The Dynamics, Causes, and Impacts of Mammalian Evolutionary Rates Revealed by the Analyses of Capybara Draft Genome Sequences

Isaac Adeyemi Babarinde[1,2] and Naruya Saitou[2,3,4,5,*]

[1]Department of Biological Sciences, Southern University of Science and Technology, Shenzhen, China

[2]Population Genetics Laboratory, National Institute of Genetics, Mishima, Japan

[3]School of Medicine, University of the Ryukyus, Okinawa, Japan

[4]Department of Genetics, School of Life Science, Graduate University for Advanced Studies, Mishima, Japan

[5]Department of Biological Sciences, Graduate School of Science, University of Tokyo, Japan

*Corresponding author: E-mail: saitounr@nig.ac.jp.

Accepted: 23 July 2020

Data deposition: The newly generated short sequence reads and the genome assembly draft have been deposited at the DNA Data Bank of Japan under the Bioproject accession PRJDB7394.

Abstract

Capybara (*Hydrochoerus hydrochaeri*) is the largest species among the extant rodents. The draft genome of capybara was sequenced with the estimated genome size of 2.6 Gb. Although capybara is about 60 times larger than guinea pig, comparative analyses revealed that the neutral evolutionary rates of the two species were not substantially different. However, analyses of 39 mammalian genomes revealed very heterogeneous evolutionary rates. The highest evolutionary rate, 8.5 times higher than the human rate, was found in the Cricetidae–Muridae common ancestor after the divergence of Spalacidae. Muridae, the family with the highest number of species among mammals, emerged after the rate acceleration. Factors responsible for the evolutionary rate heterogeneity were investigated through correlations between the evolutionary rate and longevity, gestation length, litter frequency, litter size, body weight, generation interval, age at maturity, and taxonomic order. The regression analysis of these factors showed that the model with three factors (taxonomic order, generation interval, and litter size) had the highest predictive power ($R^2 = 0.74$). These three factors determine the number of meiosis per unit time. We also conducted transcriptome analysis and found that the evolutionary rate dynamics affects the evolution of gene expression patterns.

Key words: capybara, rodents, genome sequencing, evolutionary rate, transcriptome, expression dynamics.

Significance

Capybara is the largest rodent species. We newly determined its genome sequences and compared them with those of other mammalian genomes and found that generation interval and litter size are important to determine the molecular evolutionary rates.

Introduction

Molecular clock (Zuckerkandl and Pauling 1965) is the constancy of evolutionary rates across lineages. This constancy has been widely used in determining the divergence time from nucleotide or amino acid sequences (Vawter et al. 1980; Hasegawa et al. 1985; Bromham and Penny 2003; Peterson et al. 2004), and its theoretical basis is the neutral theory of molecular evolution (Kimura 1983). As the number of genome sequences increases, the heterogeneity of nucleotide divergence among genes is now widely accepted

(Bromham 2011; Dos Reis et al. 2015; Zhu et al. 2015; Takezaki 2018). Notably, genes associated with immunity and defense tend to have higher nucleotide divergence. On the other hand, genes associated with morphogenesis and development tend to have lower divergence (Babarinde and Saitou 2013). Although differences in intensity and direction of selection pressure contributes to the differences in overall nucleotide divergence, analyses using synonymous substitution patterns have suggested that difference in the substitution rates contributes to the nucleotide diversity measured in GC contents (Duret et al. 2002; DeRose-Wilson and Gaut 2007) and noncoding sequences conserved over long evolutionary time tend to have lower GC content (Babarinde and Saitou 2013).

The heterogeneity of substitution rates is not only limited to genes. Several studies have reported the heterogeneity of evolutionary rates across lineages. An interesting example of different evolutionary rates in mammals is the difference between rodent and primate evolutionary rates (Li and Wu 1987; Li et al 1996). Using the available nucleotide sequences for mouse, rat, human, and dog, Li and Wu (1987) reported that the rodent evolutionary rate is much higher than that of primates. The observed differences were then attributed to the difference in generation interval, which causes difference in the number of replications per year. The result implies that species with short generation time would have higher evolutionary rates. When generation interval was used as the unit of time, the differences in evolutionary rates was not found (Li et al. 1996). Martin and Palumbi (1993) proposed the reconsideration of the generation time hypothesis to include physiological processes. It is now widely accepted that molecular clock does not always hold. In fact, many studies have supported differences in mouse and human evolutionary rate, not only in protein-coding sequences (Dos Reis 2015, Zhu et al. 2015) but also in conserved noncoding sequence evolution (Hiller et al. 2012; Takahashi and Saitou 2012; Babarinde and Saitou 2013; Saber et al. 2016; Hettiarachchi and Saitou 2016).

A number of hypotheses have been proposed to explain evolutionary rate differences across lineages (Martin and Palumbi 1993; Ohta 1993; Bromham et al. 1996; Huttley et al. 2007; Thomas et al. 2010; Bromham 2011). A notable hypothesis is the generation time hypothesis (Martin and Palumbi 1993; Ohta 1993; Bromham et al. 1996). The hypothesis predicts that species with long generation interval would have lower replications per unit time. Since mutations arise from replication errors, more replications would imply higher substitution rate (Bromham 2011). The generation time hypothesis has been found to hold across broad range of species, including invertebrates (Thomas et al. 2010). It should be noted that Kumar and Subramanian (2002) challenged this hypothesis. Using complete genomes of several mammalian species, Huttley et al. (2007) also reported that their observation is inconsistent with the generation time hypothesis. They reported that replication enzymology and shifts in nucleotide pools are the factors that contribute to the differences in mutation rates.

Another attribute that has been associated with substitution rate difference among species is the body size (Martin and Palumbi 1993). As body size has been suggested to be correlated to the metabolic rates and generation interval, it is logical to conceive the relationship between body size and evolutionary rate. Indeed, a number of reports have shown results suggesting the relationship (Martin and Palumbi 1993; Speakman 2005). On the contrary, other studies (Gardner et al. 2011; Clavel and Morlon 2017) have reported that body size is adaptive, suggesting higher evolutionary rates in related coding regions of large body size animals due to positive selection.

The earlier studies (Li and Wu 1987; Li et al. 1996) that reported evolutionary rate difference between primates and rodents were carried out in pregenomic era. At that time, it was not easy to analyze many species because of limited data. At the time of the study, mouse and rat were used as the representative of rodent species, whereas human was used as the representative primate species. Although the difference in evolutionary rates of human and mouse/rat is widely supported, it is still unclear whether the evolutionary rates of these species reflect the overall respective orders. Although mouse and rat were found to have higher evolutionary rates than primates (Li and Wu 1987; Li et al. 1996; Hiller et al. 2012; Takahashi and Saitou 2012; Babarinde and Saitou 2013), it is still difficult to generalize that rodents have higher evolutionary rates than primates. This is because mouse and rats are phylogenetically very close, both of them belong to subfamily Murinae (Musser and Carleton 1993; Huchon et al. 2002; Blanga-Kanfi et al. 2009). It is also possible that the high evolutionary rate is a unique attribute of Murinae. For example, Kim et al. (2011) reported that the nucleotide divergence of naked mole rat is much lower than those of mouse and rat. Similarly, the reported low evolutionary rate in primates may be unique to the human lineage and may not represent the general feature of primates. Therefore, it is important to further investigate more phylogenetically and phenotypically heterogeneous species before concrete conclusions could be drawn.

With the advancement of sequencing technology, it is now possible to investigate wider genomic regions of larger number of species. Higher number of species and higher evolutionary rates at least in some species make rodents an interesting mammalian order for the study of evolutionary rate. Determination of capybara genome sequences would be very important in investigating evolutionary rate dynamics because capybara is the largest extant rodent species with adult body weight up to 91 kg (Alvaro and Juhani 1986). Although mouse and rat are phylogenetically close and are

both murids, capybara belongs to Ctenohysterica (Musser and Carleton 1993; Huchon et al. 2002; Blanga-Kanfi et al. 2009). The genome sequencing of capybara would make it possible to further investigate heterogeneity of rodent evolutionary rates and the impact of body size.

