PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Puckett EE *et al*. 2016 Global population divergence and admixture of the brown rat (*Rattus norvegicus*). *Proc. R. Soc. B* **283**: 20161762.
http://dx.doi.org/10.1098/rspb.2016.1762

Received: 9 August 2016
Accepted: 26 September 2016

Subject Areas:
evolution, genomics

Keywords:
commensal, invasive species, population genomics, cityscapes, phylogeography, RAD-Seq

Authors for correspondence:
Emily E. Puckett
e-mail: emily.e.puckett@gmail.com
Jason Munshi-South
e-mail: jmunshisouth@fordham.edu

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3500433.

THE ROYAL SOCIETY PUBLISHING

Global population divergence and admixture of the brown rat (*Rattus norvegicus*)

Emily E. Puckett[1], Jane Park[1], Matthew Combs[1], Michael J. Blum[2], Juliet E. Bryant[3], Adalgisa Caccone[4], Federico Costa[5], Eva E. Deinum[6,7], Alexandra Esther[8], Chelsea G. Himsworth[9], Peter D. Keightley[6], Albert Ko[10], Åke Lundkvist[11], Lorraine M. McElhinney[12], Serge Morand[13], Judith Robins[14,15], James Russell[15,16], Tanja M. Strand[11], Olga Suarez[17], Lisa Yon[18] and Jason Munshi-South[1]

[1]Louis Calder Center, Biological Field Station, Fordham University, Armonk, NY 10504, USA
[2]Xavier Center for Bioenvironmental Research, Tulane University, New Orleans, LA 70112, USA
[3]Clinical Research Unit, Oxford University, Hanoi, Vietnam
[4]Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, New Haven, CT 06520-8106, USA
[5]Instituto de Saúde Coletiva, Universidade Federal da Bahia, Salvador, Brazil
[6]Ashworth Laboratories, Institute of Evolutionary Biology, University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3FL, UK
[7]Mathematical and Statistical Methods Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[8]Federal Research Centre for Cultivated Plants, Institute for Plant Protection in Horticulture and Forests, Vertebrate Research, Julius Kühn Institute, Münster, Germany
[9]Animal Health Centre, British Columbia Ministry of Agriculture, 1767 Angus Campbell Road, Abbotsford, British Columbia, Canada V3G 2M3
[10]Laboratory of Epidemiology and Public Health, Yale University, New Haven, CT, USA
[11]Department of Medical Biochemistry and Microbiology, Zoonosis Science Center, Uppsala University, Uppsala, Sweden
[12]Wildlife Zoonoses and Vector Borne Disease Research Group, Animal and Plant Health Agency (APHA), Woodham Lane, New Haw Surrey, UK
[13]CNRS-CIRAD, Centre d'Infectiologie Christophe Mérieux du Laos, Vientiane, Lao PDR
[14]Department of Anthropology, [15]School of Biological Sciences, and [16]Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand
[17]Laboratorio de Ecologia de Roedores Urbanos, IEGEBA-CONICET, EGE-Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires Pabellon II, Ciudad Universitaria (C1428EHA), Buenos Aires, Argentina
[18]School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK

EEP, 0000-0002-9325-4629; JM-S, 0000-0002-8067-4341

Native to China and Mongolia, the brown rat (*Rattus norvegicus*) now enjoys a worldwide distribution. While black rats and the house mouse tracked the regional development of human agricultural settlements, brown rats did not appear in Europe until the 1500s, suggesting their range expansion was a response to relatively recent increases in global trade. We inferred the global phylogeography of brown rats using 32 k SNPs, and detected 13 evolutionary clusters within five expansion routes. One cluster arose following a southward expansion into Southeast Asia. Three additional clusters arose from two independent eastward expansions: one expansion from Russia to the Aleutian Archipelago, and a second to western North America. Westward expansion resulted in the colonization of Europe from which subsequent rapid colonization of Africa, the Americas and Australasia occurred, and multiple evolutionary clusters were detected. An astonishing degree of fine-grained clustering between and within sampling sites underscored the extent to which urban heterogeneity shaped genetic structure of commensal rodents. Surprisingly, few individuals were recent migrants, suggesting that recruitment into established populations is limited. Understanding the global

population structure of *R. norvegicus* offers novel perspectives on the forces driving the spread of zoonotic disease, and aids in development of rat eradication programmes.

