Rates of Genome Evolution and Branching Order from Whole Genome Analysis

Gavin A. Huttley,* Matthew J. Wakefield,† and Simon Easteal*

*John Curtin School of Medical Research, The Australian National University, Canberra, Australia; and †Walter and Eliza Hall Institute, Parkville, Victoria, Australia

Accurate estimation of any phylogeny is important as a framework for evolutionary analysis of form and function at all levels of organization from sequence to whole organism. Using alignments of nonrepetitive components of opossum, human, mouse, rat, and dog genomes we evaluated two alternative tree topologies for eutherian evolution. We show with very high confidence that there is a basal split between rodents (as represented by the mouse and rat) and a branch joining primates (as represented by humans) and carnivores (as represented by dogs), consistent with some but not the most widely accepted mammalian phylogenies. The result was robust to substitution model choice with equivalent inference returned from a spectrum of models ranging from a general time reversible model, a model that treated nucleotides as either purines and pyrimidines, and variants of these that incorporated rate heterogeneity among sites. By determining this particular branching order we are able to show that the rate of molecular evolution is almost identical in rodent and carnivore lineages and that sequences evolve −11%−14% faster in these lineages than in the primate lineage. In addition by applying the chicken as outgroup the analyses suggested that the rate of evolution in all eutherian lineages is ~30% slower than in the opossum lineage. This pattern of relative rates is inconsistent with the hypothesis that generation time is an important determinant of substitution rates and, by implication, mutation rates. Possible factors causing rate differences between the lineages include differences in DNA repair and replication enzymology, and shifts in nucleotide pools. Our analysis demonstrates the importance of using multiple sequences from across the genome to estimate phylogeny and relative evolutionary rate in order to reduce the influence of distorting local effects evident even in relatively long sequences.

Introduction

The high level of interest in the biology of eutherian mammals is reflected in the number of mammalian genome projects completed or underway. Much of the value of sequencing multiple genomes comes from the comparisons that can be made among them. Comparative analysis can be used to investigate important biological processes from rates of nucleotide sequence evolution to the direction of morphological change, and to understand relationships among these different processes. However, because species arise through a series of hierarchical branching events during evolution they do not have independent evolutionary histories so that when they are analyzed comparatively their phylogenetic relationships must be taken into account. Accurate estimation of phylogeny is thus essential for most comparative evolutionary analyses (Harvey and Pagel 1991). The current release of the draft genome sequence of a marsupial is particularly significant in this context as it provides the necessary outgroup for phylogenetic analysis of the eutherian orders. It also acts as a critical reference in resolving other important questions about the evolution of form and function in eutherian mammals.

Despite the importance of reliable phylogenies for the interpretation of mammalian evolution, consensus has not been reached as to some basal divergences. Early efforts at estimating the phylogeny used comparison of the anatomy of fossil and extant taxa. Neither was a good source of phylogenetic information. Data were interpreted as indicating a radiation rather than a branching phylogeny. Very few eutherian fossils were known from the Cretaceous and many of the orders appear at approximately the same time in the Paleocene, suggesting an adaptive radiation following the extinction of the dinosaurs (Benton 1999). Sequence-based phylogenetic analysis does not support this view, indicating instead that the ordinal branches were clearly separated in time. Most authors now interpret the sequence data as indicating that much of the ordinal divergence occurred much earlier than previously thought - during the Cretaceous (Graur 1993; Easteal 1999; Penny et al. 1999; Arnason et al. 2000; Waddell, Kishino, and Ota 2001; Bininda-Emonds et al. 2007). The discovery of an increasing diversity of eutherian fossils from the Cretaceous supports this view (Hu et al. 2005).

One phylogenetic issue important in the assessment of evolutionary rates is the relative divergence order of Primates, Rodentia, Carnivora, and Artiodactyla. Easteal (1985, 1988, 1990) first suggested that among these orders, there is a basal split between rodents and a branch including carnivores and primates (rodent-first theory), and that an early rodent divergence required a revision of the prevailing theory of the evolutionary rate of nucleotide sequences. Previous analysis, assuming simultaneous divergences of the different orders, suggested that the rate of nucleotide sequence evolution was much faster in rodents than primates (Wu and Li 1985). This interpretation led to the theory that the neutral rate of nucleotide sequence evolution, and hence the rate of mutation, was affected by generation time, since rodents have much shorter generation times than primates. The rate difference was not apparent when an early rodent divergence was assumed (Easteal 1992; Kumar and Subramanian 2002). This branching order appeared to be resolved by a more extensive analysis that refuted the rodent-first theory, indicating, with high confidence, that carnivores diverged first (Murphy et al. 2001). This topology is now assumed in many comparative genomic analyses. The work has also stimulated considerable debate, with numerous authors questioning the magnitude of support for this topology (Misawa and Nei 2003; Kullberg et al. 2006). Analysis of different mixtures of the nuclear and mitochondrial genes used by Murphy et al. (2001) showed the vulnerability of support for this topology to the choice of genes.

