

Out of Southern East Asia of the Brown Rat Revealed by Large-Scale Genome Sequencing

Lin Zeng,^{†,1,2} Chen Ming,^{†,3,4} Yan Li,^{1,5} Ling-Yan Su,^{2,6} Yan-Hua Su,⁷ Newton O. Otecko,^{1,2,8} Ambroise Dalecky,^{9,10} Stephen Donnellan,¹¹ Ken Aplin,¹² Xiao-Hui Liu,¹³ Ying Song,¹³ Zhi-Bin Zhang,¹⁴ Ali Esmailzadeh,¹⁵ Saeed S. Sohrabi,¹⁵ Hojjat Asadollahpour Nanaei,¹⁵ He-Qun Liu,^{1,2} Ming-Shan Wang,^{1,2} Solimane Ag Atteynine,^{16,17} Gérard Rocamora,¹⁸ Fabrice Brescia,¹⁹ Serge Morand,²⁰ David M. Irwin,^{1,21} Ming-Sheng Peng,^{1,2,8} Yong-Gang Yao,^{2,6} Hai-Peng Li,^{*,3} Dong-Dong Wu,^{*,1,2,8} and Ya-Ping Zhang^{*,1,2,5}

¹State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

²Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China

³CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

⁴University of Chinese Academy of Sciences, Beijing, China

⁵State Key Laboratory for Conservation and Utilization of Bio-Resource in Yunnan, Yunnan University, Kunming, China

⁶Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, China

⁷College of Animal Science and Technology, Yunnan Agricultural University, Kunming, China

⁸Sino-Africa Joint Research Center, Chinese Academy of Sciences, Kunming, China

⁹Institut de Recherche pour le Développement (Ird), CBGP (UMR INRA/IRD/Cirad/Montpellier SupAgro), Montpellier sur Lez cedex, France

¹⁰Institut de Recherche pour le Développement (Ird), LPED (UMR AMU/IRD), Marseille, France

¹¹University of Adelaide and the South Australian Museum, Adelaide, Australia

¹²Division of Mammals, National Museum of Natural History, Smithsonian Institution, Washington, DC

¹³State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

¹⁴State Key Laboratory of Integrated Management on Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

¹⁵Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran

¹⁶Institut de Recherche pour le Développement (Ird), IMBE (UMR AMU/CNRS/IRD/UAPV), Bamako, Mali

¹⁷Faculté des Sciences et Techniques (FST), Université des Sciences, des Techniques et des Technologies de Bamako (USTTB), Bamako, Mali

¹⁸Island Biodiversity & Conservation Center, University of Seychelles, Mahé, Seychelles

¹⁹Diversité Biologique et Fonctionnelle des Ecosystèmes, Institut Agronomique néo-Calédonien, Port Laguerre, Paita, New Caledonia

²⁰CNRS-CIRAD, Centre d'Infectiologie Christophe Mérieux du Laos, Vientiane, Lao PDR

²¹Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mails: lihaipeng@picb.ac.cn; wudongdong@mail.kiz.ac.cn; zhangyp@mail.kiz.ac.cn.

Associate editor: Nicolas Vidal

Abstract

The geographic origin and migration of the brown rat (*Rattus norvegicus*) remain subjects of considerable debate. In this study, we sequenced whole genomes of 110 wild brown rats with a diverse world-wide representation. We reveal that brown rats migrated out of southern East Asia, rather than northern Asia as formerly suggested, into the Middle East and then to Europe and Africa, thousands of years ago. Comparison of genomes from different geographical populations reveals that many genes involved in the immune system experienced positive selection in the wild brown rat.

Key words: *Rattus norvegicus*, origin, demographic history, natural selection.

Introduction

A detailed understanding of the geographic origin of wild rodents and their subsequent dispersal routes across the globe has important implications in clarifying the spread of diseases and human migration (Matisoo-Smith and Robins 2004; Lin et al. 2012). For example, investigation on mtDNA phylogenies of the Pacific rat (*Rattus exulans*) recapitulated origins and dispersal of the Pacific people (Matisoo-Smith and Robins 2004). Study of mtDNA indicated that the house mice (*Mus musculus domesticus*) were a valuable proxy for Viking movements (Searle et al. 2009). Phylogeographical analysis on black rats (*R. rattus*) in the western Indian Ocean demonstrated that the history of black rats was correlated with human colonization history (Tollenaere et al. 2010).

The brown rat (*R. norvegicus*), one of the most common commensal rats, draws substantial public health interest, acting as a reservoir for a number of zoonotic pathogens such as Hantavirus, and disseminating many diseases (Meerburg et al. 2009; Kosoy et al. 2015). It is well-recognized that the brown rat spread out of Asia to Europe (Silver 1941; Southern 1964; Freye and Thenius 1968; Amori and Cristaldi 1999; Kosoy et al. 2015). This conclusion is supported by historical records (Suckow et al. 2006) and genetic evidences from mitochondrial and nuclear markers (Song et al. 2014; Puckett et al. 2016). However, until now, a detailed geographic origin and the dispersal routes of brown rats from Asia to Europe have remained subjects of extensive speculation. Based on historical records, it is opined that brown rats originated in Northeast China and Southeast Siberia (Wilson and Reeder 2005; Ness et al. 2012; Kosoy et al. 2015), and then dispersed westward through the Eurasian steppes into Europe (Barnett 2002; Gibbs et al. 2004). However, the earliest fossils of the species were found in today's southern China (Wu and Wang 2012), indicating a potential southern origin of brown rats.

Likewise, dating the arrival of brown rats into Europe/Africa remains elusive. Some historical records indicate that brown rats appeared in Europe during the Medieval Ages and became widespread during the Industrial Revolution (Amori and Cristaldi 1999). Suckow et al. (2006) suggested arrival dates for the rats in Ireland, England, France, Germany, and Spain to be 1722, 1730, 1735, 1750, and 1800AD, respectively. However, additional evidence suggests that brown rats might have been present in Europe as early as 1553AD (Freye and Thenius 1968), and got introduced into North America by 1750s (Armitage 1993).

In this study, we sequenced whole genomes of 110 wild brown rats drawn from diverse geographic locations and reveal that brown rats migrated out of southern East Asia, rather than northern Asia as formerly suggested, into the Middle East and then to Europe and Africa. During the migration, several adaptations for immune protection were developed.

Results and Discussion

Genome-Scale Sequencing of Wild Brown Rats Drawn from across the World

In the present study, to systematically explore the geographic origin and dispersal routes of the brown rat, a total of 117

Rattus norvegicus samples with a global representation and 1 black rat (*R. rattus*) were collected for genome sequencing. The species status of these *R. norvegicus* was initially determined by morphology and further confirmed based on cytochrome *b* (*cytb*) sequences by Sanger sequencing. Additional exploratory data analysis based on whole genomes found that the species status of seven individuals was ambiguous, as they displayed a closer relationship with the black rat, which might be attributable to genetic introgression between different species (Teng et al. 2017). These seven individuals, along with the black rat were treated as outgroup. Finally, whole genome sequences of 110 brown rats from southern East Asia (including Southeast Asia and southern China, $n = 31$), northern East Asia (including northern China and Russia, $n = 20$), Middle East ($n = 12$), Europe ($n = 26$), and Africa ($n = 21$), as well as the eight outgroup rats were generated to investigate the geographic origin of the brown rat (supplementary fig. S1 and table S1, Supplementary Material online). From the whole genome sequences, we annotated and filtered SNPs using Genome Analysis Toolkit (GATK) (McKenna et al. 2010). Following a stringent quality control, a total of 24,977,888 autosomal SNPs were obtained for the subsequent population genetic analyses.

Population Structure and Genetic Relationships among Rat Populations

To identify population structure and the genetic relationship of different rat populations, we first performed a series of classical analyses including phylogenetic assessments (fig. 1A and supplementary figs. S2–S5, Supplementary Material online), principal components analysis (PCA) (fig. 1B), and Bayesian clustering analysis by ADMIXTURE (fig. 1A and supplementary fig. S6, Supplementary Material online), using autosomal SNPs. The ADMIXTURE analysis suggested that the brown rat could be separated into East Asia and non-East Asia populations when the number of presumed ancestral population (K) is 2. East Asia individuals could further be grouped into southern East Asia and northern East Asia populations when $K = 4$ (fig. 1A). A good correspondence was found with results from the reconstructed phylogenetic trees (fig. 1A) and PCA (fig. 1B), when the East Asian rats were categorized into southern East Asia and northern East Asia populations. The grouping also fits well with the geographical distribution of samples (fig. 1C). However, when we defined more subpopulations, for example, where southern East Asian rats are further grouped into southern China and Southeast Asia, or northern East Asian rats are subdivided into northern China and northern Asia, the three classical methods gave inconsistent assignments. Therefore, we grouped East Asian individuals into two populations, that is, southern East Asia and northern East Asia.

We then assessed whether rats from Europe/Africa/Middle East were closer to rats from southern East Asia or northern East Asia. From the phylogenetic trees (fig. 1A and supplementary figs. S2–S5, Supplementary Material online), European/African/Middle East rats clearly exhibit a closer relationship with rats from southern East Asia compared with those from northern East Asia. Additionally, the PCA and

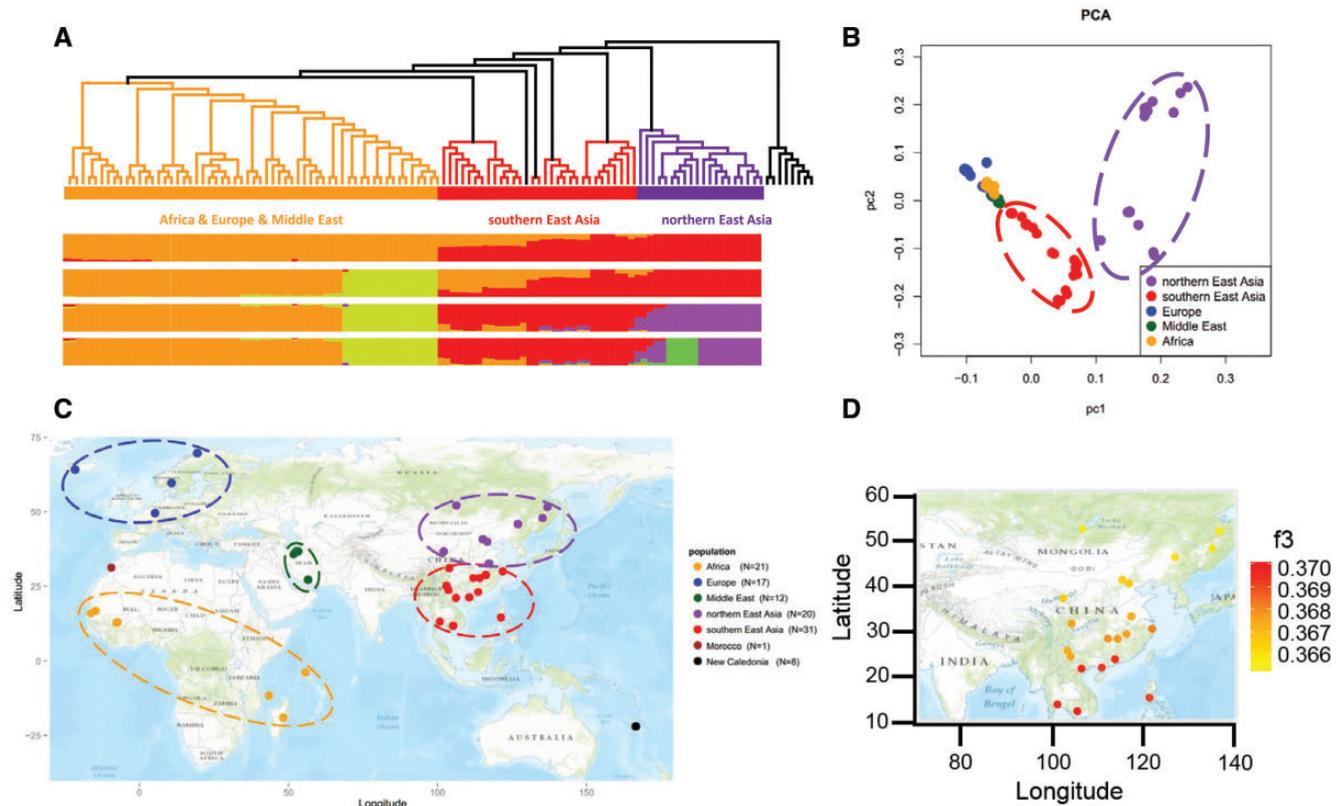


Fig. 1. Out of southern East Asia origin of wild brown rats. (A) Phylogenetic neighbor-joining tree and Bayesian clustering analysis by ADMIXTURE. Black colored lines represent Outgroup. (B) PCA analysis. From the PCA plot, two dispersal waves out of southern East Asia are presented. One wave is to non-East Asia (Middle Eastern/European/African populations), and another is to northern East Asia. The PCA pattern supports the demographic modeling result. (C) Geographic locations of 110 wild brown rat samples. (D) f_3 -outgroup statistics showing genetic proximity of Europe population to East Asian individuals.

Bayesian clustering analysis (Alexander et al. 2009) showed synonymous findings to the phylogenetic analyses, indicating a closer ancestral background between the Europe/Africa/Middle East rats and the southern East Asia rats, rather than with northern East Asia population (fig. 1A and B and supplementary fig. S6, Supplementary Material online). We additionally calculated an “outgroup f_3 -statistic” (Patterson et al. 2012; Raghavan et al. 2014) and the result was consistent with above results (fig. 1D and supplementary fig. S7, Supplementary Material online). Following (Vonholdt et al. 2010), we assessed the proportion of haplotypes of each non-East Asian population, that is, rats from Middle East, Europe, and Africa, shared with southern East Asia and northern East Asian rats in 20-kb nonoverlapping windows. This analysis revealed that haplotype sharing was consistently higher between non-East Asia and southern East Asia rats than with those from northern East Asia (supplementary fig. S8, Supplementary Material online).

Southern East Asia Origin of Wild Brown Rats

In contrast the previous hypothesis that the wild brown rat dispersed from northern Asia to Europe (Barnett 2002; Gibbs et al. 2004), all the above findings support the alternative hypothesis that wild brown rats might have dispersed from southern East Asia to Europe/Africa/Middle East. However, these classical analyses might be confounded by many factors

which may weaken their support for population history inference. For example, PCA can be influenced by technical sources of variation and complex demographic histories where interpretation of the directions of highest variability may be counterintuitive. Mismatches between projections onto PC space and geographical distribution of individuals under models of isolation by distance are common (Schraiber and Akey 2015). In addition, if substantial variations have occurred in the demographic history of different populations, particularly involving admixture among geographical regions, reconstructed relationships deduced by the above methods may not directly reflect the true history (Vonholdt et al. 2010). It is necessary to simulate and detect other hypotheses representing different demographic histories.

Therefore, to test the scenario of southern East Asia dispersal of wild brown rats into Europe/Africa/Middle East, we compared it with the alternative out of northern East Asia demographic models (supplementary fig. S9, Supplementary Material online). In the alternative models, wild brown rats migrated out of northern East Asia through different routes to colonize non-Asian regions, in line with the previous hypothesis of an out of northern Asia dispersal into Europe (Barnett 2002; Gibbs et al. 2004). These models were evaluated by a well-established maximum likelihood (ML) method based on joint site frequency spectrum (SFS) (Li and Stephan 2006; Excoffier et al. 2013). Based on the Akaike information

criterion (Akaike 1974), these alternative models displayed a poor fit compared with the out of southern East Asia model (supplementary table S2, Supplementary Material online), corroborating that Europe/Africa/Middle East rats originated from southern East Asia.

After clarifying that Europe/Africa/Middle East rats originated from southern East Asia, we further explored plausible relationship between northern East Asia and southern East Asia. In the phylogenetic trees (fig. 1A), the northern East Asian population appears closer to the outgroup than to the southern East Asia population. It could be intuitively plausible that brown rats migrated from northern East Asia to southern East Asia. However, conditional on the out of southern East Asia, we simulated the phylogenetic tree and regenerated the pattern in which the northern East Asia population is closer to the root (supplementary fig. S10, Supplementary Material online). As an alert signal, unlike the observed case in human populations (Cann et al. 1987), a closer distance to the root cannot lead the intuitive conclusion of northern East Asia origin. Therefore, we further constructed and detected plausible demographic scenarios between the two hypothetical ancestral populations (i.e., northern East Asia and southern East Asia) by the program *fastsimcoal2* (supplementary fig. S11, Supplementary Material online). We compared two mutually exclusive founder-effect dispersal models (supplementary fig. S11A and B, Supplementary Material online) and four different wide-spread East Asia models (supplementary fig. S11C–F, Supplementary Material online). Consistent with the model in supplementary figure 9A, Supplementary Material online, the origin from southern East Asia to northern East Asia was the most likely scenario (supplementary fig. S11A and table S3, Supplementary Material online). Furthermore, we found that the southern East Asia population has the largest number of private variants (supplementary fig. S12, Supplementary Material online), the highest genetic diversity among all populations (supplementary table S4, Supplementary Material online), and exhibits the fastest decay of linkage disequilibrium (supplementary fig. S13, Supplementary Material online). All these findings point toward the likelihood of southern East Asia as the cradle of wild brown rats. In this case, the rats migrated from southern East Asia to northern East Asia, which is consistent with previous fossil records (Wu and Wang 2012) and a previous study on mitochondrial DNA (Song et al. 2014).

Dating Two Migration Waves Out of Southern East Asia

We then dated the dispersal history of wild brown rats out of southern East Asia, using an Ancestral-to-Derived Hierarchical demographic search strategy to reduce the number of models (supplementary materials and methods, supplementary figs. S14–S18 and tables S5–S10, Supplementary Material online). By assuming that the newly established derived population does not affect the demography of the ancestral population, we dramatically reduced the number of models that we had to evaluate (27 vs. 4,375). Our analysis estimated that wild brown rats migrated from southern East Asia to northern East

Asia \sim 173,700 years ago (95% CI: 146,000–750,000, fig. 2), whereas wild brown rats spread from southern East Asia to the Middle East \sim 3,100 years ago (95% CI: 3,000–4,800), to Africa \sim 2,000 (95% CI: 1,900–3,400) years ago, and to Europe \sim 1,800 (95% CI: 1,700–2,900) years ago (fig. 2 and supplementary fig. S15F and table S11, Supplementary Material online), assuming two generations per year (Deinum et al. 2015) and a divergence time of 22.6 Ma between rat and mouse (Hedges et al. 2006, 2015; Kumar and Hedges 2011). We also performed robustness analyses by assuming three generations per year for rat (Anderson and Jones 1967), 15 or 30 My divergence time between rat and mouse, and different population assignment criteria (supplementary materials and methods, Supplementary Material online). The re-estimated introduction times generally agreed with the results obtained above (supplementary fig. S19 and tables S12 and S13, Supplementary Material online).

The estimated introduction times of brown rats to non-East Asia are much older than historical reports which propose migrations in the 18th century (Freye and Thenius 1968). However, it is undisputed that maritime trade has been in existence in the Indian Ocean and southern East Asia region for over 4,000 years (Forbes 1995; Miksic 2013). These early human activities could have facilitated the migration and dispersal of brown rat from southern East Asia to other regions. Such kind of human assisted migration has widely been proposed for rodents. For example, the pacific rat (*R. exulans*) and the black rat (*R. rattus*) migrated out of southern East Asia to remote areas of Oceania and Madagascar, respectively, $>$ 3,000 years ago (Matisoo-Smith and Robins 2004; Tollenaere et al. 2010), supporting the association between human adventures and the migration of rats. In clarifying the migration of brown rats, our study is definitely providing further highlights on the importance of rats as proxy for human events. Integration of multiple rodent species like mice, black rats, and brown rats, will be informative in clarifying the spread of rodent-related diseases and human migration.

It is important to point out that one caveat of our analyses is attributable to the relatively low coverage of genome sequencing which would affect the SFS that the demographic inferences were based on (Nielsen et al. 2011, 2012; Han et al. 2015), although we have used high-quality genotypes for analyses. In addition, sample size and distribution also have effect on genomic inference (Fumagalli 2013; Vieira et al. 2013). High-coverage genome sequences and large sample size will undoubtedly offer more precision and advance knowledge on the demographic history of brown rats.

Rapid Evolution of Immune Response Genes in Wild Brown Rat Populations

Wild rats have a puzzling ability to host several pathogens which can be transmitted to humans, often resulting in devastating diseases (Meerburg et al. 2009; Kosoy et al. 2015). An “arms-race” that drives rapid evolution of the immune system in a host (Van Valen 1973) might have endowed rats with this potential. We retrieved genes displaying significantly higher level of population differentiation (top 1%) between

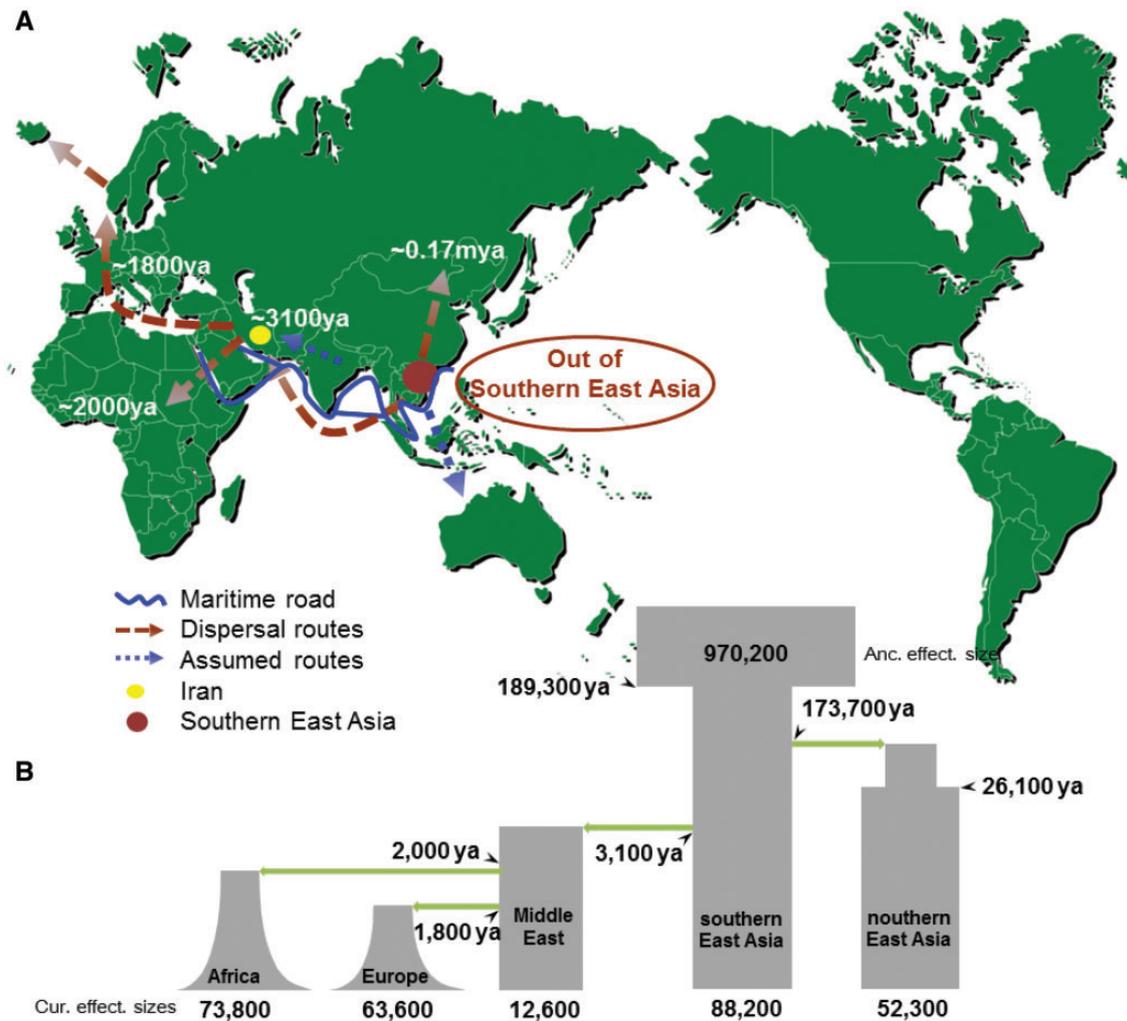


Fig. 2. Dispersal routes and demographic histories of wild brown rats. (A) Proposed dispersal routes of the wild brown rats based on our analyses. (B) The inferred joint demographic model of different wild brown rat populations based on the maximum likelihood method.

