

Independent Mammalian Genome Contractions Following the KT Boundary

Mina Rho,* Mo Zhou,* Xiang Gao,† Sun Kim,* Haixu Tang,* and Michael Lynch†

*School of Informatics, Indiana University, Bloomington; and †Department of Biology, Indiana University, Bloomington

Although it is generally accepted that major changes in the earth's history are significant drivers of phylogenetic diversification and extinction, such episodes may also have long-lasting effects on genomic architecture. Here we show that widespread reductions in genome size have occurred in multiple lineages of mammals subsequent to the Cretaceous–Tertiary (KT) boundary, whereas there is no evidence for such changes in other vertebrate, invertebrate, or land plant lineages. Although the mechanisms remain unclear, such shifts in mammalian genome evolution may be a consequence of an increase in the efficiency of selection against excess DNA resulting from post-KT population size expansions. Independent historical changes in genome architecture in diverse lineages raise a significant challenge to the idea that genome size is finely tuned to achieve adaptive phenotypic modifications and suggest that attempts to use phylogenetic analysis to infer ancestral genome sizes may be problematical.

Introduction

The evolutionary patterning of genome architecture by nonadaptive forces is supported by population genetic theory, estimates of the relative power of the major forces of evolution, and comparative analyses of whole-genome sequences (Lynch 2007). Nevertheless, some biologists still adhere to the idea that even the most arcane aspects of genome evolution, including expansions of genome size by mobile element proliferation, are direct products of natural selection (e.g., Gregory 2005; Kirschner and Gerhart 2005; Caporale 2006). Unfortunately, resolving whether the evolution of genome architecture is largely driven by variation in the forces of random genetic drift and mutation, as opposed to natural selection, is impeded by the long timescale of the underlying processes, which often necessitates comparative analyses involving extant species.

Although comparative methods are central to many endeavors in evolutionary biology, they are often burdened by assumptions regarding the equilibrium status and/or evolutionary independence of current-day taxa. For example, under most methods of comparative analysis, the estimated phenotype of an internal node surrounded by lineages with similar phenotypes will generally be interpreted as being roughly equal to the average of the descendent species, leading to a prediction of little or no change. Such an interpretation can be quite misleading if the descendent lineages have actually evolved directionally in a parallel manner.

Fortunately, some genomic features harbor internal information on the historical pattern of genomic expansion and contraction experienced by specific lineages, eliminating the uncertainties of comparative analysis. Here we utilize genome-wide surveys of mobile elements and two types of pseudogenes to suggest that diverse orders of mammals have undergone substantial, independent reductions in genome size following the Cretaceous–Tertiary (KT) boundary, a period of global ecological upheaval occurring ~65 Ma (Archibald 1996). Although the mechanisms driving such change remain unclear, these results provide a compelling example of a broad syndrome of genomic changes being driven by apparently nonadaptive events,

while also demonstrating that mammalian genome architecture is currently in a nonequilibrium state.

Materials and Methods

Acquisition and Analysis of Long Terminal Repeat Elements

We employed the *ab initio* method of Rho et al. (2007) to detect all long terminal repeat (LTR)–containing families within fully sequenced genomes, restricting our analyses to elements with paired-end sequences, which are essential for dating purposes. Although such treatment excludes solo elements resulting from intraelement recombination, which serves as one potential mechanism of element loss, elements nested within each other are included. The age of each element was determined by aligning the paired-end sequences and converting the observed sequence divergence to an estimated number of substitutions per site by the Jukes–Cantor method (Jukes and Cantor 1969).

Our analyses are specifically focused on the analysis of copies of LTR retroelements that were highly likely to have been derived from autonomous (self-replicating) parental elements, as opposed to being derivatives of potential nonautonomous relatives. In the first pass through a genome, candidate elements in this category were extracted when the length of a pair of repeats fell in the range of 130–2,000 bp and the distance between them fell in the range of 1,200–18,000 bp. In subsequent refinements, such fragments were retained as *bona fide* LTR retroelements if the interior contained a set of one or more retroelement protein domains with a combined probability of $<e^{-10}$. Use of this criterion minimizes the likelihood of misinterpreting two independent solo LTRs as a pair of LTRs associated with a single element (which would lead to a false element), while also eliminating any potential nonautonomous elements.

Because the LTR elements employed in this study were limited to those having both LTRs and at least remnants of protein domains, the total counts are substantially smaller than those within the various genome project papers, which include solo LTRs, fragments, and nonautonomous elements. As a direct comparison, we applied our search criteria to the human and *Arabidopsis* genomes. For the human genome, RepeatMasker (Smit et al. 2004) has previously estimated a total of 505,950 LTR fragments, whereas application of our search criteria yielded only 3,272 putative autonomous LTR elements (or descendants of them). This discrepancy is a simple reflection of the fact

Key words: genome evolution, genome size, KT boundary, mammalian evolution, mobile elements, pseudogenes, retrotransposons.

E-mail: milynych@indiana.edu.

Genome Biol. Evol. Vol. 2009:2–12.

doi:10.1093/gbe/evp007

Advance Access publication May 5, 2009

that LTR-associated sequences estimated by RepeatMasker do not necessarily contain protein domains and/or paired LTRs. For the *Arabidopsis* genome, we obtained a total of 297 putative pairs of autonomous LTR elements out of the 4,264 LTR fragments estimated by RebaseUpdate (Kapitonov and Jurka 2002).

To obtain reasonably accurate age distributions of LTR retrotransposons, our method must be able to identify both recent and relatively ancient LTR retrotransposons. To evaluate the power of our method, computer simulations were carried out using randomly mutated LTR pairs with sequence identities in the range of 50–100% and also employing an insertion/deletion level of 10% and 30%. The generated LTR pairs were then inserted into random genomic sequence. This analysis showed that our method is capable of identifying more than 90% of the LTR pairs having <30% divergence and essentially all such pairs with divergence <25%. Thus, because the following analyses are largely based on age distributions of LTR elements with divergences <30%, they should be unaffected by any biases in the identification of extremely old inserts.

Acquisition and Analysis of Ribosomal Protein Pseudogenes

For each genome analyzed, we ran PseudoPipe, a computational pipeline for pseudogene identification (Zhang et al. 2006), using the extracted ribosomal protein (RP) gene sequences from the same genome as the query. The full sets of annotated protein sequences from vertebrate species were downloaded from the Ensemble database, whereas those for other species were obtained from sites associated with particular genome projects. As a check on our work, we also referred to the Ribosomal Protein Gene Database (<http://ribosome.med.miyazaki-u.ac.jp/>) for animal RPs. For the genomes that are not assembled into chromosomes (e.g., the *Fugu* genome), we integrated the scaffold sequences into several large chunks of sequences before applying PseudoPipe.

Active RP genes are among the slowest evolving genes in any genome, with 99% amino acid sequence identity between mouse and human proteins (Zhang et al. 2002), and an approximation of the expected divergence rate between such genes and their pseudogenes can be obtained by assuming that all sites of the latter, but only a fraction of 0.25 sites of the former, are free to evolve at the neutral rate (0.25 being the approximate fraction of synonymous nucleotide sites in coding exons). Letting the neutral rate of base substitutional evolution be μ , the taxon-specific values of which are given below, the divergence rate is then 1.25μ , which for a given level of observed divergence then allows for a transformation to absolute time.