Key words: mutation, substitution rate, male biased evolution, eutherian, phylogeny.

E-mail: Gavin.Huttley@anu.edu.au.

doi:10.1093/molbev/msm094
Advance Access publication May 9, 2007

© 2007 The Authors.
This is an Open Access article published under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
used in the analysis, and to differences in the amount of data obtained from each of the sampled lineages (Misawa and Nei 2003). Another critical factor affecting phylogenetic estimates, which was evident at 3rd codon positions of the Murphy et al. data (Kullberg et al. 2006), is systematic variation in the composition (as distinct from the sequence) of nucleotides among compared sequences (Ho and Jermiin 2004; Phillips et al. 2004). Using amino acid sequences (i.e., avoiding synonymous 3rd codon variation) does not entirely overcome the problem as amino acids also vary in composition among lineages (Knight, Freeland, and Landweber 2001).

In support of the carnivore-first branching order have been publications that use retrotransposon insertion events as phylogenetic markers (Kriegs et al. 2006; Nishihara, Hasegawa, and Okada 2006). The use of these multi-gene families for phylogenetic reconstruction is a relatively recent development, yet known aspects of transposable element biology suggest that homoplasies probably affect these molecular markers, casting doubt on results obtained from their application. First, evidence supports ectopic exchange, in the form of nonhomologous recombination and/or gene conversion between element family members, as a major evolutionary force affecting transposable elements (Langley et al. 1988; Huttley, MacRae, and Clegg 1995; Abrusan and Krambeck 2006). Establishing unambiguously the orthology of observed elements is therefore difficult. Second, element insertions are not random with respect to sequence composition and/or chromatin status as shown by the existence of insertion hotspots (Cantrell et al. 2001). Third, lineage sorting, which can result in character distributions incongruent with the true species relationships, of element insertions has been reported (Nishihara, Hasegawa, and Okada 2006). Finally, transposable elements are increasingly being found to have important roles in genome evolution (Bejerano et al. 2006; Kamal, Xie, and Lander 2006) and their status as neutral genetic markers therefore cannot be assumed.

Recent studies strongly indicate that adding more genes is a better way of improving phylogenetic estimates than adding more taxa (see for example, Rosenberg and Kumar 2001; Rokas and Carroll 2005). Furthermore, the enormous volume of data available from analysis of whole genomes coupled with the statistical consistency of maximum likelihood (Chang 1996) suggests that selection of genome blocks and use of appropriate sequence-evolution models should provide very robust phylogenetic estimates. Here we investigate the relatively simple phylogenetic question of the branching order of Primates, Rodentia, and Carnivora, based on comparison of multiple 10kb blocks of nonrepetitive sequences from the human, rat, mouse, and dog genomes, using the marsupial opossum genome (Mikkelsen et al. 2007) as an outgroup reference. We focus on comparing only the rodent-first and carnivore-first topologies. As no previous studies have found substantial support for the third possible, primate-first, topology, we do not consider it here. Using the resulting best-supported tree topology we revisit the question of relative rates of sequence evolution in the lineages of these different eutherian orders.

Another approach to investigating factors that affect mutation and substitution rate is to compare evolutionary rates of nucleotide sequences on sex chromosomes and autosomes, since these spend different amounts of time in male and female lineages, which in turn have different properties. The most obvious of these, and the one most discussed, is generation time; in mammals there are more cell generations in males. Y-chromosomes spend all their time in male lineages and X-chromosomes spend two-thirds of their time in female lineages and one-third in male lineages, whereas autosomes spend an equal amount of time in both lineages. These differences lead to clear predictions about the relative evolutionary rates of nucleotide sequences on these chromosomes, if generation time has an important influence on mutation rate. However, rate differences between chromosomes need to be interpreted with caution as sex-biased evolution may result from other factors, such as methylation pattern, that differ between male and female lineages (Eastal, Collet, and Betty 1995; Huttley et al. 2000).

