal theropod) and branch lengths (unscaled, scaled with kappa, lambda, and delta in Continuous, all 1, etc.) were used to assess the robustness of our approach, though the analyses were robust to most changes in topology and branch lengths. While placing *Herrerasaurus* as a basal saurischian changed the phylogenetic position of evolutionary changes to genome size and architecture, it did not alter our inferences of genome size or repetitive DNA density significantly.

Regression Model

We have a dependent variable y and an independent variable x measured on a sample of n species, assumed to be related by the regression model $y_i = \beta_0 + \beta_1 x_i + e_i$, where β_0 is the y -axis intercept and β_1 is the slope of the line relating y to x . The e_i are the random errors and the residual variance of the regression is given by the variance of the $e_i = \sigma^2$. Writing y and x as vectors containing the observed data across species we study the probability of y in the familiar generalized least squares framework such that $p(y|x, m) \propto \exp\{-(y - (\beta_0 + \beta_1 x))' V^{-1} (y - (\beta_0 + \beta_1 x))\}$, where m denotes the regression model, and V is the expected variance-covariance matrix of the residual errors and is given by the phylogenetic tree that describes the relationships among species (Pagel 1997; Pagel 1999).

Markov chain Monte Carlo

We wish to study $p(y)$ accounting for uncertainty about the true regression model. The probability of the observed data, y , is given x as $p(y|x) = \int_m p(y|x,m)p(m)dm$, where $p(m)$ is the prior probability of m , and $p(y|x)$ is evaluated as given above, taking account of the phylogenetic relationships among species. The integration is over all possible values of m .

This integral is difficult to evaluate because m varies continuously. Alternatively, $p(y|x)$ can be evaluated by Markov chain Monte Carlo (MCMC) methods (Gilks, Richardson et al. 1996). A Markov chain is constructed in which new values of the parameters of the regression model are proposed on successive iterations of the Markov chain. At each step in the chain, models are accepted or rejected by the Metropolis-Hastings algorithm (Gilks, Richardson et al. 1996). The chain is allowed to run to convergence, after which it samples from the posterior distribution of $p(y|x)$. The converged chain also samples the posterior distribution of m .

By letting the converged chain run for a large number of iterations, the integral $p(y|x)$ and the posterior distribution of m can be evaluated to an arbitrary level of precision. We used uniform priors on m for all of our chains, and chains quickly reached convergence after which we sampled from widely spaced iterations to ensure independence among successive points. Chains begun from random starting points always converged to the same posterior distributions of m .

Hypothesis Testing

The contribution of the regression model to explaining the data is evaluated by Bayes Factors (Gilks, Richardson et al. 1996) in which $p(y|x)$ is compared under two scenarios: one in which β_1 is estimated as part of the model and one in which it is forced to be zero. The ratio of $p(y|x)$ when β_1 is estimated to $p(y|x)$ when it is not are compared. This is the Bayesian analogue to testing the hypothesis of $\beta_1=0$ in a conventional significance test. Bayes Factors greater than 3 are considered as positive evidence for a model. Values greater than 12 are taken as very strong evidence.

Posterior Predictive Distributions

We wish to obtain estimates for unknown values of the dependent variable, \tilde{y} , in a new sample for which x is known. If the true regression model were known it would be a simple matter to derive these new values. Instead, we know the posterior distribution of m from $p(y|x)$. Posterior predictive modeling (Gelman, Carlin et al. 2004, pp. 157-192) derives \tilde{y} from this posterior distribution. The probability of the unknown \tilde{y} is written as $p(\tilde{y}|x) = \int_m p(\tilde{y}|x,m)p(m)dm$, where the x values here correspond to species for which we wish to estimate the unknown \tilde{y} . This integral is approximated by MCMC methods in which we propose new values of m and normally distributed error, e . The values of m and e are sampled together from their joint posterior distribution as derived in the initial regression modeling step, where the errors are derived from the posterior distribution of σ^2 . These are combined to produce new values of $\tilde{y} = \beta_0 + \beta_1 x + e$.

This chain explores $p(\tilde{y} | x)$ yielding a posterior distribution of \tilde{y} . By allowing the chain to run long enough \tilde{y} 's posterior distribution can be sampled to an arbitrary level of precision. The MCMC procedure is equivalent to sampling repeatedly from a multivariate normal distribution $\sim MN(\beta_0 + \beta_1 x, \sigma^2 V)$ in which the values of the regression model and σ^2 are taken from their posterior distributions.

We used posterior predictive modeling to derive unknown values of genome size for the extinct dinosaur samples for which we had cell-size information. The posterior predictive distribution of genome sizes takes into account uncertainty in the parameters of the regression model as well as the inherent uncertainty of prediction summarized in σ^2 .

We additionally used the posterior distributions of genome size along with measured genome sizes on extant species to derive posterior predictive distributions of repetitive genetic elements. As before, let \tilde{y} denote the unknown values of these variables. The probability of these unknown values is written as $p(\tilde{y} | x) = \int_{\tilde{x}} \int_m p(\tilde{y} | x, \tilde{x}, m) p(m) p(\tilde{x}) dm d\tilde{x}$. We evaluated this integral using a Markov chain that sampled from posterior distributions of the model parameters (including residual error) derived from known data and from the posterior distribution of the genome sizes of extinct dinosaurs. The posterior distributions of the repetitive elements, therefore, take into account uncertainty in the regression model and in the values of some of the predictors.

Model/Method Checking

We evaluated the adequacy of our regression models by generating simulated species data from a multivariate normal distribution $\sim MN(\beta_0 + \beta_1 x, \sigma^2 V)$. We then calculated $p(y | x)$ for these data and compared it to that obtained for the observed data using Bayes Factors. In every case, the observed and simulated data yielded Bayes Factors < 3 , indicating that the regression model yielded data comparable to the observed data.

We also individually removed several extant species and then inferred their genome size from osteocyte data and summarized the results in Supplementary Table 2. Our inferences are within 2% to 15% of the reported average values. Estimations for interspersed repetitive DNA density for *Alligator mississippiensis*, which spanned all three datasets, is as accurate.

Supplementary Table 2: Evaluation of Bayesian GLS Regression Inferences

Taxon	Actual	Normal Prediction	Bayes GLS Prediction	Bayesian SD
<i>Iguana iguana</i> (c-value)	2.89	3.51	3.23	1.27
<i>Capra hircus</i> (c-value)	3.24	2.7	2.93	0.43
<i>Dendroica dominica</i> (c-value)	1.4	1.48	1.47	0.3
<i>Meleagris gallopavo</i> (c-value)	1.45	1.53	1.48	0.31
<i>Alligator mississippiensis</i> (c-value)	2.49	3	2.86	0.95

<i>Alligator mississippiensis</i> (repeat)	0.13	0.16	0.19	0.36
---	------	------	------	------

As a final validation of our methods, we compared our approach with a similar phylogenetically correct method to infer character traits in unsampled species (Garland and Ives 2000) using Mesquite (Maddison and Maddison 2006) and the PDAP package (Midford, Garland et al. 2005). We obtained similar results to those presented here under a variety of branch length transformations including: unscaled, all 1, Pagel 1992 (Pagel 1992), and Grafen's rho (Grafen 1989).

