



HAL
open science

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

Lin Zeng, Chen Ming, Yan Li, Ling-Yan Su, Yan-Hua Su, Newton O. Otecko,
Ambroise Dalecky, Stephen Donnellan, Ken Aplin, Xiao-Hui Liu, et al.

► **To cite this version:**

Lin Zeng, Chen Ming, Yan Li, Ling-Yan Su, Yan-Hua Su, et al.. Out of Southern East Asia of the Brown Rat Revealed by Large-Scale Genome Sequencing. *Molecular Biology and Evolution*, 2018, 35 (1), pp.149-158. 10.1093/molbev/msx276 . hal-02627003

HAL Id: hal-02627003

<https://hal.inrae.fr/hal-02627003v1>

Submitted on 14 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

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

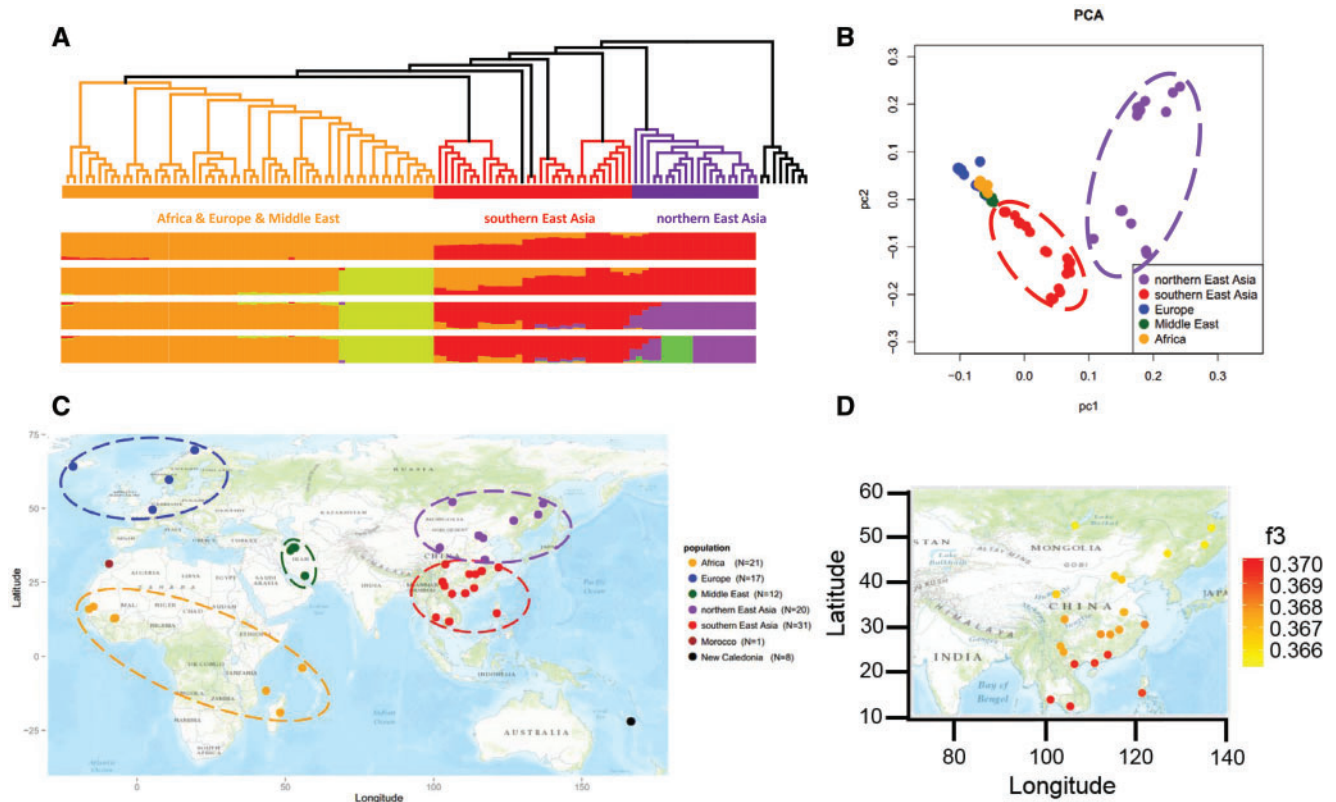


Fig. 1. Out of southern East Asia origin of wild brown rats. (A) Phylogenetic neighbor-joining tree and Bayesian clustering analysis by ADMIXTURE. Black colored lines represent Outgroup. (B) PCA analysis. From the PCA plot, two dispersal waves out of southern East Asia are presented. One wave is to non-East 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