The marsupial genome sequence reveals a 46.85 Megabase region translocated from an autosomal region to the X chromosome in the mammalian common ancestor, before the divergence of eutherian mammals (Mikkelsen et al. 2007). We use a comparative analysis of this large region and X to autosome ratios to investigate if an accurate estimate of the extent of sex-biased evolution in mammals can be made.

Materials and Methods

We use Mlagan (Brudno et al. 2003) genome sequence alignments from release 40 of Ensembl (http://aug 2006.archive.ensembl.org/). Mlagan is a global multiple aligner that requires prior knowledge of syntenic regions in which alignments can be made. The Ensembl Mlagan alignment pipeline uses WU-BLAST 2.0 (http://blast.wustl.edu) followed by a full Smith-Waterman alignment (Smith and Waterman 1981) to find similar coding exons in these genomes. Mercator (http://bio.math.berkeley.edu/mercator) was used to build a synteny map from all best-matching reciprocal local alignments between these genomes. Finally, Mlagan was used to build multiple alignments of the syntenic regions identified in this way. The assemblies included in our alignments are: Cow (Bos taurus); Chicken (Gallus gallus); WASHUC 1; Dog (Canis familiaris); CanFam 1.0; Human (Homo sapiens); NCBI 36; Mouse (Mus musculus); NCBI m36; Opossum (Monodelphis domestica); NCBI m36; and Rat (Rattus norvegicus); RGSC 3.4. The species tree used to build the alignment was: (Chicken,(Opossum,(Cow,Dog),(Human,(Mouse,Rat))))). As gaps are excluded from our analysis, this guide-tree will have little influence on phylogenetic inference.

The result was a set of alignments that contained either strictly Autosomal-linked blocks (the A set); X-linked blocks (the X set) or, Eutherian X-linked with Marsupial and Bird Autosomal-linked blocks (the A:X set). The alignments contain a mixture of coding and noncoding sequence. Only alignments longer than 10 kb were retained.

Our focus is on substitutions resulting from single nucleotide mutations and the probabilistic evolutionary
models we use are restricted to this kind of variation. We therefore eliminated from the alignments all columns with sequence differences likely to derive from other kinds of mutation such as slipped-strand-mispairing. Runs of 4 or more mono- to tetra-nucleotide repeats were replaced by Ns and columns containing gaps or consisting only of Ns were removed.

We split all alignments into 10 kb blocks and, given the ample volume of data available, discarded remaining segments < 10 kb. Splitting the data into identically sized multiple blocks in this way allowed us to assess the influence of genomic region on evolutionary rate and phylogenetic estimation. Base composition varies across mammalian genomes because of differences in mutational and or DNA repair processes, and the resulting compositional bias is likely to affect rates of substitution (Wolfe, Sharp, and Li 1989) and phylogenetic estimation (Ho and Jermiin 2004; Phillips et al. 2004).

We assessed two alternative tree topologies: the carnivore-first tree (Carnivores, (Rodents, Primates)) (Murphy et al. 2001), which places carnivores as the outgroup to rodents and primates; and the rodent-first tree (Rodents, (Carnivores, Primates)) (Eastueal 1988) which places rodents as the outgroup (fig. 1A and 1B, respectively). Because these lineages have diverged in their base composition (GC%) and the maximum-likelihood reconstruction methods employed assume compositional homogeneity, we sought to minimize the influence of this violation in branch length estimation in two ways: (1) we recoded the sequences into a 2-state alphabet of purines (R) and pyrimidines (Y) (Phillips et al. 2004); (2) any of these R/Y recoded alignment blocks from which a pair of sequences gave a nominally significant departure ($P < 0.05$) from a $\chi^2$ goodness-of-fit test of compositional homogeneity was excluded.

We specified two substitution models for modeling sequence evolution. Both models are continuous time Markov processes of the two states R and Y. Equilibrium probabilities of R and Y were estimated as the averaged frequency across the sequences from an alignment block. In one model we allowed site heterogeneity using a discretized $\Gamma$ distribution (Yang 1997) with four rate classes, as modeling rate-heterogeneity has been demonstrated to reduce the influence of long-branch attraction. All substitution models were implemented using PyEvolve 0.89.9 (Butterfield et al. 2004). For each combination of alignment, substitution model (RY or RY+\Gamma), and tree topology (rodent-first, carnivore-first), the likelihood was maximized using first a global (simulated annealing) optimizer and then a local (Powell) optimizer with default settings.