There are three major reasons why the investigation of capybara genome would shed an important light into the understanding of evolutionary rate dynamics. First, being a rodent species, analysis of the capybara genome would reveal the evolutionary patterns among rodents. Second, capybara is the extant rodent species with the highest body weight. Third, the generation interval of capybara is not especially long as would be expected from its body size. The large body size with the unexpectedly short generation interval makes capybara an important species for evolutionary studies. Starting with the comparison between capybara and guinea pig, the heterogeneity found in mammalian evolutionary rates with higher rates and dynamics in rodents are presented. The causes of these evolutionary rate dynamics are investigated. Finally, the impact of these dynamics on gene expression patterns is investigated with the link between the dynamics and the speciation.

Materials and Methods

DNA Extraction, Library Preparation, and Sequencing

Masseter muscle of an adult male capybara which died of an undisclosed cause was kindly donated by Izu Shaboten Park, Japan. The tissue sample was digested with Qiagen ATL reagent and protease and then treated with RNase. DNA was then extracted using sodium dodecyl sulphate lysis (Akinwole and Babarinde 2019) and purified with phenol isomyl chloroform. The quantity and the quality of the extracted DNA were checked using nanodrop, qubit, and Agilent 2100 Bioanalyzer. Agilent SureSelect QXT library preparation kit was used for the library preparation, and 750 bp average fragment size was selected. The library preparation was carried out in two technical replications. Three runs of sequencing were done for the pooled library set on the Illumina MiSeq platform using 600 cycle kits with 350 bp read1 and 250 bp read2. Further details of the library preparation and sequencing are presented in supplementary methods, Supplementary Material online.

Extraction of Reliable Predicted Capybara Gene Sequences

The extraction of reliable gene sequences from raw reads started with de novo genome assembly (supplementary fig. S1, Supplementary Material online). CLC workbench installed on Maser (Kinjo et al. 2018) was used for contig formation. The read lengths were set to 350 and 250 bp. SOAP de novo (Li et al. 2010) was used for scaffolding and gap closing. To get a reliable genomic region, the raw reads were strictly filtered using read cleaner, a newly written python script. The filtering was done such that all the base positions were at least Q20 and the minimum length was 50 bp. In paired reads with a read shorter than 50 bp, the other read was considered as a single read. Critically filtered reads were mapped to the assembled draft with BWA (Li and Durbin 2009). SAMtools and BCFtools (Li et al. 2009) were used to make the VCF file containing depth and mapping quality per site from the alignment involving filtered reads. Regions with depth of between 3× and 30× and with mapping quality of at least 30 were retained, whereas the remaining regions were masked to N.

For gene prediction, guinea pig transcript sequences were retrieved from Ensembl database. For each gene, the longest transcript was used for gene prediction using GMAP (Wu and Watanabe 2005). Both exon and protein sequences were predicted from the assembled capybara genome. Transcripts shorter than 100 bp were discarded. Detailed GMAP commands are presented in supplementary methods, Supplementary Material online. The genes predicted with GMAP were then used for evolutionary analyses.

Evolutionary Analyses

Protein- and nucleotide-coding sequences for evolutionary analyses were downloaded from Ensembl and UCSC databases. Initially, the genome, amino acid and coding sequences of 57 species from 17 orders (see supplementary data, Supplementary Material online) were downloaded from public databases. Species in which the percentage of undetermined positions was >5% and/or the number of annotated genes was <13,000 were discarded (supplementary fig. S2, Supplementary Material online). As another quality check, CLUSTALW2 alignment (Larkin et al. 2007) of coding nucleotide sequences was used to remove species with abnormally long phylogenetic branches (supplementary fig. S3, Supplementary Material online). Finally, 39 mammalian species (see table 1), including 37 Boreoeutherian species and two outgroup species were included. For each gene, amino acid and coding sequences of the longest transcript were retrieved.

Using guinea pig as the reference species, pairwise homology searches were conducted between genome sequences of guinea pig and all 38 other mammalian species. Guinea pig was used as the reference because it is the closest species to the newly sequenced capybara. BLASTP (Altschul et al. 1997) was used for homology search. For rodent analyses, all rodent species and rabbit (as an outgroup) were used. For mammalian analyses, all 39 species were used.

The amino acid sequences were concatenated and used for the construction of phylogenetic tree to establish phylogenetic relationship. A total of 726,242 aligned amino acids with no gaps were used in the analyses involving mammalian species. We used the neighbor-joining (NJ) method (Saitou

Table 1
Mammalian Species Used in the Evolutionary Analyses

Species Name	Common Name	Order
Hydrochoerus hydrochaeris	Capybara (genome determined in this study)	Rodentia
Mus musculus	House mouse	Rodentia
Dipodomys ordii	Kangaroo rat	Rodentia
Chinchilla lanigera	Long-tailed chinchilla	Rodentia
Mesocricetus auratus	Syrian hamster	Rodentia
Microtus ochrogaster	Prairie vole	Rodentia
Rattus norvegicus	Rat	Rodentia
Cricetulus griseus	Chinese hamster	Rodentia
Octodon degus	Common degu	Rodentia
Cavia procellus	Ginea pig	Rodentia
Nannospalax galili	Upper Galilee mountain blind mole rat	Rodentia
Ictidomys tridecemlineatus	Thirteen-lined ground squirrel	Rodentia
Heterocephalus glaber	Naked mole rat	Rodentia
Jaculus jaculus	Jerboa	Rodentia
Fukomys damarensis	Damaraland mole rat	Rodentia
Peromyscus maniculatus	Deer mouse	Rodentia
Homo sapiens	Human	Primates
Pan troglodytes	Chimpanzee	Primates
Pongo abelii	Orangutan	Primates
Gorilla gorilla	Gorilla	Primates
Nomascus leucogenys	Gibbon	Primates
Papio anubis	Baboon	Primates
Macaca mulatta	Rhesus macaque	Primates
Microcebus murinus	Mouse lemur	Primates
Callithrix jacchus	Marmoset	Primates
Otolemur garnetti	Northern greater galago	Primates
Chlorocebus sabaeus	Green monkey	Primates
Equus ferus caballus	Horse	Perisodactyla
Dasypus novemcinctus	Armadillo	Cingulata
Loxodonta africana	African elephant	Proboscidea
Oryctolagus cuniculus	Rabbit	Largomorpha
Myotis lucifugus	Microbat	Chiroptera
Ovis aries	Sheep	Cetartiodactyla
Sus scrofa	Pig	Cetartiodactyla
Bos taurus	Cow	Cetartiodactyla
Canis lupus familiaris	Dog	Carnivora
Ailuropoda melanoleuca	Panda	Carnivora
Felis catus	Cat	Carnivora
Mustela putorius furo	Ferret	Carnivora

and Nei 1987) implemented in MEGA 6 (Tamura et al. 2013) for estimating the phylogenetic relationships among the species (supplementary fig. S4, Supplementary Material online) as described in supplementary methods, Supplementary Material online. To confirm the phylogenetic relationship, we also used the first 50,000 amino acids for the maximum likelihood method (ML; supplementary fig. S5, Supplementary Material online). The two trees were very similar with very high bootstrap supports and mostly consistent with previous studies (Wildman et al. 2006; Blanga-Kanfi et al. 2009; Perelman et al. 2011; Babarinde and Saitou 2013; Hedges et al. 2015). The differences in the trees produced by the two methods were in the rodent basal group and the position of gibbons (Nomascus). Whereas the NJ tree placed squirrel as the basal group (supplementary fig. S4, Supplementary Material online), ML tree placed mouse-related group as the basal group (supplementary fig. S5, Supplementary Material online). This phylogenetic location, which might have been caused by rapid evolution in the rodent common ancestor (Blanga-Kanfi et al. 2009), has been under debate in rodent phylogeny (Montgelard et al. 2008; Blanga-Kanfi et al. 2009). The NJ tree clearly showed correct phylogenetic position of gibbons. In any case, we used the NJ method because it used more amino acid positions and many

internal branches in the NJ tree had higher bootstrap supports than in the ML tree (supplementary figs. S4 and S5, Supplementary Material online). For distance estimation, protein sequence alignments were converted into coding nucleotide sequence alignments. Using the aligned coding nucleotide sequences, we also computed separate phylogenetic trees using synonymous (supplementary fig. S6, Supplementary Material online) and nonsynonymous (supplementary fig. S7, Supplementary Material online) substitution rates. Then, the sequences were split into first, second, and third codon positions. Using guinea pig sequences as reference, all 4-fold degenerate sites (all4fold) were extracted. In addition, 4-fold degenerate sites with nucleotides at positions 1 and 2 conserved in all species (cons4fold) were also analyzed. Using the phylogenetic tree produced by the NJ method (Saitou and Nei 1987), the nucleotide distances were computed by the ML method using *baseml* package in PALM 4.8 (Yang 2007). The details are presented in supplementary methods, Supplementary Material online.