Acquisition and Analysis of Mitochondrial Protein Gene Fragment Insertions

For each nuclear genome analyzed, the DNA sequences of the protein-coding regions in mitochondrial genome from the same species were used as probes for Blast, with hits with $<e^{-5}$ being retained and merged as single frag-

ments when consecutive hits for different proteins were in close proximity (gap < 50 bp). Because of their altered genetic code, fragments of mitochondrial genomes inserted into nuclear genomes are expected to be nonfunctional and hence to evolve in a neutral fashion, but in mammals, only ~10% of substitutions at silent sites of active mitochondrial genes (in the mitochondrion) are capable of going to fixation (Lynch 2007). Thus, letting the neutral substitution rates in the nuclear and mitochondrial genomes be μ_n and μ_m , respectively, and again assuming that approximately 25% of the nucleotide sites in coding DNA are synonymous, the expected rate of divergence of a mitochondrial gene and its counterpart inserted into the nuclear genome is approximately $\mu_n + \mu_m[0.25 + (0.75 \times 0.10)] = \mu_n + 0.325\mu_m$.

Estimation of Element Birth and Death Rates

Provided the rates of gain and loss of element insertions remain constant over time, the age distribution of a family of elements during such a phase should be closely approximated by the function

$$N_t = Be^{-Dt},$$

where B is the rate of origin of new insertions per haploid genome and D is the rate of loss. This formula expresses the equilibrium age distribution that will result from any birth and death rate, that is, B need not equal D to achieve equilibrium, as birth and survivorship parameters are on fundamentally different scales (the first per total genome and the second per existing element).

From a least squares regression of $\ln(N_t)$ on t , $\ln(B)$ can be estimated from the intercept and D from the slope. By Taylor's expansion, the sampling variance (the square of the standard error) of B is estimated as

$$\text{Var}(\hat{B}) = \text{Var}(\hat{a})e^{2\hat{a}},$$

where \hat{a} is the estimated intercept of the regression.

The half-life of an element is estimated as

$$t_{0.5} = -\ln(0.5)/\hat{D},$$

where \hat{D} is the estimated slope of the regression, and its sampling variance is estimated as

$$\text{Var}(t_{0.5}) = \text{Var}(\hat{D})[\ln(0.5)/\hat{D}^2]^2.$$

We used the estimated number of substitutions per nucleotide site in paired LTRs as a surrogate measure of t and scaled B so that it represents the rate of origin of new insertions per genome during the interval required for 1% divergence of terminal repeats. Age calibrations relied on published information on magnitudes of nucleotide substitution at silent sites and estimated divergence times for various lineages (see Supplementary Material online).

For nonstable age distributions, the current birth rate (B_0) can be approximated from the count of elements in the youngest age class (N_0), assuming a particular death rate (here, we assumed the estimate for the pre-KT period). Noting that

$$N_0 = B_0 \int_0^{S_k} e^{-DS} dS,$$

then

$$\hat{B}_0 = \frac{DN_0}{1 - e^{-S_k D}},$$

where S_k is the upper value for the range of silent site divergence among LTRs for the youngest class. A lower bound estimate of B_0 can be obtained by assuming $D = 0$, which yields $\hat{B}_0 = N_0/S_k$.

Generation of Expected Age Distributions

Our results revealed substantial evidence for discontinuities in the birth/death rates of various insertion types, most notably a common appearance of bulges in the age distributions of mammalian genome insertions. For diagnostic purposes, this raised the necessity of evaluating the types of demographic events that could plausibly lead to such forms. For an age distribution to exhibit an intermediate peak, there must be an increase in the birth rate toward the past in excess of the rate of loss of elements per unit time. In other words, if elements are generally lost at an approximately constant rate per unit time, D , such that the fraction of a newborn cohort of elements remaining after t time units is e^{-Dt} , the increase in the birth rate toward the past must exceed e^D per unit time. As described in the text, if such conditions are met, it is relatively easy to generate expected age distributions with features similar to those observed in mammals. The precise position and height of the peak as well as the pattern of progression toward it will be a function of the temporal pattern of change in the birth and death rates of elements.

The expected form of an age distribution of LTR elements is a function of both the birth–death dynamics over time and the stochastic accumulation of sequence divergence between paired-end sequences. Letting u denote the rate of mutation per nucleotide site per unit time, in the absence of selection, the distribution of the number (n) of substitutions for a pair of LTRs of length L will be Poisson,

$$P(n, t) = (2uLt)^n e^{-2uLt} / n!.$$

Thus, letting $B(t)$ be the total number of new elements arising per diploid genome t time units in the past, the number of expected elements per genome with $d = n/L$ substitutions per site is

$$N(d) = \sum_{t=0}^{\infty} B(t) e^{-Dt} P(n, t)$$

assuming a constant rate of element loss per unit time (D). For a specified time-dependent function $B(t)$, this expression yields an expected age distribution of elements on the scale of substitutions per site.

Results

LTR retrotransposons provide an ideal source of information on the temporal dynamics of genome evolution as

all such elements employ replication mechanisms that lead to the production of long flanking repeats of 100% identity at the time of element insertion. Because the LTRs of resident sequences are not under postinsertion selection, they are expected to diverge neutrally over time, yielding a natural chronometer for the age of the encompassed element (Promislow et al. 1999). A genome-wide collection of all recognizable elements then provides an estimate of the age distribution of such insertions, which must reflect the historical pattern of element-copy gains and losses in the line of descent leading to the current-day genome.

For the de novo identification of all families containing autonomous LTR retrotransposons in fully sequenced genomes, we employed a computational procedure that does not rely on prior knowledge of element structure or content (Rho et al. 2007). This method is capable of harvesting all elements identified by earlier computational methods while also locating previously undetected copies, down to a level of $\sim 50\%$ divergence between paired LTRs, well beyond the point of most element survival. Using this approach, all invertebrates, aquatic vertebrates, and land plants for which data are available are found to exhibit age distributions for such elements that are approximately negative exponential, as expected under a long-term steady-state birth–death process (fig. 1). Qualitatively similar patterns have been observed with smaller sets of data for several other species, including maize (San Miguel et al. 1998), wheat (San Miguel et al. 2002), pea (Jing et al. 2005), and yeast (Promislow et al. 1999). Because the average silent site divergence of randomly sampled alleles within a population is ~ 0.008 for land plants and 0.013 for invertebrates (Lynch 2007), the low levels of terminal repeat divergence in figure 1 imply that a large fraction of the LTR elements identified in these species is unlikely to be fixed in host populations. This is consistent with the hypothesis that the vast majority of mobile element inserts have deleterious effects on host fitness (Charlesworth 1985).

In striking contrast, several mammalian lineages (primate, carnivore, and artiodactyl) exhibit dramatic recent declines of major LTR element families, with age distributions generally exhibiting peaks or shoulders more recent than the KT boundary or very close to it (fig. 2). Some element families in rodents and opossum are exceptions to this pattern, exhibiting very recent expansions, although even these lineages show evidence of an earlier (post-KT) slowdown in element proliferation. The age distributions for ancestrally shared elements in human and macaque (and in mouse and rat) only roughly correspond with each other prior to the divergence of these species pairs. Such deviations in the age distributions of elements of shared ancestry are not an artifact of our methods of LTR identification but can be expected if the loss rates of elements from two lineages deviate subsequent to their separation.