1. Introduction

The development of agriculture and resultant transition from nomadic to sedentary human societies created new ecological niches for species to evolve commensal or parasitic relationships with humans [1]. The phylogeographic history of species living in close association with people often mirrors global patterns of human exploration [2,3] and colonization [4–7]. In particular, commensal rodent distributions have been strongly influenced by the movement of humans around the world. Three rodent species, the house mouse (*Mus musculus*), black rat (*Rattus rattus*) and brown rat (*Rattus norvegicus*) are the most populous and successful invasive mammals, having colonized most of the global habitats occupied by humans [8]. The least is known about genomic diversity and patterns of colonization in brown rats, including whether a history of commensalism resulted in population divergence, and if so, at what spatial scales. Our lack of knowledge of the ecology and evolution of the brown rat is striking given that brown rats are responsible for an estimated $19 billion of damage annually [9]. Understanding the evolutionary trajectories of brown rats is also a prerequisite to elucidating the processes that resulted in a successful global invasion, including adaptations to a variety of climates and anthropogenic stressors.

We inferred the global routes of brown rat expansion, population differentiation and admixture using a dense, genome-wide nuclear dataset, a first for a commensal rodent [10]. A previous mitochondrial study identified the centre of origin [11] but did not resolve relationships among invasive populations. That work, in combination with fossil distributions [12], suggested that brown rats originated in the colder climates of northern China and Mongolia before expanding across central and western Asia, possibly through human settlements associated with Silk Road trade routes. Based on historical records, brown rats became established in Europe by the 1500s and were introduced to North America by the 1750s [13]. Brown rats now occupy nearly every major landmass (outside of polar regions), and human-assisted colonization of islands remains a constant threat to insular fauna [14].

Elucidating global brown rat phylogeographic patterns has several important implications. First, the spread of brown rats may illuminate patterns of human connectivity via trade, or unexpected movement patterns as observed in other commensal rodents [2]. Second, rats are hosts to many zoonotic diseases (e.g. *Leptospira interrogans*, Seoul hantavirus, etc.); understanding the distribution of genomic backgrounds may provide insights into differential disease susceptibilities. Additionally, an understanding of contemporary population structure in rats may elucidate source and sink areas for disease transmission. Third, brown rat eradication programmes occur in urban areas to decrease disease transmission and on islands where rats prey upon native fauna. A comprehensive understanding of global population structure will allow for better design of eradication efforts, particularly for understanding how to limit new invasions. Thus, our aim was to test biological hypotheses developed from an understanding of the historical narrative of spread using phylogeographic inference. We estimated the number of distinct clusters around the world, the genomic contribution of these clusters within invaded areas, and whether genetic drift and/or post-colonization admixture elicits evolutionary divergence from source populations.

2. Material and methods

We obtained rat tissue samples from field-trapped specimens, museum or institute collections and wildlife markets (electronic supplementary material, tables S1 and S2). As GPS coordinates for individuals were not always available, the sampling location was recorded as either the city, nearest town or island where rats were collected. Samples were genotyped using ddRAD-Seq [15], then missing genotype and relatedness filters were applied (see electronic supplementary material, Methods for details) for a final nuclear dataset containing 32 127 single nucleotide polymorphisms (SNPs) genotyped in 314 individuals, and a mitochondrial dataset with 115 SNPs (electronic supplementary material, table S3) that comprised 104 haplotypes in 144 individuals.

(a) Population genomic analyses

To describe population structure, we ran ADMIXTURE v. 1.23 [16] at each cluster from 1 to 40; given that a single SNP per RAD-tag was retained, we met the criterion for unlinked data for this analysis. Given the known effects of sampling bias on clustering analyses, we repeated this analysis with a subset of the data where four or five samples from each city were randomly selected ($n = 158$). The CV error was the lowest for $K = 14$ clusters, which supported the analysis of our full dataset. We also subdivided the full dataset into the Asian and non-Asian clusters and reran ADMIXTURE at each cluster from 1 to 25. We used the CV error values to identify the best-supported clustering patterns across the range. Using the same datasets (full, Asian and non-Asian), we ran principal components analysis (PCA) in EIGENSOFT v. 5.0.2 [17] and identified significant PCs using Tracy–Widom statistics.

We also estimated evolutionary clusters using FINESTRUCTURE v. 2.0.7 [18], which elucidates the finest grained clusters by accounting for linkage disequilibrium and allows detailed admixture inference based upon the pairwise co-ancestry coefficients. We limited this analysis to the 20 autosomes (31 489 SNPs), removing SNPs on unassembled scaffolds in the dataset. Data for each chromosome were phased and imputed using FASTPHASE v. 1.2 [19]. Initial analyses using the linked model indicated our data were effectively unlinked (c-factor 0.0104); therefore, we ran the unlinked model. We used default settings except for the following parameters: 25% of the data were used for initial EM estimation; 750 000 iterations of the MCMC were run (375 000 of which were burn-in) with 1000 samples retained, 20 000 tree comparisons and 500 000 steps of the tree maximization were run. We viewed MCMC trace files to confirm the stability of all parameters.