Divergence Time Estimation

The divergence times were estimated using Bayesian method of relaxed molecular clock implemented in MCMCTree of PAML 4.8 (Yang 2007). MCMCTree estimates species divergence times using fossil calibration by performing Bayesian estimation under various molecular clock models. The estimation of divergence times in MCMCTree involves three major steps, similar to what were used by Inoue et al. (2010). Three constraints were used (supplementary fig. S8, Supplementary Material online), including 6–8 Ma for human–chimpazee split (Brunet et al. 2002; Patterson et al. 2006; Jensen-Seaman and Hooper-Boyd 2013), 13–23 Ma for human–orangutan split (Jensen-Seaman and Hooper-Boyd 2013), and 45–65 Ma for cat–dog split (Steiper and Young 2006). The pipeline for the methods is presented in supplementary figure S9, Supplementary Material online.

Regression Analyses of the Factors

Linear regression analyses were run for each factor on evolutionary rate. For each of the individual factors, the R^2 values were high and the predictions were significant. When all the factors were combined together, although the adjusted R^2 was slightly higher, the significance of the prediction dropped, suggesting multicolinearity among the factors. R package ppcor (Kim 2015) was used to run partial correlations between evolutionary rates and each of the factors, such that every other factor was controlled. We ran the variance inflation factor (VIF) analysis to confirm multicolinearity. We then ran independent stepwise regression analyses for all the models involving all possible combinations of the factors. The model with the highest adjusted R^2 value was selected as the optimal model. To check the behavior of the model under the consideration of species-level phylogenetic relationships (Felsenstein 1985), we ran phylogenetic generalized least square analyses. First, tree correlation was computed from all4fold phylogenetic tree using the corBrownian function of the *ape* R package (Paradis and Schliep 2019). We then used the *gls* function in the *nlme* R package (Lindstrom and Bates 1990) for phylogenetic generalized least square computation.

Transcriptome Data Analyses

The short read sequences of 45 mammalian liver samples were retrieved from the short read archive (https://www.ncbi.nlm.nih.gov/sra; last accessed August 7, 2002). The species from which transcriptome data were analyzed include mouse, rat, naked mole rat, and guinea pig (Rodentia); rabbit (Lagomorpha); vervet, marmoset, rhesus macaque, and human (Primates); cow and pig (Cetartiodactyla); cat, dog, and ferret (Carnivora). For each species, the cDNA files were downloaded from Ensembl database (http://asia.ensembl.org/info/data/ftp/index.html; last accessed August 7, 2002). The downloaded cDNA files were used as references in corresponding species. RSEM (Li and Dewey 2011) was used to estimate abundance from alignments made by Bowtie (Langmead 2010). To merge the expression estimates for all the samples, the longest transcript for each gene was extracted in each species. Using mouse as the reference in BLAST (Altschul et al. 1997) searches, the reciprocal best hits between mouse and each of the species were extracted. Homologous genes in all species used were then mapped. Only genes that had homologs in all considered species were retained. The expression levels (in TPM) for each homologous gene in each sample were extracted to build the expression matrix for all samples. Genes without expression in any sample were discarded. Eventually, 9,457 genes were used. The expression levels were first log2-transformed before running principal component analyses (PCAs) with R Core Team (2018).

Results

Capybara Genome Sequencing and Assembly

The genome sequences of capybara were determined using Illumina MiSeq platform. Agillent QXT library preparation kit was used to prepare the 650-bp fragment size library. The prepared library was sequenced in three MiSeq runs. For the MiSeq sequencing, paired-end reads of 350 and 250 bp were produced. In total, ~42 Gb from 157.6 million reads of sequence data were produced. In each of the runs, an average of 75% of the total yield had the minimum quality value of Q30. The details of the reads are presented in supplementary table S1, Supplementary Material online. The reads have been deposited to DNA Data Bank of Japan Sequence Read Archive under the Bioproject Accession PRJDB7394.

For the assembly, CLC workbench and SOAP-denovo2 were used. The estimated genome size from k-mer (19-mer)

Table 2
Capybara Genome and Assembly Statistics

Parameter	Value
Estimated genome size (Gb)	2.62
Total positions (Gb)	2.48
Determined positions (Gb)	2.47
Percent gap	0.45
Percent GC content	40.2
Minimum scaffold (bp)	200
Longest scaffold (kb)	140.24
N20 length (bp)	18,744
N50 length (kb)	7,745
N80 length (kb)	2,193

distribution is 2.62 Gb. The assembled fragments cover 2.49 Gb, representing about 95% of the estimated genome size. However, 2.47 Gb (~94% of the estimated genome size) had a determined nucleotide. The longest fragment from the assembly was 140.2 kb with N50 length of ~8 kb (table 2). The average read coverage was 15× with the peak around 13× (supplementary fig. S10*A*, Supplementary Material online).

Extracting Reliable Genomic Regions

Read filtering was done by *read cleaner* that filters the reads to minimum Q20 and 50 bp lengths. After filtering to minimize sequencing error, about 24 Gb (57% of the initial nucleotides) were available. The quality of the filtered reads was examined using FastQC tool (supplementary fig. S11, Supplementary Material online). When the filtered reads were mapped back to the genome, the average depth was 10× with the peak around 8× (supplementary fig. S10*B*, Supplementary Material online). The maximum sequencing error rate per position was 10^{-6} (minimum quality and depth of Q20 and 3×, respectively). This value corresponds to one sequencing error in 1-Mb region. Setting stricter thresholds of Q30 and 50 bp kept only ~17 Gb (or 40% of the initial bases) but did not substantially improve the results. Although increasing the stringency further to Q30 and 100 bp theoretically minimizes sequencing error, the final results were poorer because the amount of mapped reads becomes much lower and the lower depth makes the call less reliable. For example, we extracted the transcripts from the Q20- and Q30-filtered reads and ran blast with guinea pig transcripts. The percent identity of the amino acid between guinea pig and Q20-filtered sequences is significantly higher (Mann–Whitney *U* test P value $< 2.2\ e{-}16$) than the percent identity between guinea pig and Q30-filtered sequences (supplementary fig. S12*A*, Supplementary Material online). Sequencing error has been reported to be nucleotide biased (Hansen et al. 2010, Schirmer et al. 2015). If there is sequencing error, huge differences would be expected in comparison with the closely related species. The nucleotide compositions of the third codon positions in capybara and guinea pig were not significantly different (supplementary fig. S12*B*, Supplementary Material online). Therefore, the adopted threshold values (Q20 and 3×) seem to be appropriate.

Rodent Evolutionary Rate Dynamics

Having established the integrity of the newly determined capybara genome, we proceeded to investigate rodent evolutionary rate dynamics. The closest rodent species to capybara with determined genome at the time of this analysis was guinea pig. Guinea pig and capybara diverged around 22 Ma (Hedges et al. 2015). The two species are very different in body weights. With adult capybara weighing up to 82 kg and adult guinea pig weighing up to 1 kg, there is more than 60-fold difference. If body size is related to evolutionary rate, one would expect that their evolutionary rates would be very different. We first took advantage of the phylogenetic relationships to estimate differences in evolutionary rates as employed by Li and Wu (1987). In this measure, outgroup with the same divergence time to the species being considered was chosen. As such, the difference between the evolutionary distances would reflect the difference in rate. Using rabbit as the outgroup of the rodent species, there was no substantial difference in various measures of neutral evolutionary distances investigated (supplementary table S2, Supplementary Material online). In fact, the evolutionary distance of 4-fold sites between rabbit and capybara was not statistically different from the value between rabbit and guinea pig (*t*-test P value = 0.65). The trends were similar when common degu, which is one of the closest species to both capybara and guinea pig used in this study, was used as an outgroup (supplementary table S3, Supplementary Material online). Using codon3, distances to guinea pig were slightly lower than distances to capybara (supplementary tables S2 and S3, Supplementary Material online). On the contrary, distances to capybara computed using cons4fold were slightly lower than distances to guinea pig. These reflect the longer capybara branches obtained in amino acid (supplementary fig. S5, Supplementary Material online) and nonsynonymous substitution (supplementary fig. S7, Supplementary Material online) trees. This is not surprising because some third codon substitutions are nonsynonymous. Anyway, the neutral evolutionary rate difference between capybara and guinea pig does not seem to be commensurate to the substantial body size difference between the two species.