When subjected to regression analysis, regions of age distributions that can be fit to a negative exponential function can be used to estimate rates of element insertion and loss during such periods. Birth rate estimates obtained in this manner denote the rate of origin of new insertions by entire element groups at the level of the host haploid genome, that is, they are not equivalent to fixation rates, as should be clear from the argument noted above. The

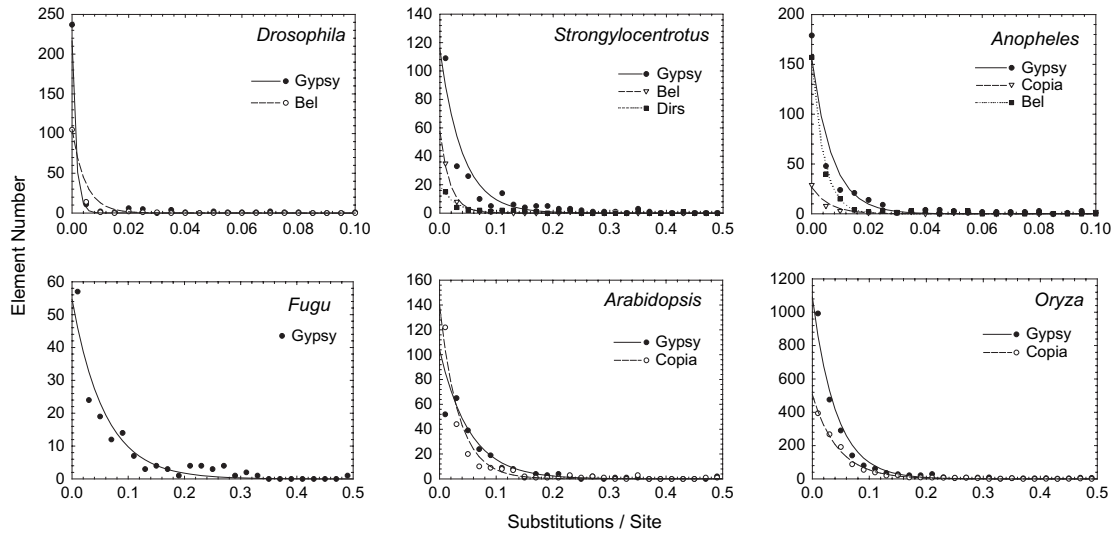


FIG. 1.—Age distributions of major LTR retrotransposon groups in the genomes of invertebrates, fish, and land plants. The curves denote exponential functions derived by weighted least squares analysis. Ages are in units of substitutions per site between flanking LTR sequences.

estimated loss rates, defined on a per-element basis, reflect physical losses by either natural selection at the host level or large-scale DNA-level excision processes, including intra- or interelement recombination (which generates solo LTRs), and do not include potential mutational inactivation of the retrotransposon machinery by point mutations that otherwise leave the element intact.

As defined by the exponential curves in figure 1, on a timescale of 1% LTR sequence divergence, the average genome-wide birth rate of LTR element families in invertebrates and fish from the deep past to the current time is 120 (standard error = 34), whereas that for land plants is slightly larger, 193 (76) (table 1). This approach can also be

used for the earliest stages of mammalian evolution because despite the obvious post-KT changes in evolutionary demography, the age distributions of the older cohorts of mammalian LTR elements (to the right of the age distribution bulges) are approximately exponential and therefore consistent with earlier steady-state phases. During these periods, the birth rates of mammalian LTR element families averaged 276 (85), again on a timescale of 1% LTR sequence divergence (equivalent to ~ 0.3 – 1.0 My for the taxa involved) (table 1). Because the per-generation mutation rate of short-lived species is up to $10\times$ lower than that for mammals (Lynch 2007), these results imply a substantially higher per-generation rate of LTR element insertion in

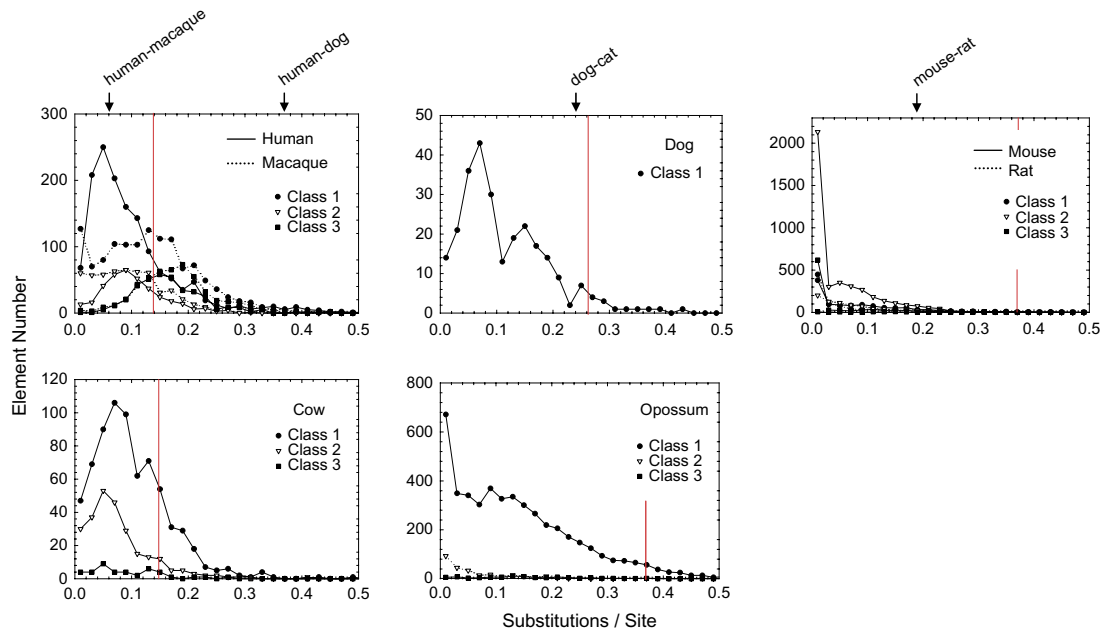


FIG. 2.—Age distributions of major mobile element families containing LTRs in seven mammalian genomes. Red lines denote the approximate locations of the KT boundary (65 Ma). The estimated points of divergence of some reference taxa are given in the margins.