As we only consider two possible tree topologies, we used a sign test from the difference in log-likelihoods to discriminate between them. Specifically, for each alignment segment we subtracted the log-likelihood estimated for the rodent-first topology ($ln_{L_\text{R}}$) from that estimated for the carnivore-first ($ln_{L_\text{C}}$) topology: $\Delta = ln_{L_\text{C}} - ln_{L_\text{R}}$. A positive $\Delta$ indicates the alignment segment supports the carnivore-first tree, and a negative $\Delta$ indicates support for the rodent-first tree. A two-sided sign test was applied (using PyEvolve) to the resulting set of signs to determine, first, whether there was a significant difference in support for the two topologies and, second, which topology was supported.

We tested for equivalence of evolutionary rate between lineages using the likelihood form of the relative rate test (Muse and Weir 1992). These tests were conducted using the RY model and four-taxon trees consisting of the chicken and opossum outgroups and two eutherian lineages. All possible combinations of eutherian lineages were considered. For assessment of evolutionary rates of opossum against a eutherian lineage, a three-taxon tree was employed with chicken the designated outgroup. Under the null hypothesis of equal rate, the branches of two lineages of interest were set to be equal, and the model parameters were estimated by maximizing the likelihood. For the alternative hypothesis, the branch lengths were allowed to differ. A likelihood ratio (LR) was then determined as $LR = 2(ln_{L_{alt}} - ln_{L_{null}})$, with one degree of freedom (df) equal to the difference in numbers of free parameters between the alternative and null hypotheses.

To test the hypothesis of rate equivalence across alignments, we used the sum of their LRs on the assumption that alignments have evolved independently, with degrees of freedom equal to the number of alignment blocks. The probability that a LR of equal or greater size than the observed LR would occur by chance was estimated from the $\chi^2$ distribution (Goldman 1993; Ota et al. 2000). We evaluated the proportion of a data set contributing to a significant LR test by determining the number of alignment segments for which the marginal LR statistic was significant, after applying a Bonferroni correction for the number of segments in the data set. We also determined the number of alignment segments for which the marginal LR statistic was significant at the nominal 0.05 level. All relative rate tests were implemented using PyEvolve and were conducted using the optimization procedure described above. On rejection of the null hypothesis of rate equivalence, a subsequent two-sided sign test was applied to assess whether there was a consistent direction to the rate differences. The number of alignment blocks for which species 1 of the eutherian pair had a longer branch length than species 2 was counted as a success, the number of trials being the total number of alignment blocks.
Results

We estimated the maximum-likelihood tree for each 10-kb alignment block. We then counted the number of alignment blocks that support each of the two alternative tree topologies and used that number as a measure of phylogenetic support. The substantial volume of aligned genome sequence data available means that this is a good estimate of the support for the topology of the whole genome. The rodent-first topology (fig. 1A) was strongly favored over the carnivore-first (fig. 1B) topology by our analysis. Of 462 compositionally homogeneous A (autosomal-linked) and X (X-linked) blocks (totaling 4.62 Mb of aligned sequences), 386 (~84%) supported the rodent-first tree, and 76 supported the carnivore-first tree (two-sided \( P = 1.1e-16 \); fig. 2). This pattern was robust because of the data set and evolutionary model used: equivalent results were obtained from the A and X data sets analyzed separately (fig. 2 A and B), from analyses without the chicken sequence (results not shown), and from applying either the RY, RY+\( \Gamma \), a nucleotide general time-reversible (GTR), or the GTR+\( \Gamma \) substitution model to these data sets (see Supplementary Material online). This consistency across evolutionary models with and without rate heterogeneity affirms strong support of the rodent-first topology and that this support is unlikely to result from long-branch attraction (Yang 1994).

Relative rate tests indicated a significant difference in evolutionary rate between marsupials and eutherians for all data sets (table 1), with eutherians having a relatively slower rate for nearly all alignment blocks. The difference in rate was greater for \( \mathbf{X} \) alignments (mean eutherian/marsupial ratio 0.6041, s.d. = 0.1199) than \( \mathbf{A} \) alignments (mean eutherian/marsupial ratio 0.6972, s.d. = 0.1342). This pattern is consistent with a reduced average rate of substitution affecting X-linked blocks arising from male-biased evolution in eutherians.