The observation above raises an important question of whether all rodents have uniformly high evolutionary rates. To answer this question, publicly available genomes of rodent and other mammalian species (supplementary data, Supplementary Material online) were retrieved. Qualities of the data were examined by the proportion of undetermined amino acid or the number of predicted genes in the

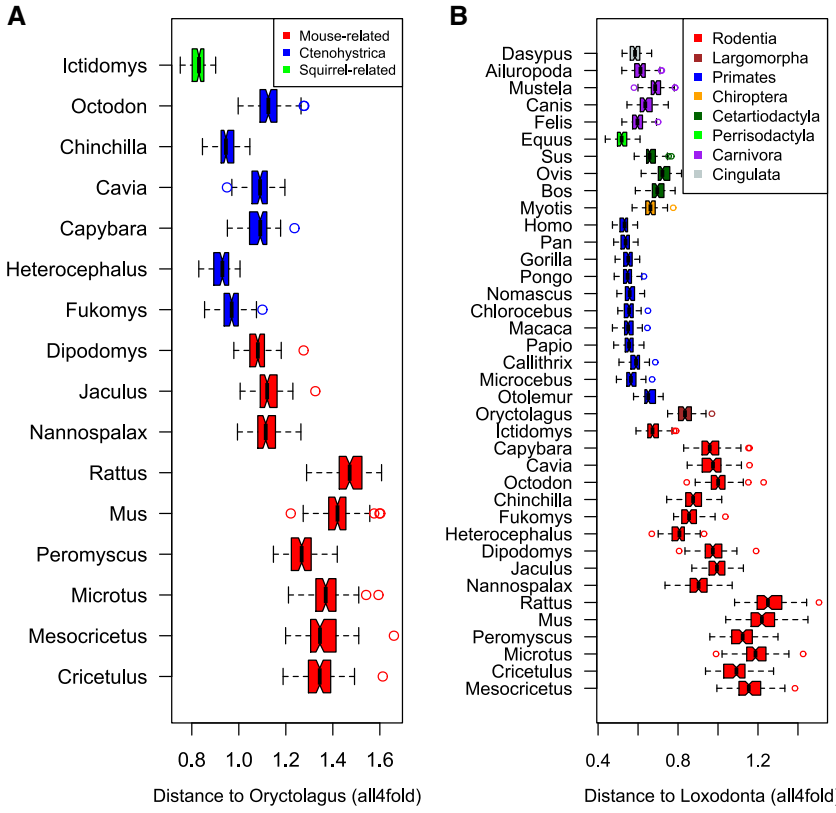

Fig. 1.—Heterogenous and higher evolutionary rates among rodents. (*A*) All4fold distances to Oryctolagus show heterogenity in rodent evolutionary rates. (*B*) All4fold distances to Loxodonta show that rodents have higher evolutionary rates compared with other examined orders.

annotation (supplementary fig. S2, Supplementary Material online). Also, eight species with unusually long external branch length were removed (supplementary fig. S3, Supplementary Material online). Sixteen rodent species were eventually used (table 1 and supplementary fig. S8, Supplementary Material online). Evolutionary distances were computed using ML method (supplementary fig. S9, Supplementary Material online). ANOVA result shows that all4fold distances of the examined rodent species to rabbit were not uniform ($F = 1{,}108.2$; P value $< 2.2 \times 10^{-16}$; supplementary data and supplementary table S4, Supplementary Material online). Figure 1*A* and supplementary figures S13*A* and S13*B*, Supplementary Material online, show the distances between rabbit and the analyzed rodent species, using all 4-fold sites (all4fold), all third codon (codon3) sites, and 4-fold degenerate sites with conserved first and second sites (cons4-fold), respectively. The trends of the results were similar. The evolutionary rates across the three rodent lineages were also not uniform ($F = 11.78$, P value $= 0.001205$, supplementary table S5, Supplementary Material online). Mouse-related lineage appeared to have higher evolutionary rate (fig. 1*A*). Although only one species from squirrel-related lineage was analyzed, the evolutionary rate in squirrel seemed to be the lowest. Even among the mouse-related clade, there seemed to be heterogeneity ($F = 474.71$, P value $< 2.2 \times 10^{-16}$, and supplementary table S6, Supplementary Material online). Although Ctenohystrica species were more uniform, species of mouse-related lineage were heterogeneous (supplementary fig. S14, Supplementary Material online). Descendants of Crecitidae-Muridae common ancestor seemed to have higher evolutionary rate, with the highest rate found in rat (Student's t-test P value $< 5 \times 10^{-06}$).

Mammalian Evolutionary Rate Heterogeneity

Having established rodent evolutionary rate dynamics, we incorporated more species from Boreoeutheria orders, including Primates, Carnivora, and Cetartiodactyla (table 1 and supplementary fig. S8, Supplementary Material online). African elephant (Loxodonta) and armadillo (Dasypus) were selected as outgroups. Initially, measures that take advantage of phylogenetic relationship without the requirement of the actual divergence times were employed and African elephant was used as outgroup. Evolutionary distances were estimated using first (codon1), second (codon2), third (codon3) codon positions, sites with 4-fold degeneracy in guinea pig (all4fold) and 4-fold degenerate sites with conserved first and second codon positions (cons4fold). In addition, we computed the synonymous (d*S*) and nonsynonymous (d*N*) substitution rates.

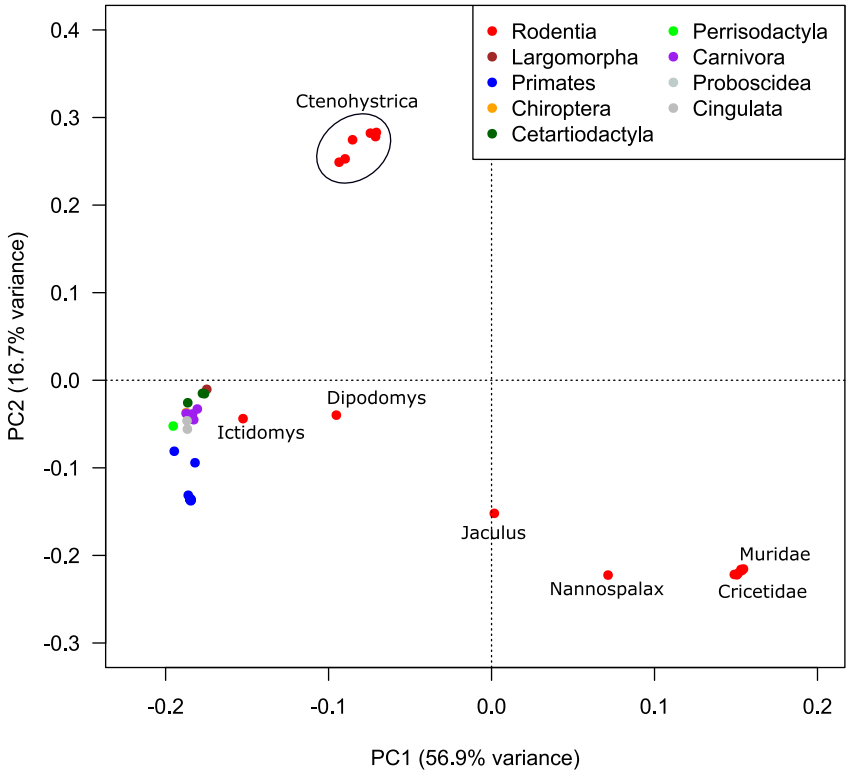

Fig. 2.—Extensive heterogeneity in rodent evolutionary rates compared with other examined mammals. All4fold pairwise distances were used for the PCAs. Mouse and rat are both in family Muridae which cludsters with family Cricetidae containing Chinese hamster.

As expected, all estimates were correlated (supplementary fig. S15, Supplementary Material online). Codon1, codon2, and d*N* are measures that include substitutions that mostly lead to amino acid changes and were highly correlated. However, codon3, all4fold, cons4fold, and d*S* that are mostly measures of neutral evolution are found to be highly correlated (supplementary fig. S15, Supplementary Material online). We focused our analyses on neutral evolutionary rates. To ensure that we compare the same sites in all species, we used all4fold for the subsequent analyses. All4fold evolutionary distances to elephant were different across orders ($F = 19.284$, P value $= 1.13 \times 10^{-9}$, supplementary table S7 and fig. S16, Supplementary Material online, and fig. 1B). In all the measures of neutral evolutionary distances, rodents have higher evolutionary rates than other orders considered (fig. 1*B* and supplementary figs. S17*A* and S17*B*, Supplementary Material online). Among all taxonomic groups studied, mouse-related rodent species, specifically rat and mouse, have the highest evolutionary rates. The lowest rates were found in Primates and Perrisodactyla.