Table 1
Estimated Demographic Parameters for LTR Elements in Various Lineages during Phases of Apparent Steady-State Birth/Death Rates (on a timescale of 1% divergence)

Species	Group	Birth Rate	Death Rate	Half-Life	r	df	Range
Mammals							
Human	Class 1	233.19 (68.49)	12.20 (0.29)	0.057 (0.002)	0.982	20	0.05–0.47
	Class 2	158.37 (55.17)	16.96 (0.35)	0.041 (0.002)	0.991	8	0.09–0.27
	Class 3	325.58 (142.61)	15.24 (0.44)	0.045 (0.005)	0.949	8	0.15–0.33
Macaque	Class 1	398.48 (123.76)	12.71 (0.31)	0.055 (0.004)	0.968	14	0.13–0.43
	Class 2	192.61 (84.00)	15.07 (0.44)	0.046 (0.005)	0.943	9	0.13–0.33
	Class 3	936.43 (394.86)	17.18 (0.42)	0.040 (0.002)	0.984	7	0.19–0.35
Cow	Class 1	161.97 (51.52)	13.63 (0.32)	0.051 (0.005)	0.950	10	0.07–0.29
	Class 2	69.90 (25.25)	17.86 (0.36)	0.039 (0.002)	0.986	9	0.05–0.25
	Class 3	5.40 (2.41)	7.99 (0.45)	0.087 (0.044)	0.602	5	0.05–0.17
Dog	Class 1	40.19 (12.67)	10.32 (0.32)	0.067 (0.009)	0.911	10	0.07–0.29
Mouse	Class 1	263.27 (91.92)	49.69 (0.35)	0.014 (0.002)	0.916	9	0.00–0.05
	Class 2	1,949.66 (945.17)	72.80 (0.48)	0.010 (0.002)	0.883	9	0.00–0.05
	Class 3	844.62 (471.40)	138.68 (0.56)	0.005 (0.001)	0.890	9	0.00–0.05
Rat	Class 1	164.13 (68.66)	14.05 (0.42)	0.049 (0.006)	0.892	18	0.01–0.39
	Class 2	97.83 (29.33)	11.28 (0.30)	0.061 (0.002)	0.984	18	0.01–0.39
Opossum	Class 1	319.15 (63.30)	6.48 (0.20)	0.107 (0.007)	0.944	27	0.01–0.57
	Class 2	43.06 (16.25)	15.35 (0.38)	0.045 (0.005)	0.937	10	0.01–0.23
	Class 3	17.95 (6.32)	9.42 (0.35)	0.074 (0.010)	0.904	10	0.13–0.35
Other animals							
Chicken	Class 2	15.16 (7.16)	33.17 (0.47)	0.021 (0.001)	0.990	3	0.01–0.09
	Class 3	12.17 (3.02)	5.71 (0.25)	0.121 (0.025)	0.782	13	0.01–0.29
<i>Fugu rubripes</i>	Gypsy	27.37 (10.75)	17.23 (0.39)	0.040 (0.004)	0.945	9	0.01–0.21
<i>Strongylocentrotus purpuratus</i>	Gypsy	57.49 (28.86)	25.02 (0.50)	0.028 (0.004)	0.927	6	0.01–0.15
	Bel	28.79 (17.94)	55.75 (0.62)	0.012 (0.001)	0.979	3	0.01–0.09
	DirS	9.75 (5.63)	37.27 (0.58)	0.019 (0.003)	0.955	3	0.01–0.09
<i>Ciona intestinalis</i>	Gypsy	31.85 (17.03)	27.65 (0.53)	0.025 (0.003)	0.940	7	0.01–0.17
	Bel	14.82 (9.47)	55.09 (0.64)	0.013 (0.002)	0.972	1	0.01–0.07
<i>Anopheles gambiae</i>	Gypsy	313.08 (167.69)	138.72 (0.54)	0.005 (0.001)	0.956	5	0.00–0.03
	Copia	53.02 (30.30)	162.39 (0.57)	0.004 (0.001)	0.952	3	0.00–0.02
	Bel	306.70 (181.10)	240.44 (0.59)	0.003 (0.000)	0.996	3	0.00–0.02
<i>Drosophila melanogaster</i>	Gypsy	433.48 (253.86)	199.18 (0.59)	0.003 (0.001)	0.870	4	0.00–0.03
	Bel	210.97 (130.96)	417.67 (0.62)	0.002 (0.000)	0.998	1	0.00–0.01
<i>Daphnia pulex</i>	Gypsy	136.27 (72.63)	37.55 (0.53)	0.018 (0.002)	0.936	9	0.01–0.21
	Copia	81.73 (49.59)	40.02 (0.61)	0.017 (0.003)	0.923	6	0.01–0.15
	Bel	55.88 (33.79)	46.04 (0.60)	0.015 (0.002)	0.947	5	0.01–0.13
<i>Caenorhabditis elegans</i>	Bel	34.02 (22.54)	117.57 (0.66)	0.006 (0.000)	1.000	0	0.01–0.03
Plants							
<i>Populus trichocarpa</i>	Gypsy	110.52 (30.69)	9.80 (0.28)	0.071 (0.003)	0.976	20	0.01–0.43
	Copia	131.86 (39.32)	9.63 (0.30)	0.072 (0.005)	0.958	20	0.01–0.43
<i>Arabidopsis thaliana</i>	Gypsy	52.03 (20.53)	18.98 (0.39)	0.037 (0.003)	0.972	9	0.03–0.23
	Copia	69.96 (36.30)	29.41 (0.52)	0.024 (0.003)	0.957	6	0.01–0.15
<i>Oryza sativa</i>	Gypsy	546.26 (244.24)	24.00 (0.45)	0.029 (0.002)	0.970	9	0.01–0.21
	Copia	253.32 (99.04)	22.42 (0.39)	0.031 (0.001)	0.991	9	0.01–0.21

NOTE.—The birth and death rates are on different scales here (the former being for the total population of elements and the latter per element), so their inequality does not imply a nonequilibrium situation. Standard errors are in parentheses. Correlation coefficients and degrees of freedom are denoted by r and df, and the range denotes the span of divergence over which the regression was evaluated (in units of substitutions per site between LTRs).

pre-KT mammals than in most modern day species (including mammals). The average half-life of insertions in invertebrates, 1.4 (0.3) in units of percent sequence divergence among paired LTR sequences, is significantly lower than that for pre-KT mammals and land plants, 5.3 (0.5) and 4.4 (0.9), respectively, although perhaps not much lower on a per-generation basis.

The recent decline in LTR element numbers in placental mammals may be a consequence of a decline in the insertion rate, an increase in the loss rate, or both. In the absence of a stable age distribution of elements, it is difficult to infer the historical record of change in element loss rates, but it is clear that major declines in insertion rates have occurred. From the numbers of elements in the youngest age class alone, estimates of current birth rates of the mammalian elements can be acquired, and these show that

across all lineages exhibiting bulges in LTR element age distributions (human, macaque, cow, and dog), current birth rates for Classes 1, 2, and 3 (Griffiths 2001) elements average just 17.7% (0.6%), 12.7% (5.5%), and 10.2% (10.0%), respectively, of those in pre-KT phases. Thus, because the split between mammalian orders predates the KT boundary (Benton and Donoghue 2007; Bininda-Emonds et al. 2007; Wible et al. 2007), there appear to have been dramatic independent reductions in LTR element insertion rates in isolated lineages of placental mammals.