To understand the patterns of population divergence, we ran TREEMIX v. 1.12 [20]. As the *R. rattus* data (see electronic supplementary material, Methods) were mapped to the *R. norvegicus* genome, we extracted SNPs at the same genomic positions for 31 black rats (we removed two samples showing admixture; electronic supplementary material, figure S1) with the SAMTOOLS v. 1.2 [21] mpileup function using a position list. We selected the sampling location with the largest sample size from each of the well-supported clusters at $K = 13$ (figure 1 and electronic supplementary material, figure S2), plus the *R. rattus* samples for the outgroup (which were not subdivided owing to lack of population structure, electronic supplementary material, figure S3). We added migration edges to the population tree sequentially by fixing the population tree to the tree with $n - 1$ migration edges, where blocks of 1000 SNPs and the sample size correction were enabled. We assessed both the proportion of variance (electronic supplementary material, figure S4*a*) and the residuals of the population tree (electronic supplementary material, figure S4*b*) and chose the model with three migration edges. We decided to thin the sampling areas owing to uneven

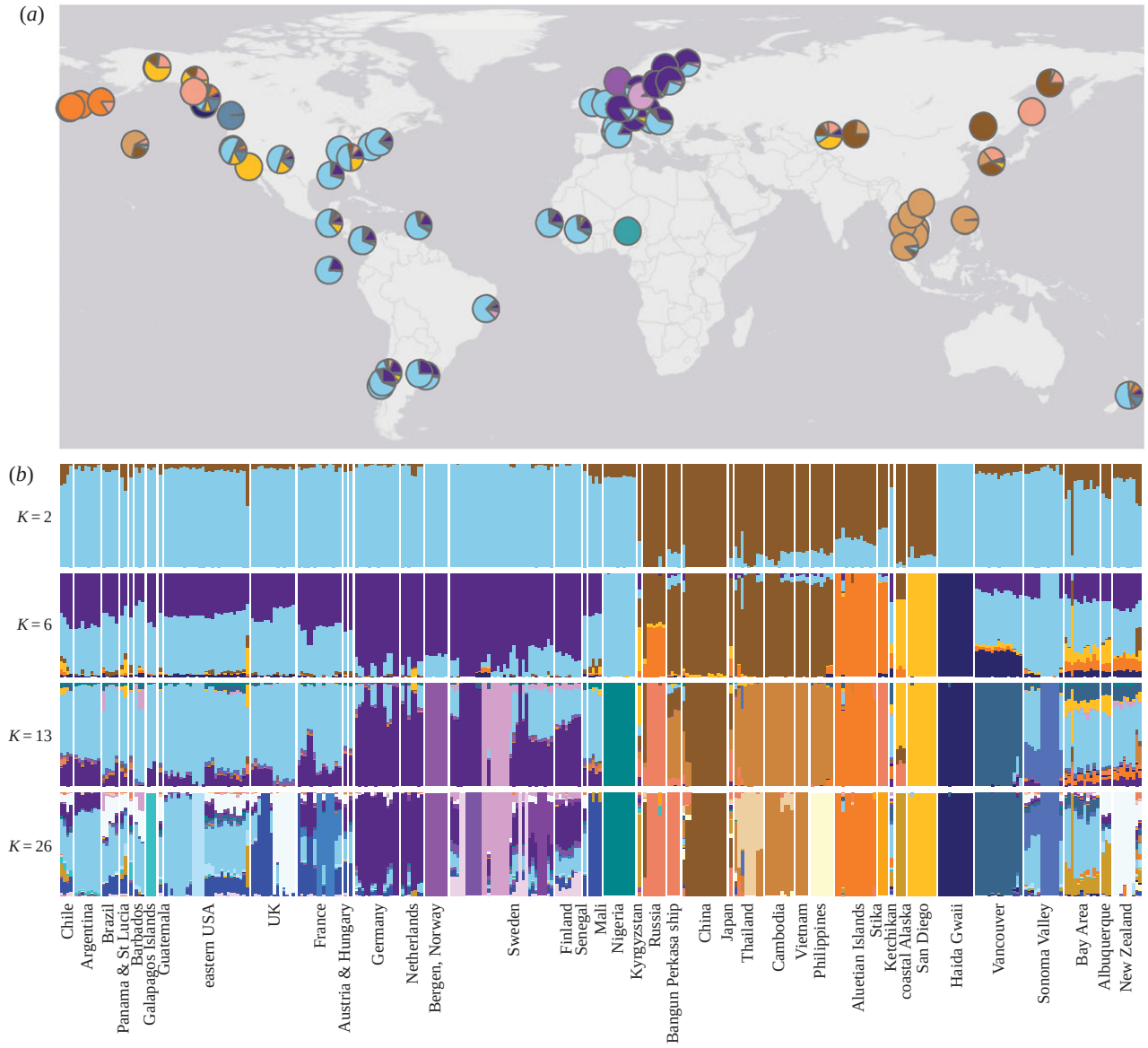

Figure 1. (*a*) Map of brown rat sampling locations with average proportion of ancestry per site inferred using 32 k nuclear SNPs. Ancestry was based on ADMIXTURE estimates from 13 clusters (China: brown; SE Asia: light brown; Russia: pink; Aleutian Archipelago: orange; western North America: gold; W Euro: light blue; N Euro: purple; Kano: turquoise; Sonoma Valley: medium blue; Haida Gwaii: dark blue; Vancouver: cerulean; Bergen: medium purple; Malmo: light purple). (*b*) Ancestry proportions from ADMIXTURE for 314 samples at two, six, 13 and 26 clusters.