PCAs (fig. 2) further highlight the heterogeneity among rodents. The first two principal components, which together explain about 77% of the total variance, failed to cluster all rodents together. In contrast, other considered orders were clustered by both principal components. Further, rodent heterogeneity was found to be pronounced in the mouse-related lineage. The heterogeneity in rodents was very different from other examined mammalian orders that were found to cluster together by both first and second principal components. In fact, the distance between Rattus and Octodon (1.295) is higher than the distance between African elephant and any other mammalian species considered (<1.26).

The above measure of evolutionary rate takes the average evolutionary distances over certain evolutionary time. When the divergence time is large, averaging the distances over time would mask certain differences. This is particularly important because the divergence time between African elephant and Boreoeutherian mammals is about 105 Ma (Hedges et al. 2015). Therefore, estimating the evolutionary rate per unit time for each branch would be more informative. We estimated the evolutionary rate per unit time for each branch by dividing the evolutionary distance by the corresponding time duration (see supplementary methods, Supplementary Material online). The evolutionary distances computed from all4fold positions were used. To calculate the evolutionary time duration for each branch, MCMCTree (Yang and Rannala 2006) was used to estimate the divergence times for every bifurcation. The time duration for each branch is simply the difference between the divergence times of the start and end bifurcations. For external branches, the time

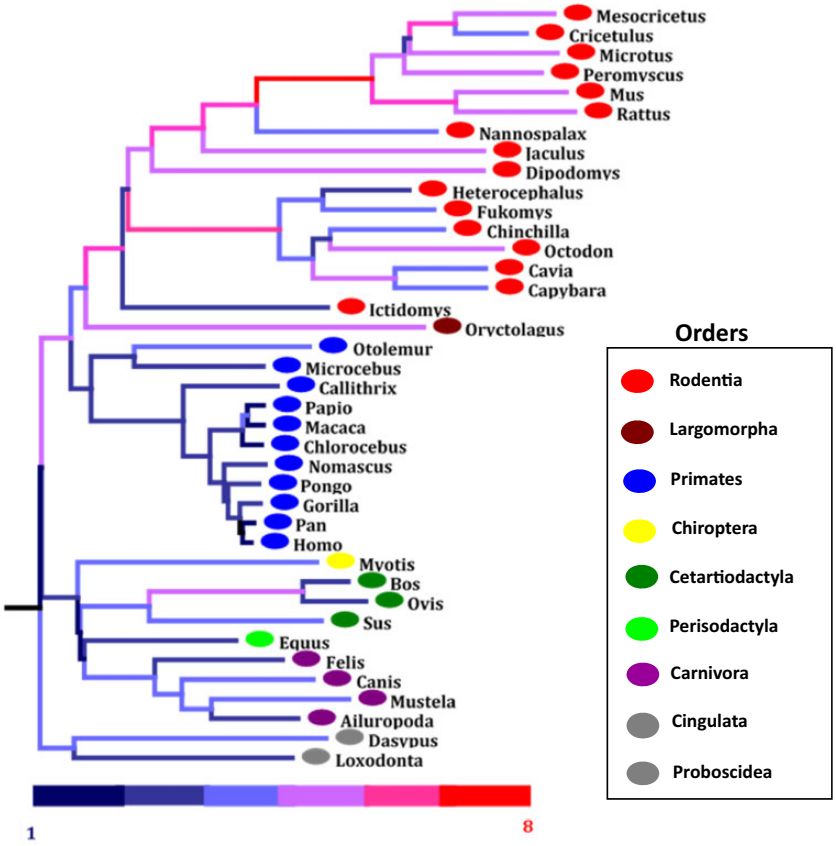

Fig. 3.—The phylogenetic tree showing the rate dynamics across phylogenetic timescale. The value on color of each branch represents the evolution rate ($\times 10^{-9}$/site/year) computed from supplementary figure S20, Supplementary Material online. The highest rate is found in the Muridae and Nannospalacidae common ancestor. The actual value on each branch is shown in supplementary figure S20, Supplementary Material online.

duration is simply the divergence time from the closest species. To get the best estimates, divergence times were estimated using first (codon1), second (codon2), third (codon3), and combination of first, second, and third (codon123) positions. In addition, divergence times were estimated from all 4-fold degenerate sites (all4fold) and 4-fold degenerate sites with conserved first and second sites (cons4fold). The divergence times of the tree of life (TOL, Hedges et al. 2015) were also retrieved. The divergence time estimates were highly correlated (Pearson correlation coefficient >0.96 and supplementary fig. S18, Supplementary Material online). We based the evolutionary rate estimates on codon2 time estimate which has the highest correlation coefficient to TOL estimate (0.99, supplementary fig. S18, Supplementary Material online).

As would be predicted from figure 1*B*, evolutionary rates were different across orders ($F = 7.3328$, P value $= 2.27 \times 10^{-5}$, supplementary table S8 and fig. S19, Supplementary Material online), with rodents having the highest rate (P value <0.001; supplementary fig. S19*A*, Supplementary Material online). Evaluation of species rate shows that the highest rate is found in rat, whereas the lowest rate is found in human (supplementary fig. S19*B*, Supplementary Material online). Figure 3 and supplementary figure S20, Supplementary Material online, are the phylogenetic trees showing the evolutionary rates for the species investigated. The rates were computed using all4fold (supplementary fig. S9, Supplementary Material online) distances and divergence times computed with codon2 (supplementary fig. S18, Supplementary Material online). It is interesting that although Rodentia and Primates are evolutionarily closer and are both Euarchontoglire, they have very divergent evolutionary rates. The evolutionary rates for Euarchontoglire common ancestor was 4.13×10^{-9}/site/year. This was followed by a slowdown of evolutionary rate in primate and glire lineages. In rodent common ancestor after the divergence of rabbit, the rate accelerated. The highest evolutionary rate (8.1×10^{-9}/site/year) was found in the Muridae-Cricetidae common ancestor after the divergence of Nannospalacidae. On the contrary, the lowest evolutionary rate (2.0×10^{-9}/site/year) among rodents was found in the Chinchilla-Octodon common ancestor. Among extant rodents considered, Heterocephalus had the lowest rate (2.0×10^{-9}/site/year), whereas the highest rate was found

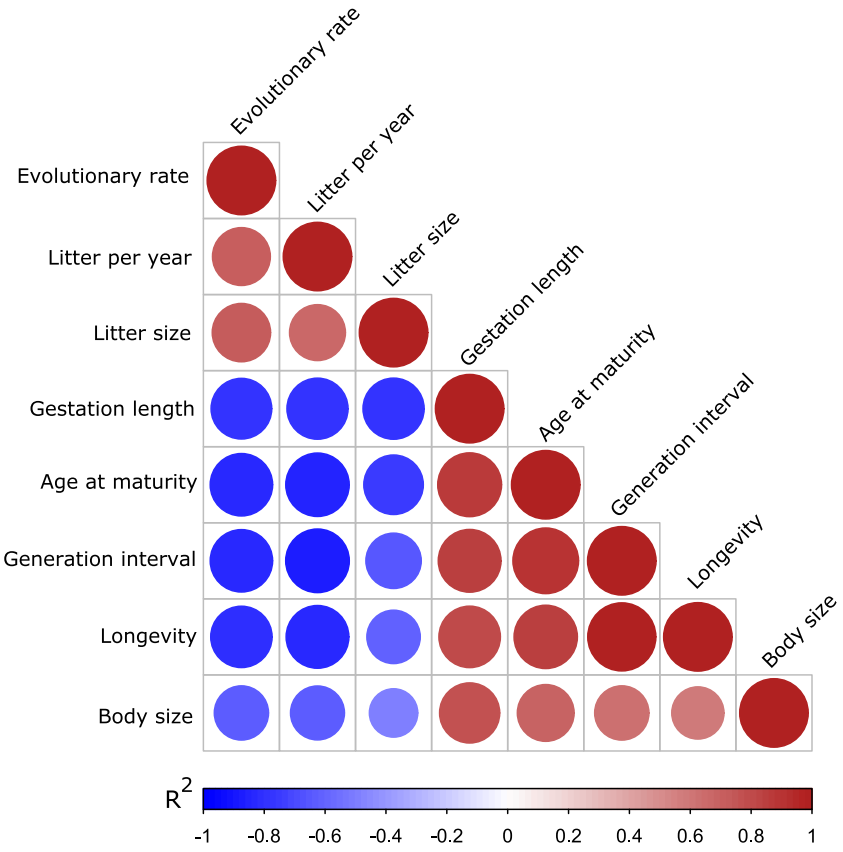

Fig. 4.—Correlations between evolutionary rates and investigated quantitative factors. Pearson's correlated coefficients are shown in heatmap. Circle sizes correspond to the absolute values of the correlation coefficients.

in Rattus (4.4×10^{-9}/site/year). Though higher in rodents, rate heterogeneity is not only limited to rodents.

