**Supplementary methods**

**Library preparation and sequencing**

The library was prepared with Agilent SureSelect QXT library preparation kit. The starting material was 30ng genomic DNA sample reconstituted in 1ul of ultrapure water. The genomic DNA sample was then fragmented and adaptor-tagged. The adaptor-tagged library was purified before PCR amplification. The amplified library was then purified with AMPure XP beads after which library DNA quantity and quality were assessed using Qubit and Agilent 2100 Bioanalyzer and DNA 1000 Assay. The average fragment size was 750bp. Two library sets were prepared to minimize random effects due to fragmentation. After the library integrity had been ascertained, the two library sets were pooled by combining 2.5ul of 4nM of each library set. Three runs of sequencing were done for the pooled library set on Miseq sequencing platform using 600 cycle kits. The set sequencing length for read1 was 350bp while the length of 250bp was set for read2.

**Sequencing output and quality assessment**

After combining the output of the three sequencing runs, a total of ~158 million reads were produced. The quality of the sequencing reads were assessed using FastQC. Because the aim of the study is to evaluate evolutionary rate, sequencing error would have a serious impact on the results. The reads were filtered and/or trimmed. To maximize the yield while ensuring the strict quality threshold, a python script, read\_cleaner.py, was written. With the python script, all base positions in any kept read would not be lower than the threshold Phred quality score and length. The python script therefore filters and/or trims the reads as the situation demands. The command for the read cleaner is as follows;

*read\_cleaner.py -i Capybara\_raw.1.fastq -j Capybara\_raw.2.fastq -1 Capybara\_filtered.1.fastq -2 /*

*Capybara\_filtered.1.fastq -u Capybara\_filtered\_unpaired.1.fastq –v / Capybara\_filtered\_unpaired.2.fastq -q 20 -l 50*

***De novo* assembly of capybara draft genome**

CLC workbench and SOAP-denovo2 were used for the assembly of the reads. In order to have abundant data for better contig formation, I used the unfiltered reads for initial de novo assembly. First, the contig formation was done using CLC workbench available in Cell Innovation project (<https://cell-innovation.nig.ac.jp>) reported by Kinjo et al. (2018). Thereafter, scaffolds were made and gaps were closed using SOAP-denovo2 (Li et al. 2010). For contig formation, scaffold making and gap closing, default settings of the respective software were used.

**Extraction of high-quality genomic regions**

Because of the limited amount of sequencing output and the purpose of this study, it was important to minimize the effect of the sequencing error. At very high coverage, such concerns would not be too serious. However, in this study, we attempted to minimize the error rate to the bare minimum. First, the sequencing depth of the raw sequence outputs was analyzed (Fig. S1A). BWA-MEM 0.7.5a (Li and Durbin 2009) was used to align the reads to the assembled draft genome using the default settings. SAMtools 0.1.19 (Li et al. 2009) was used to extract the depth. To extract high-quality regions, filtered reads (minimum Phred quality score =20; minimum length = 50bp) were mapped to the assembled draft genome. The depth per position when high-quality reads were used was computed (Fig. S1B). The average depth was about ~10×. While low-depth regions might have some erroneous positions, too high depth may also have low integrity because of repetitive sequences. Therefore, we decided to use regions between 3-30×depths (between a third of the average and three times the average). Minimum Phred quality of 20 and depth of 3× imply that the sequencing error rate would be at most 10-2×3 (i.e. 10-6) per site. This value corresponds to one erroneously determined site per megabase pair region. Nucleotide on each position was examined using SAMtools (Li et al. 2009) mpileup with –uf flag. We thereafter converted the files to vcf files to assess the variant sites. If the high-quality mapped reads have a variant nucleotide, the variant nucleotide is called for the position. In the case of heterozygous sites, the allele with the highest frequency was picked. In addition, we understand that mapping may be another source of errors and required that the mapping quality must be at least 30 for each position. Positions that do not meet these criteria were masked to “N” and treated as undetermined in the downstream analyses. The command for the script is as follows;

*depth\_masker.py -v Capybara\_20Q\_50bp\_d3D30.vcf -o Capybara\_20Q\_50bp\_d3D30.fa -d 3 /*

*–D 30 -Q 30*

**Gene extraction**

Because of the sequencing depth and coverage, we reasoned that de novo gene prediction might be difficult. I therefore relied on the homology-based gene extraction. At the time of this analysis, the closest species with annotated genome sequences to capybara was guinea pig. The availability of guinea pig gene annotation made the gene extraction feasible. GMAP (Wu and Watanabe 2005) was used to detect capybara genes using guinea pig coding sequences as the query. Because capybara and guinea pigs are not the same species, ‘cross-species’ option was used. First, the genome reference was built with 8-mer and 1bp step size. Protein and exon sequences were then predicted from the genome reference using guinea pig transcripts. The GMAP commands are as follow;

*gmap\_build -D . -d Capybara\_20Q\_50bp\_d3D30.fa -k 8 -q 1*

*gmap -n 1 -t 100 --cross-species -D . -d Capybara\_20Q\_50bp\_d3D30 transcript.fa -Q > / Capybara\_20Q\_50bp\_d3D30\_prot.fa*

*gmap -n 1 -t 100 --cross-species -D . -d Capybara\_20Q\_50bp\_d3D30 transcript.fa –E genomic > / Capybara\_20Q\_50bp\_d3D30\_exons.fa*

The extracted exons were joined to form a single coding sequence per homologous guinea pig gene.