Phylogenetically Corrected *t*-Test for Differences Among Groups

Figures 2 and 3 and the primary text discuss "phylogenetic *t*-tests" to evaluate differences of inferred parameters (genome size and repeats) among groups. Normal *t*-tests could not be implemented because of the data's nonindependence (shared evolutionary history). Instead, trees were pruned to isolate the groups of interest and a new binary variable was created to group the taxa being compared, assigning 1's to one of the groups and 0's otherwise. The modified version of the computer program Continuous (Pagel 1997; Pagel 1999) was then used to test for correlations between the inferred parameter and the new binary grouping variable. The correlation test was carried out using MCMC and the proportion of the posterior distribution of the regression coefficient that crossed zero was used to calculate *p*-values.

The significance of such a correlation between a variable and a dummy coded grouping variable is, for all intents and purposes, statistically equivalent to a standard *t*-test. But the phylogenetic *t*-test, like the phylogenetic correlation, attempts to control for the fact that the various observations (species means) that make up the two groups being compared are not independent of each other, owing to their phylogenetic relationships. A non-significant phylogenetic *t*-test merely says that, given the variability within the samples, and the fact that they probably represent far fewer independent observations than their simple counts would suggest, it is possible that the observed difference in means between the two groups could have evolved by chance.

Inference of Genome Size

We used a modified version of the computer program Continuous (Pagel 1997; Pagel 1999) to test for correlations between parameters using a Bayesian framework as described above. We estimated three parameters for scaling branch lengths to determine if their incorporation into the regression analysis significantly improved the fit of the data to the model. The branch length scaling parameters kappa, lambda, and delta were determined simultaneously. A kappa of 2.09 was calculated for the extant taxa using maximum likelihood, suggesting more change occurred in longer branches. A delta of 1.3 was estimated, suggesting species-specific evolution and a lambda of

0.67 was estimated, suggesting moderate phylogenetic signal in the data. The settings we used to detect for trait correlation were: 5010000 iterations with a burn in of 1000 and a rate deviation of 0.5. The correlation is strongly supported by a Bayes factor of 9.9. We used the parameters and settings as detailed above for generating the regression models. The results were then analyzed in the software package JMP v5.12 (SAS Institute 2005).

Supplementary Table 3: Summary of GLS Regressions for Cell Size and Genome Size

	Alpha	Beta	GLS R ²
n	50091	50091	50091
Mean	-0.87	0.36	0.32
Var	0.41	0.015	0.005
95% CI	0.006	0.001	0.0006

The posterior regression distributions above were then used in the inference phase of the analysis, which included a new character matrix and tree. Both included the extinct dinosaurs (as described above). The branch length scaling parameters determined in the first half of the analysis were used during the inference phase. However, the portion of the tree that contained the dinosaurs had two polytomies. These branches were joined into bifurcations and the new basal branches were given lengths of 0. 5010000 iterations were used with a burn in of 1000 and a rate deviation of 0.5. The results are summarized in Supplementary Table 4 below.

Supplementary Table 4: Summary of Inferences for Dinosaurs

Taxon	Genome Size		Repeats	
	mean	95% CI	mean	95% CI
<i>Allosaurus fragilis</i>	1.7	0.002	0.09	0.002
<i>Apatosaurus sp.</i>	2.22	0.0033	0.1	0.0011
<i>Brachylophosaurus canadensis</i>	2.14	0.0024	0.09	0.0008
<i>Coelophysis bauri</i>	1.7	0.0012	0.06	0.0004
<i>Confuciusornis sanctus</i>	2.04	0.0017	0.08	0.0009
<i>Deinonychus antirrhopus</i>	1.62	0.002	0.07	0.0008
<i>Diatryma gigantea</i>	1.95	0.0016	0.07	0.0006
<i>Dryosaurus altus</i>	2.47	0.0024	0.11	0.0015
<i>Einiosaurus procurvicornis</i>	2.97	0.004	0.14	0.0012
<i>Euoplocephalus tutus</i>	2.88	0.0073	0.19	0.0033
<i>Herrerasaurus ischigualastensis</i>	1.47	0.0013	0.06	0.0005
<i>Hesperornis sp.</i>	1.65	0.0022	0.07	0.0009
<i>Hypacrosaurus stebingeri</i>	2.62	0.0032	0.13	0.0018
<i>Hypsilo-phodon foxii</i>	2.74	0.0033	0.14	0.0016
Lambeosaur	2.58	0.0031	0.15	0.0049

<i>Maiasaura peeblesorum</i>	2.51	0.0029	0.13	0.0018
Dinornithid	1.7	0.0024	0.06	0.0004
Ornithomimid	1.43	0.0033	0.12	0.0066
<i>Orodromeus makelai</i>	2.47	0.0045	0.12	0.001
<i>Oviraptor philoceratops</i>	2.53	0.0047	0.12	0.0016
Pachycephalosaurid	2.17	0.0033	0.1	0.001
<i>Protoceratops andrewsi</i>	1.74	0.0021	0.07	0.0005
<i>Psittacosaurus mongoliensis</i>	1.81	0.0021	0.08	0.0006
<i>Scutellosaurus lawleri</i>	2.65	0.0025	0.12	0.0008
<i>Spinostropheus gautieri</i>	1.92	0.0036	0.08	0.001
<i>Stygimoloch spinifer</i>	2.88	0.0047	0.14	0.001
<i>Styracosaurus albertensis</i>	2.47	0.0031	0.11	0.0009
<i>Tenontosaurus tilletti</i>	1.97	0.0024	0.09	0.0008
<i>Triceratops horridus</i>	3.31	0.005	0.16	0.0013
<i>Troodon formosus</i>	1.48	0.0027	0.06	0.0008
<i>Tyrannosaurus rex</i>	1.96	0.0043	0.1	0.0016

Visualization of genome size evolution (Figure 2 in the primary text) was accomplished by stratifying genome size into color-coded bins and mapping these colors onto the phylogeny. Ancestral genome sizes were obtained as the mean of the posterior prediction distribution of genome size for each node of the tree, using the same posterior predictive model as for species' genome size.

Inference of Interspersed Repetitive Element Density

Chicken, human, dog, rat and mouse genome data were taken from published genome assembly projects (Lander, Linton et al. 2001; Waterston, Lindblad-Toh et al. 2002; Kirkness, Bafna et al. 2003; Gibbs, Weinstock et al. 2004; Hillier, Miller et al. 2004). For species in which published data were not available, we used the software program RepeatMasker (Smit, Hubley et al. 1996-2004) to summarize repeat and GC content. We used BAC clone sequences downloaded from Genbank (www.ncbi.nlm.nih.gov) and random regions along scaffolds from published genomes on the UCSC genome browser (genome.ucsc.edu). GC content percentages for dog, rat, and mouse were based on flow cytometry estimates (Vinogradov 1998). SINE and LINE values for alligator, turtle, and anole were adjusted by the proportions of lineage-specific repeats observed between queries against human vs. chicken reference libraries (adjustment not required for other repeat types). As above, genome sizes were obtained from www.genomesize.com, values were averaged for species with multiple entries and in one case (*Myotis*), the species were averaged within the genus.