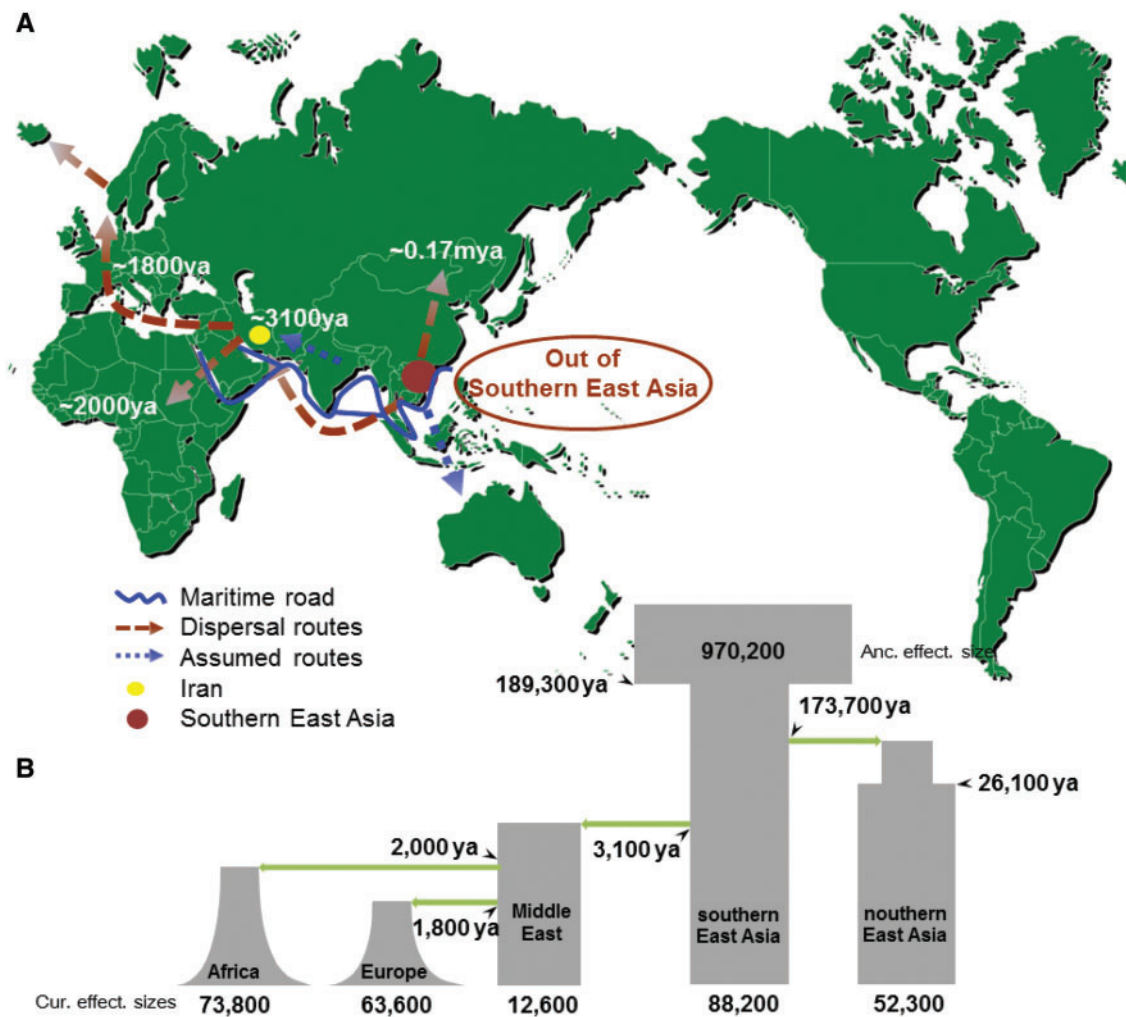


Fig. 2. Dispersal routes and demographic histories of wild brown rats. (A) Proposed dispersal routes of the wild brown rats based on our analyses. (B) The inferred joint demographic model of different wild brown rat populations based on the maximum likelihood method.

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

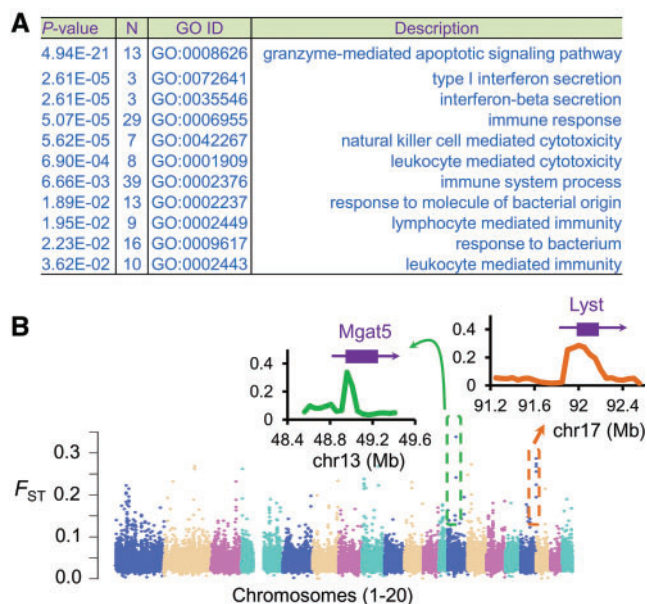


FIG. 3. Immune response genes are under selection in the wild brown rats. (A) List of genes with significantly high level of population differentiation between European and Chinese rats are enriched for immune system related functions. Gene Ontology analysis of the protein-coding genes was conducted 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 $f_3(\text{outgroup}; A, B)$ 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

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

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-Asian 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.bigac.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, Kernysky 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.