**Acquisition of gene sequences**

The genome, coding and protein sequences of 59 mammalian species (see Supplementary Data) were downloaded from either Ensembl or UCSC databases. As important as the number of species included is, the quality of the genome sequences is also very important. Species with greater than 5% undetermined sequences in the assembled genome were discarded. For species with no gene annotation, the transcripts of the closest annotated species were used for gene prediction using GMAP. Also, species with fewer than 14,000 annotated genes were excluded. Finally, the homologous sequences of the available species were aligned. The species with unusually longer branch were excluded. At the end, 39 species including 37 boreoeutherian and two outgroup species were included (Supplementary Data).

**Extraction of homologous gene sets**

The protein-coding genes of the selected species were retrieved from Ensembl database. We then extracted the longest transcript per gene. Reciprocal BLASTP (Altschul et al. 1997) searches were conducted using guinea pig and every other species (Fig. S2). For the blast searches, the threshold e-value was set to 0.00001. For each gene in each in each species pair, we extracted the reciprocal best hit using the bit scores of the alignments. Any alignment with the bit score of less than 100 was discarded. Thus, we retrieved the homologous gene pairs between guinea pig and every other species. For each guinea pig gene with identifiable homologs in all other species, we extracted the amino acid sequences. Each gene cluster was written into different file.

**Multiple sequence alignment**

The extracted multiple sequences for each gene cluster were aligned using CLUSTALW2 (Larkin et al. 2007). To make the codon alignment, we retrieved the coding sequences for all the genes from Ensembl database. The amino acid sequence alignments were then converted to coding sequence alignments using the retrieved coding sequences. The aligned coding sequences of all gene clusters were concatenated to produce single multiple alignment sequences. From the multiple alignment sequences, the first, second and third codon positions were extracted independently. Any position with gap in any of the species used was discarded.

**Phylogenetic and distance computation**

The phylogenetic relationship was first established using the gapless amino acid alignment. The Neighbor-Joining (NJ) tree (Saitou and Nei 1987) was computed using MEGA6 (Tamura et al. 2013). For the initial NJ tree, the phylogeny was tested with bootstrap method. The nucleotide substitution model used was maximum composite likelihood. Gamma distribution (k =5) was used for rates among sites. Also, the patterns among lineages were set to be heterogeneous. Almost all phylogenetic relationships were supported with 100% bootstrap value. Using the phylogenetic relationship from the first NJ tree, we computed the nucleotide substitution using General Time Reversible (GTR) model implemented in BASEML of PAML 4.8 (Yang 2007).

To investigate the precision of the distances, we split the concatenated sequences into 100 fragments. The splitting was done in a step-wise manner such that there is not overrepresentation of any fragments in a region. For example, the first fragment contains the 1st, 101st, 201st, 301st ... sites of the original alignment. Likewise, the second fragment contains the 2nd, 102nd, 202nd, 302nd... sites of the original alignment. This procedure checks the precisions of the distances and reduces overestimation due to regional variations.

**Estimation of divergence times using MCMCtree**

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. We used the steps similar to what were used by Inoue et al. (Inoue 2010). First, overall substitution rate was estimated assuming molecular clock. Next, the gradient and Hessian are estimated using the estimated substitution rate. Finally, the actual MCMC analyses was carried out using the gradient and Hessian as inputs.

For the estimation of the overall substitution rate, BASEML of PAML 4.8 was used with the concatenated multiple alignment file of second codon positions as the input. GTR model was used and strict molecular clock was assumed. After the estimation of the overall substitution rate, gradient and Hessian matrix was produced by running MCMCtree on the original alignment and tree files. For gradient and Hessian estimation mode, usedata=3 was used. GTR model of nucleotide substitution was used. In addition, relaxed clock model (clock=2) was used for independent rate. Finally, the divergence times were calculated by running MCMCtree. This time, the gradient and Hessian matrix file from the previous run was used by setting usedata=2. The prior substitution rate (rgene\_gamma) was set to 1 15.8 based on the overall substitution rate. The sigma2-gamma was set to 1 1.05 based on the reported divergence time of 105 MYA for elephant and human (Hedges 2015). Relaxed clock model (clock=2) was used. Root age was restricted to less than 110 MYA.

Divergence times were also computed with first codon positions, third codon positions, all codons (codon 1, 2 and 3 positions), all four-fold degenerate sites (all4fold) and four-fold degenerate sites with conserved first and second positions (cons4fold). In addition, the divergence times for all the species were extracted from tree of life (Hedges 2015).