The species examined, proportions of repeat types, and other RepeatMasker results, as well as genome size are summarized in Supplementary Table 5. Genbank accession numbers and the amount of DNA sequence examined per species are

listed in appendix of this Supplementary Information. We acknowledge that RepeatMasker is likely providing a conservative estimate of repeat diversity in our sequences given the limited size of our data set, incomplete nature of the reference database, and extensive level of sequence divergence between some species compared. We used sequence from randomly selected BAC clones where available, but recognize that sampling a limited number of BAC clones may yield repeat and GC estimates that are biased upwards or downwards from whole-genome counts for both bioinformatic and experimental reasons.

Correlations between genome size and subcategories of interspersed repetitive elements (such as CR1 LINEs and MIR SINEs) would allow more nuanced reconstructions of genome evolution. Most intermediate categories (such as all LINEs or all SINEs) do show positive relationships with genome size. Interspersed repetitive elements were analyzed collectively because taken together they withstood GLS phylogenetic correction better, and perhaps more importantly, because they collectively have a major impact on modulating genome size (Petrov 2001). Other parameters, such as simple sequence repeats (SSRs), GC content, and chromosome number displayed no correlative relationship with genome size.

Supplementary Table 5: Summary of Repetitive Data

Taxon	G+C	LINEs	SINEs	LTRs	DNA TEs	All TEs	SSR	C-Val
<i>Alligator mississippiensis</i>	0.4240	0.1210	0.0050	0.0024	0.0010	0.1294	0.0031	2.49
<i>Anolis smaragdinus</i>	0.4150	0.0264	0.0010	<0.1%	0.0033	0.0307	0.0098	2.29
<i>Bos taurus</i>	0.4550	0.2350	0.0930	0.0270	0.0120	0.3670	0.0059	3.59
<i>Dromaius novaehollandiae</i>	0.4275	0.0110	0.0000	0.0001	0.0000	0.0111	0.0050	1.63
<i>Macaca mulatta</i>	0.4002	0.2070	0.1000	0.0850	0.0350	0.4270	0.0080	3.36
<i>Rhinolophus ferrumequinum</i>	0.4420	0.1790	0.0390	0.0350	0.0227	0.2757	0.0080	2.22
<i>Myotis</i>	0.4177	0.1227	0.0205	0.0183	0.0145	0.1760	0.0100	2.29
<i>Monodelphis domestica</i>	0.3713	0.1485	0.0925	0.0006	0.0068	0.2484	0.0158	4.17
<i>Ornithorhynchus anatinus</i>	0.4669	0.0966	0.2609	0.0000	0.0000	0.3575	0.0095	3.06
<i>Oryctolagus cuniculus</i>	0.4698	0.1453	0.1805	0.0168	0.0095	0.3521	0.0093	3.18
<i>Sphenodon punctatus</i>	0.4820	0.0680	0.0070	<0.1%	0.0002	0.0752	0.0063	5.00
<i>Meleagris gallopavo</i>	0.4284	0.0332	0.0028	0.0028	0.0000	0.0388	0.0061	1.54
<i>Chrysemys picta</i>	0.4360	0.0580	0.0100	0.0010	0.0020	0.0710	0.0038	2.80
<i>Xenopus tropicalis</i>	0.4200	0.0158	0.0005	<0.1%	0.0290	0.0453	0.0111	3.30
<i>Taeniopygia guttata</i>	0.4003	0.0296	0.0000	0.0041	0.0000	0.0337	0.0055	1.25
<i>Homo sapiens</i>	0.4000	0.2160	0.1350	0.0890	0.0310	0.4710	0.0090	3.50
<i>Pan troglodytes</i>	0.4150	0.1900	0.1470	0.0920	0.0250	0.4540	0.0090	3.75
<i>Gallus gallus</i>	0.4020	0.0640	0.0010	0.0130	0.0080	0.0860	0.0070	1.25
<i>Canis familiaris</i>	0.4170	0.1649	0.0910	0.0330	0.0190	0.3079	0.0150	3.12

<i>Rattus norvegicus</i>	0.4230	0.2310	0.0705	0.0904	0.0080	0.3999	0.0240	3.36
<i>Mus musculus</i>	0.4150	0.1740	0.0750	0.0900	0.0080	0.3470	0.0200	3.28
<i>Danio rerio</i>	0.3612	0.0169	0.0418	0.0102	0.2384	0.3073	0.0360	1.90
<i>Tetraodon nigroviridis</i>	0.4578	0.0038	0.0007	0.0001	0.0022	0.0087	0.0154	0.41
<i>Fugu rubripes</i>	0.4535	0.0032	0.0017	0.0019	0.0029	0.0117	0.0101	0.41

The Bayesian method implemented by the program Continuous described above for cell size and genome size was also used to test for correlation between the percent of the genome formed by intersperse repetitive elements and genome size. The parameters kappa (1.78), lambda (0.79), and delta (1.53) were determined as detailed above. 5010000 iterations were used with a burn in of 1000 and a rate deviation of 4.0. The correlation is strongly supported by a Bayes factor of 9.4. We used the settings and scaling parameters as detailed above for generating the regression models.

Supplementary Table 6: Summary of Regressions for Genome Size and Repetitive Element Density

	Alpha	Beta	GLS R ²
n	50091	50091	50091
Mean	-3.39	1.32	0.34
Var	0.45	0.19	0.005
95% CI	0.006	0.004	0.001

The posterior regression distributions above were then used in the Inference phase of the analysis as detailed above with cell size and genome size. We accounted for error generated from inferring intersperse repetitive elements from genome size, which itself was inferred as described above. Cell size was not used to infer repeat density directly because much larger datasets could be employed using genome size. However, given the correspondence between cell size and genome size we would expect results similar to those present here. 5010000 iterations were used with a burn in of 1000 and a rate deviation of 2.0. The inference results are summarized in Supplementary Table 4 above.

Trees and Character Matrices

The cell size – genome size tree (tree 1) for extant taxa is:

(1:346.000000,(2:312.784363,((3:200.185181,(4:130.185715,5:130.185715):70.000000):57.700001,((6:69.500000,7:69.500000):183.385269,((8:70.000000,9:70.000000):31.002556,(10:85.000000,(11:73.688950,12:73.691406):11.400000):16.000000):151.843277):5.000000):54.799999):10.000000,(13:192.000000,(14:129.687134,(((15:43.688313,16:43.686695):23.500000,(17:28.195805,18:28.205883):39.000000):33.500000,((19:45.196507,(20:28.688742,21:28.699301):16.500000):39.000000,((22:42.000000,(23:4.000000,24:4.000000):38.000000):26.000000,(25:34.000000,26:34.000000):34.000000):16.435715):16.500000):29.000000):62.000000):131.000000):23.000000);

The cell size – genome size tree (tree 2) for extant and extinct taxa is:

(1:346.000000,(2:312.784363,((3:200.185181,(4:130.185715,5:130.185715):70.000000):57.700001,((6:69.500000,7:69.500000):183.385269,(((8:8.800000,9:138.699997):3.000000,(((10:72.000000,11:71.500000):48.000000,(12:49.000000,(13:19.600000,(14:33.000000,(15:24.299999,16:24.400000):4.000000):4.000000):46.750000):36.000000):14.500000,((17:99.500000,18:47.200001):8.700000,(19:62.000000,(20:16.400000,((21:28.000000,22:30.875000):31.000000,(23:28.000000,24:28.044001):31.000000):31.799999):6.700000):8.500000):23.000000):7.714286):35.954887,(25:92.199997,(26:9.500000,((27:11.030000,28:111.000000):4.400000,(29:66.000000,((30:130.000000,31:130.000000):0.000000,(32:108.000000,((33:106.000000,34:65.699997):0.000000,(35:37.000000,(36:73.500000,((37:75.800003,(38:70.000000,39:70.000000):8.445113):22.557444,(40:85.000000,(41:26.000000,(42:73.688950,43:73.691406):6.000000):5.400000):16.000000):37.500000):23.600000):9.000000):10.816389):13.100000):14.700000):15.200000):5.300000):6.600000):5.700000):10.326881):5.000000):54.799999):10.000000,(44:192.000000,(45:129.687134,(((46:43.688313,47:43.686695):23.500000,(48:28.195805,49:28.205883):39.000000):33.500000,((50:45.196507,(51:28.688742,52:28.699301):16.500000):39.000000,((53:42.000000,(54:4.000000,55:4.000000):38.000000):26.000000,(56:34.000000,57:34.000000):34.000000):16.435715):16.500000):29.000000):62.000000):131.000000):23.000000);

Supplementary Table 7: Data and taxa for trees 1 and 2.