European and Chinese wild brown rats to investigate whether genes involved in the immune system might have evolved rapidly under positive selection in wild brown rats during their dispersal. Gene enrichment analysis revealed a rapid evolution of genes in the immune system through overrepresentation of many immunological categories such as: “leukocyte mediated immunity,” “response to bacterium,” and “leukocyte mediated cytotoxicity” (fig. 3A and supplementary table S14, Supplementary Material online). A comparison between African and Chinese rat genomes yielded a similar finding (supplementary table S15, Supplementary Material online). In particular, the gene *Mgat5* exhibited the highest level of population differentiation between Chinese and European rats (fig. 3B). *Mgat5* participates in the synthesis of galectins, cell-surface ligands involved in T-cell proliferation. *Mgat5*-knock-out mice display an autoimmune phenotype, and loss of *Mgat5* lowers the threshold for T-cells activation (Demetriou et al. 2001). We did not observe any nonsynonymous SNP exhibiting high level of population differentiation in *Mgat5*, suggesting that expressional changes might drive the adaptive evolution of this gene. The window with the second highest level of

differentiation contained a single gene, *Lyst* (fig. 3B). Mutations in *Lyst* cause Chediak–Higashi Syndrome in human, a genetic immunodeficiency disease characterized by defective T-cell and natural killer cell cytotoxicity (Trantow et al. 2010). Six nonsynonymous SNPs in *Lyst* exhibited high level of population differentiation ($F_{ST} > 0.6$). Further assessments will help clarify their functional consequences. The differentiation of immune genes suggests that wild rats from different regions of the world might differ in their susceptibility to specific pathogens, a hypothesis that needs experimental validation.

In conclusion, this study provides evidence for a southern East Asia origin of the brown rat and two historical migration waves out of its cradle. Wild brown rats dispersed to the Middle East, Europe, and Africa thousands of years ago. Along with the migration, many genes involved in immune response have adaptively evolved under natural selection in wild brown rats.

Materials and Methods

DNA Samples for Genome Sequencing

All animal handling required for this study was carried out in accordance with the animal experimentation guidelines and

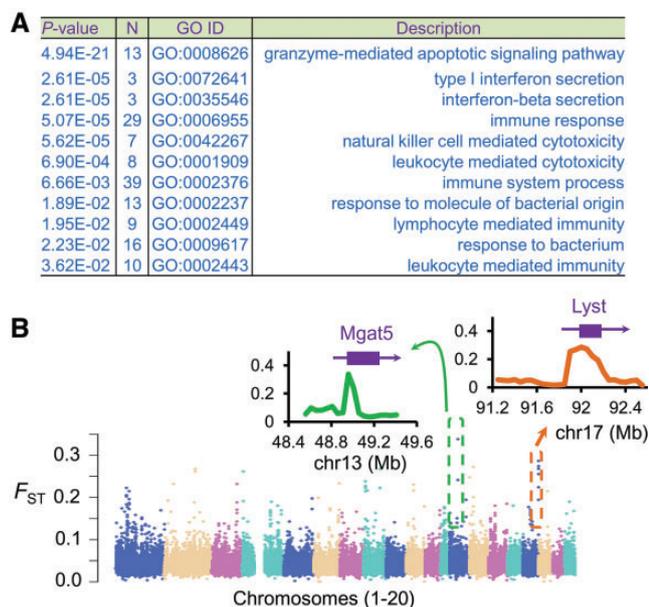


Fig. 3. Immune response genes are under selection in the wild brown rats. (A) List of genes with significantly high level of population differentiation between European and Chinese rats are enriched for immune system related functions. Gene Ontology analysis of the protein-coding genes was conducted using an online annotation tool g:Profiler and *P* values corrected by Benjamini–Hochberg FDR (Reimand et al. 2011). (B) Genomic landscape of the population differentiation (F_{ST}) between European and Chinese rats. Top two clusters with high level of population differentiation across genes *Mgat5* and *Lyst* are presented.

regulations of the Kunming Institute of Zoology. This research was approved by the Institutional Animal Care and Use Committee of the Kunming Institute of Zoology.

A total of 117 putative *R. norvegicus* samples from Russia, China, Southeast Asia, Europe, Africa, and Middle East (supplementary fig. S1 and table S1, Supplementary Material online) were collected for genome sequencing, along with 1 black rat (*R. rattus*). The species status of putative *R. norvegicus* individuals was assessed via morphology and cytb sequences by Sanger sequencing.

Genome Sequencing

Tissues for DNA extraction were stored in alcohol at -80°C . A 10- μg sample of genomic DNA (gDNA), prepared using the standard phenol chloroform extraction protocol, was used to construct libraries with a 350 base pair (bp) insertion size. Sequencing libraries were generated using the NEB Next Ultra DNA Library Prep Kit for Illumina (NEB), following the manufacturer's recommendations. Index codes were added to attribute sequences to each sample. DNA was purified using the AMPure XP system (Beckman Coulter, Beverly). Following adenylation of 3' ends of DNA fragments, NEB Next Adaptors with hairpin loop structures were ligated to prepare for hybridization. Electrophoresis was then used to select DNA fragments of a specified length, whereas 3 μl of USER enzyme (NEB) was used with size-selected, adaptor-ligated DNA at 37°C for 15 min, followed by 5 min at 95°C prior to PCR. This reaction was carried out using Phusion high-fidelity DNA

polymerase, universal PCR primers, and the index primer. Finally, PCR products were purified (AMPure XP system) and their library quality assessed using the Agilent Bioanalyzer 2100 system. Clustering of index-coded samples was performed using a cBot cluster generation system and the HiSeq 2500 PE Cluster Kit (Illumina) following the manufacturer's instructions. Subsequent to cluster generation, library preparations were sequenced using an Illumina HiSeq 2500 platform and 125-bp paired-end reads were generated. However, because raw sequencing data contain numerous low-quality reads with adaptors, we applied filtering strategies to obtain high-quality data. This procedure entailed: 1) Removal of read pairs containing adaptors; 2) Removal of read pairs generating a single sequence where the N content is $>10\%$ of the read length; and 3) Read pairs generating a single sequence where the number of low-quality bases (i.e., $Q < 5$) is greater than half the reads.

Read Mapping, SNPs Calling, Filtering, and Imputation

After filtering of reads using Btrim (Kong 2011), qualified reads were mapped onto the reference *R. norvegicus* genome (rn5) (ENSEMBL version 72) (Gibbs et al. 2004) using the program BWA-MEM (Li 2013). SNPs were detected and filtered using the Genome Analysis Toolkit (GATK) (McKenna et al. 2010). Duplicate read pairs were first identified using Picard tools (<http://picard.sourceforge.net/>) before being realigned and recalibrated around putative variants downloaded from dbSNP (ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/rat_10116/VCF/). We chose high-quality genotypes by filtering the SNPs data set with $\text{QUAL} > 30$, and removing sites where more than or equal to ten individuals were not called. This criteria ($\text{QUAL} > 30$) narrows the false calling rate of each SNP to $<0.1\%$. Further, since the genome sequences included in our data set were of low depth, we performed ungenotyped marker imputation using the software BEAGLE (Browning and Browning 2009), after removing triallelic sites and filtering the raw vcf file. We finally generated a total of 24,977,888 autosomal SNPs.

Phylogenetic Relationships and Population Structure Analysis

We constructed a phylogenetic assessment based on the entire SNP data set using neighbor-joining approach via rapidNJ method (fig. 1A) (Simonsen et al. 2008). However, taking into account the computation memory demands and possible effects of linkage disequilibrium (LD), we also constructed neighbor-joining or maximum-likelihood (ML) phylogenetic trees that included bootstrap support values (supplementary figs. S2–S5, Supplementary Material online) after thinning SNP data sets by vcftools (Danecek et al. 2011) at 50 k ($-\text{thin } 50,000$ parameter and 49,196 SNPs remains), 10 k ($-\text{thin } 10,000$ parameter and 225,012 SNPs remains), and 1 k ($-\text{thin } 1,000$ parameter and 1,628,064 SNPs remains) distances. The trees were constructed using MEGA, raxml, fasttree, and rapidNJ (Simonsen et al. 2008; Price et al. 2010; Stamatakis 2014; Kumar et al. 2016).

To evaluate the observed phylogenetic branching pattern (fig. 1A), we used msms program (Ewing and Hermisson 2010)

to simulate phylogenetic trees under the given demographic scenario (fig. 2B). The command line is shown below:

```
java -Xmx2000m -jar msms.jar 110 100 -t 583.68 -r 0 1000000
-1 5 31 20 12 26 21 1 -N 88158 -n 2 0.5929468 -n 3 0.143447 -n
4 0.7214206 -n 5 0.8375757 -g 5 206.4481 -g 4 445.8592 -ma x
5.119047 0.02974563 0 0.004104413 4.670567 x 0.5479121
0.6976758 0 0.00357659 0.001643775 x 0.0005432679
0.02944141 0 2.369616 1.160949 x 3.95137 0.01954961 0
0.004955953 0.02744081 x -en 0.01153037 5 0.07748544 -ej
0.01153037 5 3 -en 0.01027127 4 0.007401442 -ej 0.01027127 4
3 -ej 0.01781167 3 1 -en 0.1482897 2 0.08392838 -ej 0.9853474
2 1 -en 1.073493 1 11 -T
```

From the 100 simulations, all northern East Asia samples clustered together as a branch and closer to the root than other samples in the simulated phylogenetic trees. We randomly drew one simulated phylogenetic tree to display as an example (supplementary fig. S10, Supplementary Material online). This indicates that two migration waves out-of-southern East Asia can also cause the observed branching pattern in figure 1A.

In order to understand the relationships between different geographic population, we computed a PCA (fig. 1B) using the software GCTA (Yang et al. 2011) after pruning the SNPs by plink with `-indep-pairwise 50 10 0.1` parameter to get the relatively independent sites (Purcell et al. 2007). Population structure was then deduced using the software ADMIXTURE (supplementary fig. S6, Supplementary Material online), a tool for ML estimation of individual ancestries from multi locus SNP genotype data sets (Alexander et al. 2009). We evaluated different K values (from 2 to 10). The suitable value of K (K = 5) exhibited a low cross-validation (CV) error compared with other K values.

Outgroup f3 Statistic

To obtain a statistic that is informative of the genetic relatedness between a particular population and each East Asia individual, we implemented the “outgroup f3-statistic” (Patterson et al. 2012; Raghavan et al. 2014). We used black rat as the outgroup and computed the statistic f3(outgroup; A, B) with non-East Asia (Africa/Europe/Middle East) population as A and each one of the 51 East Asian individuals as B. Non-East Asia population displayed the closest relationship with the southern East Asia rats (supplementary fig. S7, Supplementary Material online).

Haplotype Sharing Analysis

Haplotype sharing analysis between different populations was performed as described by (Vonholdt et al. 2010). Here, phased haplotype were inferred using SHAPIT (Delaneau et al. 2013). The genome was divided into 20-kb nonoverlapping windows for the haplotype analyses. For windows with greater than or equal to five SNPs, we selected a random subset of five SNPs, which were used for all individuals. Windows with fewer than five SNPs were discarded. East Asia was assumed to be the center of origin for the brown rat. Therefore, we assessed the proportion of haplotypes for

each local population (Middle East, Europe, and Africa) shared with the two potential ancestral population (northern East Asia, and southern East Asia) following the method described elsewhere (Vonholdt et al. 2010).

To minimize the effects from difference in population sample sizes, we selected a random subset of 12 individuals from each population for analysis. Specifically, we tabulated the number of haplotypes within a local population that were present in only one of the two East Asia populations (supplementary fig. S8, Supplementary Material online). For instance, taking the Middle East population to explain the haplotype sharing analysis: at a window i , let MN_i denote the number of haplotypes present both in Middle East and northern East Asia (but absent from Europe, Africa, and southern East Asia), MS_i denote the number of haplotypes present both in Middle East and southern East Asia (but absent from Europe, Africa, and northern East Asia), and let P_{MN} denote the proportion of haplotypes across the genome shared between Middle East and northern East Asia. Then

$$P_{MN} = \frac{\sum_{all\ i} MN_i}{\sum_{all\ i} MN_i + MS_i}$$

Calculation of Private Variants and LD Decay Rate of Each Population

We counted the private variants within each of the five populations (supplementary fig. S12, Supplementary Material online). Considering the genetic background of the five rat populations, we calculated the private variants in two parts: two Asian populations (southern East Asia and northern East Asia) and three non-Asia populations (Europe, Africa, and Middle East). Between the two Asia populations, if a site was polymorphic in southern East Asia, but nonpolymorphic in northern East Asia, we defined the site to be a southern East Asia private variant, and vice versa. Among the three non-Asian populations, if a site was polymorphic in Europe, but nonpolymorphic in both African and Middle East populations, we defined the site to be a Europe private variant, and vice versa.

The linkage disequilibrium value r^2 (supplementary fig. S13, Supplementary Material online) of all pairwise SNPs within 1000 Kb distance for each population was calculated using PLINK (Purcell et al. 2007) with parameters `-maf 0.2 -r2 -ld-window 9999 -ld-window-r2 0.2`.

Composite Maximum Likelihood Inference for Demographic History of the Brown Rat Based on SFS

In order to calculate the joint SFS (Li and Stephan 2006; Gutenkunst et al. 2009), we first filtered raw SNPs with `QUAL > 30`, and removed sites which were not called in more than ten individuals. The genotypes were then imputed by Beagle. We inferred the ancestral state of each allele using the house mouse reference genome (mm10) (Chiaromonte et al. 2002; Kent et al. 2003; Schwartz et al. 2003). We only used biallelic SNPs with known ancestral alleles to build the joint SFS. Subsequently, we extracted the joint SFS based on imputed genotypes to infer demographic history by fastsimcoal2 (supplementary fig. S20, Supplementary Material

online). To avoid bias from imputation, we also compared the SFS before and after genotype imputation, and got very similar SFS (supplementary fig. S21, Supplementary Material online).

We calculated the likelihood function for different demographic scenarios using the software *fastsimcoal2* (Excoffier et al. 2013). For each scenario, 100,000 coalescent simulations per likelihood estimation (i.e., -n 100,000 -N 100,000) and at least 20 expectation-conditional maximization (ECM) cycles (-l20), up to a maximum of 40 (-L40), were used as the command line parameters for each run. At the same time, to avoid getting stuck on local optimum, 400 runs to 2,000 runs were carried out, whereas the Akaike information criterion (AIC) (Akaike 1974) was used to compare different models. In this case, $AIC = 2k - 2\ln(\text{MaxEstLhood})$, where k is the number of parameters estimated by each model, and *MaxEstLhood* is the ML function value for each model. Moreover, when searching for a ML value, *fastsimcoal2* may reach a local optimum instead of a global optimum. Thus, we repeated each step at least twice, to ensure we were not ending in a local optimum, thereby getting better estimates of the global optimum.

In order to obtain confidence intervals (CIs) for final estimates, 100 independent DNA polymorphism data sets were simulated as joint SFSs conditional on estimated demographic parameters. ML analysis was then applied to each joint SFS over 30 ECM cycles and 30 runs. Overall, 100,000 coalescent simulations were used to calculate likelihoods, giving empirical estimate distributions and 95% CIs.

The 100 simulated polymorphism data sets were further used to generate averaged simulated SFSs and to calculate simulated genetic diversity under the given demographic parameters. We compared the observed SFSs and observed genetic diversity to tell how well our estimation could explain the observed data set (supplementary fig. S20 and table S4, Supplementary Material online).

Because of a large number of parameters to be estimated, and many demographic models to be compared, it is difficult to infer the demographic history of all populations simultaneously. Therefore, we extended our previous approach (Li and Stephan 2006) and introduced an Ancestral-to-Derived Hierarchical Search strategy (supplementary materials and methods and supplementary figs. S14 and S15, Supplementary Material online). This strategy assumes that the newly established derived populations do not affect the demography of the ancestral population. By so doing, we could dramatically reduce the number of models that we had to evaluate (27 vs. 4,375).

Calculation of Genome-Wide Substitution Rate and Robustness Analysis of Rat and Mouse Generation and Divergence Times

Because substitution rates in the rodent lineage are generally faster than they are in many other mammal lineages (Wu and Li 1985), we re-estimated the substitution rate based on pairwise alignment between the rat (*R. norvegicus*, rn5) and house mouse (*M. musculus*, mm10) reference genomes

(Chiaromonte et al. 2002; Kent et al. 2003; Schwartz et al. 2003). On this basis, a total of 1,720,780,766 sites were included within the alignment, and indels and sites containing ambiguous nucleotides, N, were excluded. Consequently, 257,482,102 substitutions occurred since the divergence between rat and house mouse. The mean divergence time was estimated at ~22.6 Ma by TIMETREE (Hedges et al. 2006, 2015; Kumar and Hedges 2011). Assuming that rats have two generations per year (Ness et al. 2012; Halligan et al. 2013; Deinum et al. 2015), the estimated genome-wide nucleotide substitution rate (μ) was estimated to be 1.655×10^{-9} per generation per base pair. For three generations per year (Anderson and Jones 1967), the estimated substitution rate fell to 1.103×10^{-9} , which could be used to evaluate the robustness of our estimated migration times. Evaluation of 47 studies in the TIMETREE (Hedges et al. 2006, 2015; Kumar and Hedges 2011) showed that estimated divergence times between rat and mouse varied (supplementary table S13, Supplementary Material online). Therefore, we also re-estimated introduction times using rat–mouse divergence times ranging between 15 Ma ($\mu = 2.494 \times 10^{-9}$) and 30 Ma ($\mu = 1.247 \times 10^{-9}$).

Analysis of Positive Selection Signatures

F_{ST} value were calculated as described previously for each SNP (Akey et al. 2002). Sliding window analysis was performed with a window size of 100 kb, and a step size of 50 kb. F_{ST} value for each sliding window was calculated by averaging the values of all SNPs in the window. We employed an outlier approach based on genome-wide empirical data to retrieve the top 1% of windows showing high-level F_{ST} values, indicating candidate regions under positive selection.

Analysis of Functional Term Enrichment

We performed GO analysis of protein-coding genes using the online annotation tool g:Profiler, whereas P values were corrected using the Benjamini–Hochberg FDR (Reimand et al. 2011).

Accession Number

All the sequences reported in this study are deposited in the Genome Sequence Archive database, <http://gsa.big.ac.cn/> under Accession ID (PRJCA000251).

Supplementary Material

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

Acknowledgments

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences, Grant No XDB13020600, and the National Natural Science Foundation of China (91731304). N.O.O. thanks the support of CAS-TWAS President's Fellowship Program for Doctoral Candidates. We thank the collaborators who kindly shared brown rat samples: J.-M. Duplantier, L. Granjon, K. Bâ, C. Brouat, M. Diallo, M. Kane, A. Sow

(IRD, CBGP), B. Sicard (IRD, IMBE), S. Traoré (Institut d'Économie Rurale), C. Goarant (Institut Pasteur de Nouvelle Calédonie), S. Piry, Y. Chaval, N. Charbonnel (INRA, CBGP), C. Gotteland (CNRS/Université Lyon 1), Alexey P. Kryukov, Mr Erwan Lagadec, and Dr Pablo Tortosa (CRVOI-PIMIT). Authorization to use rat samples from the Seychelles collected during a CRVOI-IRD mission (SBS authorization dated 24.05.11) was provided by the Ministry of the Environment, Energy and Climate Change of Seychelles (special thanks to M. Alain de Comarmond, Principal Secretary, and M. Ronley Fanchette, Director of Conservation). Samples hosted at CBGP are stored at the collection platform (http://www6.montpellier.inra.fr/cbcp_eng/Platforms/Collections-platform), and included in the small mammal database (<http://vminfontron-dev.mpl.ird.fr/bdrss/index.php>).

References

- Akaike H. 1974. A new look at the statistical model identification. *Automatic Control IEEE Trans.* 19(6):716–723.
- Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12(12):1805–1814.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Amori G, Cristaldi M. 1999. The atlas of European mammals. London: Academic Press.
- Anderson S, Jones JK. 1967. Recent mammals of the world; a synopsis of families. New York, USA: Ronald Press.
- Armitage P. 1993. Commensal rats in the New World, 1492–1992. *Biologist* 40:174–178.
- Barnett SA. 2002. The story of rats: their impact on us, and our impact on them. Allen & Unwin.
- Browning BL, Browning SR. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet.* 84(2):210–223.
- Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325(6099):31–36.
- Chiaromonte F, Yap VB, Miller W. 2002. Scoring pairwise genomic sequence alignments. *Pac Symp Biocomput* 7:115–126.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
- Deinum EE, Halligan DL, Ness RW, Zhang Y-H, Cong L, Zhang J-X, Keightley PD. 2015. Recent evolution in *Rattus norvegicus* is shaped by declining effective population size. *Mol Biol Evol.* 32(10):2547–2558.
- Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 10(1):5–6.
- Demetriou M, Granovsky M, Quaggin S, Dennis JW. 2001. Negative regulation of T-cell activation and autoimmunity by Mgat5N-glycosylation. *Nature* 409(6821):733–739.
- Ewing G, Hermisson J. 2010. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 26(16):2064–2065.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet.* 9(10):e1003905.
- Forbes VL. 1995. The maritime boundaries of the Indian Ocean region. Singapore: National University of Singapore Press.
- Freye HA, Thenius E. 1968. Die Nagetiere. *Grzimeks Tierleben* 11:204.
- Fumagalli M. 2013. Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS One* 8(11):e79667.
- Gibbs RA, Weinstock GM, Metzker ML, et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5(10):e1000695.
- Halligan DL, Kousathanas A, Ness RW, Harr B, Eöry L, Keane TM, Adams DJ, Keightley PD. 2013. Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet.* 9(12):e1003995.
- Han E, Sinsheimer JS, Novembre J. 2015. Fast and accurate site frequency spectrum estimation from low coverage sequence data. *Bioinformatics* 31(5):720–727.
- Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23):2971–2972.
- 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.
- Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A.* 100(20):11484–11489.
- Kong Y. 2011. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* 98(2):152–153.
- Kosoy M, Khlyap L, Cosson J-F, Morand S. 2015. Aboriginal and invasive rats of genus *Rattus* as hosts of infectious agents. *Vector Borne Zoonotic Dis.* 15(1):3–12.
- Kumar S, Hedges SB. 2011. TimeTree2: species divergence times on the iPhone. *Bioinformatics* 27(14):2023–2024.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33(7):1870–1874.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Bioinformatics* arXiv:1303.3997v2.
- Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet.* 2(10):e166.
- Lin X-D, Guo W-P, Wang W, Zou Y, Hao Z-Y, Zhou D-J, Dong X, Qu Y-G, Li M-H, Tian H-F, et al. 2012. Migration of Norway rats resulted in the worldwide distribution of Seoul hantavirus today. *J Virol.* 86(2):972–981.
- Matisoo-Smith E, Robins JH. 2004. Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat. *Proc Natl Acad Sci U S A.* 101(24):9167–9172.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297–1303.
- Meerburg BG, Singleton GR, Kijlstra A. 2009. Rodent-borne diseases and their risks for public health. *Crit Rev Microbiol.* 35(3):221–270.
- Miksic JN. 2013. Singapore and the Silk Road of the Sea, 1300_1800. Singapore: National University of Singapore Press.
- Ness RW, Zhang Y-H, Cong L, et al. 2012. Nuclear gene variation in wild brown rats. *G3 Genes Genomes Genet.* 2:1661–1664.
- Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J. 2012. SNP calling, genotype calling, and sample allele frequency estimation from next-generation sequencing data. *PLoS One* 7(7):e37558.
- Nielsen R, Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet.* 12(6):443–451.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* 192(3):1065–1093.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.
- Puckett EE, Park J, Combs M, et al. 2016. Global population divergence and admixture of the brown rat. *Proc Biol Sci.* 283:20161762.

- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Orlando L, Metspalu E, et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91.
- Reimand J, Arak T, Vilo J. 2011. g: Profiler – a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res.* 39:W307–W315.
- Schraiber JG, Akey JM. 2015. Methods and models for unravelling human evolutionary history. *Nat Rev Genet.* 16(12):727–740.
- Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human-mouse alignments with BLASTZ. *Genome Res.* 13(1):103–107.
- Searle JB, Jones CS, Gündüz I, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry RJ, Giménez MD, et al. 2009. Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proc Biol Sci.* 276(1655):201–207.
- Silver J. 1941. The house rat. *Wildl Circ.* 6:1–18.
- Simonsen M, Mailund T, Pedersen CNS. 2008. Rapid neighbour-joining. In: Crandall KA, Lagergren J. (eds) Algorithms in Bioinformatics. WABI 2008. Lecture Notes in Computer Science, vol 5251. Berlin, Heidelberg: Springer.
- Song Y, Lan Z, Kohn MH. 2014. Mitochondrial DNA phylogeography of the Norway rat. *PLoS One* 9(2):e88425.
- Southern HN. 1964. The handbook of the British mammals. Oxford: Blackwell Scientific Publications.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Suckow AM, Weisbroth SH, et al. 2006. The laboratory rat. 2nd ed. San Diego (CA): Academic Press.
- Teng H, Zhang Y, Shi C, Mao F, Cai W, Lu L, Zhao F, Sun Z, Zhang J. 2017. Population genomics reveals speciation and introgression between brown Norway rats and their sibling species. *Mol Biol Evol.* 34(9):2214–2228.
- Tollenaere C, Brouat C, Duplantier J-M, et al. 2010. Phylogeography of the introduced species *Rattus rattus* in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. *J Biogeography* 37:398–410.
- Trantow CM, Hedberg-Buenz A, Iwashita S, Moore SA, Anderson MG. 2010. Elevated oxidative membrane damage associated with genetic modifiers of Lyst-mutant phenotypes. *PLoS Genet.* 6(7):e1001008.
- Van Valen L. 1973. A new evolutionary law. *Evol Theory* 1:1–30.
- Vieira FG, Fumagalli M, Albrechtsen A, Nielsen R. 2013. Estimating inbreeding coefficients from NGS data: impact on genotype calling and allele frequency estimation. *Genome Res.* 23(11):1852–1861.
- Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A, et al. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464(7290):898–902.
- Wilson DE, Reeder DM. 2005. Mammal species of the world: a taxonomic and geographic reference. 3rd ed. Baltimore, USA: Johns Hopkins University Press.
- Wu C-I, Li W-H. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci U S A.* 82(6):1741–1745.
- Wu X, Wang Y. 2012. Fossil materials and migrations of *Mus musculus* and *Rattus norvegicus*. *Res China's Front Archaeol.* 1:1–9.
- Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 88(1):76–82.

Supplementary Materials

Criteria and justification of sample stratification for inferring demographic history

In the ADMIXTURE analysis, wild rats could be separated into East Asian and non-East Asian populations when the number of presumed ancestral population (K) = 2. East Asian individuals could further be grouped into southern East Asia and northern East Asia sub-populations, when $K=4$. Phylogenetic tree topology, result of population structure by ADMIXTURE, as well as principal component analysis clearly separated the East Asian individuals into southern East Asian and northern East Asian (**Figure 1**). We found a good correspondence among the two sub-populations assignment of East Asian rats by the three classical analyses. The grouping also fits well with geographical distribution (**Figure 1C**). However, when we defined more sub-populations, for example, where southern East Asian rats are further grouped into southern China and Southeast Asia, or northern East Asian rats sub-divided into northern China and northern Asia, the three classical methods gave inconsistent assignments. We further inferred population history by fastsimcoal2, and found a very high level of genetic migrations between southern China and Southeast Asia, or between northern China and northern Asia (data not shown). It indicated that southern China and Southeast Asia could be the same population while northern China and northern Asia could be signed as a single population. Therefore, we grouped East Asian individuals into two populations, i.e. southern East Asia and northern East Asia.

Since non-East Asia individuals mix together into a large population, the three classical methods, i.e. phylogenetic tree, ADMIXTURE, and PCA, did not obtain consistent grouping. For the purpose of inferring migration routes, we grouped the non-East Asian individuals according to the geographical location information, i.e. Middle East, Europe and Africa. Considering that the sample from Morocco ($N=1$, an African country) cluster with European samples, and the special location of samples from New Caledonia ($N=8$, French colony), we dropped the 9 samples from the model testing to avoid possible biased signals of very recent migration events. Finally, we categorized the 101 brown rat samples to be 5 sub-populations (i.e., southern East Asia (SEA, $n=31$), northern East Asia (NEA, $n=20$), Africa (AF, $n=21$), Europe (EU, $n=17$), Middle East (ME, $n=12$)) to infer demographic history (**Supplementary Table S1**).

We also tested different grouping strategies to evaluate potential effects on inferred migration times. For example, we only used rats without admixture signals, which are from Southeast Asia ($n=14$), and from northern Asia ($n=10$) (**Supplementary Table S1**), to re-estimate migration times (the total sample size of all five populations used here is 74). The re-estimated times to migrate to the Middle East/Europe/Africa remained similar with the previous estimates (**Supplementary Table S12**).

Since ADMIXTURE analysis indicated that the Norway and Iceland individuals (Northern Europe) were separated from other non-East Asian samples when $K \geq 3$ (**Supplementary Figure S6**), we re-estimated migration time to Northern Europe by replacing original Europe samples with only northern Europe samples (including

Norway and Iceland, n=15) in the final demographic model (i.e., using populations of SEA, NEA, ME, AF and northern Europe). The estimated times were closely similar to the initial estimates above (**Supplementary Table S12**).

These analyses justified the sample grouping and lent credence to the applicability of our conclusions to different population grouping strategies.

Composite maximum likelihood (ML) inference for demographic history of the brown rat based on site frequency spectrum (SFS)

The joint site frequency spectra among brown rat populations

In order to calculate the joint site frequency spectrum (SFS) (Gutenkunst et al. 2009; Li and Stephan 2006), we first filtered raw SNPs with $QUAL > 30$, and removed sites which were not called in more than 10 samples. The genotypes were then imputed by Beagle. We inferred the ancestral state of each allele using the house mouse reference genome (mm10) (Chiaromonte et al. 2002; Kent et al. 2003; Schwartz et al. 2003). We only used bi-allelic SNPs with known ancestral alleles to build the joint SFS. Subsequently, we used an in-house perl script to extract the joint SFS based on imputed genotypes to infer demographic history by fastsimcoal2 (**Supplementary Figure S20**). To avoid bias from imputation, we also compared the SFS before and after genotype imputation, and got very similar SFS (**Supplementary Figure S21**).

Running fastsimcoal2

We calculated the likelihood function for different demographic scenarios using the software fastsimcoal2 (Excoffier et al. 2013). For each scenario, 100,000 coalescent simulations per likelihood estimation (i.e. -n 100,000 -N 100,000) and at least 20 expectation-conditional maximization (ECM) cycles (-l20), up to a maximum of 40 (-L40), were used as the command line parameters for each run. At the same time, to avoid getting stuck on local optimum, 400 runs to 2,000 runs were carried out, while the Akaike information criterion (AIC) (Akaike 1974) was used to compare different models. In this case, $AIC = 2k - 2\ln(\text{MaxEstLhood})$, where k is the number of parameters estimated by each model, and MaxEstLhood is the ML function value for each model. Moreover, when searching for a maximum likelihood value, fastsimcoal2 may reach a local optimum instead of a global optimum. Thus we repeated each step at least twice (results not shown), to ensure we were not ending in a local optimum, thereby getting better estimates of the global optimum.

In order to obtain confidence intervals (CIs) for final estimates, 100 independent DNA polymorphism datasets were simulated as joint SFSs conditional on estimated demographic parameters. ML analysis was then applied to each joint SFS over 30 ECM cycles and 30 runs. Overall, 100,000 coalescent simulations were used to calculate likelihoods, giving empirical estimate distributions and 95% CIs.

The 100 simulated polymorphism datasets were further used to generate averaged simulated SFSs and to calculate simulated genetic diversity under the given demographic parameters. By comparing with the observed SFSs and observed genetic diversity, we can see how well our estimation can explain the observed data set (**Supplementary Figure S20, Supplementary Table S4**).

An Ancestral-to-Derived Hierarchical Search strategy

As there are a large number of parameters to be estimated, and many demographic models to be compared, so it is difficult to infer the demographic history of all populations simultaneously. Therefore, we extended our previous approach (Li and Stephan 2006) and introduced an Ancestral-to-Derived Hierarchical Search strategy (**Supplementary Materials and Methods, Supplementary Figure S14-S15**). This strategy assumes that the newly established derived population do not affect the demography of the ancestral population. By doing so, we could dramatically reduce the number of models that we had to evaluate (27 vs 4,375).

Inferring demographic history of East Asian rats

Firstly, we evaluated preferred demographic models of the two hypothetical ancestral populations (i.e. southern East Asia, and northern East Asia) in East Asia from seven possible scenarios (**Supplementary Figure S16**). The instantaneous population size reduction model gave the smallest AIC for both populations, which was used in subsequent inference (**Supplementary Figure S16G, Supplementary Table S5-S6**).

Secondly, to clarify the detailed origin of wild brown rats within East Asia, we constructed two mutually exclusive founder-effect dispersal models (**Supplementary Figure S11A-S11B**) and four different widespread East Asia models (**Supplementary Figure S11C-S11F**). To increase accuracy, a larger number of simulations per likelihood estimation ($-n$ 1,000,000 $-N$ 1,000,000) were used. The ECM cycles ($-l$ 20 $-L$ 40) were kept constant. To get a global optimum, 1000 independent runs per model were simulated. The southern East Asia origin to northern East Asia gave the most likely scenario (**Supplementary Figure S11A, Supplementary Table S3**).

To further explore the possibility of northern East Asia origin, we constructed different out-of-northern East Asia demographic models (**Supplementary Figure S9B-S9D**). Model B and D can represent the previous hypothesis that the wild brown rat dispersed from northern Asia to Europe (Barnett 2002; Gibbs et al. 2004). Based on AIC, two migration waves out-of-southern East Asia model was most likely (**Supplementary Table S2**). Our results indicated that southern East Asia as the more likely cradle for the wild brown rats, which is consistent with previous fossil records (Wu and Wang 2012) and the previous study on mitochondrial DNA (Song et al. 2014).

Inferring demographic history of Non-East Asian populations

To explore how population sizes of derived populations (non-East Asia) might have changed with time, five possible demographic models were considered for each Non-East Asian population (**Supplementary Figure S17**). We expanded the number of considered populations step by step until all five populations were covered. In each step, all parameters were free when searching for the maximum likelihood value, unless otherwise noted.

An instantaneous expansion after out-of-SEA with a small founder population had the best fit for northern East Asia population (**Supplementary Figure S17C**,

Supplementary Table S7). A constant population size model since being derived from SEA was the best for Middle East population (**Supplementary Figure S17A, Supplementary Table S8**). On the other hand, the exponential growth model out of ME was the best for both Europe and African populations (**Supplementary Figure S18B, Supplementary Table S9, S11**).

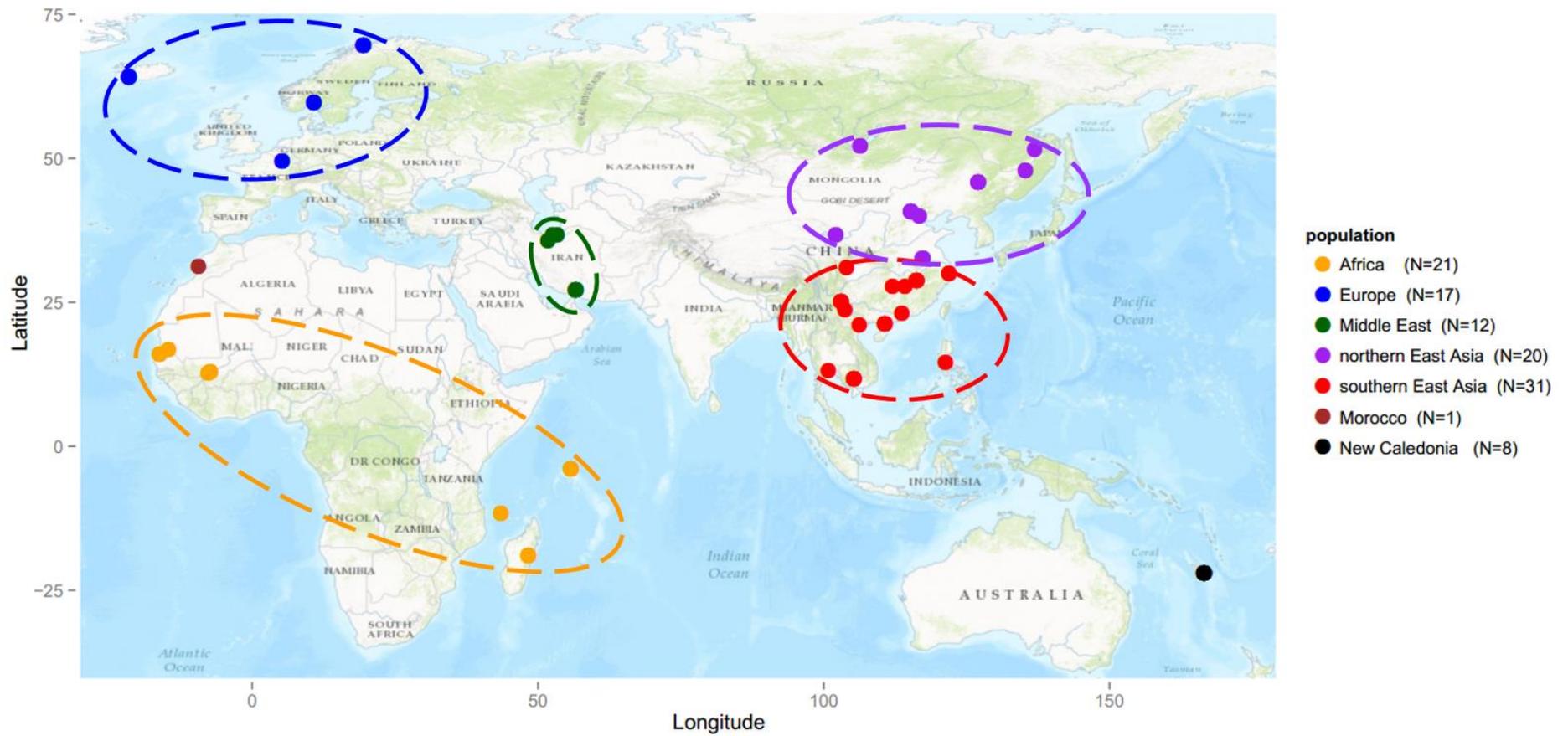
Final inference for the joint demographic history of five populations

The final overall demographic parameters were estimated for the joint demographic model for all five populations (**Figure 2, Supplementary Figure S15F**). For accuracy, we calibrated the lower range limit of the migration time of northern East Asian population out of Southern East Asia (T_{NEA}) to be the fossil time estimate of 140,000 years ago (Wu and Wang 2012). This takes into consideration the colonization time of northern East Asian population which should be larger than the fossil record. The number of parameters was 31, and we set fastsimcoal2 to run 3600 data sets and chose the best scenario, which had a log (MaxL) = -29,270,845 and AIC = 134,797,284 (**Figure 2, Supplementary Figure S15F, Supplementary Table S11**). We reviewed the results by re-running the estimation process (results not shown). We also compared the simulated SFS with the observed ones (**Supplementary Figure S20**), and found that our estimated scenario fits the observations well.

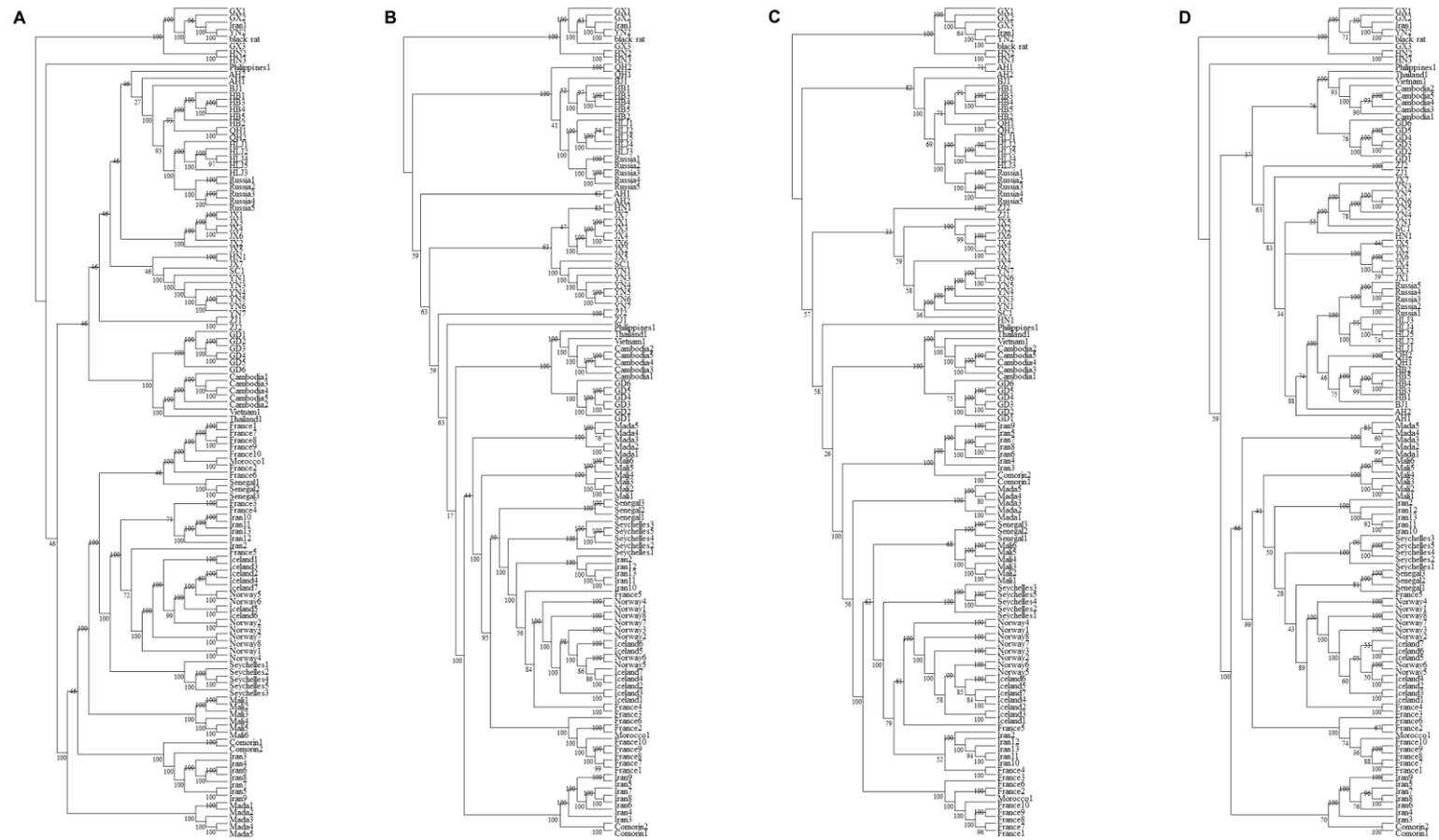
Robust analysis under different generation and divergence times

In the analysis above, we followed the procedures laid out in a previous study (Deinum et al. 2015), assuming 2 generations per year for brown rats. The uncertainty of the estimated generation time may affect the estimated split time. To test the robustness of the estimated demographic parameters, especially the introduction times of the derived populations, we set 3 generations per year and re-performed the analysis (Anderson et al. 1967). The introduction times of the derived populations varies very slightly (2,900 vs 3,100; 1,500 vs 1,800; 2,100 vs 2,000), while the confidence interval remained robust (**Supplementary Figure S19**). This is likely due to the joint SFS containing information on divergence among the populations (i.e. the fixed polymorphic sites within populations) (Li and Stephan 2006). The estimated introduction times are likely mainly determined by the ratio between population divergence and the rat-mouse divergence time. Therefore, our conclusions are robust to the uncertainty of the estimated generation time for brown rats.