Because the equilibrium number of elements within a genome is equal to the ratio of birth and death rates, our results can be used to estimate the numbers of such elements prior to the KT boundary (assuming long-term stasis during this period, as supported by the data in fig. 2) as well as to project the expected numbers well into the future

Table 2
Estimates of the Total Numbers of Insertions per Genome at the Current Time, in the Future, and during Times prior to the KT Boundary

Species	Future	Current	Pre-KT
LTR elements			
Human	363	2,156	4,982
Macaque	815	2,314	9,865
Cow	328	1,001	1,647
Dog	86	304	1,168
Mouse	3,817	6,184	3,817
Rat	2,059	1,870	3,067
Opossum	5,517	5,002	5,397
RP pseudogenes			
Human	332	1,790	3,404
Macaque	188	601	1,768
Cow	53	351	1,655
Dog	43	157	511
Mouse	143	864	10,514
Rat	182	1,120	1,804
Inserts of mitochondrial protein-coding genes			
Human	3	430	252,374
Macaque	17	481	2,299
Cow	0	112	1,544
Dog	1	132	62,957
Mouse	11	77	49,125
Opossum	101	805	6,085

NOTE.—Current estimates are simply the direct counts of insertions detected. Pre-KT estimates are derived from the ratios of birth and death rates given in tables 1, 3, and 4. Future estimates are obtained by assuming that the death rate remains constant, using the current birth rate as estimated from the number of elements in the youngest age class. Note that these estimates apply only to the expected subsets of element copies that have structural forms that are identifiable by the methods employed in this study, that is, they are not the full populations of all insertional remnants.

(assuming that current birth rates continue to hold and the death rates are the same as in the distant past). For placental mammalian lineages (excluding rodents), such analyses suggest that the numbers of elements of the types included in this study were on average 3.0 (0.6) times as abundant prior to the KT boundary as they are today and that they would eventually decline to between 17% and 35%, 28% (4%) on average, of current levels if the present birth/death parameters continued into the future (table 2).

Two previous observations are consistent with a contraction in genome size on the primate lineage following the KT boundary. First, pseudogenes in the human genome exhibit a peak at ~ 58 My in the past, followed by a negative exponential age distribution (Zhang et al. 2002). Similar to the situation noted above for LTR elements, the pre-KT birth rate of such pseudogenes appears to be $\sim 8\times$ the current rate and the estimated loss rate is not greatly different from that observed here for LTR elements (Lynch 2007). Second, a dramatic slowdown in the rate of insertion of mitochondrial DNA fragments into the human nuclear genome (numts) is thought to have occurred 25–40 Ma (Bensasson et al. 2003; Gherman et al. 2007).

To determine the generality of these additional observations, we evaluated the age distributions of RP pseudogenes in a wide spectrum of lineages. Although the exact pattern differs among species, all mammalian genomes with adequate data exhibit the hallmarks of a recent reduction in the accumulation of such pseudogenes—either a pronounced intermediate bulge or a recent shoulder in the age distribution (fig. 3). In all cases, the shift in activity appears more recent than or close to the estimated position of the KT boundary. Analysis of the demographic parameters of such inserts during the early steady-state phase suggests that RP pseudogenes were about 4.4 (1.6) times as abundant prior to the KT boundary as they are today and that they are destined to eventually decline to $\sim 21\%$ (3%) of their current levels, assuming the maintenance of current birth/death parameters (tables 2 and 3). In all other species examined (land plants, invertebrates, and fish), the total numbers of detectable RP pseudogenes were generally fewer than 10, so as in the case of LTR retrotransposons, aberrant non-stable age distributions of these insertions are only recognizable in mammalian lineages.

A broad survey of numts yielded a very similar conclusion. For all mammalian species (except rat, where the data are insufficient), there is a clear peak in the age distribution of numts at a level of divergence of 0.2–0.3 substitutions per site (fig. 3). As the estimated position of the KT boundary is at the point of 0.65–1.06 divergence for all lineages, the transitions in demographic behavior for numts

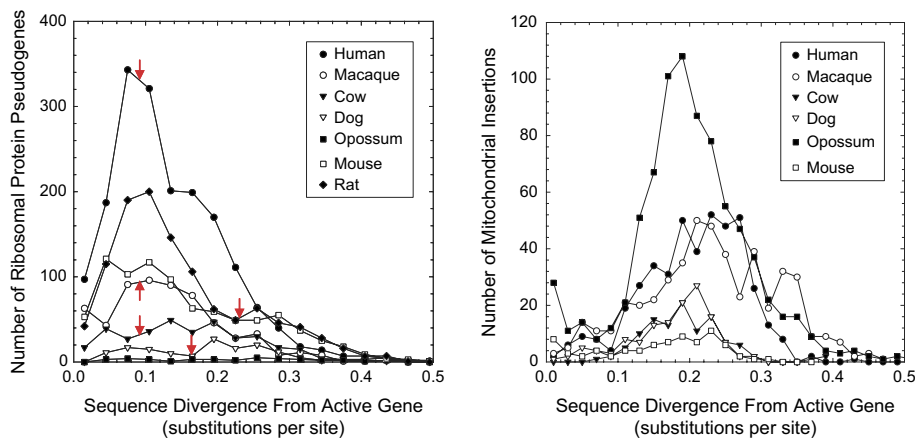


FIG. 3.—Age distributions of processed pseudogenes derived from RP genes (left) and of insertions of mitochondrial protein-coding gene fragments into the nuclear genome (right). Red arrows on the left denote the approximate positions of the KT boundary for various lineages (nearest lines that they touch); for mitochondrial insertions, which diverge more rapidly, these positions are beyond the right side of the graph.

Table 3
Estimated Demographic Parameters for Processed Pseudogene Insertions of RP in Mammalian Lineages with Adequate Data during Phases of Apparent Steady-State Birth/Death Rates (on a timescale of 1% divergence)

Species	Birth Rate	Death Rate	Half-Life	<i>r</i>	df	Range	
Human	391.95 (133.34)	11.51 (0.34)	0.060 (0.005)	0.960	11	0.11	0.50
Macaque	240.92 (104.49)	13.62 (0.43)	0.051 (0.004)	0.960	10	0.17	0.50
Cow	216.83 (103.91)	13.10 (0.48)	0.053 (0.006)	0.949	8	0.20	0.47
Dog	50.75 (31.93)	9.92 (0.63)	0.070 (0.012)	0.883	7	0.23	0.47
Mouse	1,633.34 (792.83)	15.53 (0.49)	0.045 (0.003)	0.985	6	0.29	0.50
Rat	158.10 (50.10)	8.76 (0.32)	0.079 (0.006)	0.966	11	0.11	0.47

NOTE.—Standard errors are in parentheses. Correlation coefficients and degrees of freedom are denoted by *r* and df, and the range denotes the span of sequence divergence over which the regression was run (in units of total substitutions per site between pseudogene and active gene).

appear to be more recent than that for retrotransposons and RP pseudogenes. Again, there is a dramatic disparity in the current number of numts per genome and the expected abundance prior to the KT boundary, with the latter being ~ 288 (127) \times the former, and an expected decline to just 7% (4%) of current levels into the future (tables 2 and 4). And again, the dramatic instabilities in age structure noted for mammalian numts are absent from all other taxa, with the few lineages with adequate numbers of such insertions for analysis all exhibiting an approximately exponential decline with age (fig. 4).

The preceding analyses were confined to placental mammals and opossum, raising the question as to whether monotremes also exhibit ancient discontinuities in the evolutionary dynamics of insertions. An overview of the recently released platypus genome (Warren et al. 2008) suggests that they do not. Although we located too few RP pseudogenes in platypus to perform a demographic analysis, and only a few LTR elements (73 copies with paired ends), the overall distribution of the latter is not discernably different from a negative exponential (fig. 5). The age distribution of the very large number of numts in the platypus genome is also consistent with a roughly steady-state process (fig. 5) and is therefore strikingly different from that seen in other mammals. Although the total fraction of the platypus genome associated with LTR elements is quite low, 0.15% compared with an average of 7.6% (0.9%) for the other mammals in this study, the fraction of the platypus genome associated with all types of mobile elements is comparable to other mammals (44.2% vs. 41.6% (2.7%)). By comparison, the chicken genome contains only 1.3% LTR-associated DNA and 8.6% associated with the full pool of mobile elements (International Chicken Genome Sequencing Consortium 2004), so if birds also ex-

perienced an early history of genomic bloating, they must have been more successful at eradicating the remnants of this earlier period (at least in the lineage leading to chicken).