Codes Tree 1	Codes Tree 2	Taxa	Ln Genome Size	Ln Cell Size
1	1	<i>Rana pipiens</i>	1.91102289	6.02197429
2	2	<i>Chrysemys picta</i>	1.029619417	5.022039213
3	3	<i>Iguana iguana</i>	1.061256502	5.503064053
4	4	<i>Gekko sp.</i>	0.940007258	5.69643967
5	5	<i>Varanus sp.</i>	0.717839793	4.875327066
6	6	<i>Crocodylus niloticus</i>	1.166270937	5.418603417
7	7	<i>Alligator mississippiensis</i>	0.91228271	5.217735709
-	8	<i>Scutellosaurus lawleri</i>	?	5.342890058
-	9	<i>Euoplocephalus tutus</i>	?	5.470694675
-	10	<i>Stygimoloch spinifer</i>	?	5.535345796
-	11	Pachycephalosaurid	?	4.811186501
-	12	<i>Psittacosaurus mongoliensis</i>	?	4.364164234
-	13	<i>Protoceratops andrewsi</i>	?	4.277516784
-	14	<i>Triceratops horridus</i>	?	5.88424658
-	15	<i>Einiosaurus procurvicornis</i>	?	5.624180798
-	16	<i>Styracosaurus albertensis</i>	?	5.154202594
-	17	<i>Orodromeus makelai</i>	?	5.104970128
-	18	<i>Hypsilophodon foxii</i>	?	5.421431901
-	19	<i>Tenontosaurus tilletti</i>	?	4.573002654
-	20	<i>Dryosaurus altus</i>	?	5.165823536
-	21	Lambeosaur	?	5.269148002
-	22	<i>Hypacrosaurus stebingeri</i>	?	5.299755275
-	23	<i>Maiasaura peeblesorum</i>	?	5.191113147
-	24	<i>Brachylophosaurus canadensis</i>	?	4.797490418
-	25	<i>Apatosaurus sp.</i>	?	4.871284852
-	26	<i>Herrerasaurus ischigualastensis</i>	?	3.846698494
-	27	<i>Coelophysis bauri</i>	?	4.222667359
-	28	<i>Spinostropheus gautieri</i>	?	4.489186139
-	29	<i>Allosaurus fragilis</i>	?	4.227640976
-	30	<i>Tyrannosaurus rex</i>	?	4.54958573
-	31	Ornithomimid	?	3.7487167
-	32	<i>Oviraptor philoceratops</i>	?	5.210568671
-	33	<i>Troodon formosus</i>	?	3.8669717
-	34	<i>Deinonychus antirrhopus</i>	?	4.112368062
-	35	<i>Confuciusornis sanctus</i>	?	4.717020801
-	36	<i>Hesperornis sp.</i>	?	4.185395591
-	37	<i>Mo sp.</i>	?	4.340828291
8	38	<i>Dromaius novaehollandiae</i>	0.488580015	4.461979164
9	39	<i>Struthio camelus</i>	0.770108222	4.361030854

10	40	<i>Meleagris gallopavo</i>	0.431782416	4.012885761
-	41	<i>Diatryma gigantea</i>	?	4.72022725
11	42	<i>Corvus corax</i>	0.19062036	4.528748612
12	43	<i>Dendroica dominica</i>	0.336472237	3.949544749
13	44	<i>Ornithorhynchus anatinus</i>	1.118414916	5.032045203
14	45	<i>Didelphis virginiana</i>	1.415853163	4.663415345
15	46	<i>Oryctolagus cuniculus</i>	1.156881197	5.229971798
16	47	<i>Rattus norvegicus</i>	1.211940974	5.411823534
17	48	<i>Ateles paniscus</i>	1.244154594	5.596973875
18	49	<i>Homo sapiens</i>	1.252762968	5.565250818
19	50	<i>Pteropus giganteus</i>	0.797507196	4.990635837
20	51	<i>Eptesicus fuscus</i>	0.862889955	4.693013267
21	52	<i>Phylloderma stenops</i>	0.90825856	4.903015152
22	53	<i>Ursus arctos</i>	1.011600912	5.089527923
23	54	<i>Canis familiaris</i>	1.137833002	4.865258982
24	55	<i>Canis latrans</i>	1.036736885	4.93742932
25	56	<i>Bos taurus</i>	1.278152203	5.287199953
26	57	<i>Capra hircus</i>	1.17557333	5.034416017

The genome size – interspersed repetitive elements tree (tree 3) for extant taxa is:
 ((1:145.642685,(2:95.706322,3:95.782043):49.838047):347.888947,(4:345.811584,((5:312.48
 5809,((6:223.265793,7:223.456268):34.616989,(8:252.647385,(9:100.829147,(12:84.441803,(
 10:42.291897,11:42.291801):42.291897):16.158226):152.235474):5.000000):54.799999):10.0
 00000,(13:190.953690,(14:130.019012,(((15:53.523335,(16:5.464286,17:5.662869):48.09841
 9):33.864605,(18:63.583363,(19:49.726471,20:49.642971):13.883577):23.920790):13.553921
 ,((21:48.768749,22:48.647182):36.159203,(23:39.742176,24:39.744057):44.751495):16.4357
 15):29.000000):60.868362):131.746429):23.372326):147.594299);

The genome size – interspersed repetitive elements tree (tree 4) for extant and extinct
 taxa is:
 ((1:145.642685,(2:95.706322,3:95.782043):49.838047):347.888947,(4:345.811584,((5:312.48
 5809,((6:223.265793,7:223.456268):34.616989,(8:252.647385,(((9:8.842105,10:138.710449):
 3.248426,(((11:71.981964,12:71.591667):48.000000,(13:48.811409,(14:19.608805,(15:33.058
 525,(16:24.281919,17:24.399538):4.062500):4.000000):46.750000):36.000000):14.500000,((
 18:99.459602,19:47.185635):8.701825,(20:62.011383,(21:16.391737,((22:27.952953,23:30.8
 75000):30.921879,(24:27.950138,25:28.044001):30.963362):31.750744):6.666667):8.532610)
 :23.445332):7.714286):35.954887,(26:92.157532,(27:9.500000,((28:11.027027,29:111.13073
 0):4.444668,(30:65.567360,((31:129.897247,32:130.191650):0.000000,(33:107.801300,((34:1
 06.359123,35:65.693207):0.000000,(36:37.204895,(37:73.547668,((38:75.830803,39:78.2717
 06):22.557444,((40:42.291897,41:42.291801):42.291897,(42:26.256865,43:79.041801):5.400
 000):16.158226):37.532467):23.601532):9.212121):10.816389):13.080746):14.745370):15.19
 6906):5.325107):6.612245):5.785714):10.326881):5.000000):54.799999):10.000000,(44:190.