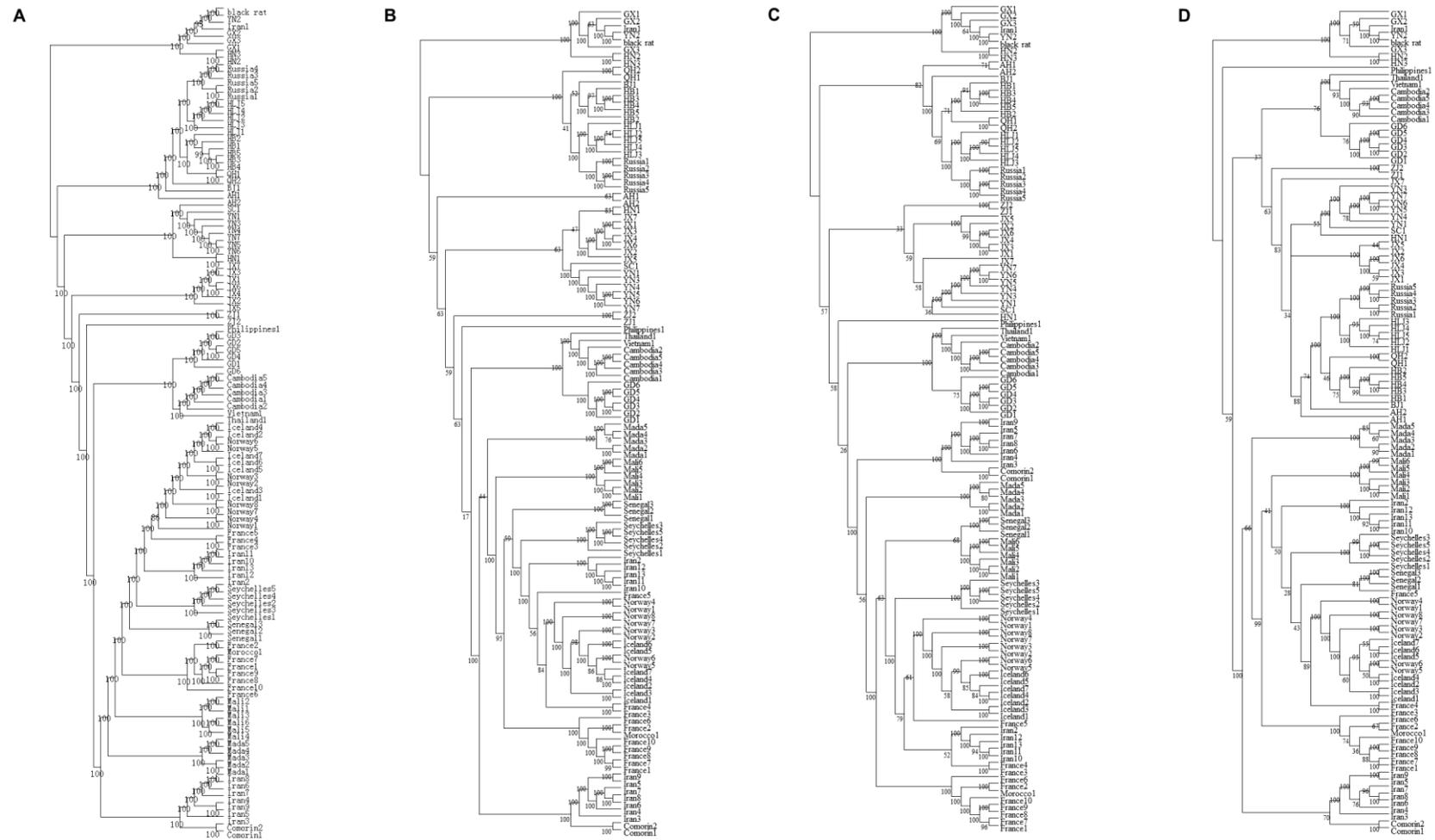
In above analysis, the mean divergence time for rats and mice from TIMETREE (22.6 million years ago) was used. In consideration of the different coalescent times for mitochondria and nuclear genes, and to exclude bias from the estimates from mitochondrial data, we first examined the effect of only using divergence time estimates based on nuclear data. This left us with 47 records in TIMETREE (**Supplementary Table S13**). Secondly, considering that the divergence time records from TIMETREE vary over a large range (**Supplementary Table S13**), we used two time points (15 million years ago and 30 million years ago) to validate the final demographic model (**Supplementary Figure S19**). The estimates from these two divergence times vary with consistency within a relative narrow range (**Supplementary Figure S19**). Thus our conclusions still hold after considering different divergence times.



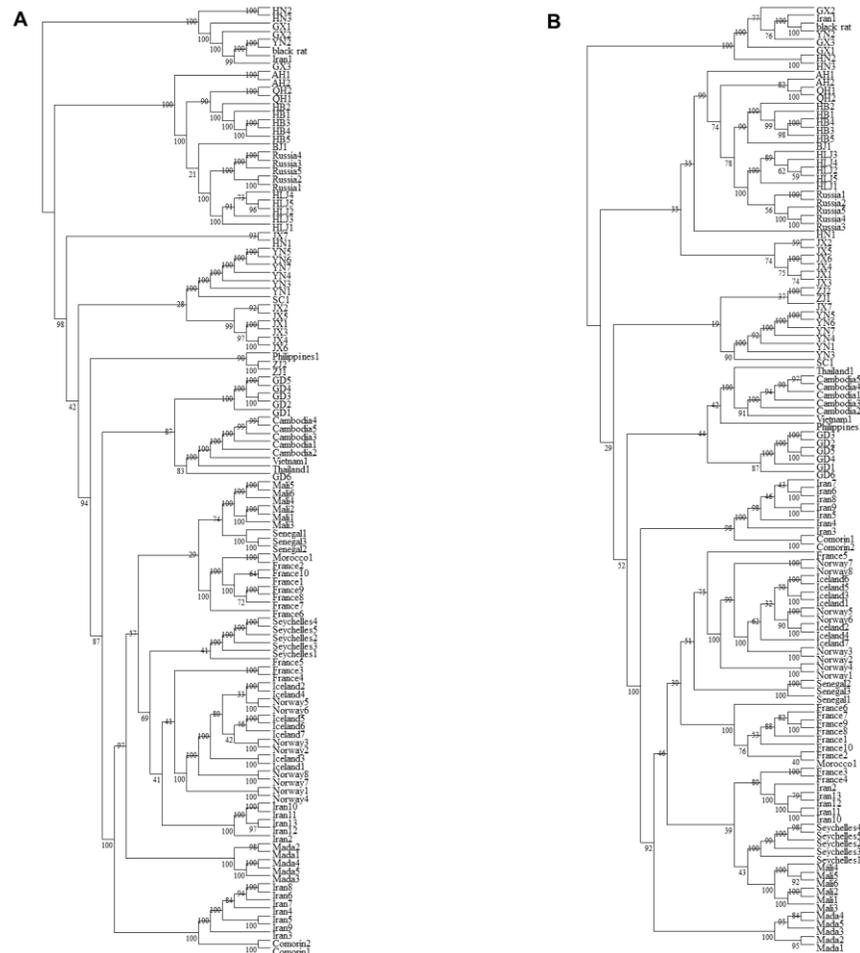
Supplementary Figure S1. Geographic locations of the wild brown rat samples.



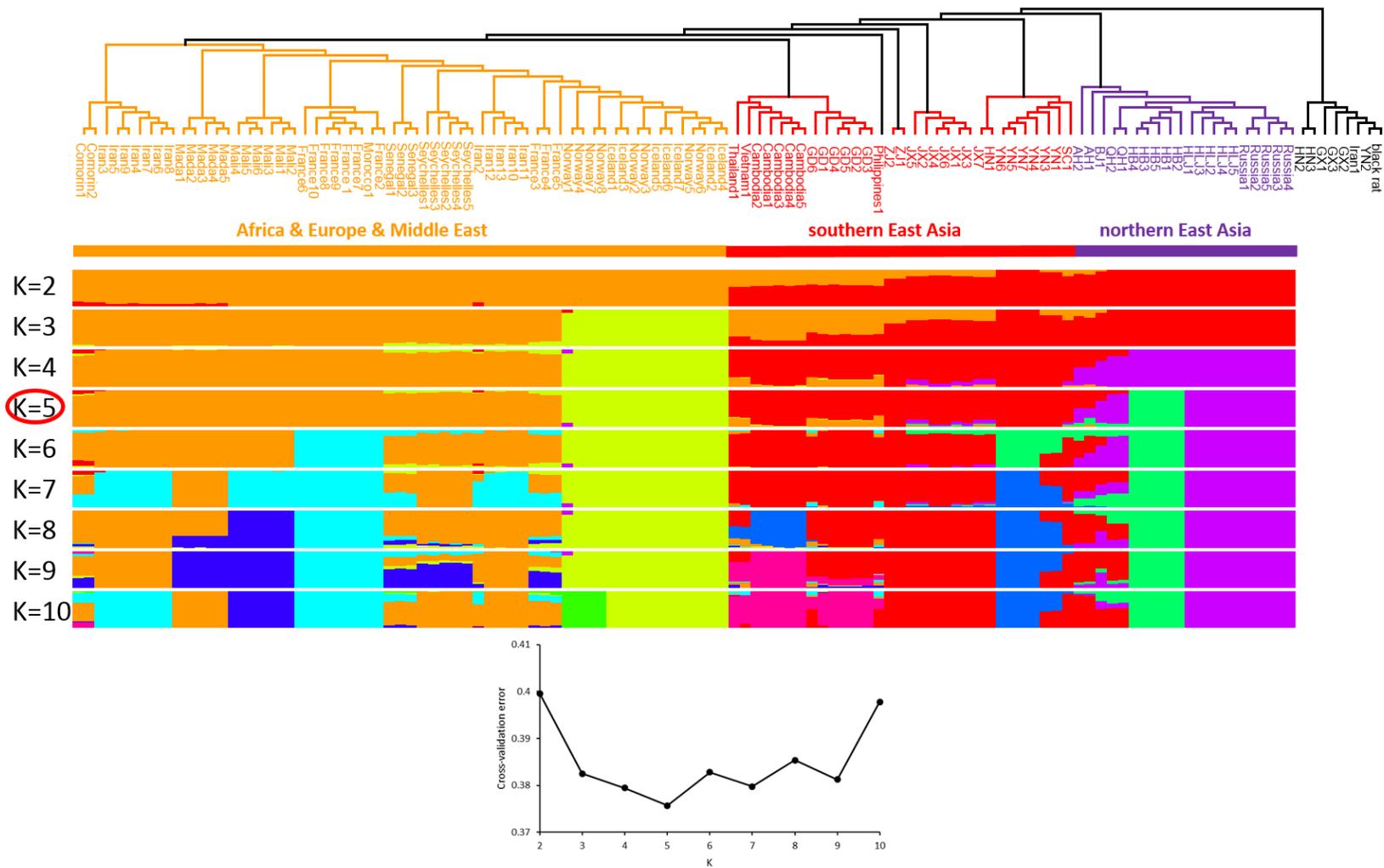
Supplementary Figure S3. Neighbor-joining phylogenetic tree of wild rats constructed by MEGA7 based on (A) 24,977,888 SNPs, (B) 1,628,064 SNPs, (C) 225,012 SNPs, and (D) 49,196 SNPs. Sample IDs are described in supplementary table S1.



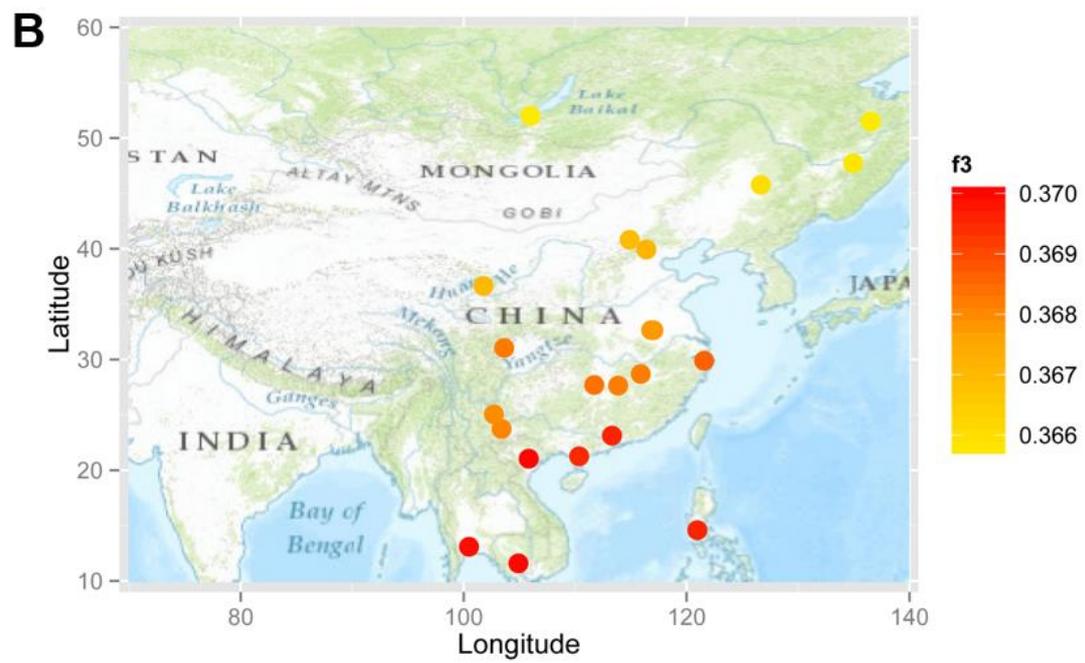
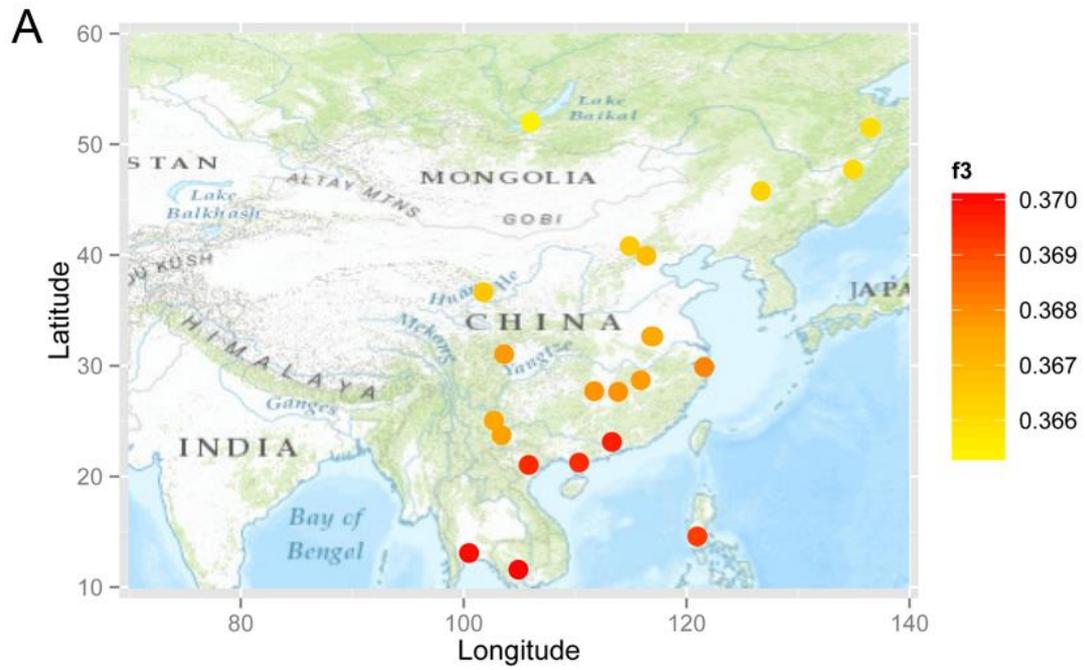
Supplementary Figure S4. Neighbor-joining phylogenetic tree of wild rats constructed by rapidNJ with (A) 24,977,888 SNPs, (B) 1,628,064 SNPs, (C) 225,012 SNPs, and (D) 49,196 SNPs. Sample IDs are described in supplementary table S1.



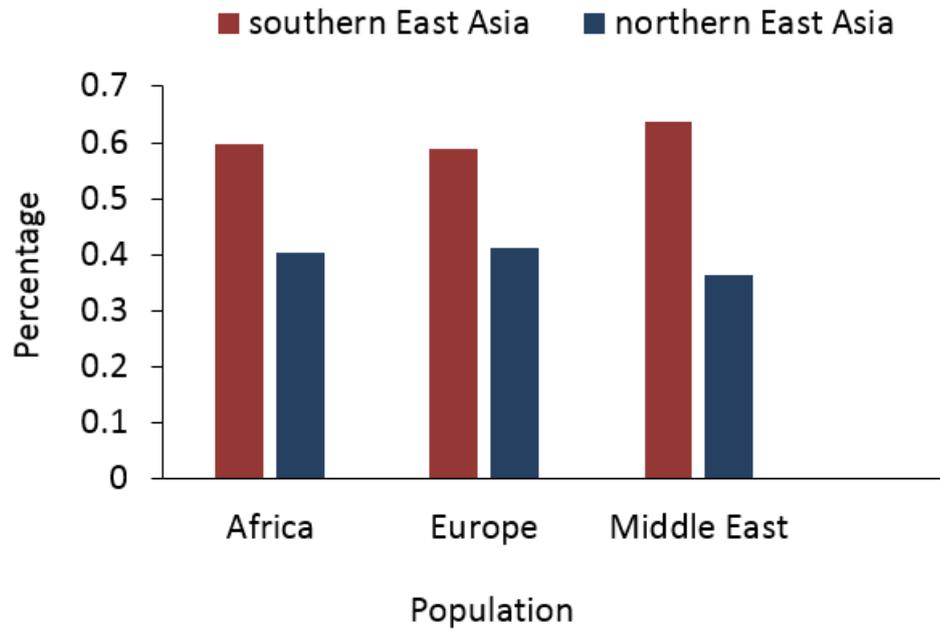
Supplementary Figure S5. Maximum-likelihood phylogenetic tree of wild rats by RAXML with (A) 225,012 SNPs and (B) 49,196 SNPs. Sample IDs are described in supplementary table S1.



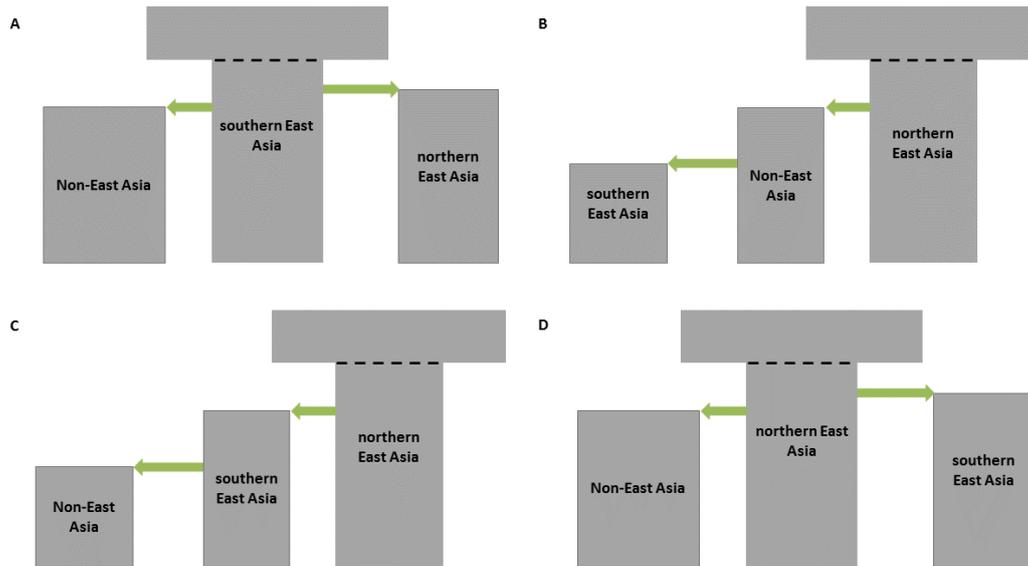
Supplementary Figure S6. Phylogenetic neighbor-joining tree and Bayesian clustering analysis by ADMIXTURE. Black colored lines represent Outgroup.



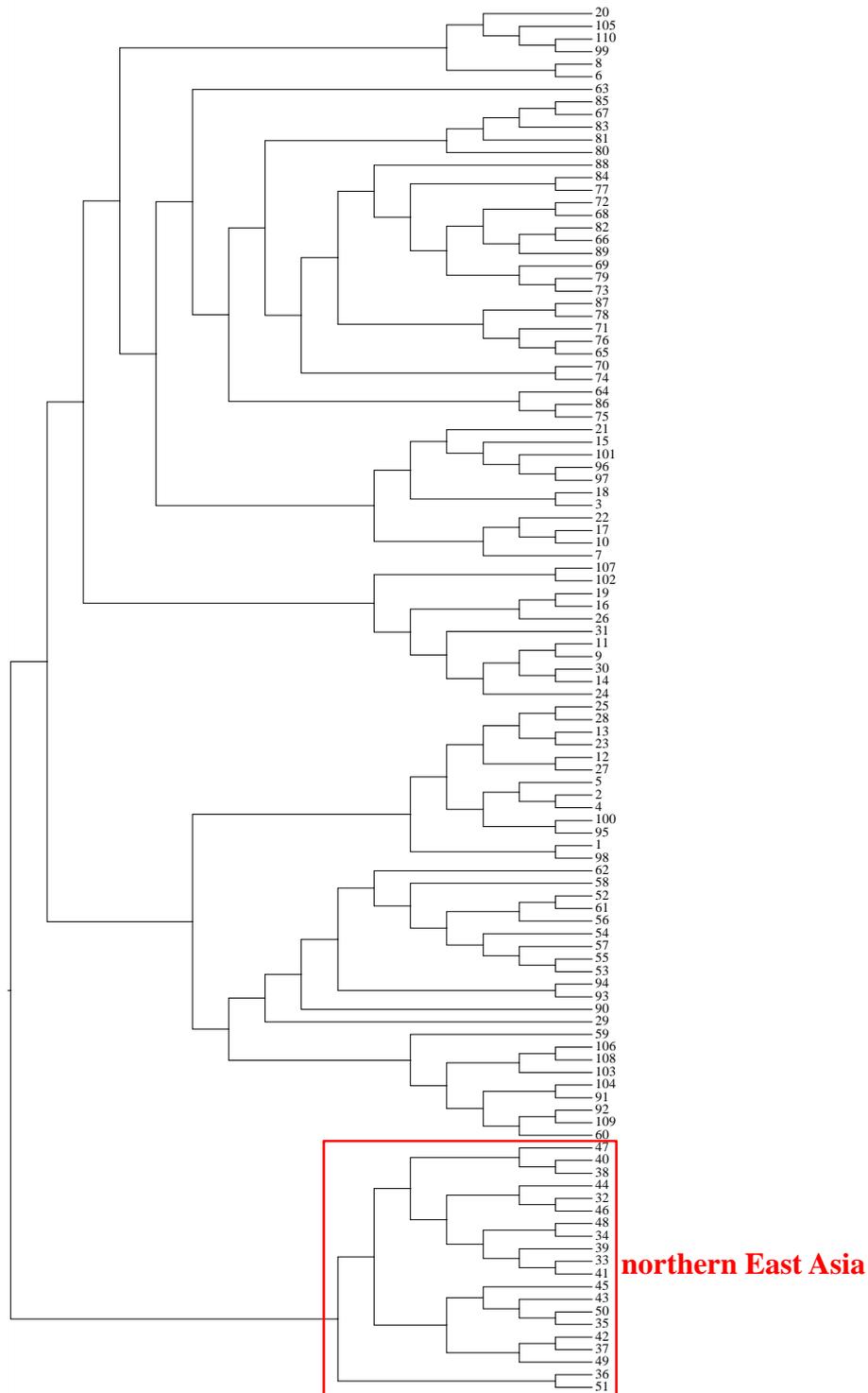
Supplementary Figure S7. Genetic proximity of Africa (A) and Middle-East (B) population to East Asia individuals evaluated by outgroup f_3 statistic.