Discussion

The broad set of observations outlined above is generally consistent with ongoing episodes of genome size reduction having initiated in independent mammalian lineages over the past 50 or so My. Although our results rely on three special categories of insertions that lend themselves to internal age distributional analyses, other classes of mammalian mobile elements appear to have experienced similar reductions in proliferation following the KT boundary. For example, previous analyses of transposons and non-LTR retrotransposons led to the inference of age distributional bulges in the human genome dating to ~ 35 – 50 Ma (International Human Genome Sequencing Consortium 2001; Ohshima et al. 2003; Khan et al. 2006; Gherman et al. 2007; Pace and Feschotte 2007), and similar post-KT behavior of these elements is apparent in the mouse genome (Mouse Genome Sequencing Consortium 2002).

It is seductive to interpret age distributional bulges as outcomes of prior episodes of elevated birth rates rather than as simple consequences of recent reductions in element activity, and indeed most studies on primate and rodent LINES and SINES both (non-LTR elements), have invoked periods of massive increases in insertion rates to explain the past history of these retrotransposons. However, without an explicit model for the temporal dynamics of element birth and death rates like that provided here, it is impossible to evaluate the demographic forces underlying discontinuities in insertion age distributions. This problem is particularly acute with previous studies of mammalian transposons and non-LTR

Table 4
Estimated Demographic Parameters for Nuclear Insertions of Mitochondrial Protein-Coding Genes (numts) in Mammalian Lineages with Adequate Data during Phases of Apparent Steady-State Birth/Death Rates (on a timescale of 1% divergence)

Species	Birth Rate	Death Rate	Half-Life	<i>r</i>	df	Range	
Human	75,086.69 (50,278.18)	29.75 (0.67)	0.023 (0.001)	0.990	4	0.27	0.43
Macaque	208.89 (127.65)	9.09 (0.61)	0.076 (0.016)	0.797	11	0.23	0.47
Cow	258.48 (190.57)	16.74 (0.74)	0.041 (0.007)	0.905	5	0.19	0.37
Dog	22,035.50 (17,705.25)	35.00 (0.80)	0.020 (0.002)	0.983	3	0.21	0.29
Mouse	17,182.13 (15,624.46)	34.98 (0.91)	0.020 (0.002)	0.981	3	0.23	0.31
Opossum	842.09 (269.22)	13.84 (0.32)	0.050 (0.002)	0.984	15	0.19	0.51

NOTE.—Standard errors are in parentheses. The correlation coefficients and degrees of freedom are denoted by *r* and df, and the range denotes the span of sequence divergence over which the regression was run (in units of total substitutions per site between numt and active gene).

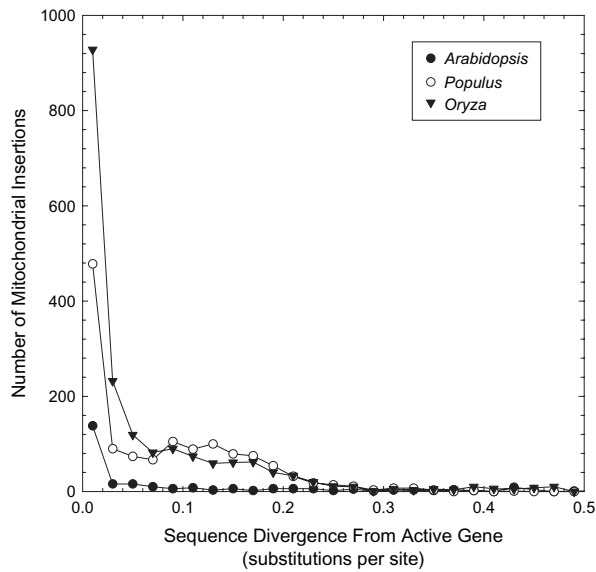


FIG. 4.—Age distributions of numt insertions into land plant genomes, with age being measured as divergence of the numt from the cognate mitochondrial gene.

retrotransposons that have relied upon measures of sequence divergence between extant elements and their ancestral consensus sequences as estimates of element age. As a consensus sequence will be a function of the most common sublineages of inserts, such comparisons are inherently biased with respect to actual age distributions. Thus, it is entirely feasible that the past evolutionary demographics of mammalian transposons and non-LTR retrotransposons reflect the same syndrome of events outlined above for LTR elements, pseudogenes, and numts, that is, simple slowdowns in recent birth rates rather than ancient spikes in element activity.

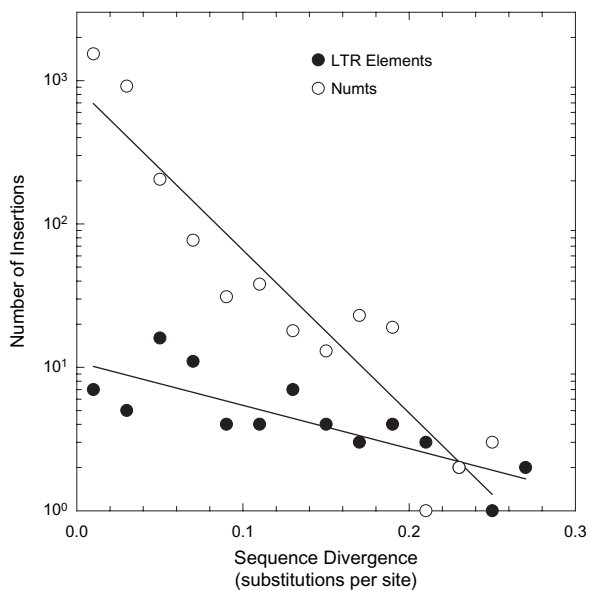


FIG. 5.—Age distributions of total LTR elements and numt insertions in the platypus genome fitted with least squares regressions, with sequence divergences defined as in previous figures ($r^2 = 0.66$ and 0.86 , respectively).

The plausibility of this scenario is supported by three additional observations. First, as noted above, all nonmammalian species observed to date exhibit age distributions of LTR elements that are approximately exponential in form. Under the argument that peaks in element age distributions reflect bursts of insertion activity, one must then postulate that all nonmammalian species are currently experiencing independent episodes of unprecedented activity, for which there is no evidence. Second, one of the most peculiar features of LINES in mammalian genomes is the paucity, and in some cases complete absence, of active elements (Grahn et al. 2005; Cantrell et al. 2008), despite the massive numbers that have obviously resulted from ancient insertions. Third, it might be argued that simple stochastic variation in the rate of base substitutions in LTRs can lead to a gradual decline in element numbers (on the timescale of sequence divergence) to the right of an age distribution peak, yielding a false impression of an earlier steady-state birth–death process. However, under the burst model, this type of statistical artifact leads to much more precipitous declines than those actually observed, whereas when incorporated into a model that allows for an early steady-state process, there is virtually no influence on the age distribution (fig. 6).