All data sets indicate that the primate lineage has evolved significantly more slowly than the other sampled eutherian lineages (table 1; fig. 3). This result is robust to whether a 3, 4, or (when possible) 6 taxon tree was used (results not shown). The mean proportion of primate to rodent substitution rates was 88.5% (s.d. = 17%), and the mean proportion of primate to carnivore was 86% (s.d. = 13%).

There is little evidence supporting a systematic rate difference between the carnivore and rodent lineages or between the rodent lineages. For the \( \mathbf{A} \) and \( \mathbf{X} \) data sets, the rates in the rodent lineages are marginally higher than in the carnivore lineage (table 1), whereas the reverse is true for the \( \mathbf{A}:\mathbf{X} \) (eutherian X-linked, marsupial autosomal) data set. Performing a sign test on these combined results did not allow rejection of the null hypothesis of rate equivalence (two-sided \( P \approx 0.09, 0.25 \) for mouse and rat, respectively). Although the \( \mathbf{X} \) data set supports the previously reported (Cooper et al. 2004) overall significantly higher evolutionary rate in the rat than in the mouse lineage (table 1), this was not supported by the other data sets. Performing a sign test on the combined mouse–rat results also did not allow rejection of the null hypothesis (two-sided \( P \approx 0.13 \)).

All analyses demonstrated significant correlations between branch lengths (substitution rates) for the same alignment segment in different lineages (fig. 4). Thus, for example, the fastest evolving blocks between human and dog tend also to be the fastest evolving between mouse and rat. This is consistent with previous observations of local correlations in evolutionary rate that persist across lineages (Webster et al. 2004).

Discussion

This work demonstrates the value of a marsupial genome sequence as an outgroup in resolving key issues about the evolution of eutherian mammals. We have demonstrated with very high confidence that the rodents diverged before carnivores and primates. This analysis contributes to the development of a eutherian phylogeny based on whole genome sequences. Extension of our approach will provide an increasingly detailed and accurate picture of the evolutionary history and relationship between mammals, and a detailed understanding of the different molecular
evolutionary processes and how they vary across different lineages. Our approach, using large-scale data, allows aggressive filtering of alignment positions influenced by processes other than point mutation and sites that have sequence ambiguities or missing data in the alignments. The scale of the resulting data set is large enough to dilute out regional effects and to give a strong indication of the central trend of the evolution of the entire genome. That scale further enables alternate approaches to measuring support for a tree topology. By using identically sized segments sampled from the genome alignments, it was possible to attain an estimate of the proportion of alignments among the sampled genomes that support the alternative trees. The scale of the resulting data set is large enough to dilute out regional effects and to give a strong indication of the central trend of the evolution of the entire genome.

An important aspect of our results is the finding that a minority of the 10-kb alignment blocks we analyzed in- 

cludes (Opossum was the 3-taxon outgroup); and, A:X.(number) is X-linked eutherian, Autosomal other, a (number)-taxon tree (Chicken was the 3-taxon outgroup).