Estimation of overall substitution rate;

seqfile = mammal\_aligned\_codon2.fa

treefile = raw\_mammals\_number

outfile = mammal\_codon2\_mlb \* main result file

noisy = 2 \* 0,1,2,3: how much rubbish on the screen

verbose = 0 \* 1: detailed output, 0: concise output

runmode = 0 \* 0: user tree; 1: semi-automatic; 2: automatic

\* 3: StepwiseAddition; (4,5):PerturbationNNI

model = 7 \* 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85

\* 5:T92, 6:TN93, 7:REV, 8:UNREST, 9:REVu; 10:UNRESTu

Mgene = 0 \* 0:rates, 1:separate; 2:diff pi, 3:diff kapa, 4:all diff

\* ndata = 100

clock = 1 \* 0:no clock, 1:clock; 2:local clock; 3:CombinedAnalysis

fix\_kappa = 0 \* 0: estimate kappa; 1: fix kappa at value below

kappa = 5 \* initial or fixed kappa

fix\_alpha = 0 \* 0: estimate alpha; 1: fix alpha at value below

alpha = 0.5 \* initial or fixed alpha, 0:infinity (constant rate)

Malpha = 0 \* 1: different alpha's for genes, 0: one alpha

ncatG = 5 \* # of categories in the dG, AdG, or nparK models of rates

nparK = 0 \* rate-class models. 1:rK, 2:rK&fK, 3:rK&MK(1/K), 4:rK&MK

nhomo = 0 \* 0 & 1: homogeneous, 2: kappa for branches, 3: N1, 4: N2

getSE = 1 \* 0: don't want them, 1: want S.E.s of estimates

RateAncestor = 0 \* (0,1,2): rates (alpha>0) or ancestral states

Small\_Diff = 7e-6

cleandata = 1 \* remove sites with ambiguity data (1:yes, 0:no)?

\* icode = 0 \* (with RateAncestor=1. try "GC" in data,model=4,Mgene=4)

\* fix\_blength = -1 \* 0: ignore, -1: random, 1: initial, 2: fixed

method = 0 \* Optimization method 0: simultaneous; 1: one branch a time

Gradient and Hessian estimation;

seed = -1

seqfile = mammal\_aligned\_codon2.fa

treefile = constrained\_mammal\_tree.nwk

outfile = out\_usedata3

ndata = 1

seqtype = 0 \* 0: nucleotides; 1:codons; 2:AAs

usedata = 3 \* 0: no data; 1:seq like; 2:use in.BV; 3: out.BV

clock = 2 \* 1: global clock; 2: independent rates; 3: correlated rates

RootAge = '<1.1' \* safe constraint on root age, used if no fossil for root.

model = 7 \* 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85

alpha = 0.5 \* alpha for gamma rates at sites

ncatG = 5 \* No. categories in discrete gamma

cleandata = 0 \* remove sites with ambiguity data (1:yes, 0:no)?

BDparas = 1 1 0 \* birth, death, sampling

kappa\_gamma = 6 2 \* gamma prior for kappa

alpha\_gamma = 1 1 \* gamma prior for alpha

rgene\_gamma = 1 29 \* gamma prior for overall rates for genes

sigma2\_gamma = 1 1.1 \* gamma prior for sigma^2 (for clock=2 or 3)

finetune = 1: .05 0.1 0.12 0.1 .3 \* auto (0 or 1) : times, rates, mixing, paras, RateParas, FossilErr

print = 1

burnin = 50000

sampfreq = 50

nsample = 20000

Divergence time estimate;

seed = -1

seqfile = mammal\_aligned\_codon2.fa

treefile = constrained\_mammal\_tree.nwk

outfile = out\_usedata2

ndata = 1

seqtype = 0 \* 0: nucleotides; 1:codons; 2:AAs

usedata = 2 \* 0: no data; 1:seq like; 2:use in.BV; 3: out.BV

clock = 2 \* 1: global clock; 2: independent rates; 3: correlated rates

RootAge = '<1.1' \* safe constraint on root age, used if no fossil for root.

model = 7 \* 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85

alpha = 0.5 \* alpha for gamma rates at sites

ncatG = 5 \* No. categories in discrete gamma

cleandata = 0 \* remove sites with ambiguity data (1:yes, 0:no)?

BDparas = 1 1 0 \* birth, death, sampling

kappa\_gamma = 6 2 \* gamma prior for kappa

alpha\_gamma = 1 1 \* gamma prior for alpha

rgene\_gamma = 1 29 \* gamma prior for overall rates for genes

sigma2\_gamma = 1 1.1 \* gamma prior for sigma^2 (for clock=2 or 3)

finetune = 1: .05 0.1 0.12 0.1 .3 \* auto (0 or 1) : times, rates, mixing, paras, RateParas, FossilErr

print = 1

burnin = 50000

sampfreq = 50

nsample = 20000