Factors Associated with Evolutionary Rate

Having established that evolutionary rates are heterogeneous, it became necessary to evaluate factors affecting the rate heterogeneity. In addition to evolutionary rates, age at maturity, body weight, generation interval, gestation length, longevity, litter size, and number of litter per year were also different across orders (P value <0.001, supplementary fig. S21 and table S9, Supplementary Material online). Order and the seven factors considered were significantly related to evolutionary rates (fig. 4). Whereas litter size and litter number per year were found to be positively correlated and order was categorical, all other factors were found to be negatively correlated with evolutionary rates. However, the factors were also significantly correlated to one another (fig. 4). Linear regression for each factor is shown in supplementary figure S22, Supplementary Material online. The inclusion of all the eight factors investigated to be associated with the evolutionary rate in linear regression model gave the adjusted R^2 value of 0.7191 (supplementary table S10, Supplementary Material online). We ran partial correlations between the evolutionary rate and each of the factors considered controlling other factors. We found that, when controlled for other factors, none of the factors was significantly correlated to evolutionary rate (supplementary table S11, Supplementary Material online). The correlation between the factors and the insignificant partial correlations raise the possibility of multicolinearity. As expected, the VIF of some of the factors was very high (supplementary table S12, Supplementary Material online). To get the most predictive model, we ran the regression analyses for all possible combinations of models starting from models involving single factors to the model involving all factors. Our regression analyses involved stepwise regression analyses with focus on predictive power rather than statistical significance. Regression analyses involving all possible combinations of the factors showed that the highest adjusted R^2 value (0.7411) was found in a model consisting order, generation interval and litter size (supplementary table S13, Supplementary Material online). The addition of more factors did not improve the adjusted R^2 value (supplementary fig. S23, Supplementary Material online). Our analyses did not consider phylogenetic relatedness among species (Felsenstein 1985) because such would mask the effect of taxonomic orders on the

evolutionary rate. Interestingly, phylogenetic generalized least square revealed that the inclusion of phylogenetic information in the most predictive model did not significantly affect the result (supplementary table S14, Supplementary Material online). The obvious and expected difference is the general decrease in statistical power leading to the loss of statistical significance of taxonomic order when species-level phylogenetic information is included in the final model (supplementary tables S13 and S14, Supplementary Material online).

Impacts of Rodent Evolutionary Rate Dynamics on Gene Expression Data

We previously analyzed conserved noncoding sequences and found higher dynamics in rodents (Babarinde and Saitou 2013) suggesting higher expression dynamics (Babarinde and Saitou 2016). We therefore checked if the heterogeneous evolutionary rates in rodents have impacts on gene expression patterns. To investigate the impact of higher rodent evolutionary rate on gene expression dynamics, we obtained and uniformly analyzed 45 short reads RNA samples of 14 species from five orders (see Materials and Methods). The numbers of samples for each species are listed in supplementary table S15, Supplementary Material online. Liver sample was chosen because of the availability of RNA short read samples in a large number of species (supplementary table S15, Supplementary Material online). RSEM (Li and Dewey 2011) was used for abundance estimation employing alignment made by Bowtie (Langmead 2010). We then conducted pricipal component analysis on the 45 liver RNA samples using the expression levels (TPM) estimated from RSEM (supplementary data, Supplementary Material online). PC1 that represents 75.9% of the total variants failed to classify the samples (supplementary fig. S24, Supplementary Material online). However, PC2 reveals rodents heterogeneity (fig. 5 and supplementary fig. S24, Supplementary Material online) in the manner that resembles the PCA of evolutionary rates in figure 2. Specifically, PC2 separates mouse and rat (Muridae) from other species. By contrast, guinea pig and naked mole rat (Ctenohystrica) were clustered with other mammlian samples. Interestingly, PC3 clusters all investigated primate samples, demonstrating lower heterogeneity among primates. It is important to note that samples from the same species, even though ocassionally retrieved from different studies, tend to cluster together. This suggests that the patterns observed in figure 5 and supplementary figure S24, Supplementary Material online, are less likely to be artifacts. Therefore, the evolutionary rate dynamics impact the gene expression dynamics in mammals.

Discussion

The draft genome sequences of capybara were determined using next generation sequencing technology. A new pipeline for extracting genomic regions from de novo genome sequencing project is also reported. This pipeline makes use of reliable read sequences and mapping for extracting highly reliable genomic regions. The pipeline takes care of the error that may arise due to sequencing or assembling. To minimize alignment error, mapping quality was set to 30. In fact, too stringent conditions, involving fewer reads would increase alignment error. The comparison of genome statistics showed that the structures and compositions of capybara and guinea pig genomes were very similar. The newly determined genome sequences of capybara and 38 other mammalian species were analyzed to investigate evolutionary rate dynamics.

The quality of the genomes is very crucial for the analyses. For the reliability of the analyses, certain steps were taken in this study. First, a pipeline was developed to minimize the error from capybara de novo genome sequences. The pipeline ensured that the maximum sequencing error rate was 10^{-6}. Several steps were taken to ensure the quality of the publicly available genome sequences analyzed. Species having low-quality genome sequences were not included. Although the procedure did not completely eliminate errors, the magnitude of neutral evolutionary rates estimated from substitution rates, relative to the sequencing and mapping errors, is so large that the impact of the errors in the analyses is minimal. To further investigate the impact of errors, 4-fold degenerate third codon positions with conserved first and second positions were investigated. The requirement for conservation at the first and second codon positions ensures minimal error. Although all4fold rates were higher than the cons4fold rate, the correlation was nearly perfect (supplementary fig. S18, Supplementary Material online). Furthermore, when cons4-fold were used, the patterns of the results remained essentially the same. Therefore, even with genomes with much better qualities, the results are very likely to be similar. The sequencing technology is improving and the cost is rapidly falling. Indeed, another effort to determine capybara genome sequences with the accession number PRJNA399400 has been publicly announced (https://www.ncbi.nlm.nih.gov/assembly/GCA_004027455.1/; last accessed August 7, 2002). Therefore, the impact of the genome qualities on the analyses and further studies of the evolutionary rate dynamics on the morphological effects can be further investigated.

As reported before (Li and Wu 1987; Li et al. 1996; Hiller et al. 2012; Takahashi and Saitou 2012; Babarinde and Saitou 2013), molecular clock does not hold among the examined species. We investigated the evolutionary rates among rodent lineages and found a high rate of heterogeneity, with mouse-related lineage having the highest evolutionary rate. When other mammalian lineages were incorporated, rodents were indeed found to have higher evolutionary rates (fig. 1*B*). On the other hand, primates were found to have the lowest evolutionary rates. Also, rodents were found to be more heterogeneous than other investigated mammalian orders (fig. 2). Investigation of evolutionary rates across the phylogenetic