Table 1
Comparison of Evolutionary Rates Between Different Mammalian Kinages

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Species 1</th>
<th>Species 2</th>
<th>LR</th>
<th>df</th>
<th>P</th>
<th>Bonf.</th>
<th>SigAlns</th>
<th>Greater</th>
<th>Less</th>
<th>Sign P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Dog</td>
<td>Human</td>
<td>2271</td>
<td>373</td>
<td>0.00</td>
<td>46</td>
<td>153</td>
<td>322</td>
<td>51</td>
<td>0.000</td>
</tr>
<tr>
<td>A</td>
<td>Dog</td>
<td>Mouse</td>
<td>2914</td>
<td>373</td>
<td>0.00</td>
<td>61</td>
<td>161</td>
<td>167</td>
<td>206</td>
<td>0.049</td>
</tr>
<tr>
<td>A</td>
<td>Dog</td>
<td>Rat</td>
<td>2931</td>
<td>373</td>
<td>0.00</td>
<td>59</td>
<td>160</td>
<td>175</td>
<td>198</td>
<td>0.255</td>
</tr>
<tr>
<td>A</td>
<td>Human</td>
<td>Mouse</td>
<td>2670</td>
<td>373</td>
<td>0.00</td>
<td>52</td>
<td>182</td>
<td>81</td>
<td>292</td>
<td>0.000</td>
</tr>
<tr>
<td>A</td>
<td>Human</td>
<td>Rat</td>
<td>2596</td>
<td>373</td>
<td>0.00</td>
<td>48</td>
<td>172</td>
<td>90</td>
<td>283</td>
<td>0.000</td>
</tr>
<tr>
<td>A</td>
<td>Mouse</td>
<td>Rat</td>
<td>498</td>
<td>373</td>
<td>0.00</td>
<td>1</td>
<td>32</td>
<td>185</td>
<td>188</td>
<td>0.918</td>
</tr>
<tr>
<td>A:X.4</td>
<td>Dog</td>
<td>Human</td>
<td>296</td>
<td>37</td>
<td>0.00</td>
<td>10</td>
<td>20</td>
<td>35</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>A:X.4</td>
<td>Dog</td>
<td>Mouse</td>
<td>251</td>
<td>37</td>
<td>0.00</td>
<td>9</td>
<td>18</td>
<td>24</td>
<td>13</td>
<td>0.099</td>
</tr>
<tr>
<td>A:X.4</td>
<td>Dog</td>
<td>Rat</td>
<td>274</td>
<td>37</td>
<td>0.00</td>
<td>10</td>
<td>18</td>
<td>24</td>
<td>13</td>
<td>0.099</td>
</tr>
<tr>
<td>A:X.4</td>
<td>Human</td>
<td>Mouse</td>
<td>181</td>
<td>37</td>
<td>0.00</td>
<td>5</td>
<td>16</td>
<td>14</td>
<td>23</td>
<td>0.188</td>
</tr>
<tr>
<td>A:X.4</td>
<td>Human</td>
<td>Rat</td>
<td>213</td>
<td>37</td>
<td>0.00</td>
<td>6</td>
<td>19</td>
<td>13</td>
<td>24</td>
<td>0.099</td>
</tr>
<tr>
<td>A:X.4</td>
<td>Mouse</td>
<td>Rat</td>
<td>52</td>
<td>37</td>
<td>0.05</td>
<td>0</td>
<td>2</td>
<td>16</td>
<td>21</td>
<td>0.511</td>
</tr>
<tr>
<td>A:3</td>
<td>Dog</td>
<td>Opossum</td>
<td>10836</td>
<td>373</td>
<td>0.00</td>
<td>256</td>
<td>331</td>
<td>18</td>
<td>355</td>
<td>0.000</td>
</tr>
<tr>
<td>A:3</td>
<td>Human</td>
<td>Opossum</td>
<td>14331</td>
<td>373</td>
<td>0.00</td>
<td>303</td>
<td>356</td>
<td>5</td>
<td>368</td>
<td>0.000</td>
</tr>
<tr>
<td>A:3</td>
<td>Opossum</td>
<td>Rat</td>
<td>12970</td>
<td>373</td>
<td>0.00</td>
<td>274</td>
<td>341</td>
<td>365</td>
<td>8</td>
<td>0.000</td>
</tr>
<tr>
<td>A:3</td>
<td>Opossum</td>
<td>Rat</td>
<td>13209</td>
<td>373</td>
<td>0.00</td>
<td>274</td>
<td>341</td>
<td>364</td>
<td>9</td>
<td>0.000</td>
</tr>
<tr>
<td>X</td>
<td>Dog</td>
<td>Human</td>
<td>381</td>
<td>89</td>
<td>0.00</td>
<td>9</td>
<td>30</td>
<td>73</td>
<td>16</td>
<td>0.000</td>
</tr>
<tr>
<td>X</td>
<td>Dog</td>
<td>Mouse</td>
<td>437</td>
<td>89</td>
<td>0.00</td>
<td>12</td>
<td>37</td>
<td>39</td>
<td>50</td>
<td>0.289</td>
</tr>
<tr>
<td>X</td>
<td>Dog</td>
<td>Rat</td>
<td>485</td>
<td>89</td>
<td>0.00</td>
<td>13</td>
<td>37</td>
<td>37</td>
<td>52</td>
<td>0.137</td>
</tr>
<tr>
<td>X</td>
<td>Human</td>
<td>Mouse</td>
<td>524</td>
<td>89</td>
<td>0.00</td>
<td>15</td>
<td>41</td>
<td>16</td>
<td>73</td>
<td>0.000</td>
</tr>
<tr>
<td>X</td>
<td>Human</td>
<td>Rat</td>
<td>637</td>
<td>89</td>
<td>0.00</td>
<td>23</td>
<td>48</td>
<td>13</td>
<td>76</td>
<td>0.000</td>
</tr>
<tr>
<td>X</td>
<td>Mouse</td>
<td>Rat</td>
<td>154</td>
<td>89</td>
<td>0.00</td>
<td>1</td>
<td>11</td>
<td>31</td>
<td>58</td>
<td>0.006</td>
</tr>
<tr>
<td>A:X.3</td>
<td>Dog</td>
<td>Opossum</td>
<td>1592</td>
<td>37</td>
<td>0.00</td>
<td>35</td>
<td>35</td>
<td>1</td>
<td>36</td>
<td>0.000</td>
</tr>
<tr>
<td>A:X.3</td>
<td>Human</td>
<td>Opossum</td>
<td>2252</td>
<td>37</td>
<td>0.00</td>
<td>35</td>
<td>35</td>
<td>0</td>
<td>37</td>
<td>0.000</td>
</tr>
<tr>
<td>A:X.3</td>
<td>Opossum</td>
<td>Mouse</td>
<td>2305</td>
<td>37</td>
<td>0.00</td>
<td>35</td>
<td>36</td>
<td>36</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>A:X.3</td>
<td>Opossum</td>
<td>Rat</td>
<td>2268</td>
<td>37</td>
<td>0.00</td>
<td>35</td>
<td>36</td>
<td>36</td>
<td>1</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note. — LR, df, P are likelihood ratio statistics; degrees of freedom and corresponding probability estimated from the $\chi^2$ distribution; Bonf. is the number of alignments significant after a Bonferroni correction for the number of alignments; SigAlns is the number of nominally significant (marginal LR $P < 0.05$) alignments; Greater, Less are the number of alignments for which branch length from Species 1 was greater or less than, respectively, the branch length from Species 2; Sign $P$ is the probability from the two-sided sign test; A is the autosomal-linked data, a 4-taxon tree ((Species 1, Species 2), Opossum, Chicken); X represents the X-linked data, a 3-taxon tree (Opossum was the 3-taxon outgroup); and, A:X.(number) is X-linked eutherian, Autosomal other, a (number)-taxon tree (Chicken was the 3-taxon outgroup).
argue that our analysis is conservative. We would expect accuracy of synteny (and thus orthology) to increase with increasing sequence coverage of a genome, because of fewer errors in assembly. The impact of assembly errors on genome scale analyses may be a bias toward clustering of genomes that are more extensively sequenced, because orthologous regions will have evolutionary distances less than or equal to those between paralogous regions. As human and mouse are the most densely sequenced genomes in our study, such a bias would favor the carnivore-first topology, which is not what we observed.