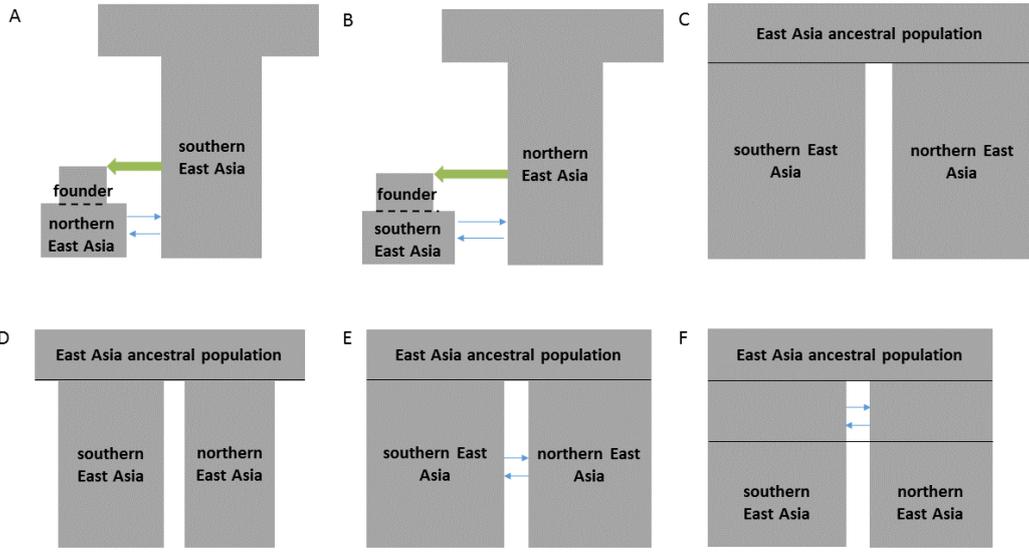
Supplementary Figure S8. Fraction of unique haplotypes shared between 2 East Asia populations and other 3 populations for 5 SNP windows.



Supplementary Figure S9. To further exclude the possibility of northern East Asia origin, we constructed different out-of-northern East Asia demographic models to compare with two migration waves out-of-southern East Asia model. (A) Two independent migration waves out of southern East Asia to Non-Asia and northern East Asia. (B) Wild brown rats migrated out of northern East Asia to colonize Non-Asia. Then they migrated back to southern East Asia from Non-Asia. (C) Concatenated migration routes from northern East Asia to southern East Asia, then to Non-Asia. (D) Two independent migration waves out of northern East Asia to southern East Asia and Non-Asia.



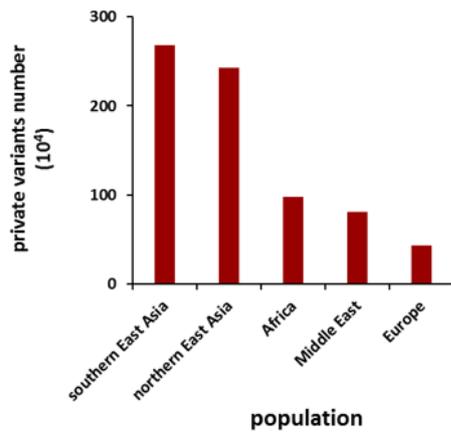
Supplementary Figure S10. A phylogenetic tree of wild brown rats based on the simulated data using msms considering the southern East Asia-origin demographic scenario. Numbers 1~31 represent southern East Asia, 32~51 represent north East Asia, 52~63 represent Middle East, 64~89 represent Europe, and 90~110 represent Africa. A total of 100 independent data sets were simulated. All simulated phylogenetic trees were rooted at the branch of the northern East Asia population, which indicates that two migration waves can cause the observed phylogenetic tree.



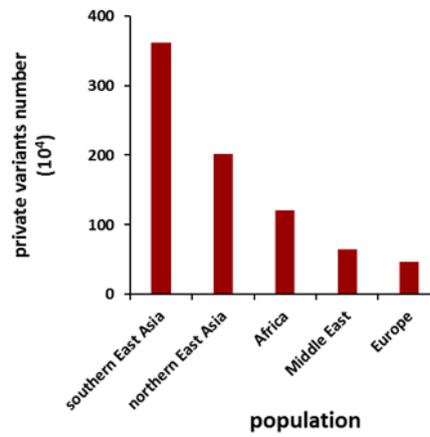
Supplementary Figure S11. Plausible demographic models between southern East Asia population and northern East Asia population, to clarify the birthplace of wild brown rats.

(A) The southern East Asia population was treated as the ancestral population, while the northern East Asia population migrated from it with a small founder size. (B) The northern East Asia population was treated as the ancestral population, while the southern East Asia population migrated from it with a small founder size. (C) A widespread East Asia model. The ancestral East Asia population split into southern East Asia population and northern East Asia population at some time in the past, and no genetic communication since. (D) A similar East Asia population split scenario like Model-C, but with a flexible ancestral population size. (E) East Asia population split model with continuous genetic communication. (F) East Asia population split model with continuous genetic communication at first and a genetic barrier occurring at reassume point in the past.

choosing randomly the same number of individuals (n=12) within each population



considering all individuals within each population

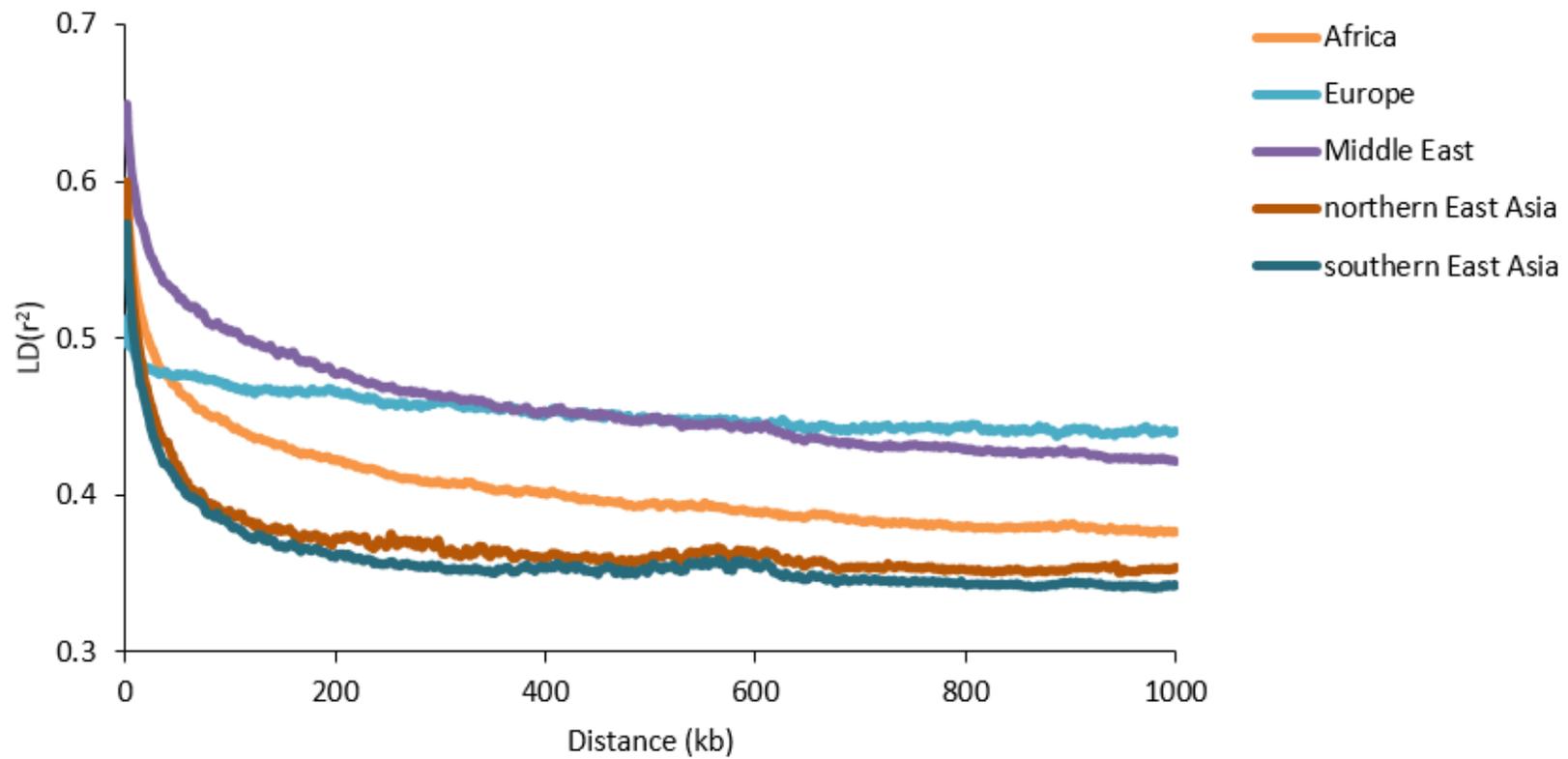


population	private variants number ¹	private variants number ²
southern East Asia	2682476	3612559
northern East Asia	2428134	2015423
Africa	977947	1209748
Middle East	812434	647428
Europe	428534	465684

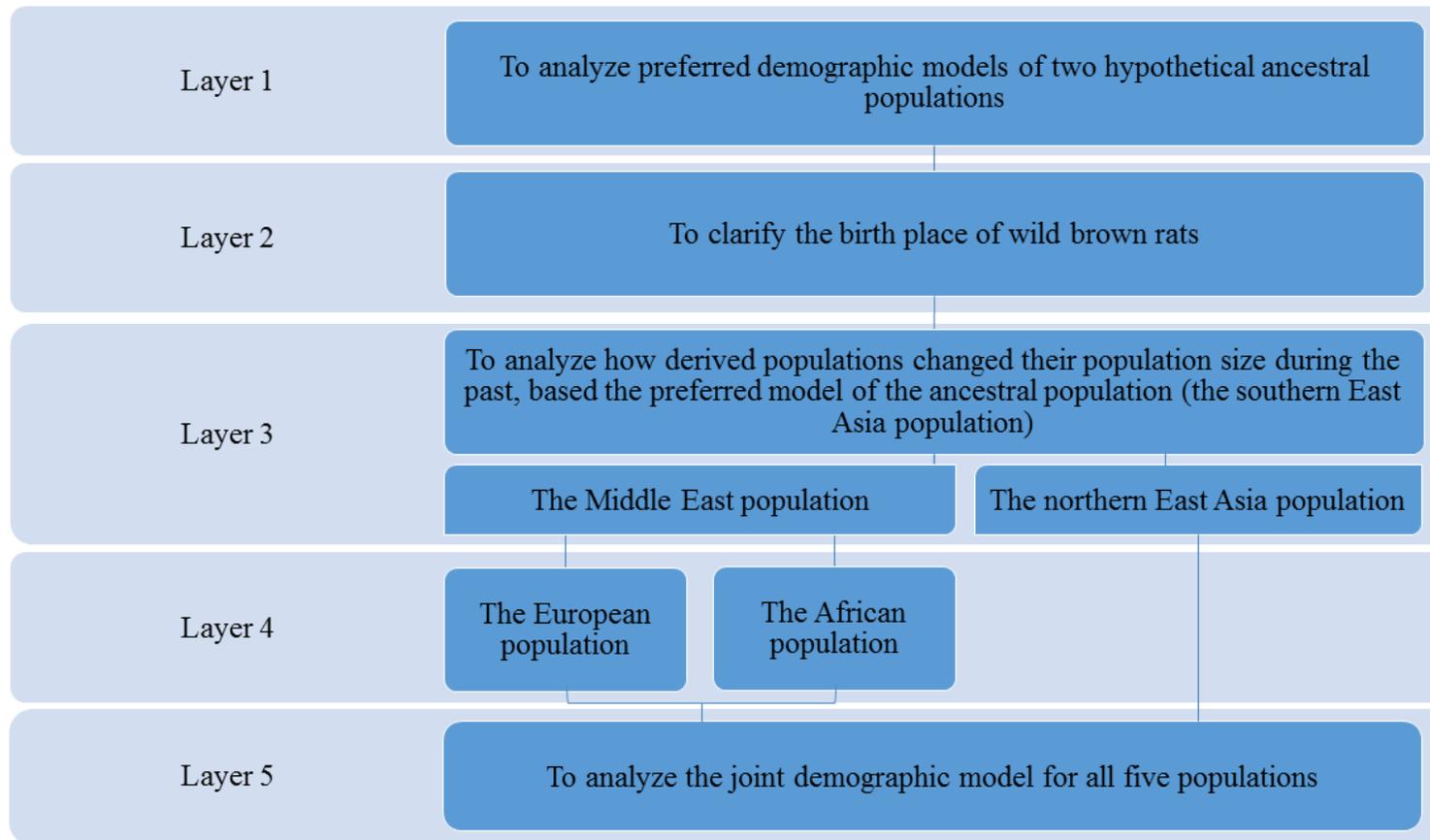
Note:

- 1) The data was calculated by considering all individuals within each population.
- 2) The data was calculated after choosing randomly the same number of individuals (n=12) within each population.

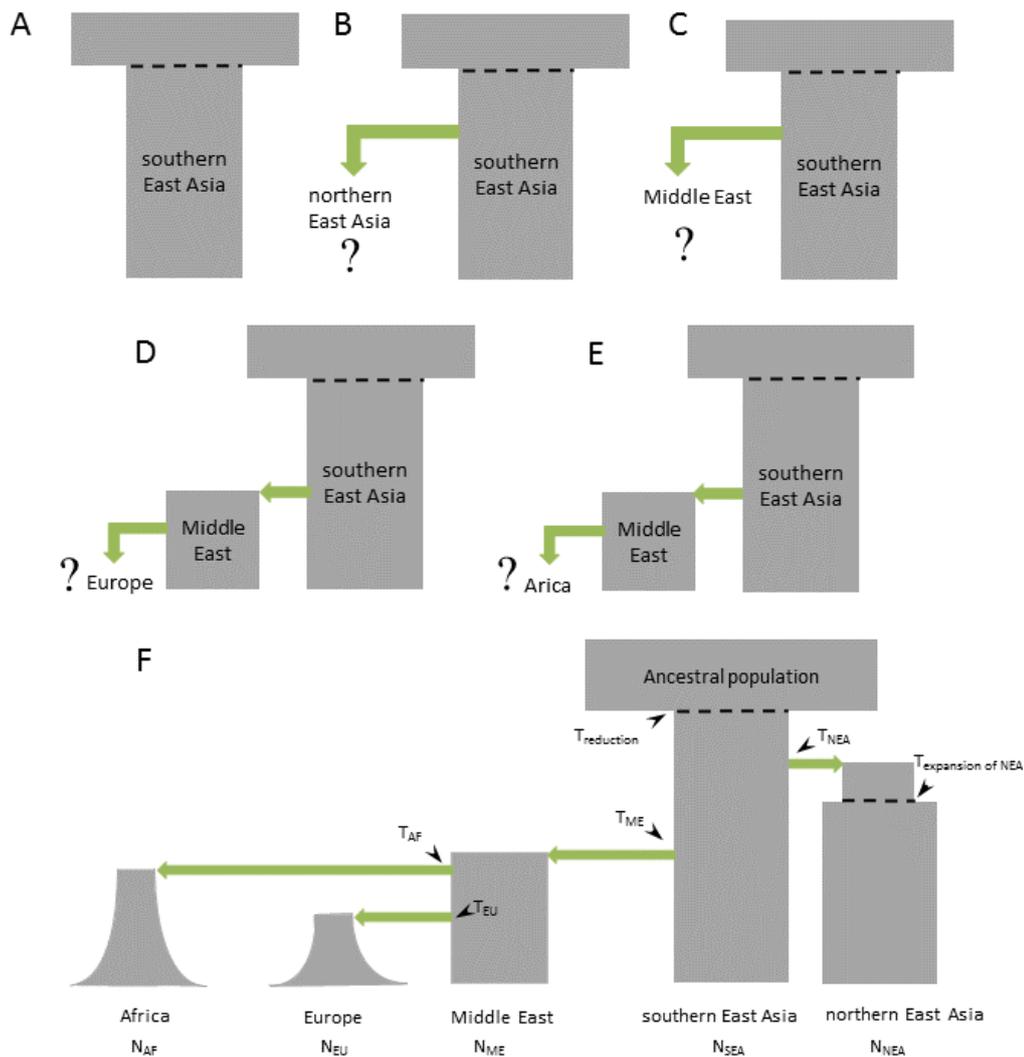
Supplementary Figure S12. Numbers of private variants of each population.



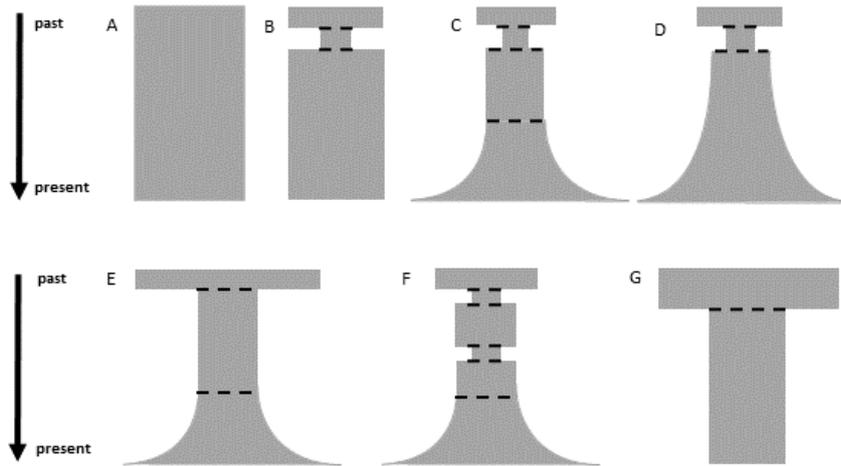
Supplementary Figure S13. LD decay of each population. The same numbers of individuals were chosen randomly for each population to calculate r^2 .



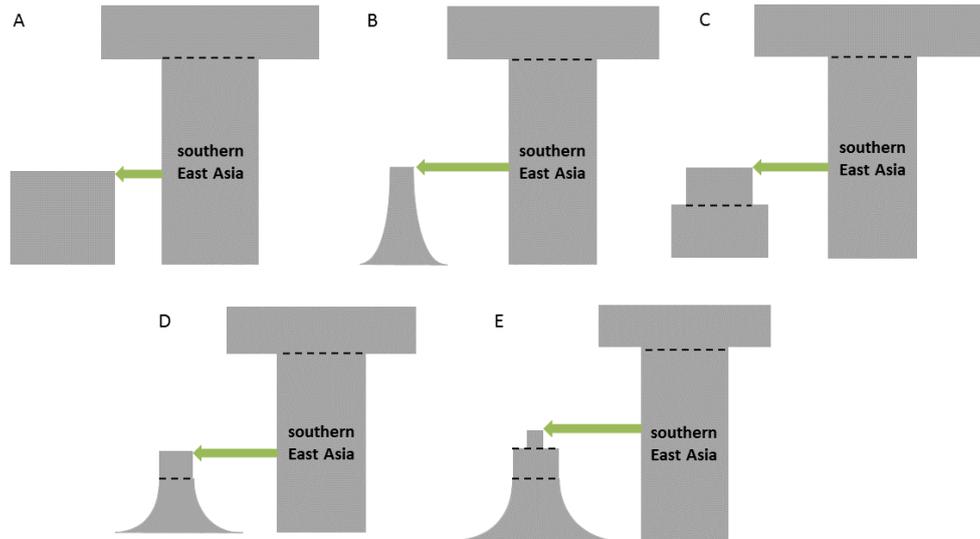
Supplementary Figure S14. Illustration of the Ancestral-to-Derived Hierarchical Search strategy for estimating the demographic history of wild brown rat. This analysis process is illustrated in detail in Supplementary Figure S15.



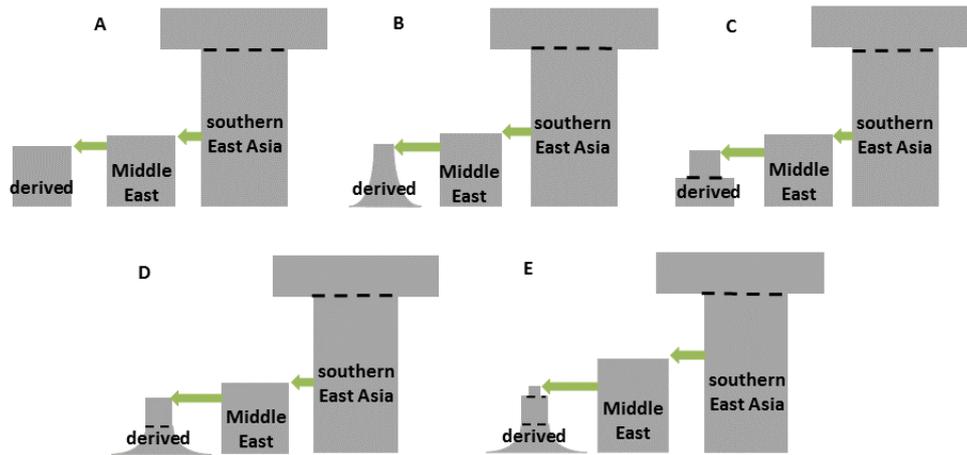
Supplementary Figure S15. Illustration of the models implemented in the Ancestral-to-Derived Hierarchical Search process. (A) The putative ancestral population (the Southern East Asia population) experienced one reduction event. (B, C) The joint demographic model for the northern East Asia and Middle East populations. The question mark represents where one of the demographic modules/models (**Supplementary Figure S17**) will be plugged-in as the demographic model for the derived northern East Asia or Middle East population. (D, E) The joint demographic model for the Europe/Africa population and its ancestral populations (**Supplementary Figure S18**). (F) The final joint demographic model of all five populations. Green arrows represent the introduction events of each derived population.