953690,(45:130.019012,(((46:53.523335,(47:5.464286,48:5.662869):48.098419):33.864605,(49:63.583363,(50:49.726471,51:49.642971):13.883577):23.920790):13.553921,((52:48.768749,53:48.647182):36.159203,(54:39.742176,55:39.744057):44.751495):16.435715):29.000000):60.868362):131.746429):23.372326):147.594299);

Supplementary Table 8: Data and taxa for trees 3 and 4.

Codes Tree 3	Codes Tree 4	Taxa	Ln Repeats	Ln Genome Size
1	1	<i>Danio rerio</i>	-1.17993081	0.641853886
2	2	<i>Fugu rubripes</i>	-4.744432253	- 0.891598119
3	3	<i>Tetraodon nigroviridis</i>	-4.448166437	- 0.891598119
4	4	<i>Xenopus tropicalis</i>	-3.094448246	1.193922468
5	5	<i>Chrysemys picta</i>	-2.645075402	1.029619417
6	6	<i>Anolis smaragdinus</i>	-3.483492624	0.828551818
7	7	<i>Sphenodon punctatus</i>	-2.587604048	1.609437912
8	8	<i>Alligator mississippiensis</i>	-2.044846897	0.91228271
-	9	<i>Scutellosaurus lawleri</i>	?	1.154996565
-	10	<i>Euoplocephalus tutus</i>	?	1.221672706
-	11	<i>Stygimoloch spinifer</i>	?	1.254806363
-	12	Pachycephalosaurid	?	0.870322145
-	13	<i>Psittacosaurus mongoliensis</i>	?	0.632997088
-	14	<i>Protoceratops andrewsi</i>	?	0.58879109
-	15	<i>Triceratops horridus</i>	?	1.440446227
-	16	<i>Einiosaurus procurvicornis</i>	?	1.300816024
-	17	<i>Styracosaurus albertensis</i>	?	1.051754221
-	18	<i>Orodromeus makelai</i>	?	1.027297241
-	19	<i>Hypsilophodon foxii</i>	?	1.194900601
-	20	<i>Tenontosaurus tilletti</i>	?	0.743259868
-	21	<i>Dryosaurus altus</i>	?	1.058867076
-	22	Lambeosaur	?	1.114517845
-	23	<i>Hypacrosaurus stebingeri</i>	?	1.127264163
-	24	<i>Maiasaura peeblesorum</i>	?	1.075461598
-	25	<i>Brachylophosaurus canadensis</i>	?	0.864441815
-	26	<i>Apatosaurus sp.</i>	?	0.903741733
-	27	<i>Herrerasaurus ischigualastensis</i>	?	0.354875577
-	28	<i>Coelophysis bauri</i>	?	0.555155597
-	29	<i>Spinostropheus gautieri</i>	?	0.698507795
-	30	<i>Allosaurus fragilis</i>	?	0.560835348
-	31	<i>Tyrannosaurus rex</i>	?	0.733118106
-	32	Ornithomimid	?	0.304880731

-	33	<i>Oviraptor philoceratops</i>	?	1.083665127
-	34	<i>Troodon formosus</i>	?	0.367229865
-	35	<i>Deinonychus antirrhopus</i>	?	0.499942171
-	36	<i>Confuciusornis sanctus</i>	?	0.820647424
-	37	<i>Hesperornis sp.</i>	?	0.536677525
-	38	Dinornithid	?	0.619514878
9	39	<i>Dromaius novaehollandiae</i>	-4.500810171	0.488580015
10	40	<i>Meleagris gallopavo</i>	-3.249335032	0.431782416
11	41	<i>Gallus gallus</i>	-2.453407983	0.223143551
-	42	<i>Diatryma gigantea</i>	?	0.82343152
12	43	<i>Taeniopygia guttata</i>	-3.390257442	0.223143551
13	44	<i>Ornithorhynchus anatinus</i>	-1.028619917	1.118414916
14	45	<i>Monodelphis domestica</i>	-1.392714929	1.427916036
15	46	<i>Macaca mulatta</i>	-0.850971266	1.211940974
16	47	<i>Pan troglodytes</i>	-0.789658081	1.32175584
17	48	<i>Homo sapiens</i>	-0.752897185	1.252762968
18	49	<i>Oryctolagus cuniculus</i>	-1.043840053	1.156881197
19	50	<i>Mus musculus</i>	-1.058430499	1.187843422
20	51	<i>Rattus norvegicus</i>	-0.916540763	1.211940974
21	52	<i>Rhinolophus ferrumequinum</i>	-1.288441961	0.797507196
22	53	<i>Myotis lucifugus</i>	-1.737271284	0.828551818
23	54	<i>Canis familiaris</i>	-1.002393431	1.278152203
24	55	<i>Bos taurus</i>	-1.177980224	1.137833002

Reference Information for Sequence Data Used in Repeat Analysis

FORMAT: Species common name, *Latin binomial*; Total sequence per species in base pairs (bp); Genbank sequence ID numbers; Scaffold number (*Danio*, *Tetraodon*, *Fugu* only).

American Alligator, *Alligator mississippiensis*; 2,519,551 bp; 250707-25398.

Painted Turtle, *Chrysemys picta*; 2,432,811 bp; 253983-257443.

Bahamian Anole, *Anolis smaragdinus*; 1,358,158 bp; 390731-389174.

Emu, *Dromaius novaehollandiae*; 2,405,186 bp; 68533374, 65133587, 63055134, 62751227, 63028194, 62868028, 62860364, 62732662, 62123031, 62123030, 62123029, 60279832, 60115736, 58866123, 58866122, 58801763, 58801762, 58082206, 57900846.

Greater Horseshoe Bat, *Rhinolophus ferrumequinum*; 8,084,283 bp; 93204891, 92866910, 92866909, 92866908, 92866907, 92110077, 92110076, 92110075, 91932892, 91064949, 91064948, 90819309, 90819308, 90704938, 90093371, 90093370, 89572688, 89572686, 89257414, 89257413, 89001183, 87240303, 86610393, 86610392, 86361368, 85861299, 85740571, 85719345, 85719344, 85719343, 85662479, 85060526, 85060525, 84310130, 84095212, 83817104, 83817103, 83423623, 83415296, 83415295, 82830617, 82830615, 82779754, 82779753, 82465560, 81777655, 80751234, 78271995, 78190047, 78190046.

Cow, *Bos taurus*; 8,869,078 bp; 82734381, 77416341, 68268124, 67514666, 66275414, 62945624, 62945622, 62238041, 60223274, 60223271, 55667698, 34849950, 34849949, 34849947, 34849946, 20429385, 88759681, 88759507, 88759504, 88759467, 88759460, 88759452, 88759451, 88759450, 88759446, 88759437, 88759436, 88759435, 88759434, 88759433, 88759432, 88759431, 88759430, 88759429, 88759428, 88759427, 88759426, 88759425, 88759424, 88759423, 88759422, 88759421, 88759420, 88759419, 88759418, 88759417.