We have found that the rate of nucleotide substitution in the primate lineage has been \( \sim 11\% - 14\% \) slower than the rates in the rodent and carnivore lineages, consistent with the \( \sim 9\% \) slowdown estimated from fourfold degenerate sites of protein coding genes (Kumar and Subramanian 2002). Our estimates from recoded sequences should closely reflect the underlying true substitution rate differences, as they are less likely to be affected by compositional shifts between genomes. A slower rate of molecular evolution has previously been explained as resulting from a longer generation time (Graur 1993). However, we also show that the nucleotide substitution rates in the rodent and carnivore lineages have been very similar and that the eutherian lineages are \( \sim 30\% \) slower than the marsupials, as judged from analysis applying the chicken as outgroup. The rate differences between these taxa are inconsistent with a global molecular clock, but they are also inconsistent with a generation time effect, calling into question the role of generation time in the primate slowdown. A more likely explanation is that there are differences in the biochemical processes associated with DNA metabolism between the lineages. Attributes of DNA metabolism have been demonstrated to have the capacity to evolve, and there are known differences between the sampled lineages. For instance, the nucleotide excision repair (NER) system deals with a diverse spectrum of DNA lesions via one of two main pathways: global genomic repair or transcription coupled repair. An important lesion type targeted by NER is cyclobutane pyrimidine dimers (CPDs). It has been demonstrated that rodents lack the ability to repair CPDs via the global genomic repair pathway (Tang et al. 2000). As a slow rate of substitutions affecting the time to peak tension (Tpt) has been attributed to the influence of NER (Arndt, Hwa, and Petrov 2005, H. Lindsay, P. Maxwell and G. A. Huttley, unpublished data), the global genomic operation of this repair process could contribute to the slower rate of primate substitutions. It is also the case that repair of CPDs and pyrimidine (6–4) pyrimidone photoproducts differs between eutherian mammals and marsupials. As the latter possess a highly efficient photolyase (Schul et al. 2002), this seems an unlikely contributor to the elevated substitution rate in marsupials.