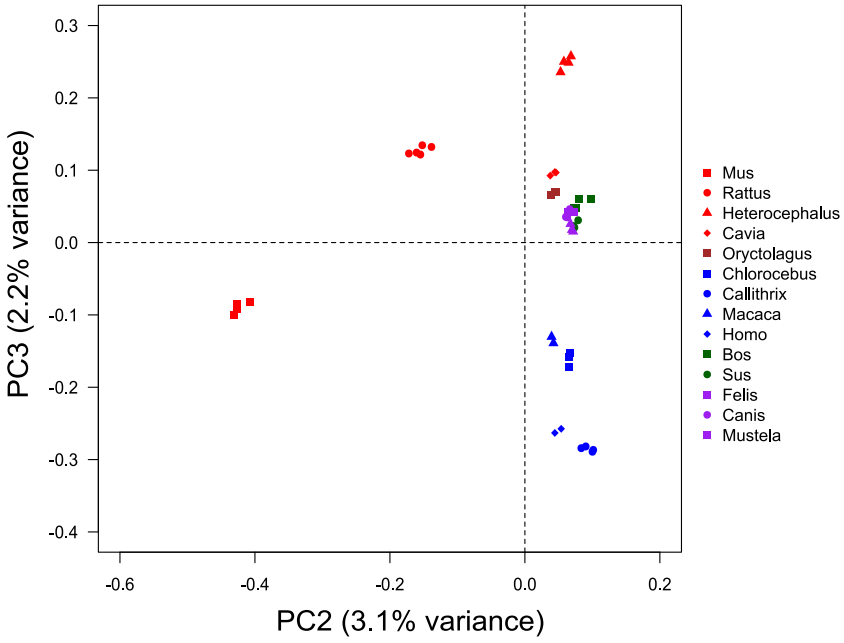

Fig. 5.—Transcriptome analyses reveal higher expression dynamics among rodent species. The expression values of 9,457 genes with homologs among the 14 species from five orders were computed in 45 liver samples. The complete list of the samples is presented in supplementary table S15, Supplementary Material online.

timescale revealed that evolutionary rates were very dynamic (fig. 4). Evolutionary rate slowdown in hominoid lineage reported by Goodman (1985) and Yi (2013) was confirmed. In addition, acceleration was found in the rodent common ancestor and the tempo reached the highest (8.0814×10^{-9}/site/year) at the common ancestor of Muridae and Crecitidae after the divergence of Spalacidae. The accelerations of evolutionary rates at two different branches (fig. 3) are captured in PCAs (fig. 2 and supplementary fig. S14, Supplementary Material online). PC1 separates the descendants of the highest evolutionary rates (8.1×10^{-9}/site/year found in Cricetidae–Muridae common ancestor) from others, whereas PC2 separates the descendants of the second highest evolutionary rate (5.5×10^{-9}/site/year found in Ctenohystrica common ancestor) from others (figs. 2 and 3 and supplementary fig. S20, Supplementary Material online). Interestingly, these two highest rates are found in rodents (fig. 3 and supplementary fig. S20, Supplementary Material online). The lowest rate (9.56×10^{-10}) was found in the human lineage. The heterogeneity of evolutionary rates across orders and species raises an important question about the likely factor that might contribute to differences in evolutionary rates.

Eight factors, including order, age at maturity, body weight, generation interval, gestation length, longevity, litter size, and number of litter per year were investigated. All the factors were found to be significantly associated with evolutionary rates (P value <0.001). However, multiple regression analyses showed that the model consisting of order, generation interval, and litter size had the highest predictive power. Including other factors reduced the predictive and statistical power of the model. Whereas generation interval has been long reported to be associated with evolutionary rates (Martin and Palumbi 1993; Ohta 1993; Bromham et al. 1996), this study reports the contribution of litter size to evolutionary rates. The significant correlations between the factors and the evolution rate and insignificant partial correlations (supplementary table S11, Supplementary Material online) might be related to the multicollinearity among the factors (fig. 4) as demonstrated by the variance inflation factors (supplementary table S12, Supplementary Material online). Therefore, these factors could not be definitely described as having direct influence on evolutionary rates. This explains why the capybara evolutionary rate was not substantially different from that of guinea pig despite capybara's huge body size.

There is a possibility that the factors investigated might have changed recently. If the changes are very recent, they might not have had a significant effect on the evolutionary rates. For example, although capybara and guinea pig diverged around 22 Ma (Hedges et al. 2015), the time when the body size of capybara evolved is not known. Assuming the body size enlargement was <1 Ma, the evolutionary effects might not be so obvious. To support this assumption, Cambrian explosion is believed to have happened within a short time (McMenamin 2013) and artificial selections have led to multiple fold size increase in domestic animals. However, it is unlikely that all the species included in this study had very recent body size change. In fact, the computation of Evans et al. (2012) suggested that it took 0.3 and 1.6 million

generations, respectively, for 10- and 100-fold increase in terrestrial mammal body mass. Using this estimate, such a huge difference in the body sizes of capybara and guinea pig is not likely to be very recent. Therefore, the behavior of the body size in regression analyses is not likely because of recent body size evolution. Even if it was relatively recent, the effect is not so strong. Nonetheless, further studies would be needed to clarify this issue.

The results showed that generation interval, order, and litter size are the direct factors associated with evolutionary rates. Order is a categorical variable that is related to phylogenetic history. In fact, species with shorter divergence times tend to have similar evolutionary rates and other factors consider (Spearman's $\rho = 0.1242133$, P value $= 0.001318$). Indeed, the phylogenetic generalized least square analyses showed that when species-level relatedness is considered, the statistical significance of order is lost. The two continuous variables in the model with the highest predictive power are factors that reflect number of effective meiotic events. Effective meiotic events are meiotic events contributing to the generation and population size. The effects of meiotic events that do not contribute to the generation or population are similar to those of mitosis. Whereas generation interval reflects the frequency of meiosis per unit time, litter size reflects the extent of meiosis as each individually is potentially a product of a meiotic event. Generally speaking, species with higher effective meiotic events arising from shorter generation interval and/or higher litter size tend to have higher evolutionary rates. Therefore, errors due to replication at effective meiosis contribute significantly to evolutionary rate dynamics.

The impact of the evolutionary rate dynamics on gene expression patterns is an important biological question. We have previously found that rodents have higher conserved noncoding sequence turnover rates (Takahashi and Saitou 2012; Babarinde and Saitou 2013). Although some studies have reported that regulatory elements might not necessarily be conserved (Schmidt et al. 2010; Shen et al. 2012), we showed that conserved noncoding sequences are often associated with conserved expression patterns although in a tissue-specific manner (Babarinde and Saitou 2016). Therefore, it is important to investigate the importance of evolutionary rate dynamics on expression patterns. Using liver tissues in 14 species (see supplementary table S15, Supplementary Material online), we actually found that the heterogeneity in rodent liver expression patterns was higher than in primates. Specifically, mouse and rat (Muridae) were separated from other species by PC2. This suggests that the higher evolutionary rates in Muridae indeed affect the expression patterns, at least in liver. Further, the transcriptome analyses further capture rodent heterogeneity with slowly evolving Ctenohysterica species having expression patterns similar to those of other mammalian species.

Finally, our analyses have focused on sites that are largely neutrally evolving. The correlations between codon3, all4fold, and cons4fold are very high (supplementary fig. S15, Supplementary Material online), suggesting that any of the measures would produce similar results. As measures of neutral evolution, the estimates are highly correlated with synonymous substitution rates. On the other hand, codon1 and codon2, which are mostly associated with amino acid changes, are highly correlated with nonsynonymous substitution rates (supplementary fig. S15, Supplementary Material online). A careful look at the trees showing neutral rates of evolution (supplementary figs. S6 and S20, Supplementary Material online) and those reflecting amino acid changes (supplementary figs. S4, S5, and S7, Supplementary Material online) show that the branch lengths are different. Specifically, the differences in branch lengths of phylogenetically related species tend to be higher in amino acid changing mutations. The differences between the estimates of neutral evolution and those that affect amino acids might be related to the strength of selection. As the synonymous substitution rates (supplementary fig. S6, Supplementary Material online) are independent of selection and tend to be folds higher than nonsynonymous substitution rates (supplementary fig. S7, Supplementary Material online), the general evolutionary rates would be robustly captured by measures of neutral evolution.

Data Availability

Some intermediate files and codes have been deposited to https://github.com/iababarinde/Mammalian-evolutionary-rates.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The study was supported by the grant-in-aid from MEXT, Japan, given to S.N. The authors appreciate Izu Shaboten Park for providing capybara sample, Professor Ituro Inoue and Dr Yoshikazu Hozomichi for their supports in library preparation and Miseq sequencing, and Dr Kirill Kryukov for proofreading. The efforts of the handling editor and anonymous reviewers in improving the quality of the manuscript are also appreciated. The supercomputer facilities of National Institute of Genetics and MEXT Cell Innovation Project were used for computations.