Opossum, *Monodelphis domestica*; 7,246,587 bp; 42539032; 85861298, 76496351, 76096431, 73476567, 71274314, 71143224, 71061567, 68159024, 63108497, 58372188, 62954883, 62868032, 60418112, 62629972, 62198320, 62175062, 62123091, 61740786, 61740785, 61740784, 61740783, 61180163, 61180161, 60301588, 59891485, 59276133, 59276132, 58418427, 58418426, 58372189, 60418111, 60223322, 60115764, 60115763, 60115738, 59933350, 59933348, 59797094, 59797093, 54969748.

Rabbit, *Oryctolagus cuniculus*; 9,430,806 bp; 93204890, 93102427, 92866501, 92866466, 92866463, 92866434, 92866431, 92110067, 92110066, 92110065, 92087087, 91754212, 91754211, 90992618, 90855821, 90855820, 90855819, 90704888, 90704887, 90704885, 90403628, 90265836, 89994752, 89994751, 89941481, 89886503, 89886502, 89886501,

89572640, 89572639, 89572637, 89572636, 89572635, 89572633, 89572632, 89572631, 89337293, 89179329, 89145444, 89001150, 89001148, 89001147, 89001146, 89001142, 88687788, 88687786, 88608896, 87299679, 87253070, 87253067, 87240296, 87116712, 86990569, 86990568.

Rhesus Monkey, *Macaca mulatta*; 6,048,336 bp; 91064889, 91064888, 91064887, 91064886, 90855786, 82779663, 82734248, 82734250, 82734249, 82734247, 82734245, 82734242, 82734240, 82581300, 82524867, 82491696, 82491695, 81295433, 81295213, 78600827, 78600826, 78214379, 77797858, 77797857, 77797856, 77797855, 77797854, 77797851, 77627799, 77539162, 77416190, 77416189, 77404881, 77404879, 77404878.

Turkey, *Meleagris gallopavo*; 5,995,831 bp; 88608895; 85861282, 63094052, 63094051, 63055116, 62954877, 62867997, 62867992, 62860352, 60279827, 60223163, 59858671, 58866109, 58082203, 57470553, 56787683, 56542311, 51571986, 47498122, 47026445, 46559449, 46559448, 46402334, 45917386, 45825161, 45825159, 45504191, 45504190,

African Clawed Frog, *Xenopus tropicalis*; 9,387,739 bp; 92110057; 91932820, 91932819, 91932818, 91932817, 91932816, 91932815, 91932814, 91932813, 91932812, 89111978, 89111977, 89111976, 89111975, 89111974, 89111973, 89111972, 89111971, 89111970, 89111969, 89111968, 89111967, 89111966, 89111965, 89111964, 89111963, 89111962, 89111961, 89111960, 85362829, 85362828, 85362827, 85362826, 85362825, 85362824, 85362822, 85362821, 85362820, 85362819, 85362818, 85362817, 85362816, 85362815, 85362814, 85362813, 85362812, 85362811, 85362810, 85362809, 85362808, 85362807, 85362806, 85362805, 85362804, 85362803, 85362802, 85362801, 85362800, 85362799, 85362798, 85362797, 85362796, 85362795, 85362794, 85362793, 85362792, 85362791, 85362790, 85362789, 85362788, 85362787, 85362786, 85362785, 85362784, 85362783, 85362782, 78771691, 78771690, 78771689, 78771688, 78771687, 78771686, 78771685, 78771684, 78771683, 78771682, 78771681, 78771680, 78771679, 78771678, 78771677, 78771676, 78771675, 78771674.

Little Brown Bat, *Myotis lucifugus*; 7,169,971 bp; 84310077, 91932821, 91754200, 91754199, 91064890, 90265833, 90186281, 89572610, 89365808, 89337288, 89337287, 89274192, 89274191, 89274190, 89111983, 89111982, 89111981, 89001129, 89001126, 88900537, 88759692, 88687779, 87162494, 87162493, 87162492, 87162491, 87162490, 87116709, 87116708, 87116707, 87116706, 86990550, 86610381, 86476176, 86476175, 86361313, 86203340, 85662454, 85662453, 85540498, 84872332.

Tuatara, *Sphenodon punctatus*; 1,588,389 bp; 86990586, 71725473, 70980627, 66864970, 66793037, 62868024, 61806831, 61562506, 58372183, 57164473, 56806696.

Duck-billed Platypus, *Ornithorhynchus anatinus*; 4,780,327 bp; 90963048, 90186388, 89941533, 89886531, 89572708, 89365860, 85815480, 85663311, 85815385, 85740601, 85719387, 85680204, 85680202, 84872423, 84095262, 84782075, 84580066, 84580065, 84039719, 84000508, 83699517, 83627857, 77415732, 77362608, 55666491, 51870449,

50080102, 49292283, 47522519, 45598044, 62175063, 54019462, 54019461, 29149951, 60603476, 60603475, 60603474, 60603473, 60603472, 60603471, 60603470, 60603469, 60603468, 60603467.

Zebra Finch, *Taeniopygia guttata*; 3,798,616 bp; 65133496, 62954879, 62945492, 62945491, 61562495, 60279830, 58801757, 58801756, 58801755, 58082205, 57470559, 56542316, 47026458, 46309616, 46309615, 46275888, 46063611, 45860979, 45773966, 45752782, 45752781, 45544662, 45504195, 45476589, 45476588, 45476587, 44151434, 66279523, 66279522, 66279521, 66279520, 66279519, 66279518, 66279517, 66279516, 66279515, 66279514, 66279513, 66279512, 66279511, 66279510, 66279509, 66279508, 66279507, 66279506, 66279505, 66279504, 66279503, 66279502, 66279501, 66279500, 66279499, 66279498, 66279497, 66279496, 66279495, 66279494, 66279493, 66279492, 66279491, 66279490, 66279489, 66279488, 66279487, 66279486, 66279485, 66279484, 66279483, 66279482, 66279481, 66279480, 66279479, 66279478, 66279477, 66279476, 66279475, 66279474, 66279473, 66279472, 66279471, 66279470, 66279469, 66279468, 66279467, 66279466, 66279465, 66279464, 66279463, 66279462, 66279461, 66279460, 66279459, 66279458, 66279457, 66279456, 66279455, 66279454, 66279453, 66279452, 66279451, 66279450, 66279449, 66279448, 66279447, 66279446, 66279445, 66279444, 66279443, 66279442, 66279441, 66279440, 66279439, 66279438, 66279437, 66279436, 66279435, 66279434, 66279433, 66279432, 66279431, 66279430, 66279429, 66279428, 66279427, 66279426, 66279425, 66279424, 66279423, 66279422, 66279421, 66279420, 66279419, 66279418, 66279417, 66279416, 66279415, 66279414, 66279413, 66279412, 66279411, 66279410, 66279409, 66279408, 66279407, 66279406, 66279405, 66279404, 66279403, 66279402, 66279401, 66279400, 66279399, 66279398, 66279397, 66279396, 66279395, 66279394, 66279393, 66279392, 66279391, 66279390, 66279389, 66279388, 66279387, 66279386, 66279385, 66279384, 66279383, 66279382, 66279381, 66279380, 66279379, 66279378, 66279377, 66279376, 66279375, 66279374, 66279373, 66279372, 66279371, 66279370, 66279369, 66279368, 66279367, 66279366, 66279365, 66279364, 66279363, 66279362, 66279361, 66279360, 66279359, 66279358, 66279357, 57806351, 57806349, 57021183, 57021182, 57021181, 57021180, 57021179, 57021178, 57021177, 57021176, 57021175, 57021174, 57021173, 57021172, 57021171, 57021170, 57021169, 57021168, 57021167, 57021166, 57021165, 57021164, 57021163, 57021162, 57021161, 57021160, 57021159, 57021158, 57021157, 57021156, 57021155, 57021154, 57021153, 57021152, 57021151, 57021150, 57021149, 57021148, 57021147, 57021146, 57021145, 57021144, 57021143, 57021142, 57021141, 57021140, 57021139, 57021138, 57021137, 57021136, 57021135, 57021134, 57021133, 57021132, 57021131, 57021130, 57021129, 57021128, 57021127, 57021126, 57021125, 57021124, 57021123, 57021122, 57021121, 57021120.