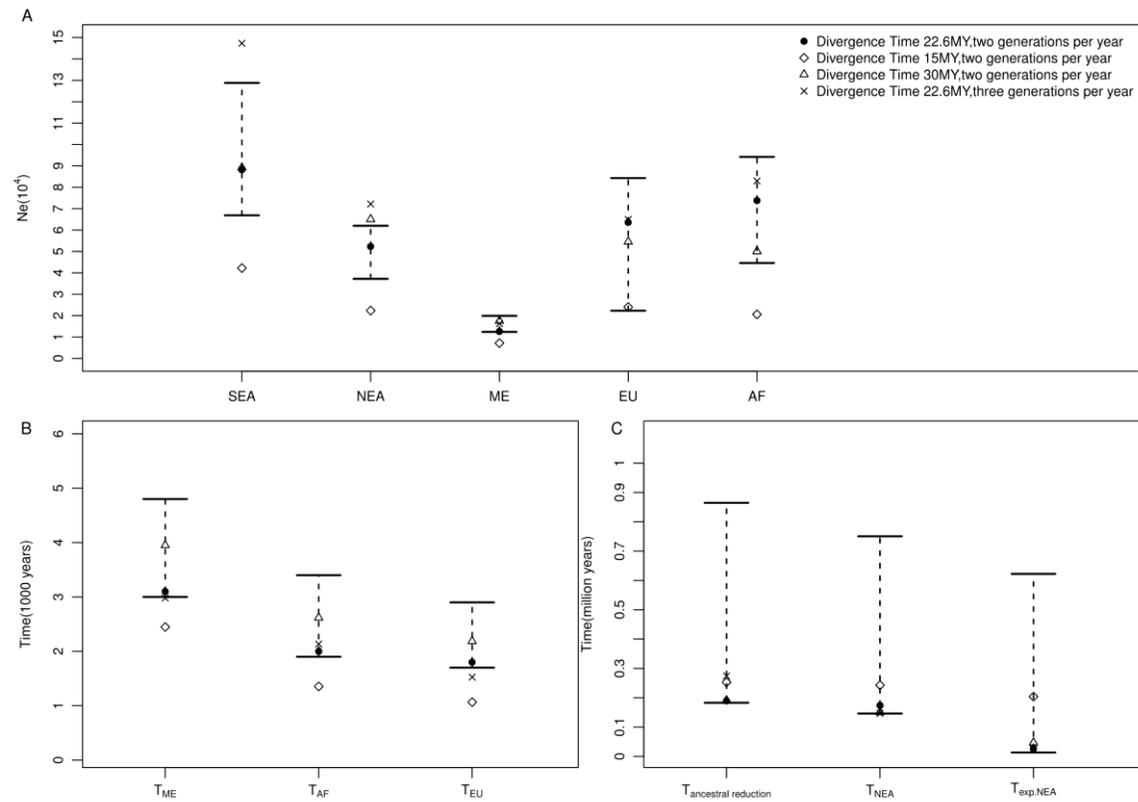
Supplementary Figure S16. Demographic models for two hypothetical ancestral populations (southern East Asia, or northern East Asia). (A) Constant population size model. (B) Instantaneous bottleneck model. (C) Instantaneous bottleneck accompanied by a delayed/recent exponential growth. (D) Instantaneous bottleneck accompanied by an immediate exponential growth. (E) Instantaneous population size reduction accompanied by a delayed/recent exponential growth. (F) Two sequential bottlenecks accompanied by a delayed/recent exponential growth. (G) Instantaneous population size reduction model.



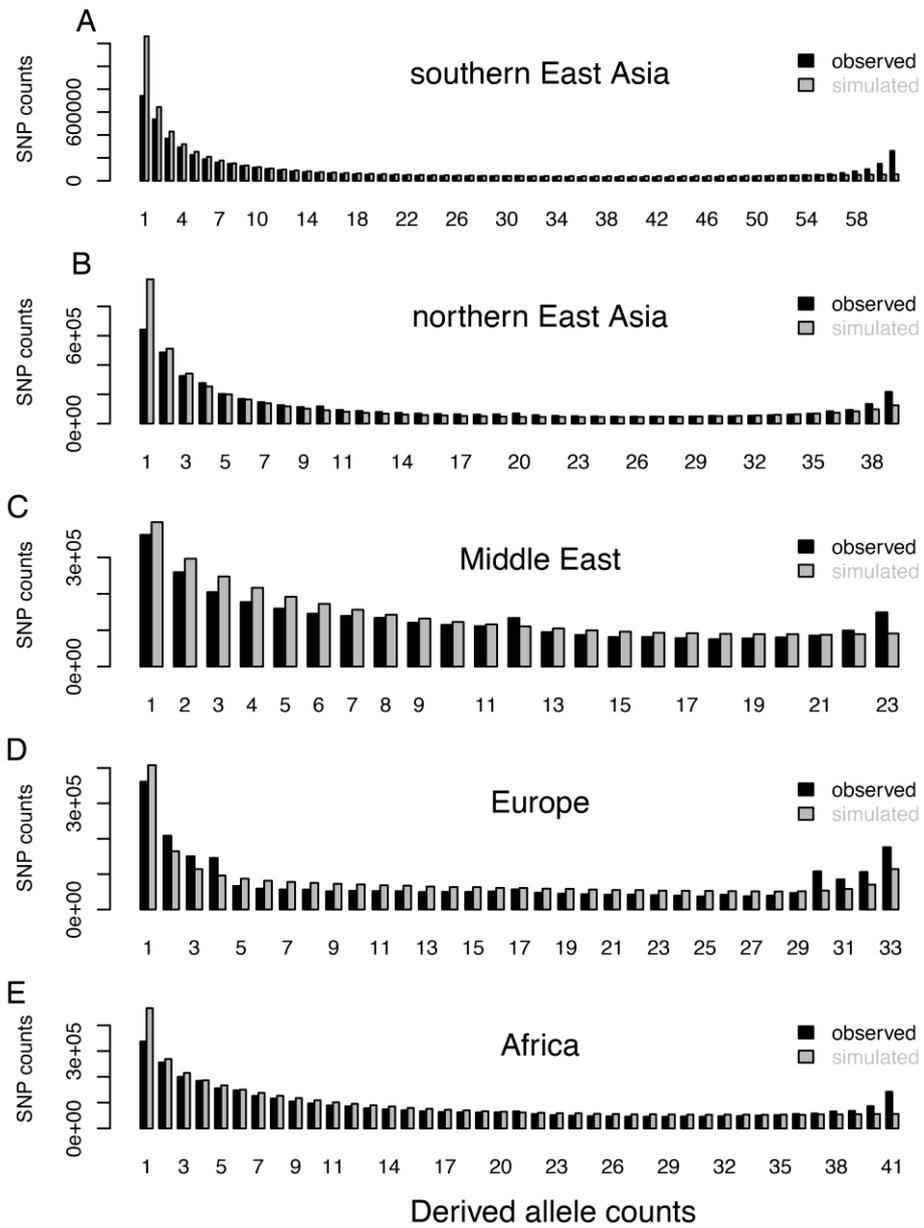
Supplementary Figure S17. Five joint demographic models for the southern East Asia population and a population directly derived from it (the northern East Asia population or the Middle East population). The derived population may follow the constant population size model (A), exponential growth model (B), instantaneous expansion model (C), delayed exponential growth model (D), or two-stage growth model (E).



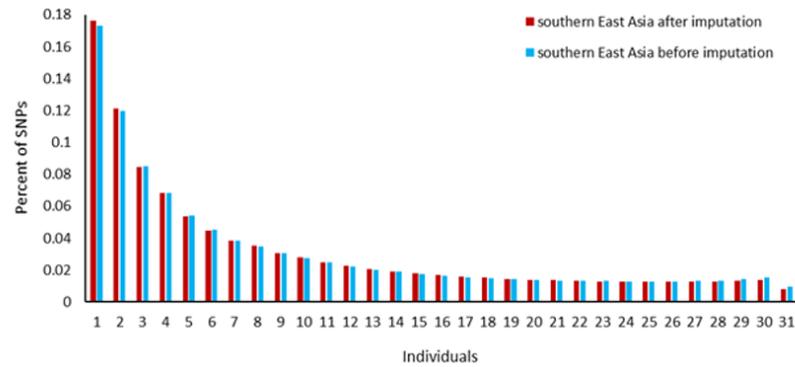
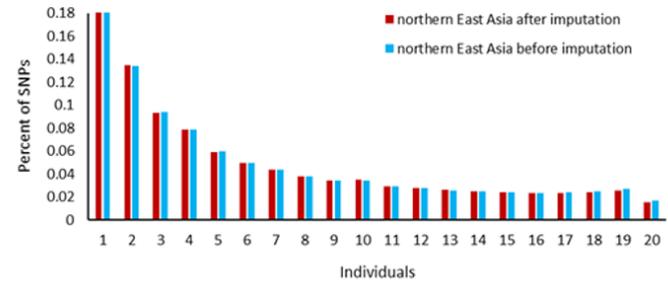
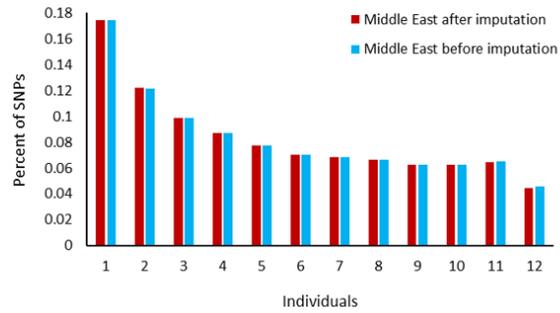
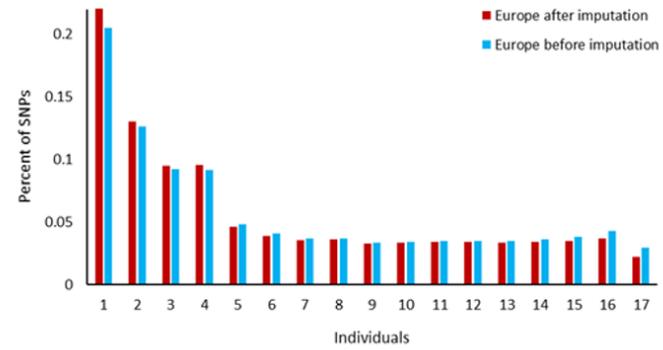
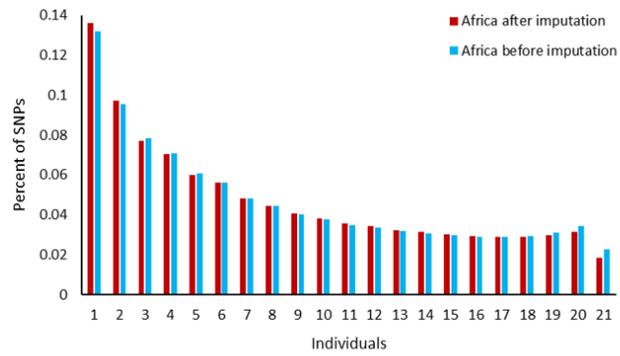
Supplementary Figure S18. Five joint demographic models for derived populations (Europe or Africa populations) out of Middle East. The derived population may follow the constant population size (A), exponential growth (B), instantaneous expansion (C), delayed exponential growth (D) or two-stage growth (E).



Supplementary Figure S19. Influence of different generation times and different divergence times on the estimation of parameters in the final joint model. (A) Comparison of the estimates of N_e . (B, C) Comparison of estimates of the migration events and population-size changing events. The black dots represent point estimates in the final model under 2 generation per year and 22.6 mya as divergence time. Vertical lines extending from black dots represent 95% confidence interval of each parameter. Empty triangles represent point estimates when a divergence time of 30 million years ago is used, with 2 generations per year. Empty rhombuses represent point estimates with a divergence time of 15 million years ago, with 2 generations per year. Crosses represent point estimates with a divergence time of 22.6 mya, and 3 generations per year.



Supplementary Figure S20. Comparison between observed and simulated site frequency spectra. (A) southern East Asia population, (B) northern East Asia population, (C) Middle East population, (D) Europe and (E) Africa population.



Supplementary Figure S21. SFS of 5 sub-populations before and after genotype imputation.

Supplementary Table S1. Sample information and genome data.

Population	Region	Country	Province/city	Individual	Longitude	Latitude	Runs	Clean Base (bp)	Data size(G)	depth	coverage
northern East Asia	northern Asia	Russia		#Russia1*	136.52	51.5	1	16,496,780,500	15.36	3.58	0.81
northern East Asia	northern Asia	Russia		#Russia2*	134.94	47.73	1	16,691,478,250	15.55	3.23	0.77
northern East Asia	northern Asia	Russia		#Russia3*	134.94	47.73	1	18,738,343,750	17.45	5.40	0.83
northern East Asia	northern Asia	Russia		#Russia4*	106	52	1	10,663,715,000	9.93	5.32	0.84
northern East Asia	northern Asia	Russia		#Russia5*	106	52	3	10,130,684,500	9.43	5.90	0.84
northern East Asia	northern Asia	China	Heilongjiang	#HLJ1*	126.6577	45.77323	2	10,454,914,000	9.74	3.34	0.78
northern East Asia	northern Asia	China	Heilongjiang	#HLJ2*	126.6577	45.77323	1	10,315,260,500	9.61	3.45	0.79
northern East Asia	northern Asia	China	Heilongjiang	#HLJ3*	126.6577	45.77323	1	10,524,870,500	9.8	3.13	0.77
northern East Asia	northern Asia	China	Heilongjiang	#HLJ4*	126.6577	45.77323	1	9,622,474,750	8.96	3.27	0.78
northern East Asia	northern Asia	China	Heilongjiang	#HLJ5*	126.6577	45.77323	2	9,465,122,250	8.82	2.99	0.75
northern East Asia	northern China	China	Hebei	HB1*	114.8938	40.81119	1	14,969,597,500	13.94	4.83	0.83
northern East Asia	northern China	China	Hebei	HB2*	114.8938	40.81119	2	14,055,669,750	13.09	4.60	0.83
northern East Asia	northern China	China	Hebei	HB3*	114.8938	40.81119	2	16,896,686,250	15.74	5.24	0.84
northern East Asia	northern China	China	Hebei	HB4*	114.8938	40.81119	2	15,980,597,250	14.88	5.22	0.80
northern East Asia	northern China	China	Hebei	HB5*	114.8938	40.81119	1	16,847,686,250	15.69	5.41	0.83
northern East Asia	northern China	China	Beijing	BJ1*	116.3956	39.92999	1	17,565,412,000	16.36	5.46	0.85
northern East Asia	northern China	China	Qinghai	QH1*	101.7679	36.64074	2	17,097,478,000	15.92	5.42	0.84
northern East Asia	northern China	China	Qinghai	QH2*	101.7679	36.64074	1	19,459,133,750	18.12	6.09	0.85
northern East Asia	northern China	China	Anhui	AH1*	116.8255	32.65239	1	15,907,410,250	14.81	5.16	0.81
northern East Asia	northern China	China	Anhui	AH2*	117.0186	32.64281	1	14,379,205,750	13.39	4.42	0.80
southern East Asia	southern China	China	Jiangxi	JX1*	115.8935	28.68958	2	16,603,182,250	15.46	5.34	0.84
southern East Asia	southern China	China	Jiangxi	JX2*	115.8935	28.68958	1	18,086,911,750	16.84	5.69	0.85
southern East Asia	southern China	China	Jiangxi	JX3*	115.8935	28.68958	2	17,583,520,000	16.38	5.50	0.82
southern East Asia	southern China	China	Jiangxi	JX4*	115.8935	28.68958	2	19,203,552,500	17.88	6.06	0.81
southern East Asia	southern China	China	Jiangxi	JX5*	115.8935	28.68958	1	18,645,720,250	17.37	5.78	0.85

southern East Asia	southern China	China	Jiangxi	JX6*	115.8935	28.68958	1	16,385,024,000	15.26	5.17	0.82
southern East Asia	southern China	China	Jiangxi	JX7*	113.8599	27.63954	2	18,253,137,750	17	5.66	0.84
southern East Asia	southern China	China	Sichuan	SC1*	103.6373	31.03912	1	17,550,349,250	16.35	5.47	0.84
southern East Asia	southern China	China	Yunnan	YN1*	103.4002	23.72906	1	15,445,808,250	14.39	4.72	0.83
southern East Asia	southern China	China	Yunnan	YN3*	102.7146	25.04915	2	17,857,407,250	16.63	5.34	0.84
southern East Asia	southern China	China	Yunnan	YN4*	102.7146	25.04915	2	15,967,136,000	14.87	5.03	0.84
southern East Asia	southern China	China	Yunnan	YN5*	102.7146	25.04915	1	16,695,266,000	15.55	5.33	0.83
southern East Asia	southern China	China	Yunnan	YN6*	102.7146	25.04915	2	17,304,409,500	16.12	5.38	0.84
southern East Asia	southern China	China	Yunnan	YN7*	102.7146	25.04915	2	19,268,580,000	17.95	6.01	0.84
southern East Asia	southern China	China	Zhejiang	ZJ1*	121.579	29.88526	2	17,175,334,000	16	5.47	0.84
southern East Asia	southern China	China	Zhejiang	ZJ2*	121.579	29.88526	1	15,282,638,250	14.23	5.06	0.80
southern East Asia	southern China	China	Hunan	HN1*	111.7207	27.69586	1	16,066,003,000	14.96	5.23	0.83
southern East Asia	southern China	China	Guangdong	#GD1*	113.3077	23.12005	1	18,286,204,750	17.03	3.25	0.77
southern East Asia	southern China	China	Guangdong	#GD2*	110.3651	21.25746	1	10,287,610,250	9.58	3.33	0.75
southern East Asia	southern China	China	Guangdong	#GD3*	110.3651	21.25746	1	11,596,887,250	10.8	3.49	0.74
southern East Asia	southern China	China	Guangdong	#GD4*	110.3651	21.25746	1	13,244,356,250	12.33	3.43	0.75
southern East Asia	southern China	China	Guangdong	#GD5*	110.3651	21.25746	2	10,748,558,500	10.01	2.83	0.70
southern East Asia	southern China	China	Guangdong	#GD6*	110.3651	21.25746	2	8,735,435,750	8.14	5.92	0.83
southern East Asia	Southeast Asia	Cambodia	Veal Renh	#Cambodia1*	103.8066	10.68800	1	13,709,900,500	12.77	4.28	0.83
southern East Asia	Southeast Asia	Cambodia	Veal Renh	#Cambodia2*	103.8654	10.71667	2	10,801,988,000	10.06	3.28	0.77
southern East Asia	Southeast Asia	Cambodia	Veal Renh	#Cambodia3*	103.8086	10.68804	1	12,928,487,250	12.04	4.23	0.83
southern East Asia	Southeast Asia	Cambodia	Veal Renh	#Cambodia4*	103.8086	10.68804	2	10,600,767,500	9.87	3.15	0.75
southern East Asia	Southeast Asia	Cambodia	Veal Renh	#Cambodia5*	103.8157	10.71174	2	15,652,823,750	14.58	8.93	0.84
southern East Asia	Southeast Asia	Thailand		#Thailand1*	100.0625	13.82056	2	12,148,290,750	11.31	3.70	0.79
southern East Asia	Southeast Asia	Vietnam		#Vietnam1*	105.8401	21.03438	1	15,049,025,500	14.02	5.13	0.75
southern East Asia	Southeast Asia	Philippines		#Philippines1*	120.9667	14.58333	1	16,269,668,250	15.15	5.20	0.78
Africa	Africa	Comorin	Moroni	#Comorin1*	43.253	-11.698	3	10,141,600,000	9.45	3.10	0.76

Africa	Africa	Comorin	Moroni	#Comorin2*	43.253	-11.698	1	10,277,050,750	9.57	3.16	0.76
Africa	Africa	Madagascar	Moramanga	#Mada1*	48.2780	-18.6338	2	10,453,407,750	9.74	2.95	0.72
Africa	Africa	Madagascar	Moramanga	#Mada2*	48.2780	-18.6338	2	9,874,723,750	9.2	3.13	0.78
Africa	Africa	Madagascar	Moramanga	#Mada3*	48.2027	-18.5351	2	10,085,910,250	9.39	2.85	0.71
Africa	Africa	Madagascar	Moramanga	#Mada4*	48.2023	-18.5351	1	12,167,245,250	11.33	3.66	0.80
Africa	Africa	Madagascar	Moramanga	#Mada5*	48.2621	-18.7504	2	10,300,548,500	9.59	2.79	0.69
Africa	Africa	Mali	Bamako	#Mali1*	-7.9825	12.61667	1	16,862,436,750	15.7	4.72	0.84
Africa	Africa	Mali	Bamako	#Mali2*	-7.9825	12.61667	1	11,594,019,250	10.8	3.49	0.79
Africa	Africa	Mali	Bamako	#Mali3*	-7.9825	12.61667	3	10,618,266,500	9.89	3.17	0.76
Africa	Africa	Mali	Massala	#Mali4*	-7.64389	12.82	1	13,272,118,250	12.36	4.01	0.82
Africa	Africa	Mali	Massala	#Mali5*	-7.64389	12.82	2	9,136,399,500	8.51	2.95	0.76
Africa	Africa	Mali	Massala	#Mali6*	-7.64389	12.82	1	13,060,967,000	12.16	3.93	0.81
Africa	Africa	Senegal	Podor	#Senegal1*	-14.9569	16.65249	2	12,024,843,500	11.2	3.75	0.79
Africa	Africa	Senegal	St-Louis	#Senegal2*	-16.5034	16.029	1	10,251,319,750	9.55	3.20	0.75
Africa	Africa	Senegal	St-Louis	#Senegal3*	-16.5035	16.02693	1	13,355,467,500	12.44	3.93	0.80
Africa	Africa	Seychelles	Sainte Anne	#Seychelles1*	55.49862	-4.61186	1	10,491,253,500	9.77	2.98	0.71
Africa	Africa	Seychelles	Grande Anse	#Seychelles2*	55.45479	-4.67672	1	13,550,527,500	12.62	3.94	0.81
Africa	Africa	Seychelles	Anse aux Pins	#Seychelles3*	55.51700	-4.69433	1	10,652,886,250	9.92	3.35	0.78
Africa	Africa	Seychelles	La Misère	#Seychelles4*	55.47711	-4.67323	1	11,100,227,250	10.34	3.42	0.79
Africa	Africa	Seychelles	La Misère	#Seychelles5*	55.47896	-4.67204	2	9,235,754,250	8.6	3.19	0.78
Europe	Europe	Morocco	Essaouira	Morocco1	-9.7583	31.52222	1	13,901,335,750	12.95	5.08	0.83
Europe	Europe	France (New Caledonia)		France1	165.5325	-21.4456	2	16,049,994,250	14.95	3.95	0.81
Europe	Europe	France (New Caledonia)		France2	165.5325	-21.4456	2	15,756,694,000	14.67	4.01	0.82
Europe	Europe	France (New Caledonia)		France3	165.5329	-21.4450	2	16,178,785,000	15.07	3.59	0.81
Europe	Europe	France (New Caledonia)		France4	165.5329	-21.4450	2	15,718,344,250	14.64	3.32	0.78