sampling between the broad Asian and non-Asian clusters; both factors should affect the variance in the model, thus we presented a potentially underfit versus overfit model. We ran f_3 tests within TREEMIX and observed no significant relationships, likely owing to highly complex admixture patterns [17].

For the nuclear dataset, we calculated expected heterozygosity (H_E) and F_{IS} within each of the 13 clusters using ARLEQUIN v. 3.5.1.3 [22], and pairwise F_{ST} using VCFTOOLS v. 0.1.13 and the Weir and Cockerham estimator [23,24]. For the mitochondrial dataset, we calculated pairwise F_{ST} between the clusters identified in the nuclear dataset in ARLEQUIN.

3. Results and discussion

(a) Evolutionary clustering

(i) Nuclear genome

Our analyses of 314 rats using 32 127 SNPs identified multiple hierarchical levels of evolutionary clustering (K). PCA distinguished two clusters along the first PC, an Asian cluster that extended to western North America, and a non-Asian cluster found in Europe, Africa, the Americas and New Zealand (electronic supplementary material, figure S5). Higher dimension PCA axes distinguished subclusters (electronic supplementary material, figure S6), then individual sampling sites; in total, 58 axes of variation were significant using Tracy–Widom statistics (20 and 37 axes were significant for PCAs with only Asian or non-Asian samples, respectively). Using the model-based clustering program ADMIXTURE, the Asian and non-Asian clusters divided into five and eight subclusters, respectively (figures 1 and 2; electronic supplementary material, figures S2, S7, S8). Higher numbers of clusters ($K = 18$, 20 and 26) were also supported by ADMIXTURE (electronic supplementary material, figures S2*a* and S7), distinguishing ever finer spatial scales from subcontinents to cities.

The subclusters in the Asian cluster reflect underlying geography and hierarchical differentiation (electronic supplementary material, figure S2*b*). The predominant four clusters reflected differentiation between: China, Southeast (SE) Asia, the Aleutian Archipelago and Western North