The precise mechanisms leading to independent reductions in the proliferation of insertions in mammalian genomes are uncertain, but they must have involved a decline in the physical rate of insertion and/or an increase in the efficiency of selective removal. The dramatic declines in the birth rates of most insertion types in mammals could be mutual consequences of the reduction in the full range of mobile element activities coincident with a decline in numbers of autonomous elements per genome. For example, processed pseudogenes are known to be inadvertently produced and inserted by promiscuous activities of retrotransposons (Esnault et al. 2000), and fragments of mitochondrial DNA are known to be captured during double-strand break repair (Ricchetti et al. 1999). What, however, might have induced the declines in subpopulations of autonomous elements essential to high genome-wide rates of insertion activities? Although post-KT increases in the average deleterious effects of insertions might have such an effect, there is no obvious reason why such changes would be incurred in parallel mammalian lineages but not in other organisms.

In principle, all the above observations might be explained by global increases in the average effective population sizes of mammals following the extinction of the dinosaur megafauna. Under this hypothesis, it is the efficiency, not the strength, of selection that would increase post-KT as the reduced power of genetic drift would have improved the ability of natural selection to eliminate aggressive mobile elements and other mildly deleterious insertions. The fact that the dynamics of insertions in the platypus genome are similar to those found in nonmammalian species is qualitatively consistent with the population size expansion hypothesis, in that such population size increases may have never been possible for this geographically confined lineage.

It is also notable that the average positions of the age distribution peaks for the three classes of mammalian genome insertions are quite different: ~ 56 (13) My for RP

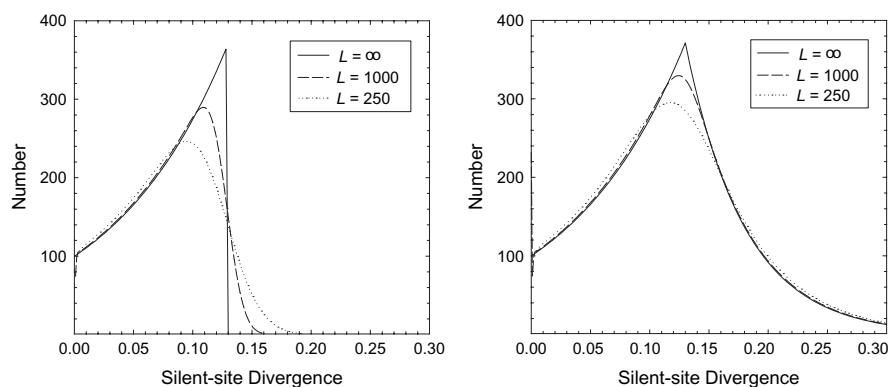


FIG. 6.—Expected age distributions of LTR element insertions assuming an expected level of sequence divergence of 0.002 per site per My in LTRs (in rough accordance with mammalian rates), a birth rate of 100 per haploid genome per My at the current time, rising exponentially into the past to a rate of 5,000 at 65 My. Base substitutional mutations are assumed to accumulate in a Poisson fashion, with L being the number of sites per LTR. The left panel represents a sudden burst of element proliferation, with the birth rate being zero beyond 65 My; the spillover in the age distribution with finite L is a consequence of stochastic variance in realized sequence divergence. The right panel represents a situation in which the birth rate remains at a constant level of 5,000 at all times prior to 65 Ma, in which case the right tail is nearly identical in form regardless of LTR length. In both cases, the rate of element loss is assumed to be a constant 0.04 per My throughout all periods. In our actual genomic analyses, average LTR lengths ranged from 290 to 630 bp, with a mean of 390, so that stochastic variation in LTR divergence estimates will shift the age distribution to a degree that is intermediate between the dashed and dotted curves.

pseudogenes, 27 (9) My for LTR retrotransposons, and 19 (2) My for numts, although errors in element age estimates may cause such peaks to underestimate the position of actual shifts in insertion demography by as much as 10% and perhaps more for certain lineages (fig. 6). Thus, the timing of the proposed initiation of mammalian genome contractions is statistically indistinguishable from the KT boundary, although it may have begun as recently as the Eocene (34–65 Ma) and may have taken as long as 30 My to take full effect. Under the population size expansion hypothesis, the types of insertions to respond first would be those with the highest average deleterious effects, thus implying diminishing average insertion effects from RP pseudogenes to LTR retrotransposons to numts.

Because increases in effective population sizes should influence aspects of sequence evolution as well as genome structural evolution, our hypothesis generates additional testable predictions, for example, that rates of amino acid replacement substitutions (relative to rates at silent sites) should be elevated on internal branches of the mammalian phylogeny prior to the KT boundary. Unfortunately, most of the deep branches that can be confidently ascribed to such periods in mammalian history (e.g., those separating the monotremes, metatherians, and eutherians) are old enough to have incurred substantial saturation effects at silent sites, essentially eliminating the possibility of such an analysis. It is notable, however, that the pressure from biased gene conversion toward G/C content in mammalian genomes is thought to have been weaker deep in mammalian history than it is today (Belle et al. 2004). Because biased gene conversion operates like selection (Nagyaki 1983), with increased efficiency in larger populations, this observation provides an independent line of evidence for increased average effective population sizes in post-KT mammalian lineages, although the specific timing of such changes cannot be estimated with nucleotide usage data.

Although the KT boundary marks the dawn of the age of mammals in terms of global dominance, direct paleon-

ological estimates of long-term population size changes are lacking and will likely be difficult to achieve. Nevertheless, at least three features of the early to mid Tertiary render the hypothesis of mammalian population size expansions plausible. First, 55 My marks the Eocene global thermal maximum, a time when there was no discernible temperature gradient at the poles and tropical forests extended into Greenland (Prothero 1994). During this period, many species had continent-wide geographic ranges and in a number of cases spanned North America, Europe, and Asia. By contrast, the ranges of most late Cretaceous species seem to have been limited to parts of single continents. Substantial evidence indicates that geographic ranges are correlated with global effective population sizes of vertebrate species (Frankham 1996; Gaston et al. 1997). Second, prior to the KT boundary, most placental mammals were insectivorous or carnivorous, with expansions into herbivory and granivory occurring post-KT. Based on ecological considerations, this radiation down the trophic pyramid may have further promoted increases in population sizes, perhaps as much as 10-fold. Third, there was a general tendency for the body sizes of mammals to increase in the early to mid Tertiary (Janis and Damuth 1990). Although mammalian body size is generally negatively correlated with population size per unit area (Damuth 1981), increases in body size are also accompanied by geographic range expansions (Gaston and Blackburn 1996; Diniz-Filho and Torres 2002), so such an effect may have been neutral or even positive with respect to effective population size change during this nonequilibrium period of mammalian diversification.