Europe	Europe	France (New Caledonia)		France5	165.5320	-21.4446 4	2	14,945,728,500	13.92	4.12	0.82
Europe	Europe	France (New Caledonia)		France6	165.5327	-21.4454 3	1	12,989,871,000	12.1	3.71	0.55
Europe	Europe	France (New Caledonia)		France7	165.5329	-21.4450 1	1	13,334,000,500	12.42	5.07	0.83
Europe	Europe	France		#France8*	4.9061	49.39547	1	11,594,442,250	10.8	5.02	0.83
Europe	Europe	France		#France9*	4.8782	49.40425	1	10,520,988,750	9.8	5.15	0.83
Europe	Europe	France (New Caledonia)		France10	165.4907	-21.4569 7	2	8,747,947,000	8.15	4.85	0.82
Europe	Europe	Iceland		#Iceland1*	-21.9333	64.15	1	12,454,364,000	11.6	3.36	0.78
Europe	Europe	Iceland		#Iceland2*	-21.9333	64.15	1	10,724,466,750	9.99	3.06	0.76
Europe	Europe	Iceland		#Iceland3*	-21.9333	64.15	2	10,584,483,750	9.86	3.35	0.77
Europe	Europe	Iceland		#Iceland4*	-21.9333	64.15	2	10,425,915,750	9.71	3.05	0.74
Europe	Europe	Iceland		#Iceland5*	-21.9333	64.15	2	8,556,728,000	7.97	2.53	0.69
Europe	Europe	Iceland		#Iceland6*	-21.9333	64.15	2	9,648,847,500	8.99	3.12	0.78
Europe	Europe	Iceland		#Iceland7*	-21.9333	64.15	2	17,011,201,000	15.84	4.61	0.79
Europe	Europe	Norway	Oslo	#Norway1*	10.45	59.56	2	9,792,185,250	9.12	3.11	0.76
Europe	Europe	Norway	Oslo	#Norway2*	10.45	59.56	1	6,845,680,500	6.38	2.32	0.64
Europe	Europe	Norway	Tromsoe	#Norway3*	19.1	69.65	1	13,388,157,000	12.47	3.90	0.80
Europe	Europe	Norway	Tromsoe	#Norway4*	19.1	69.65	1	12,953,849,250	12.06	3.95	0.82
Europe	Europe	Norway	Oslo	#Norway5*	10.45	59.56	1	6,289,878,750	5.86	2.15	0.61
Europe	Europe	Norway	Tromsoe	#Norway6*	19.1	69.65	1	9,674,172,500	9.01	3.07	0.77
Europe	Europe	Norway	Tromsoe	#Norway7*	19.1	69.65	1	10,232,988,000	9.53	3.25	0.76
Europe	Europe	Norway	Tromsoe	#Norway8*	19.1	69.65	1	13,613,082,250	12.68	4.26	0.83
Middle East	Middle East	Iran		#Iran2*	51.40528	35.72028	2	14,931,983,250	13.91	4.79	0.82
Middle East	Middle East	Iran		#Iran3*	56.26667	27.18333	2	14,511,898,500	13.52	4.60	0.82
Middle East	Middle East	Iran		#Iran4*	56.26667	27.18333	2	14,185,602,500	13.21	4.54	0.82
Middle East	Middle East	Iran		#Iran5*	56.26667	27.18333	2	17,745,789,250	16.53	5.70	0.85

Middle East	Middle East	Iran		#Iran6*	56.26667	27.18333	2	17,119,060,500	15.94	5.61	0.84
Middle East	Middle East	Iran		#Iran7*	56.26667	27.18333	2	16,313,374,250	15.19	5.37	0.84
Middle East	Middle East	Iran		#Iran8*	56.26667	27.18333	2	15,564,283,500	14.5	5.08	0.84
Middle East	Middle East	Iran		#Iran9*	56.26667	27.18333	2	18,117,490,250	16.87	5.88	0.85
Middle East	Middle East	Iran		#Iran10*	52.24528	36.6225	2	16,133,354,750	15.03	5.22	0.84
Middle East	Middle East	Iran		#Iran11*	52.24528	36.6225	2	17,549,548,250	16.34	5.65	0.84
Middle East	Middle East	Iran		#Iran12*	52.90583	36.64417	2	16,606,871,250	15.47	5.42	0.83
Middle East	Middle East	Iran		#Iran13*	52.90583	36.64417	2	15,417,471,000	14.36	4.88	0.82
Outgroup		China	Guangxi	GX1	108.6388	21.97335	2	15,175,575,000	14.13	4.88	0.76
Outgroup		China	Guangxi	GX2	108.6388	21.97335	1	17,083,242,750	15.91	5.34	0.77
Outgroup		China	Guangxi	GX3	108.6388	21.97335	1	15,263,410,750	14.22	4.92	0.77
Outgroup		China	Hunan	HN2	113.6777	25.55514	1	14,767,265,500	13.75	4.79	0.82
Outgroup		China	Hunan	HN3	113.6777	25.55514	1	13,603,852,250	12.67	4.34	0.79
Outgroup		Iran		Iran1	52.90583	36.64417	1	18,484,544,750	17.22	5.67	0.79
Outgroup		China	Yunnan	YN2	104.2463	23.37409	1	16,202,797,250	15.09	4.83	0.76
Outgroup		China	Yunnan	black_rat	102.7146	25.04915	3	82,128,029,204	76.49	19.62	0.83

Note: samples with * (n=101) were used in the demographic analysis.
samples with # (n=74) were used in the robust analysis.

Supplementary Table S2. Likelihood comparison of demographic models between different out-of-northern East Asia models and two-migration-waves out-of-southern East Asia model to further exclude out-of-northern East Asia hypothesis shown in Supplementary Figure S9.

Model	Parameters (k)	$\log_{10}(\text{MaxL})$	AIC
Model A	13	-10,995,478	50,636,073
Model B	13	-11,095,690	51,097,567
Model C	13	-11,049,657	50,885,577
Model D	13	-11,082,931	51,038,809

Supplementary Table S3. Likelihood comparison of demographic models between southern East Asia population and northern East Asia population to clarify the birthplace of wild brown rats in Supplementary Figure S11.

Model	Parameters (k)	log₁₀(MaxL)	AIC
Model A	9	-3,743,105	17,237,654
Model B	9	-3,751,881	17,278,069
Model C	3	-3,783,704	17,424,607
Model D	4	-3,774,887	17,384,005
Model E	6	-3,776,256	17,390,314
Model F	5	-3,780,600	17,410,316

Supplementary Table S4. Comparison of genetic diversity between observed and simulated datasets of the joint demographic scenario for all five populations.

Population	K_{observed}	$K_{\text{simulated}}$	Difference (%)	Sample Size	$\theta_{\text{observed}}(\%)$	$\theta_{\text{simulated}}$	$\pi_{\text{observed}}(\%)$	$\pi_{\text{simulated}}$
southern East Asia	5,665,747	6,294,286	11.09	62	0.778	0.741	0.701	0.775
northern East Asia	4,671,338	4,651,474	-0.425	40	0.635	0.635	0.672	0.615
Middle East	3,048,820	3,230,369	5.95	24	0.474	0.503	0.5693	0.622
Europe	2,548,657	2,528,540	-0.79	34	0.362	0.359	0.388	0.445
Africa	3,668,492	3,924,042	6.97	42	0.495	0.529	0.601	0.658

Supplementary Table S5. Likelihood comparison of the seven models of the southern East Asia population in Supplementary Figure S16.

Model	Parameters (k)	log10(MaxL)	AIC
Model A	1	-2,582,242	11,891,666
Model B	5	-2,557,818	11,779,197
Model C	7	-2,559,264	11,785,860
Model D	5	-2,560,437	11,791,258
Model E	5	-2,559,252	11,785,801
Model F	11	-2,557,035	11,775,603
Model G	3	-2,557,016	11,775,500

Supplementary Table S6. Likelihood comparison of the seven models of the northern East Asia population in Supplementary Figure S16.

Model	Parameters (k)	$\log_{10}(\text{MaxL})$	AIC
Model A	1	-2,100,790	9,674,497
Model B	5	-2,081,931	9,587,657
Model C	7	-2,083,226	9,593,624
Model D	5	-2,083,841	9,596,452
Model E	5	-2,082,950	9,592,349
Model F	11	-2,082,184	9,588,834
Model G	3	-2,081,220	9,584,378

Supplementary Table S7. Likelihood comparison of the five joint demographic models for the derived northern East Asia and the ancestral southern East Asia populations in Supplementary Figure S17.

Model	Parameters (k)	$\log_{10}(\text{MaxL})$	AIC
Model A	7	-3,757,919	17,305,871
Model B	8	-3,747,670	17,258,674
Model C	9	-3,732,896	17,190,639
Model D	9	-3,751,179	17,274,836
Model E	11	-3,746,419	17,252,919

Supplementary Table S8. Likelihood comparison of the five joint demographic models for derived Middle East and ancestral southern East Asia populations in Supplementary Figure S17.

Model	Parameters (k)	$\log_{10}(\text{MaxL})$	AIC
Model A	7	-3,212,223	14,792,845
Model B	8	-3,212,504	14,794,144
Model C	9	-3,213,249	14,797,578
Model D	9	-3,213,491	14,798,690
Model E	11	-3,213,398	14,798,265

Supplementary Table S9. Likelihood comparison of the five joint demographic models for Europe, Middle East and southern East Asia populations in Supplementary Figure S18.

Model	Parameters (k)	$\log_{10}(\text{MaxL})$	AIC
Model A	13	-8,510,902	39,194,178
Model B	14	-8,482,630	39,063,983
Model C	15	-8,495,734	39,124,331
Model D	15	-8,495,647	39,123,930
Model E	17	-8,494,877	39,120,388

Supplementary Table S10. Likelihood comparison of the five joint demographic models for Africa, Middle East and southern East Asia populations in Supplementary Figure S18.

Model	Parameters (k)	$\log_{10}(\text{MaxL})$	AIC
Model A	13	-8,995,876	41,427,566
Model B	14	-8,984,151	41,373,572
Model C	15	-8,997,566	41,435,353
Model D	15	-8,987,911	41,390,890
Model E	17	-8,999,317	41,443,420

Supplementary Table S11. Inferred parameters under the final joint model for all five populations in Supplementary Figure S15F.

Parameters	Point estimation	95% Confidence Interval		Range	Unit
		Lower bound	Upper bound		
N_{SEA}	88,200	66,900	128,800	1E4 -- 1E6	individual
N_{NEA}	52,300	37,200	62,000		
N_{ME}	12,600	12,400	19,900		
N_{EU}	63,600	22,300	84,300		
N_{AF}	73,800	44,600	94,200		
$N_{\text{Founder of NEA}}$	7,400	1,000	9,600	100 -- 1E4	
$N_{\text{Founder of EU}}$	650	620	1,800		
$N_{\text{Founder of AF}}$	6,800	6,700	9,800		
Reduction ratio of the ancestral population	1/11	1/35	2/5	1/100 -- 1	
$T_{\text{reduction}}$	189,300	183,000	864,550	140,000 -- 1E6	year
T_{NEA}	173,700	146,400	750,600	140,000 -- 1E6	
$T_{\text{expansion of NEA}}$	26,100	13,400	622,200	6000 -- 1E6	
T_{ME}	3,100	3,000	4,800	2000 -- 6000	
T_{EU}	1,800	1,700	2,900	1000 -- 4000	
T_{AF}	2,000	1,900	3,400	1000 -- 5000	
$m_{SEA \text{ to } NEA}^*$	1.45E-05	6.00E-06	1.63E-05	1E-9 -- 1E-4	
$m_{NEA \text{ to } SEA}^*$	1.32E-05	3.25E-06	1.68E-05		
$m_{SEA \text{ to } ME}^*$	8.44E-08	1.00E-09	5.50E-06		
$m_{ME \text{ to } SEA}^*$	1.01E-08	1.00E-09	5.50E-06		
$m_{SEA \text{ to } AF}^*$	1.16E-08	1.00E-09	6.50E-06		
$m_{AF \text{ to } SEA}^*$	5.54E-08	1.00E-09	3.75E-06		
$m_{NEA \text{ to } ME}^*$	1.55E-06	2.51E-07	1.68E-05		
$m_{ME \text{ to } NEA}^*$	4.66E-09	1.00E-09	4.25E-06		
$m_{NEA \text{ to } EU}^*$	1.98E-06	1.00E-09	7.25E-06		
$m_{EU \text{ to } NEA}^*$	6.72E-06	2.50E-06	1.23E-05		
$m_{ME \text{ to } EU}^*$	1.54E-09	1.00E-09	3.50E-06		
$m_{EU \text{ to } ME}^*$	3.29E-06	1.00E-09	1.00E-05		
$m_{ME \text{ to } AF}^*$	8.35E-08	1.00E-09	5.50E-06		
$m_{AF \text{ to } ME}^*$	1.41E-08	1.00E-09	6.25E-06		
$m_{EU \text{ to } AF}^*$	1.12E-05	2.51E-07	1.38E-05		
$m_{AF \text{ to } EU}^*$	7.78E-08	1.00E-09	8.25E-06		

Notes:

*: The migration rate (m) is the proportion of migrated individuals in each generation.

Supplementary Table S12. Comparison of estimated migration times under different population separation criteria.

	101 samples ¹	74 samples ²	North Europe samples used ³
T_{reduction}	189,300	217,900	253,400
T_{NEA}	173,700	201,800	164,500
T_{ME}	3,100	3,600	3,500
T_{EU}	1,800	1,800	2,400
T_{AF}	2,000	2,600	2,600

Notes:

The estimated time is in the unit of years. Two generations per year is used. The detailed information of samples are listed in Supplementary Table S1.

1) $n_{SEA}=31, n_{NEA}=20, n_{ME}=12, n_{EU}=17, n_{AF}=21$.

2) $n_{SEA}=14, n_{NEA}=10, n_{ME}=12, n_{EU}=17, n_{AF}=21$.

3) Population separation based on Bayesian clustering when $K=4$. $n_{SEA}=31, n_{NEA}=20, n_{ME}=12, n_{northern\ EU}=15, n_{AF}=21$.

Supplementary Table S13. Forty-seven studies on the divergence time between mouse and rat based on nuclear data*.

Divergence Time (Million years ago)	Reference	Divergence Time (Million years ago)	Reference
8.8	(Alhajeri et al. 2015)	16.4	(Huchon et al. 2007)
9.5	(Wu et al. 2012)	19	(Poux et al. 2006)
9.6	(Steppan et al. 2004)	20	(Springer et al. 2003)
9.7	(Rowe et al. 2008)	21.1	(Adkins et al. 2003)
10.6	(Fabre et al. 2013)	21.1	(Fabre et al. 2012)
11.2	(Douzery and Huchon 2004)	22.2	(Nyakatura and Bininda-Emonds 2012)
11.3	(Lecompte et al. 2008)	22.8	(Adkins et al. 2001)
11.4	(Pages et al. 2012)	23.9	(Crottini et al. 2012)
11.7	(Hallström et al. 2007)	24.2	(Babb et al. 2010)
11.9	(CHEVRET et al. 2005)	24.5	(O'hUigin and Li 1992)
11.9	(Michaux et al. 2002)	27	(Hugall et al. 2007)
12	(Schenk et al. 2013)	28.9	(Pisano et al. 2015)
12.3	(Hallström and Janke 2008)	32	(Kitano et al. 1999)
12.9	(Neumann et al. 2006)	32.5	(Alfaro et al. 2009)
13.2	(Renaud et al. 2007)	32.9	(Nei and Glazko 2002)
13.4	(Zhang et al. 2013)	32.9	(Nei et al. 2001)
13.4	(Jameson et al. 2011)	33.8	(Holmes 1991)
13.6	(Fabre et al. 2015)	40.7	(Kumar and Hedges 1998)
14	(dos Reis et al. 2012)	41	(Misawa and Janke 2003)
15	(Delsuc et al. 2004)	42.3	(Blair et al. 2005)
15.9	(Murphy et al. 2007)	41.6	(Pyron 2010)
16	(Douady and Douzery 2003)	43	(Kullberg et al. 2006)
16.2	(Hasegawa et al. 2003)	47	(Easteal and Herbert 1997)
16.3	(Douzery et al. 2003)		

*Note: The estimates and the references are obtained from TIMETREE (Hedges et al. 2006; Hedges et al. 2015; Kumar and Hedges 2011).

Supplementary Table S14. Functional categories with high level gene differentiation between Chinese and Europe wild brown rats detected by F_{ST} (except OR and VOM)

P-value	N	GO ID	term ID	Description
2.61E-05	3	GO:0072641	BP	type I interferon secretion
2.61E-05	3	GO:0035546	BP	interferon-beta secretion
2.61E-05	3	GO:0035547	BP	regulation of interferon-beta secretion
1.31E-02	3	GO:0032480	BP	negative regulation of type I interferon production
6.87E-03	3	GO:0032688	BP	negative regulation of interferon-beta production
2.61E-05	3	GO:0035548	BP	negative regulation of interferon-beta secretion
1.69E-02	13	GO:0051346	BP	negative regulation of hydrolase activity
2.21E-02	10	GO:0010466	BP	negative regulation of peptidase activity
1.58E-02	10	GO:0010951	BP	negative regulation of endopeptidase activity
9.43E-03	6	GO:1901655	BP	cellular response to ketone
3.40E-02	7	GO:0097306	BP	cellular response to alcohol
1.99E-03	6	GO:0033574	BP	response to testosterone
3.28E-05	5	GO:0071361	BP	cellular response to ethanol
7.15E-09	6	GO:0071394	BP	cellular response to testosterone stimulus
2.29E-02	2	GO:0036289	BP	peptidyl-serine autophosphorylation
2.23E-03	8	GO:0001906	BP	cell killing
3.62E-02	10	GO:0002443	BP	leukocyte mediated immunity
1.95E-02	9	GO:0002449	BP	lymphocyte mediated immunity
6.90E-04	8	GO:0001909	BP	leukocyte mediated cytotoxicity
5.00E-02	3	GO:0031342	BP	negative regulation of cell killing
3.91E-02	3	GO:0001911	BP	negative regulation of leukocyte mediated cytotoxicity
2.35E-02	9	GO:0002460	BP	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
2.25E-02	6	GO:0002456	BP	T cell mediated immunity
1.76E-02	5	GO:0001913	BP	T cell mediated cytotoxicity
1.31E-02	3	GO:0002710	BP	negative regulation of T cell mediated immunity
6.35E-05	7	GO:0002228	BP	natural killer cell mediated immunity
2.21E-02	3	GO:0002716	BP	negative regulation of natural killer cell mediated immunity
5.62E-05	7	GO:0042267	BP	natural killer cell mediated cytotoxicity
2.21E-02	3	GO:0045953	BP	negative regulation of natural killer cell mediated cytotoxicity
8.84E-04	3	GO:0001915	BP	negative regulation of T cell mediated cytotoxicity
2.58E-02	3	GO:0002719	BP	negative regulation of cytokine production involved in immune response
1.58E-02	3	GO:0002374	BP	cytokine secretion involved in immune response
6.87E-03	3	GO:0002739	BP	regulation of cytokine secretion involved in immune response
1.40E-03	3	GO:0002740	BP	negative regulation of cytokine secretion involved in immune response
4.44E-02	3	GO:0010939	BP	regulation of necrotic cell death
1.40E-03	3	GO:0010940	BP	positive regulation of necrotic cell death
1.07E-02	3	GO:0060544	BP	regulation of necroptotic process
1.04E-04	3	GO:0060545	BP	positive regulation of necroptotic process
1.76E-02	5	GO:0017144	BP	drug metabolic process
3.50E-05	4	GO:0002309	BP	T cell proliferation involved in immune response
1.84E-02	6	GO:0032609	BP	interferon-gamma production
1.06E-02	6	GO:0032649	BP	regulation of interferon-gamma production
7.19E-03	4	GO:0032689	BP	negative regulation of interferon-gamma production
1.58E-02	3	GO:0072643	BP	interferon-gamma secretion
2.14E-02	6	GO:0042107	BP	cytokine metabolic process
1.75E-02	6	GO:0042089	BP	cytokine biosynthetic process
2.21E-02	3	GO:0042533	BP	tumor necrosis factor biosynthetic process

1.31E-02	3	GO:0042095	BP	interferon-gamma biosynthetic process
1.07E-02	3	GO:0045072	BP	regulation of interferon-gamma biosynthetic process
2.57E-04	3	GO:0045077	BP	negative regulation of interferon-gamma biosynthetic process
2.21E-02	3	GO:0042534	BP	regulation of tumor necrosis factor biosynthetic process
1.40E-03	3	GO:0042536	BP	negative regulation of tumor necrosis factor biosynthetic process
4.23E-02	15	GO:0097190	BP	apoptotic signaling pathway
4.94E-21	13	GO:0008626	BP	granzyme-mediated apoptotic signaling pathway
5.09E-04	3	GO:0002765	BP	immune response-inhibiting signal transduction
2.57E-04	3	GO:0002767	BP	immune response-inhibiting cell surface receptor signaling pathway
1.04E-04	3	GO:0002774	BP	Fc receptor mediated inhibitory signaling pathway
7.22E-05	6	GO:0060416	BP	response to growth hormone
4.34E-05	5	GO:0071378	BP	cellular response to growth hormone stimulus
1.19E-02	6	GO:0071384	BP	cellular response to corticosteroid stimulus
8.88E-03	6	GO:0071385	BP	cellular response to glucocorticoid stimulus
6.66E-03	39	GO:0002376	BP	immune system process
5.07E-05	29	GO:0006955	BP	immune response
2.23E-02	16	GO:0009617	BP	response to bacterium
1.89E-02	13	GO:0002237	BP	response to molecule of bacterial origin
1.35E-02	13	GO:0032496	BP	response to lipopolysaccharide
1.88E-02	3	GO:0046629	BP	gamma-delta T cell activation
2.61E-05	3	GO:0002290	BP	gamma-delta T cell activation involved in immune response
6.87E-03	3	GO:0046643	BP	regulation of gamma-delta T cell activation
2.61E-05	3	GO:2001191	BP	regulation of gamma-delta T cell activation involved in immune response
2.95E-03	3	GO:0046645	BP	positive regulation of gamma-delta T cell activation
2.61E-05	3	GO:2001193	BP	positive regulation of gamma-delta T cell activation involved in immune response
1.07E-02	3	GO:0036037	BP	CD8-positive, alpha-beta T cell activation
2.09E-03	3	GO:2001185	BP	regulation of CD8-positive, alpha-beta T cell activation
3.91E-02	3	GO:0046636	BP	negative regulation of alpha-beta T cell activation
2.61E-05	3	GO:2001186	BP	negative regulation of CD8-positive, alpha-beta T cell activation
3.27E-02	4	GO:0032613	BP	interleukin-10 production
2.54E-02	4	GO:0032653	BP	regulation of interleukin-10 production
9.85E-04	4	GO:0032693	BP	negative regulation of interleukin-10 production
2.95E-03	3	GO:0072608	BP	interleukin-10 secretion
2.09E-03	3	GO:2001179	BP	regulation of interleukin-10 secretion
2.57E-04	3	GO:2001180	BP	negative regulation of interleukin-10 secretion
3.83E-02	4	GO:0032615	BP	interleukin-12 production
3.27E-02	4	GO:0032655	BP	regulation of interleukin-12 production
1.88E-02	3	GO:0032695	BP	negative regulation of interleukin-12 production
5.09E-04	3	GO:0072610	BP	interleukin-12 secretion
2.57E-04	3	GO:2001182	BP	regulation of interleukin-12 secretion
1.04E-04	3	GO:2001183	BP	negative regulation of interleukin-12 secretion
5.00E-02	3	GO:0002446	BP	neutrophil mediated immunity
1.40E-03	3	GO:0097048	BP	dendritic cell apoptotic process
1.40E-03	3	GO:2000668	BP	regulation of dendritic cell apoptotic process
2.57E-04	3	GO:2000669	BP	negative regulation of dendritic cell apoptotic process
2.09E-03	3	GO:0002291	BP	T cell activation via T cell receptor contact with antigen bound to MHC molecule on antigen presenting cell
5.09E-04	3	GO:2001188	BP	regulation of T cell activation via T cell receptor contact with antigen bound to MHC molecule on antigen presenting cell
1.04E-04	3	GO:2001189	BP	negative regulation of T cell activation via T cell receptor contact with antigen bound to MHC molecule on antigen presenting cell
1.38E-02	2	GO:1902969	BP	mitotic DNA replication