Zebrafish, *Danio rerio*; 14,080,684 bp; CAAK03000000; 3, 4, 5, 64, 65, 66, 119, 120, 176, 177, 178, 206, 207, 208, 209, 210, 211, 286, 287, 288.

Freshwater Pufferfish, *Tetraodon nigroviridis*; 13,004,394 bp; CAAE01000000; 14573, 14005, 1208, 18978, 9656, 10204, 22846, 15765, 10205, 15816, 24802, 19725, 18518, 17771,

12163, 18863, 11974, 20256, 14059, 15015, 16981, 7599, 14604, 14705, 11279, 3563, 6380, 11278, 11354, 9007, 13644, 23781, 13635, 6761, 3984, 921, 922, 3761, 7940, 10298, 21704, 9505, 9875, 2065, 9852, 7013, 13201, 10565, 12567, 12450, 7072, 10571, 12197, 10211, 16547, 25427, 9504, 14738, 14614, 10865, 20004, 13806, 13745, 14553, 15084, 14497, 20102, 14756, 14575, 11972, 7226, 19482, 11943, 23785, 9447, 20484, 13998, 14117, 24887, 14533, 14495, 2153, 20270, 1945, 11659, 13768, 18467, 19021, 14468, 2106, 12300, 3970, 12301, 19949, 14709, 15049, 21762, 15028, 21664, 18819, 9532, 14685, 15974, 2120, 2112, 10152, 11706, 20468, 11707, 15026.

Marine Pufferfish; *Fugu rubripes*; 11,231,546 bp; CAAB01000000; 1, 2, 6, 13, 16, 18, 11, 19, 3, 4, 5, 7, 8, 9, 10, 14, 15.

Supplementary Information References

- Alifanov, V. R. (1989). "The oldest gecko (Lacertilia: Gekkonidae) from the Lower Cretaceous of Mongolia." Paleontological Journal **23**: 128–131.
- Archibald, J. D. (2003). "Timing and biogeography of the eutherian radiation: fossils and molecules compared." Molecular Phylogenetics and Evolution **28**: 350–359.
- Archibald, J. D., A. O. Averianov, et al. (2001). "Late Cretaceous relatives of rabbits, rodents, and other extant eutherian mammals." Nature **414**: 62–65.
- Asher, R. J., J. Meng, et al. (2005). "Stem Lagomorpha and the antiquity of Glires." Science **307**: 1091–1094.
- Benton, M. J. (2004). Origin and Relationships of Dinosauria. The Dinosauria. D. B. Weishampel, P. Dodson and H. Osmólska. Berkeley, University of California Press: 7–19.
- Brochu, C. A. (1997). "Morphology, fossils, divergence timing, and the phylogenetic relationships of *Gavialis*." Systematic Biology **46**: 479–521.
- Canè, V., G. Marotti, et al. (1982). "Size and density of osteocyte lacunae in different regions of long bones." Calcified Tissue International **34**: 558–563.
- Carroll, R. (1964). "The earliest reptiles." Zoological Journal of the Linnean Society **45**: 61–83.
- Carroll, R. (2001). "The origin and early radiation of terrestrial vertebrates." Journal of Paleontology **75**(6): 1202–1213.
- Carroll, R. L. (1988). Vertebrate Paleontology and Evolution. New York, W. H. Freeman and Company.
- Cracraft, J. and M. J. Donoghue (2004). Assembling the Tree of Life. New York, Oxford University Press.
- Currey, J. D. (2002). Bones: Structure and Mechanics. New Jersey, Princeton University Press.
- de Ricqlès, A., F. J. Meunier, et al. (1991). Comparative microstructure of bone. Bone, **V.3**. B. K. Hall. London, CRC Press: 1–78.
- de Ricqlès, A. J. (1980). Tissue structure of dinosaur bone: functional significance and possible relation to dinosaur physiology. A Cold Look at the Warm-Blooded Dinosaurs. R. D. K. Thomas and C. Olson. Westview, Boulder, CO, American Association for the Advancement of Science, Selected Symposium 28: 103–139.
- de Ricqlès, A. J., K. Padian, et al. (2003). "Osteohistology of *Confuciusornis sanctus* (Theropoda: Aves)." Journal of Vertebrate Paleontology **23**(2): 373–386.
- deBraga, M. and O. O Rieppel (1997). "Reptile phylogeny and the interrelationships of turtles." Zoological Journal of the Linnean Society **120**: 281–354.

- Evans, S. E. (2003). "At the feet of the dinosaurs: the early history and radiation of lizards." Biological Reviews of the Cambridge Philosophical Society **78**: 513–551.
- Francillon-Vieillot, H., V. de Buffrénil, et al. (1990). Microstructure and mineralization of vertebrate skeletal tissues. Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends. J. G. Carter. New York, Van Nostrand Reinhold. **1**: 471-530.
- Franz-Odenaal, T. A., B. K. Hall, et al. (2005). "Buried alive: How osteoblasts become osteocytes." Developmental Dynamics **235**(1): 176 - 190.
- Garland, T., Jr. and A. R. Ives (2000). "Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods." American Naturalist **155**: 346-364.
- Gelman, A., J. B. Carlin, et al. (2004). Bayesian Data Analysis. London, Chapman Hall/CRC press.
- Gibbs, R. A., G. M. Weinstock, et al. (2004). "Genome sequence of the Brown Norway rat yields insights into mammalian evolution." Nature **428**: 493-521.
- Gilks, W. R., S. Richardson, et al. (1996). Introducing Markov chain Monte Carlo. Markov Chain Monte Carlo in Practice. W. R. Gilks, S. Richardson and D. J. Spiegelhalter. London, Chapman and Hall: 1-19.
- Gradstein, F. M., J. G. Ogg, et al. (2004). A Geologic Time Scale 2004. Cambridge, Cambridge University Press.
- Grafen, A. (1989). "The phylogenetic regression." Philosophical Transactions of the Royal Society of London Series B **326**: 119-157.
- Gregory, T. R. (2000). "Nucleotypic effects without nuclei: Genome size and erythrocyte size in mammals." Genome **43**: 895-901.
- Gregory, T. R. (2001). "The bigger the C-value, the larger the cell: Genome size and red blood cell size in vertebrates." Blood Cells, Molecules, and Diseases **27**(5): 830-843.
- Gregory, T. R. (2005). Genome Size Evolution in Animals. The Evolution of the Genome. T. R. Gregory. Boston, Elsevier Academic Press: 4-71.
- Hall, B. K. (2005). Osteoblast and osteocyte diversity. Bones and Cartilage: Development and Evolutionary Skeletal Development. Amsterdam, Elsevier Academic Press: 328-337.
- Härlid, A., A. Janke, et al. (1998). "The complete mitochondrial genome of *Rhea americana* and early avian divergences." Journal of Molecular Evolution **46**(6): 669-679.
- Hillier, L. W., W. Miller, et al. (2004). "Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution." Nature **432**(7018): 695-716.
- Hoffstetter, R. (1969). "Un primate de l'Oligocene inferieur sud-Americain: *Branisella boliviana* gen. et sp. nov." Comptes rendus de l'Academie des Sciences. Serie D **269**: 434–437.