Figure 4 shows that the evolutionary rates of some sequence blocks are substantially greater than others, and that these differences in rate are correlated across lineages. Local differences in evolutionary rate have been known for a long time and may be caused by a number of factors, including levels of transcription (Brudno et al. 2003), rates of recombination and/or gene conversion (Fullerton, Bernardo Carvalho, and Clark 2001), or differences in replication timing coupled with shifts in the composition and size of the dNTP pools (Wolfe, Sharp, and Li 1999; Martomo and Mathews 2002).

Despite these rate correlations across lineages, it is evident from figure 3 that some sequence blocks have

**FIG. 3.**—Extent of branch length difference between lineage pairs in relative rate tests of autosomal-linked (A) and X-linked (B) alignments. Y-axis is the number of 10kb alignment segments. The vertical grey line is the expectation under a molecular clock—no difference in substitution rate (branch length) between the lineages. Shifts in the distribution to the left of that line indicate the numerator species is evolving slower than the denominator species.
evolved substantially faster in primates, even though there is an overall primate slowdown. As with phylogenetic estimation, this finding demonstrates the limitations of comparisons of relative evolutionary rates based on alignments from one or a small number of regions. We conclude that alignments from many regions are needed to obtain reliable estimates of molecular evolutionary rates between lineages.

The translocation of a 46.85 Megabase autosomal region from the mammalian common ancestor onto the X chromosome prior to the eutherian radiation theoretically provides a large data set from which to address the question of male-biased evolution. Our analyses of this region are consistent with a reduced average rate of substitution affecting X-linked blocks arising from male-biased evolution in eutherians. However, this data set is confounded by variation in rates between species and by the fact that the X-chromosome has lower gene density than the autosomes (Ross et al. 2005). The inability to disentangle the male bias influence from the rate effect on a per alignment basis invalidates the mean of this distribution as valid data for calculating the magnitude of male bias. We do not estimate a distribution of the male-biased evolution statistic ($\alpha_m$) from individual alignments, because these confounding effects generate invalid negative or infinite values for many of the individual alignments.

Estimates of the ratio of mutation rates in male and female lineages, obtained from comparing substitution rates on auto- and sex chromosomes, range from 1.53 to 7.58 (Berlin et al. 2006; Goetting-Minesky and Makova 2006). Male-biased evolution may be due to any factor affecting mutation rate that differs systematically between male and female germlines, including the number of rounds of DNA replication and the pattern and extent of DNA methylation (Easteal, Collet, and Betty 1995; Huttley et al. 2000). From our results we cannot distinguish between these different causes. Our results do, however, confirm the observations of Berlin et al. (2006) that regional variation is a significant confounder of estimates of male bias. This is critically important in many studies that rely on gametologous genes (nonrecombining orthologs that spend differing amounts of time in the male and female germline), as this is a highly restricted data set and limits the amount of data that can be used.

The present study underlines the value of the opossum genome as a reference for addressing important issues about the evolution of eutherian mammals. Our results address a long-standing issue in eutherian systematics, providing robust support for rodents as an outgroup to carnivores and primates. We have reliably estimated that the rate of nucleotide substitution has been $\sim14\%$ slower in the primate lineage than in the rodent and carnivore lineages, but that there has been no appreciable rate difference between rodent and carnivore lineages, suggesting that generation time is not an important determinant of substitution rate differences. The acceleration of opossum relative to all eutherians can also not be accounted for by a generation time effect. Finally we have confirmed the general finding of greater substitution rate in male lineages, but the limitation of the data prevents us from extensive exploration of this issue.

Acknowledgments

We thank the Opossum Genome Sequencing Consortium and the Broad Institute of Harvard and MIT for providing the Monodelphis domestica genome sequence, and Javier Herrero for performing the sequence alignments. This project was supported by a time allocation from the Australian Partnership for Advanced Computing, Australian Research Council grants DP0450066 and LP0347613, and National Health and Medical Research Council grant 366739.

Literature Cited


Sudhir Kumar, Associate Editor

Accepted May 4, 2007