4.73E-02	2	GO:0006273	BP	lagging strand elongation
3.40E-02	2	GO:0033567	BP	DNA replication, Okazaki fragment processing
2.34E-03	2	GO:1903461	BP	Okazaki fragment processing involved in mitotic DNA replication
3.40E-02	2	GO:0043137	BP	DNA replication, removal of RNA primer
2.34E-03	2	GO:1903469	BP	removal of RNA primer involved in mitotic DNA replication
9.41E-11	8	GO:0006063	BP	uronic acid metabolic process
9.41E-11	8	GO:0019585	BP	glucuronate metabolic process
5.03E-11	8	GO:0052695	BP	cellular glucuronidation
1.18E-09	8	GO:0009812	BP	flavonoid metabolic process
2.54E-11	8	GO:0009813	BP	flavonoid biosynthetic process
2.54E-11	8	GO:0052696	BP	flavonoid glucuronidation
2.32E-02	4	GO:0045806	BP	negative regulation of endocytosis
5.04E-04	5	GO:0019835	BP	cytolysis
4.03E-03	3	GO:0042268	BP	regulation of cytolysis
5.09E-04	3	GO:0045919	BP	positive regulation of cytolysis
2.03E-04	37	GO:0006508	BP	proteolysis
1.43E-08	19	GO:0051604	BP	protein maturation
4.82E-09	19	GO:0016485	BP	protein processing
1.31E-02	3	GO:1903010	BP	regulation of bone development
8.84E-04	3	GO:1903011	BP	negative regulation of bone development
6.32E-03	4	GO:0098751	BP	bone cell development
3.48E-02	5	GO:1902106	BP	negative regulation of leukocyte differentiation
2.77E-02	4	GO:0097028	BP	dendritic cell differentiation
1.40E-03	3	GO:2001198	BP	regulation of dendritic cell differentiation
5.09E-04	3	GO:2001199	BP	negative regulation of dendritic cell differentiation
1.00E-02	6	GO:0030316	BP	osteoclast differentiation
1.07E-02	3	GO:0036035	BP	osteoclast development
3.00E-02	6	GO:0002761	BP	regulation of myeloid leukocyte differentiation
1.06E-02	5	GO:0045670	BP	regulation of osteoclast differentiation
1.93E-02	4	GO:0002762	BP	negative regulation of myeloid leukocyte differentiation
9.85E-04	4	GO:0045671	BP	negative regulation of osteoclast differentiation
2.09E-03	3	GO:2001204	BP	regulation of osteoclast development
2.61E-05	3	GO:2001205	BP	negative regulation of osteoclast development
2.61E-05	3	GO:0038044	BP	transforming growth factor-beta secretion
2.61E-05	3	GO:2001201	BP	regulation of transforming growth factor-beta secretion
2.61E-05	3	GO:2001202	BP	negative regulation of transforming growth factor-beta secretion
4.73E-02	2	GO:0002002	BP	regulation of angiotensin levels in blood
5.40E-03	8	GO:0044070	BP	regulation of anion transport
1.94E-03	5	GO:1903792	BP	negative regulation of anion transport
3.91E-02	3	GO:0006837	BP	serotonin transport
1.58E-02	3	GO:0001820	BP	serotonin secretion
5.33E-03	3	GO:0014062	BP	regulation of serotonin secretion
1.40E-03	3	GO:0014063	BP	negative regulation of serotonin secretion
6.87E-03	3	GO:0018879	BP	biphenyl metabolic process
1.88E-02	3	GO:0042178	BP	xenobiotic catabolic process
1.04E-04	3	GO:0070980	BP	biphenyl catabolic process
2.29E-02	2	GO:0006069	BP	ethanol oxidation
3.40E-02	2	GO:0044194	CC	cytolytic granule
7.32E-03	10	GO:0061134	MF	peptidase regulator activity
2.30E-03	10	GO:0061135	MF	endopeptidase regulator activity
3.47E-02	11	GO:0004857	MF	enzyme inhibitor activity

2.49E-03	10	GO:0030414	MF	peptidase inhibitor activity
1.59E-03	10	GO:0004866	MF	endopeptidase inhibitor activity
2.54E-03	6	GO:0004869	MF	cysteine-type endopeptidase inhibitor activity
1.31E-02	3	GO:0008157	MF	protein phosphatase 1 binding
2.21E-02	3	GO:0042287	MF	MHC protein binding
6.87E-03	3	GO:0042288	MF	MHC class I protein binding
1.31E-02	3	GO:0030553	MF	cGMP binding
2.29E-02	2	GO:0035276	MF	ethanol binding
1.38E-02	2	GO:0004027	MF	alcohol sulfotransferase activity
1.89E-03	11	GO:0016757	MF	transferase activity, transferring glycosyl groups
6.72E-06	10	GO:0008194	MF	UDP-glycosyltransferase activity
3.21E-04	10	GO:0016758	MF	transferase activity, transferring hexosyl groups
3.87E-09	8	GO:0015020	MF	glucuronosyltransferase activity
2.31E-03	85	GO:0003824	MF	catalytic activity
3.40E-02	2	GO:0004022	MF	alcohol dehydrogenase (NAD) activity
1.38E-02	2	GO:0004024	MF	alcohol dehydrogenase activity, zinc-dependent
3.40E-02	2	GO:0047760	MF	butyrate-CoA ligase activity
2.57E-04	3	GO:0032393	MF	MHC class I receptor activity
2.61E-05	3	GO:0032396	MF	inhibitory MHC class I receptor activity
2.61E-05	3	GO:0030109	MF	HLA-B specific inhibitory MHC class I receptor activity
2.61E-05	3	GO:0030107	MF	HLA-A specific inhibitory MHC class I receptor activity
2.60E-05	23	GO:0008233	MF	peptidase activity
1.20E-05	23	GO:0070011	MF	peptidase activity, acting on L-amino acid peptides
2.46E-07	22	GO:0004175	MF	endopeptidase activity
2.65E-13	22	GO:0017171	MF	serine hydrolase activity
1.83E-13	22	GO:0008236	MF	serine-type peptidase activity
2.83E-14	22	GO:0004252	MF	serine-type endopeptidase activity
1.38E-02	2	GO:0008240	MF	tripeptidyl-peptidase activity
3.40E-02	2	GO:0004690	MF	cyclic nucleotide-dependent protein kinase activity
2.34E-03	2	GO:0004692	MF	cGMP-dependent protein kinase activity
2.34E-03	2	GO:0016427	MF	tRNA (cytosine) methyltransferase activity
2.34E-03	2	GO:0016428	MF	tRNA (cytosine-5-)-methyltransferase activity
3.40E-02	2	GO:0016886	MF	ligase activity, forming phosphoric ester bonds
1.38E-02	2	GO:0003909	MF	DNA ligase activity
1.38E-02	2	GO:0003910	MF	DNA ligase (ATP) activity
3.77E-02	3	HP:0003401	hp	Paresthesia
2.62E-02	1	HP:0012147	hp	Reduced quantity of Von Willebrand factor
2.62E-02	1	HP:0004910	hp	Bicarbonate-wasting renal tubular acidosis
2.62E-02	1	HP:0005429	hp	Recurrent systemic pyogenic infections
5.00E-02	5	HP:0002815	hp	Abnormality of the knees
8.36E-03	4	HP:0002857	hp	Genu valgum
2.62E-02	1	HP:0005546	hp	Increased red cell osmotic resistance
2.62E-02	1	HP:0007868	hp	Age-related macular degeneration
4.78E-02	2	HP:0002999	hp	Patellar dislocation
2.62E-02	1	HP:0004953	hp	Abdominal aortic aneurysm
2.62E-02	1	HP:0011872	hp	Impaired thrombin-induced platelet aggregation
2.62E-02	1	HP:0200071	hp	Peripheral vitreoretinal degeneration
1.66E-02	3	KEGG:04614	ke	Renin-angiotensin system
7.06E-06	8	KEGG:00830	ke	Retinol metabolism
2.10E-06	9	KEGG:05204	ke	Chemical carcinogenesis
4.68E-02	4	KEGG:05332	ke	Graft-versus-host disease

5.98E-06	6	KEGG:00040	ke	Pentose and glucuronate interconversions
1.44E-05	6	KEGG:00860	ke	Porphyrin and chlorophyll metabolism
2.93E-02	3	KEGG:03430	ke	Mismatch repair
3.70E-06	8	KEGG:00982	ke	Drug metabolism - cytochrome P450
1.05E-04	7	KEGG:00140	ke	Steroid hormone biosynthesis
4.07E-07	6	KEGG:00053	ke	Ascorbate and aldarate metabolism
1.85E-07	9	KEGG:00980	ke	Metabolism of xenobiotics by cytochrome P450
6.90E-05	6	KEGG:00500	ke	Starch and sucrose metabolism
8.01E-05	6	KEGG:00983	ke	Drug metabolism - other enzymes
5.00E-02	4	KEGG:05330	ke	Allograft rejection

Notes:

OR: Olfactory receptor

VOM: Vomeronasal receptor

Supplementary Table S15. Supplementary Table S15. Functional categories of high level gene differentiation between Chinese and Africa wild Brown rats detected by F_{ST} (except OR and VOM)

P-value	N	GO ID	term ID	Description
8.95E-03	3	GO:0071394	BP	cellular response to testosterone stimulus
6.83E-03	2	GO:1903469	BP	removal of RNA primer involved in mitotic DNA replication
6.83E-03	2	GO:1903461	BP	Okazaki fragment processing involved in mitotic DNA replication
1.82E-07	11	GO:0019882	BP	antigen processing and presentation
2.46E-03	6	GO:0048002	BP	antigen processing and presentation of peptide antigen
2.26E-04	6	GO:0002474	BP	antigen processing and presentation of peptide antigen via MHC class I
2.32E-04	25	GO:0006955	BP	immune response
2.51E-02	4	GO:0060416	BP	response to growth hormone
5.00E-02	3	GO:0071378	BP	cellular response to growth hormone stimulus
1.49E-04	4	GO:0006063	BP	uronic acid metabolic process
1.49E-04	4	GO:0019585	BP	glucuronate metabolic process
9.54E-05	4	GO:0052695	BP	cellular glucuronidation
1.58E-05	5	GO:0009812	BP	flavonoid metabolic process
5.76E-05	4	GO:0009813	BP	flavonoid biosynthetic process
5.76E-05	4	GO:0052696	BP	flavonoid glucuronidation
2.04E-02	2	GO:0070980	BP	biphenyl catabolic process
3.55E-02	3	GO:0071361	BP	cellular response to ethanol
1.24E-02	5	GO:0002228	BP	natural killer cell mediated immunity
4.51E-02	4	GO:0002715	BP	regulation of natural killer cell mediated immunity
1.05E-02	5	GO:0042267	BP	natural killer cell mediated cytotoxicity
4.51E-02	4	GO:0042269	BP	regulation of natural killer cell mediated cytotoxicity
2.26E-04	6	GO:0042611	CC	MHC protein complex
2.41E-05	6	GO:0042612	CC	MHC class I protein complex
1.66E-03	4	GO:0015020	MF	glucuronosyltransferase activity
1.11E-02	6	GO:0003823	MF	antigen binding
1.50E-04	6	GO:0042605	MF	peptide antigen binding
2.14E-07	5	GO:0046703	MF	natural killer cell lectin-like receptor binding
4.05E-02	2	GO:0003909	MF	DNA ligase activity
4.05E-02	2	GO:0003910	MF	DNA ligase (ATP) activity
5.00E-02	2	HP:0009063	hp	Progressive distal muscle weakness
1.51E-02	2	HP:0008376	hp	Nasal, dysarthric speech
5.00E-02	2	HP:0003438	hp	Absent Achilles reflex
4.90E-02	3	HP:0002136	hp	Broad-based gait
5.00E-02	3	KEGG:00980	ke	Metabolism of xenobiotics by cytochrome P450
1.52E-02	3	KEGG:00500	ke	Starch and sucrose metabolism
3.92E-04	5	KEGG:04940	ke	Type I diabetes mellitus
1.39E-03	3	KEGG:00053	ke	Ascorbate and aldarate metabolism
1.04E-02	3	KEGG:00860	ke	Porphyrin and chlorophyll metabolism
1.44E-03	5	KEGG:04612	ke	Antigen processing and presentation
3.05E-02	5	KEGG:04514	ke	Cell adhesion molecules (CAMs)
9.48E-04	5	KEGG:05416	ke	Viral myocarditis
8.00E-03	4	KEGG:00830	ke	Retinol metabolism
3.92E-02	6	KEGG:04144	ke	Endocytosis
1.02E-03	7	KEGG:04145	ke	Phagosome
1.62E-04	5	KEGG:05332	ke	Graft-versus-host disease
1.96E-02	3	KEGG:00983	ke	Drug metabolism - other enzymes

4.53E-03	3	KEGG:00040	ke	Pentose and glucuronate interconversions
2.39E-05	6	KEGG:05320	ke	Autoimmune thyroid disease
4.30E-04	8	KEGG:05168	ke	Herpes simplex infection
9.44E-06	6	KEGG:05330	ke	Allograft rejection

Notes:

OR: Olfactory receptor

VOM: Vomeronasal receptor

Reference

- Adkins RM, Gelke EL, Rowe D, et al. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. *Molecular biology and evolution* 18: 777-791
- Adkins RM, Walton AH, Honeycutt RL 2003. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Molecular phylogenetics and evolution* 26: 409-420
- Akaike H 1974. A new look at the statistical model identification. *Automatic Control IEEE Transactions on* 19: 716-723
- Alfaro ME, Santini F, Brock C, et al. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences* 106: 13410-13414
- Alhajeri BH, Hunt OJ, Steppan SJ 2015. Molecular systematics of gerbils and deomyines (Rodentia: Gerbillinae, Deomyinae) and a test of desert adaptation in the tympanic bulla. *Journal of Zoological Systematics and Evolutionary Research* 53: 312-330
- Anderson S, Jones JK, Others 1967. Recent mammals of the world: a synopsis of families. : 453
- Babb PL, Fernandez-Duque E, Schurr TG 2010. AVPR1A sequence variation in monogamous owl monkeys (*Aotus azarai*) and its implications for the evolution of platyrrhine social behavior. *Journal of molecular evolution* 71: 279-297
- Barnett SA 2002. The Story of Rats: Their Impact on Us, and Our Impact on Them. *Allen & Unwin Press*
- Blair JE, Shah P, Hedges SB 2005. Evolutionary sequence analysis of complete eukaryote genomes. *BMC bioinformatics* 6: 1
- CHEVRET P, VEYRUNES F, BRITTON-DAVIDIAN J 2005. Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. *Biological Journal of the Linnean Society* 84: 417-427
- Chiaromonte F, Yap VB, Miller W 2002. Scoring pairwise genomic sequence alignments. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*: 115-126
- Crottini A, Madsen O, Poux C, et al. 2012. Vertebrate time-tree elucidates the biogeographic pattern of a major biotic change around the K-T boundary in Madagascar. *Proceedings of the National Academy of Sciences* 109: 5358-5363
- Deinum EE, Halligan DL, Ness RW, et al. 2015. Recent Evolution in *Rattus norvegicus* Is Shaped by Declining Effective Population Size. *Molecular biology and evolution* 32: 2547-2558
- Delsuc F, Vizcaíno SF, Douzery EJP 2004. Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. *BMC Evolutionary Biology* 4: 1
- dos Reis M, Inoue J, Hasegawa M, et al. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society of London B: Biological Sciences* 279: 3491-3500
- Douady CJ, Douzery EJP 2003. Molecular estimation of eulipotyphlan divergence times and the

- evolution of “Insectivora”. *Molecular phylogenetics and evolution* 28: 285-296
- Douzery EJP, Delsuc F, Stanhope MJ, et al. 2003. Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals and incompatibility among fossil calibrations. *Journal of molecular evolution* 57: S201--S213
- Douzery EJP, Huchon D 2004. Rabbits, if anything, are likely Glires. *Molecular phylogenetics and evolution* 33: 922-935
- Easteal S, Herbert G 1997. Molecular evidence from the nuclear genome for the time frame of human evolution. *Journal of molecular evolution* 44: S121--S132
- Excoffier L, Dupanloup I, Huerta-Sanchez E, et al. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet* 9: e1003905
- Fabre P-H, Chaval Y, Mortelliti A, et al. 2015. Molecular phylogeny of South-East Asian arboreal murine rodents. *Zoologica Scripta* 45: 349-364
- Fabre P-H, Hautier L, Dimitrov D, et al. 2012. A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evolutionary Biology* 12: 1
- Fabre P-H, Musser GG, Fitriana YS, et al. 2013. A new genus of rodent from Wallacea (Rodentia: Muridae: Murinae: Rattini), and its implication for biogeography and Indo-Pacific Rattini systematics. *Zoological Journal of the Linnean Society* 169: 408-447
- Gibbs RA, Weinstock GM, Metzker ML, et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428: 493-521
- Gutenkunst RN, Hernandez RD, Williamson SH, et al. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS genetics* 5: e1000695
- Hallström BM, Janke A 2008. Resolution among major placental mammal interordinal relationships with genome data imply that speciation influenced their earliest radiations. *BMC evolutionary biology* 8: 1
- Hallström BM, Kullberg M, Nilsson MA, et al. 2007. Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. *Molecular biology and evolution* 24: 2059-2068
- Hasegawa M, Thorne JL, Kishino H 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes & genetic systems* 78: 267-283
- Hedges SB, Dudley J, Kumar S 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22: 2971-2972
- Hedges SB, Marin J, Suleski M, et al. 2015. Tree of life reveals clock-like speciation and diversification. *Molecular biology and evolution* 32: 835-845
- Holmes EC 1991. Different rates of substitution may produce different phylogenies of the eutherian mammals. *Journal of molecular evolution* 33: 209-215
- Huchon D, Chevret P, Jordan U, et al. 2007. Multiple molecular evidences for a living mammalian fossil. *Proceedings of the National Academy of Sciences* 104: 7495-7499
- Hugall AF, Foster R, Lee MSY 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic biology* 56: 543-563
- Jameson NM, Hou Z-C, Sterner KN, et al. 2011. Genomic data reject the hypothesis of a prosimian primate clade. *Journal of Human Evolution* 61: 295-305

- Kent WJ, Baertsch R, Hinrichs A, et al. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci USA* 100: 11484-11489
- Kitano T, Oota S, Saitou N 1999. Molecular evolutionary analyses of the Rh blood group genes and Rh50 genes in mammals. *ZOOLOGICAL STUDIES-TAIPEI* 38: 379-386
- Kullberg M, Nilsson MA, Arnason U, et al. 2006. Housekeeping genes for phylogenetic analysis of eutherian relationships. *Molecular biology and evolution* 23: 1493-1503
- Kumar S, Hedges SB 1998. A molecular timescale for vertebrate evolution. *Nature* 392: 917-920
- Kumar S, Hedges SB 2011. TimeTree2: species divergence times on the iPhone. *Bioinformatics* 27: 2023-2024
- Lecompte E, Aplin K, Denys C, et al. 2008. Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. *BMC evolutionary biology* 8: 199
- Li H, Stephan W 2006. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet* 2: e166
- Michaux JR, Chevret P, Filippucci M-G, et al. 2002. Phylogeny of the genus *Apodemus* with a special emphasis on the subgenus *Sylvaemus* using the nuclear IRBP gene and two mitochondrial markers: cytochrome b and 12S rRNA. *Molecular phylogenetics and evolution* 23: 123-136
- Misawa K, Janke A 2003. Revisiting the Glires concept—phylogenetic analysis of nuclear sequences. *Molecular phylogenetics and evolution* 28: 320-327
- Murphy WJ, Pringle TH, Crider TA, et al. 2007. Using genomic data to unravel the root of the placental mammal. *Genome research* 17: 413-421
- Nei M, Glazko GV 2002. The Wilhelmine E. Key 2001 Invitational Lecture. Estimation of divergence times for a few mammalian and several primate species. *Journal of Heredity* 93: 157-164
- Nei M, Xu P, Glazko G 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proceedings of the National Academy of Sciences* 98: 2497-2502
- Neumann K, Michaux J, Lebedev V, et al. 2006. Molecular phylogeny of the Cricetinae subfamily based on the mitochondrial cytochrome b and 12S rRNA genes and the nuclear vWF gene. *Molecular phylogenetics and evolution* 39: 135-148
- Nyakatura K, Bininda-Emonds ORP 2012. Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. *BMC biology* 10: 1
- O'hUigin C, Li W-H 1992. The molecular clock ticks regularly in muroid rodents and hamsters. *Journal of Molecular Evolution* 35: 377-384
- Pages M, Chevret P, Gros-Balthazard M, et al. 2012. Paleogenetic analyses reveal unsuspected phylogenetic affinities between mice and the extinct *Malpaisomys insularis*, an endemic rodent of the Canaries. *PloS one* 7: e31123
- Pisano J, Condamine FL, Lebedev V, et al. 2015. Out of Himalaya: the impact of past Asian environmental changes on the evolutionary and biogeographical history of Dipodoidea (Rodentia). *Journal of Biogeography* 42: 856-870
- Poux C, Chevret P, Huchon D, et al. 2006. Arrival and diversification of caviomorph rodents and

- platyrrhine primates in South America. *Systematic Biology* 55: 228-244
- Pyron RA 2010. A likelihood method for assessing molecular divergence time estimates and the placement of fossil calibrations. *Systematic Biology* 59: 185-194
- Renaud S, Chevret P, Michaux J 2007. Morphological vs. molecular evolution: ecology and phylogeny both shape the mandible of rodents. *Zoologica Scripta* 36: 525-535
- Rowe KC, Reno ML, Richmond DM, et al. 2008. Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): multilocus systematics of the old endemic rodents (Muroidea: Murinae). *Molecular phylogenetics and evolution* 47: 84-101
- Schenk JJ, Rowe KC, Stepan SJ 2013. Ecological opportunity and incumbency in the diversification of repeated continental colonizations by muroid rodents. *Systematic Biology* 62: 837-864
- Schwartz S, Kent WJ, Smit A, et al. 2003. Human-mouse alignments with BLASTZ. *Genome research* 13: 103-107
- Song Y, Lan Z, Kohn MH 2014. Mitochondrial DNA phylogeography of the Norway rat. *PLoS One* 9: e88425
- Springer MS, Murphy WJ, Eizirik E, et al. 2003. Placental mammal diversification and the Cretaceous--Tertiary boundary. *Proceedings of the National Academy of Sciences* 100: 1056-1061
- Stepan SJ, Adkins RM, Anderson J 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Systematic biology* 53: 533-553
- Wu, Wang 2012. Fossil materials and migrations of *Mus musculus* and *Rattus norvegicus*. *Research of China's Frontier Archaeology 1: 1-9*
- Wu S, Wu W, Zhang F, et al. 2012. Molecular and paleontological evidence for a post-Cretaceous origin of rodents. *PloS one* 7: e46445
- Zhang Q, Xia L, Kimura Y, et al. 2013. Tracing the origin and diversification of Dipodoidea (Order: Rodentia): Evidence from fossil record and molecular phylogeny. *Evolutionary Biology* 40: 32-44