- Horner, J. R., A. J. de Ricqlès, et al. (2000). "Long bone histology of the hadrosaurid dinosaur *Maiasaura peeblesorum*: growth dynamics and physiology based on an ontogenetic series of skeletal elements." Journal of Vertebrate Paleontology **20**(1): 115-129.
- Horner, J. R., K. Padian, et al. (2001). "Comparative osteology of some embryonic and perinatal archosaurs: developmental and behavioral implications for dinosaurs." Paleobiology **27**(1): 39-58.
- Ji, Q., Z. X. Luo, et al. (2002). "The earliest known eutherian mammal." Nature **416**(6883): 816-822.
- Kirkness, E. F., V. Bafna, et al. (2003). "The dog genome: survey sequencing and comparative analysis." Science **301**.
- Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
- Laurin, M. and R. R. Reisz (1995). "A reevaluation of early amniote phylogeny." Zoological Journal of the Linnean Society **113**(2): 165-223.
- Luo, Z. X., A. W. Crompton, et al. (2001). "A new mammaliaform from the early Jurassic and evolution of mammalian characteristics." Science **292**(5521): 1535-1540.
- Maddison, W. P. and D. R. Maddison (2006). "Mesquite: a modular system for evolutionary analysis. Version 1.11 <http://mesquiteproject.org>."
- Midford, P. E., T. J. Garland, et al. (2005). "PDAP Package of Mesquite. Version 1.07."
- Moyà-Solà, S., M. Köhler, et al. (2004). "*Pierolapithecus catalaunicus*, a New Middle Miocene Great Ape from Spain." Science **306**(5700): 1339 -1344.
- Nowak, R. M. (1979). "North American Quaternary Canis." Monograph of the Museum of Natural History, University of Kansas **6**: 1 - 154.
- Olmo, E. (1983). "Nucleotype and cell size in vertebrates: a review." Basic and Applied Histochemistry **27**: 227-256.
- Olmo, E. (2003). "Reptiles: a group of transition in the evolution of genome size and of the nucleotypic effect." Cytogenetic and Genome Research **101**(2): 166-171.
- Padian, K., J. R. Horner, et al. (2004). "Growth in small dinosaurs and pterosaurs: the evolution of archosaurian growth strategies." Journal of Vertebrate Paleontology **24**: 555-571.
- Pagel, M. D. (1992). "A method for the analysis of comparative data." Journal of Theoretical Biology **156**: 431-442.
- Pagel, M. D. (1997). "Inferring evolutionary processes from phylogenies." Zoologica Scripta **26**: 331-348.
- Pagel, M. D. (1999). "Inferring the historical patterns of biological evolution." Nature **401**: 877-884.

- Pascual, R., M. Archer, et al. (1992). "First discovery of monotremes in South America." Nature **356**: 704-706.
- Petrov, D. A. (2001). "Evolution of genome size: new approaches to an old problem." Trends in Genetics **17**: 23-28.
- Pisani, D., A. M. Yates, et al. (2002). "A genus-level supertree of the Dinosauria." Proceedings of the Royal Society of London, Series B **269**(1494): 915-921.
- Reid, R. E. H. (1996). "Bone histology of the Cleveland-Llyod dinosaurs and of dinosaurs in general, part I: introduction to bone tissues." Brigham Young University Geology Studies **41**: 25-72.
- Reisz, R. R. (1977). "*Petrolacosaurus*, the oldest known diapsid reptile." Science **196**: 1091-1093.
- Rest, J. S., J. C. Ast, et al. (2003). "Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome." Molecular Phylogenetics and Evolution **29**: 289-297.
- Rose, K. D. (1996). "February 1996 On the origin of the order Artiodactyla." Proceedings of the National Academy of Sciences **93**: 1705-1709.
- Rowe, T. (2004). Chordate phylogeny and development. Assembling the Tree of Life. J. Cracraft and M. J. Donoghue. New York, Oxford University Press: 384-409.
- Santini, F. and J. C. Tyler (2003). "A phylogeny of the families of fossil and extant tetraodontiform fishes (Acanthomorpha, Tetraodontiformes), Upper Cretaceous to recent." Zoological Journal of the Linnean Society **139**: 565-617.
- SAS Institute (2005). JMP, Version 5.1.2. Cary, NC, SAS Institute Inc.
- Sereno, P. C., J. A. Wilson, et al. (2004). "New dinosaurs link southern landmasses in the Mid-Cretaceous." Proceedings of the Royal Society of London, Series B **271**(1546): 1325-1330.
- Smit, A. F., R. Hubley, et al. (1996-2004). "RepeatMasker Open-3.0. <http://www.repeatmasker.org>."
- Smith, A. B., K. J. Peterson, et al. (2004). From bilateral symmetry to pentaradiality: The phylogeny of hemichordates and echinoderms. Assembling the Tree of Life. J. Cracraft and M. J. Donoghue. New York, Oxford University Press: 365-383.
- Springer, M. S. S., M. J., O. Madsen, et al. (2004). "Molecules consolidate the placental mammal tree." Trends in Ecology and Evolution **19**(8): 430-438.
- Stiassny, M. L., E. O. Wiley, et al. (2004). Gnathostome fishes. Assembling the Tree of Life. J. Cracraft and M. J. Donoghue. New York, Oxford University Press: 410-429.
- Tazawa, K., K. Hoshi, et al. (2004). "Osteocytic osteolysis observed in rats to which parathyroid hormone was continuously administered." Journal of Bone and Mineral Metabolism **22**: 524-529.

- van Tuinen, M. and S. B. Hedges (2001). "Calibration of avian molecular clocks." Molecular Biology and Evolution **18**(2): 206–213.
- Vinogradov, A. E. (1998). "Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship." Cytometry **31**(2): 100-109.
- Vinogradov, A. E. and O. V. Anatskaya (2006). "Genome size and metabolic intensity in tetrapods: a tale of two lines." Proceedings: Biological sciences **273**(1582): 27-32.
- Wang, X. and T. R. H. (1996). Canidae. The Terrestrial Eocene-Oligocene Transition in North America, Part II: Common Vertebrates of the White River Chronofauna. . D. R. E. Prothero, R. J. Cambridge, Cambridge University Press: 433-452.
- Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-562.
- Weishampel, D. B., P. Dodson, et al. (2004). The Dinosauria. Berkeley, University of California Press.
- Xijun, N., Y. Wang, et al. (2004). "A euprimate skull from the early Eocene of China." Nature **427**: 65-68