Literature Cited

Akinwole MT, Babarinde IA. 2019. Assessing tissue lysis with sodium dodecyl sulphate for DNA extraction from frozen animal tissue. J Forensic Res. 10:446.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389–3402.
Alvaro M, Juhani O. 1986. *Hydrochoerus hydrochaeris*. Brisson, 1762. Mammalian Species. 264:1–7.
Babarinde IA, Saitou N. 2013. Heterogeneous tempo and mode of conserved noncoding sequence evolution among four mammalian orders. Genome Biol Evol. 5(12):2330–2343.
Babarinde IA, Saitou N. 2016. Genomic locations of conserved noncoding sequences and their proximal protein-coding genes in mammalian expression dynamics. Mol Biol Evol. 33 (7):1807–1817.
Blanga-Kanfi S, et al. 2009. Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. BMC Evol Biol. 9(1):71.
Bromham L. 2011. The genome as a life-history character: why rate of molecular evolution varies between mammal species. Philos Trans R Soc B. 366(1577):2503–2513.
Bromham L, Penny D. 2003. The modern molecular clock. Nat Rev Genet. 4(3):216–224.
Bromham L, Rambaut A, Harvey PH. 1996. Determinants of rate variation in mammalian DNA sequence evolution. J Mol Evol. 43(6):610–621.
Brunet M, et al. 2002. A new hominid from the Upper Miocene of Chad, Central Africa. Nature 418(6894):145–151.
Clavel J, Morlon H. 2017. Accelerated body size evolution during cold climatic periods in the Cenozoic. Proc Natl Acad Sci U S A. 114 (16):4183–4188.
DeRose-Wilson LJ, Gaut BS. 2007. Transcription-related mutations and GC content drive variation in nucleotide substitution rates across the genomes of *Arabidopsis thaliana* and *Arabidopsis lyrata*. BMC Evol Biol. 7(1):66.
Dos Reis M, et al. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr Biol. 25(22):2939–2950.
Duret L, Semon M, Piganeau G, Mouchiroud D, Galtier N. 2002. Vanishing GC-rich isochores in mammalian genomes. Genetics 162(4):1837–1847.
Evans AR, et al. 2012. The maximum rate of mammal evolution. Proc Natl Acad Sci U S A. 109(11):4187–4190.
Felsenstein J. 1985. Phylogenies and the comparative method. Am Nat. 125(1):1–15.
Gardner JL, Peters A, Kearney MR, Joseph L, Heinsohn R. 2011. Declining body size: a third universal response to warming? Trends Ecol Evol. 26(6):285–291.
Goodman M. 1985. Rates of molecular evolution: the hominoid slowdown. BioEssays 3(1):9–14.
Hansen KD, Brenner SE, Dudoit S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res. 38(12):e131.
Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 22(2):160–174.
Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. Mol Biol Evol. 32 (4):835–845.
Hettiarachchi N, Saitou N. 2016. GC content heterogeneity transition of conserved noncoding sequences occurred at the emergence of vertebrates. Genome Biol Evol. 8(11):3377–3392.
Hiller M, Schaar BT, Bejerano G. 2012. Hundreds of conserved noncoding genomic regions are independently lost in mammals. Nucleic Acids Res. 40(22):11463–11476.
Huchon D, et al. 2002. Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. Mol Biol Evol. 19(7):1053–1065.
Huttley GA, Wakefield MJ, Easteal S. 2007. Rates of genome evolution and branching order from whole genome analysis. Mol Biol Evol. 24(8):1722–1730.
Inoue JG, et al. 2010. Evolutionary origin and phylogeny of the modern holocephalans (chondrichthyes: chimaeriformes): a mitogenomic perspective. Mol Biol Evol. 27(11):2576–2586.
Jensen-Seaman MI, Hooper-Boyd KA. 2013. Molecular clocks: determining the age of the human–chimpanzee divergence. Chichester (United Kingdom): John Wiley & Sons.
Kim EB, et al. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature 479(7372):223–227.
Kim S. 2015. ppcor: an R Package for a fast calculation to semi-partial correlation coefficients. Csam 22(6):665–674.
Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.
Kinjo S, et al. 2018. Maser: one-stop platform for NGS big data from analysis to visualization. Database 2018:bay027. doi:10.1093/database/bay027.
Kumar S, Subramanian S. 2002. Mutation rates in mammalian genomes. Proc Natl Acad Sci U S A. 99(2):803–808.
Langmead B. 2010. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics. 11(7). doi:10.1002/0471250953.bi1107s32.
Larkin MA, et al. 2007. ClustalW and ClustalX version 2.0. Bioinformatics 23(21):2947–2948.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12(1):323.
Li H, et al. 2009. The sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25(14):1754–1760.
Li R, et al. 2010. De novo assembly of human genomes with massively parallel short References 187 read sequencing. Genome Res. 20(2):265–272.
Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol Phylogenet Evol. 5(1):182–187.
Li WH, Wu CI. 1987. Rates of nucleotide substitution are evidently higher in rodents than in man. Mol Biol Evol. 4(1):74–82.
Lindstrom MJ, Bates DM. 1990. Nonlinear mixed effects models for repeated measures data. Biometrics 46(3):673–687.
Martin AP, Palumbi S. 1993. Body size, metabolic rate, generation time and the molecular clock. Proc Natl Acad Sci U S A. 90(9):4087–4091.
McMenamin MAS. 2013. The Cambrian explosion: the construction of animal biodiversity. BioScience 63(10):834–835.
Montgelard C, Forty E, Arnal V, Matthee CA. 2008. Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. BMC Evol Biol. 8:321.
Musser GG, Carleton MD. 1993. Mammal species of the world. In: Wilson DE, Reeder DM, editors. Family Muridae. Washington (DC): Smithsonian Institution Press.
Ohta T. 1993. An examination of the generation-time effect on molecular evolution. Proc Natl Acad Sci U S A. 90(22):10676–10680.
Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35(3):526–528.
Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature 441(7097):1103–1108.
Perelman P, et al. 2011. A molecular phylogeny of living primates. PLoS Genet. 7(3):e1001342.
Peterson KJ, et al. 2004. Estimating metazoan divergence times with a molecular clock. Proc Natl Acad Sci U S A. 101(17):6536–6541.
R Core Team. 2018. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/. Accessed August 7, 2002.

Saber MM, Babarinde IA, Hettiarachchi N, Saitou N. 2016. Emergence and evolution of Hominidae-specific coding and noncoding genomic sequences. Genome Biol Evol. 8(7):2076–2092.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4(4):406–425.
Schirmer M, et al. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res. 43(6):e37.
Schmidt D, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding dominic. Science 328(5981):1036–1040.
Shen Y, et al. 2012. A map of the cis-regulatory sequences in the mouse genome. Nature 488(7409):116–120.
Speakman JR. 2005. Body size, energy metabolism and lifespan. J Exp Biol. 208(9):1717–1730.
Steiper ME, Young NM. 2006. Primate molecular divergence dates. Mol Phylogenet Evol. 41(2):384–394.
Takahashi M, Saitou N. 2012. Identification and characterization of lineage-specific highly conserved noncoding sequences in mammalian genomes. Genome Biol Evol. 4(5):641–657.
Takezaki N. 2018. Global rate variation in bony vertebrates. Genome Biol Evol. 10(7):1803–1815.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30(12):2725–2729.
Thomas JA, Welch JJ, Lanfear R, Bromham L. 2010. A generation time effect on the rate of molecular evolution in invertebrates. Mol Biol Evol. 27(5):1173–1180.
Vawter AT, Rosenblatt R, Gorman GC. 1980. Genetic divergence among fishes of the eastern pacific and the Caribbean: support for the molecular clock. Evolution 34(4):705–711.
Wildman DE, et al. 2006. Evolution of the mammalian placenta revealed by phylogenetic analysis. Proc Natl Acad Sci U S A. 103(9):3203–3208.
Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21(9):1859–1875.
Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.
Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol. 23(1):212–226.
Yi SV. 2013. Morris Goodman's hominoid rate slowdown: the importance of being neutral. Mol Phylogenet Evol. 66(2):569–574.
Zhu T, Dos Reis M, Yang Z. 2015. Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. Syst Biol. 64(2):267–280.
Zuckerkandl E, Pauling L. 1965. Evolving genes and proteins. In: Bryson V, Vogel HJ, editors. Evolutionary divergence and convergence in proteins. New York: Academic Press. p. 97–166.

Associate editor: Wen-Hsiung Li