In summary, although the specific mechanisms remain unclear, some type of global event operating specifically on therian mammals appears necessary to explain the patterns that we have observed. Because most mammalian genomes contain 30–40% readily identified mobile element-associated DNA (Thomas et al. 2003), even after accounting for conserved noncoding DNA (Lynch 2007), well over half of current mammalian DNA appears to be

nonfunctional and subject to the types of fluctuations that we have observed for retrotransposons and pseudogenes. The above results then imply pre-KT mammalian genome sizes at least double those of today, as well as an ongoing decline toward future sizes approximately one-third of the current state. Unless there was a parallel advantage of decreased noncoding DNA in multiple, independent lineages of mammals following the KT boundary, and none associated with invertebrates and land plants, our results challenge the notion that genome size reflects a finely tuned structural determinant of the adaptive phenotypes of organisms (Cavalier-Smith 1978; Hughes AL and Hughes MK 1995; Gregory 2005). In addition, parallel phylogenetic changes in genomic attributes raise significant caveats with respect to attempts to estimate ancestral states from observations on current-day species (Organ et al. 2007).

Many questions remain to be answered with respect to the observations we have made. For example, given the potentially large differences in rates of molecular evolution in various mammalian lineages, the absolute degree of synchrony of lineage-specific shifts in the evolutionary demography of genomic elements is uncertain, as is the precise post-KT timing of such events. Nevertheless, the general patterns outlined above suggest a previously underappreciated aspect of genome evolution—a close connection with ancient historical events, the study of which to this point has largely been in the domain of paleontology.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online (http://www.oxfordjournals.org/our_journals/gbe/)

Acknowledgments

This work was supported by the National Institutes of Health and the National Science Foundation grants to M.L. and Lilly Foundation support to Indiana University via the MetaCytte Initiative. We are greatly appreciative to the members of the Bovine Genome Sequencing Project consortium for access to data prior to publication and to S. Edwards, C. Feschotte, M. Hahn, C. Organ, D. Polly, E. Pritham, and two anonymous reviewers for very helpful comments during the development of this work.

Literature Cited

- Archibald JD. 1996. Dinosaur extinction and the end of an era. New York: Columbia University Press.
- Belle EM, Duret L, Galtier N, Eyre-Walker A. 2004. The decline of isochores in mammals: an assessment of the GC content variation along the mammalian phylogeny. *J Mol Evol*. 58:653–660.
- Bensasson D, Feldman MW, Petrov DA. 2003. Rates of DNA duplication and mitochondrial DNA insertion in the human genome. *J Mol Evol*. 57:343–354.
- Benton MJ, Donoghue PC. 2007. Paleontological evidence to date the tree of life. *Mol Biol Evol*. 24:26–53.
- Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2007. The delayed rise of present-day mammals. *Nature*. 446:507–512.
- Cantrell MA, Scott L, Brown CJ, Martinez AR, Wichman HA. 2008. Loss of LINE-1 activity in the megabats. *Genetics*. 178:393–404.
- Caporale LH, editor. 2006. The implicit genome. New York: Oxford University Press.
- Cavalier-Smith T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *J Cell Sci*. 34:247–278.
- Charlesworth B. 1985. The population genetics of transposable elements. In: Ohta T, Aoki K, editors. Population genetics and molecular evolution. New York: Springer-Verlag. p. 213–232.
- Damuth J. 1981. Population density and body size in mammals. *Nature*. 290:699–700.
- Diniz-Filho JAF, Torres M. 2002. Phylogenetic comparative methods and the geographic range size–body size relationship in new world terrestrial Carnivora. *Evol Ecol*. 16:351–367.
- Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet*. 24:363–367.
- Frankham R. 1996. Relationship of genetic variation to population size in wildlife. *Conserv Biol*. 10:1500–1508.
- Gaston KJ, Blackburn TM. 1996. Conservation implications of geographic range size–body size relationships. *Conserv Biol*. 10:638–646.
- Gaston KJ, Blackburn TM, Lawton JH. 1997. Interspecific abundance–range size relationships: an appraisal of mechanisms. *J Anim Ecol*. 66:579–601.
- Gherman A, Chen PE, Teslovich TM, Stankiewicz P, Withers M, Kashuk CS, Chakravarti A, Lupski JR, Cutler DJ, Katsanis N. 2007. Population bottlenecks as a potential major shaping force of human genome architecture. *PLoS Genet*. 3:e119.
- Grahn RA, Rinehart TA, Cantrell MA, Wichman HA. 2005. Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents. *Cytogenet Genome Res*. 110:407–415.
- Gregory TR, editor. 2005. The evolution of the genome. Boston: Elsevier Academic Press.
- Griffiths DJ. 2001. Endogenous retroviruses in the human genome sequence. *Genome Biol*. 2:REVIEWS1017.
- Hughes AL, Hughes MK. 1995. Small genomes for better flyers. *Nature*. 377:391.
- International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 432:695–716.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature*. 409:860–891.
- Janis CM, Damuth J. 1990. Mammals. In: McNamara KJ, editor. Evolutionary trends. Tucson (AZ): University of Arizona Press. p. 301–346.
- Jing R, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis TH, Flavell AJ. 2005. Insertional polymorphism and antiquity of *PDR1* retrotransposon insertions in *Pisum* species. *Genetics*. 171:741–752.
- Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro NH, editor. Mammalian protein metabolism. New York: Academic Press. p. 21–123.
- Kapitonov VV, Jurka J. 2002. Distribution of transposable and repetitive elements in the *A. thaliana* chromosomes. [Internet] [cited Aug 2007]. Available from: www.girinst.org/server/TransPub/ATinfo.html.
- Khan H, Smit A, Boissinot S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res*. 16:78–87.
- Kirschner M, Gerhart J. 2005. The plausibility of life. New Haven (CT): Yale University Press.

- Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates, Inc.
- Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 420:520–562.
- Nagylaki T. 1983. Evolution of a finite population under gene conversion. *Proc Natl Acad Sci USA*. 80:6278–6281.
- Ohshima K, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N. 2003. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol*. 4:R74.
- Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV. 2007. Origin of avian genome size and structure in non-avian dinosaurs. *Nature*. 446:180–184.
- Pace JK 2nd, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res*. 17:422–432.
- Promislow DE, Jordan IK, McDonald JF. 1999. Genomic demography: a life-history analysis of transposable element evolution. *Proc R Soc Lond B Biol Sci*. 266:1555–1560.
- Prothero DR. 1994. The Eocene-Oligocene transition. New York: Columbia University Press.
- Rho M, Choi JH, Kim S, Lynch M, Tang H. 2007. *De novo* identification of LTR retrotransposons in eukaryotic genomes. *BMC Genomics*. 8:90.
- Ricchetti M, Fairhead C, Dujon B. 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature*. 402:96–100.
- San Miguel PJ, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. 1998. The paleontology of intergene retrotransposons of maize. *Nat Genet*. 20:43–45.
- San Miguel PJ, Ramakrishna W, Bennetzen JL, Busso CS, Dubcovsky J. 2002. Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). *Funct Integr Genomics*. 2:70–80.
- Smit AFA, Hubley R, Green P. 2004. RepeatMasker Open 3.0. [Internet] [cited Jul 2004]. Available from: <http://www.repeatmasker.org>
- Thomas JW, et al. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*. 424:788–793.
- Warren WC, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature*. 453:175–183.
- Wible JR, Rougier GW, Novacek MJ, Asher RJ. 2007. Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary. *Nature*. 447:1003–1006.
- Zhang Z, Carriero N, Zheng D, Karro J, Harrison PM, Gerstein M. 2006. PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics*. 22:1437–1439.
- Zhang Z, Harrison P, Gerstein M. 2002. Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. *Genome Res*. 12:1466–1482.

Marta Wayne, Associate Editor

Accepted April 24, 2009