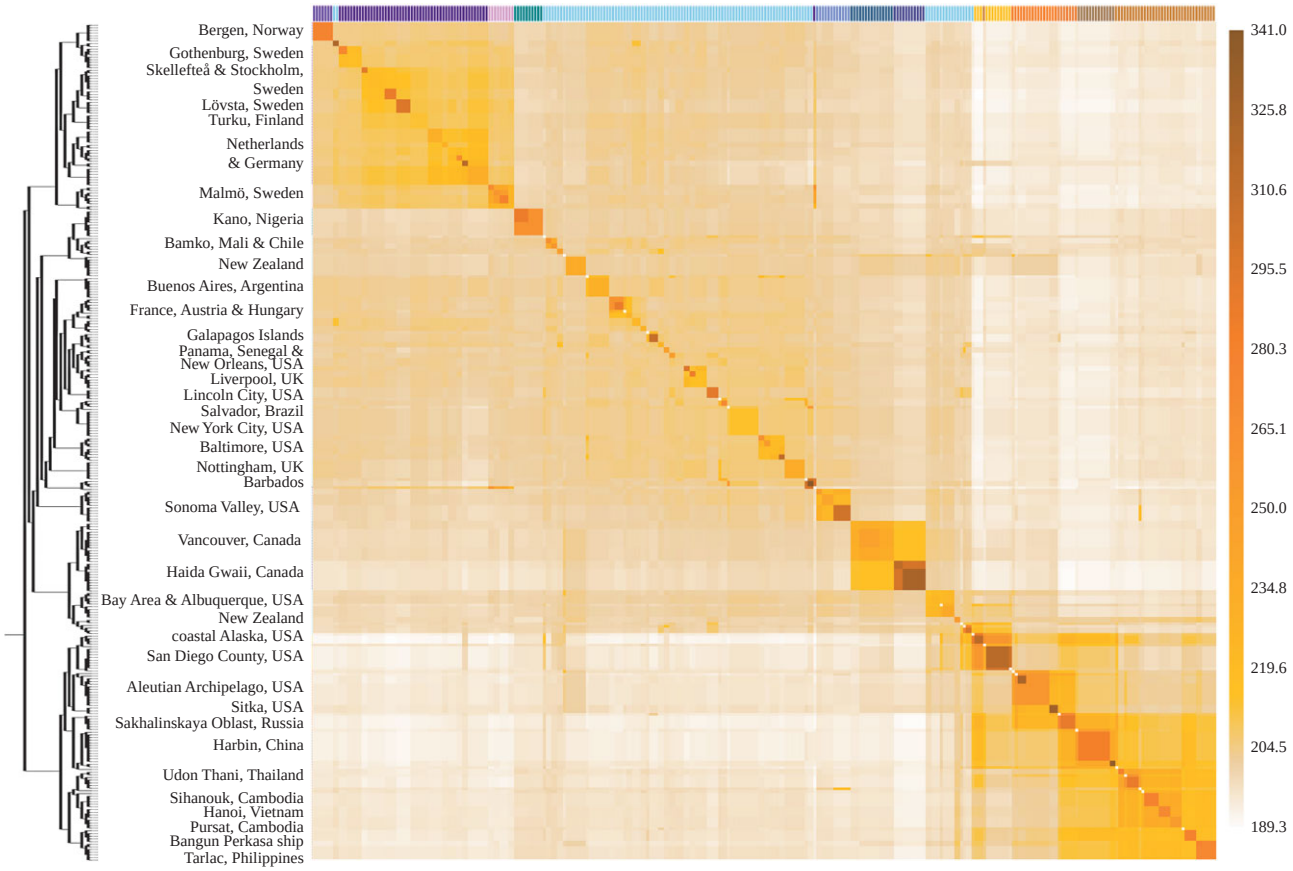

Figure 2. Co-ancestry heat map of brown rats, where light and dark brown, respectively, denote lower and higher co-ancestry. The 101 populations identified by FINESTRUCTURE appear along the diagonal. A bifurcating tree and select sampling locations are shown on the left, and assignment to one of the 13 clusters from figure 1 shown on top.

America (electronic supplementary material, figures S9 and S10). Within the SE Asia cluster, further subdivision was observed for both the Philippines and Thailand (figure 1 and electronic supplementary material, figure S10). Within the Aleutian Archipelago cluster, samples from the city of Sitka (in the Alexander Archipelago) formed a subcluster. Rats from the Russian city of Sakhalinskaya Oblast and four rats aboard the Bangun Perkasa ship each formed a subcluster (electronic supplementary material, figure S10). The Bangun Perkasa was a nationless vessel seized in the Pacific Ocean by the US government in 2012 for illegal fishing. Our analyses identified that the rats aboard were of SE Asian origin and likely represented a city in that region, probably one bordering the South China Sea, at which the ship originated or docked.

We detected greater hierarchical differentiation in the non-Asian cluster (electronic supplementary material, figure S2*c*). At $K = 3$, we observed divergence between the Western Europe (W Euro) and Northern Europe (N Euro) clusters (electronic supplementary material, figure S12). The W Euro cluster contained rats from Europe (UK, France, Austria and Hungary), Central and South America (Argentina, Brazil, Chile, Galapagos Islands, Honduras, Guatemala and Panama), the Caribbean (Barbados, Saint Lucia), North America (eastern, central and western USA and Canada), New Zealand and Africa (Senegal and Mali); and the N Euro cluster included Norway, Sweden, Finland, Germany and the Netherlands (figure 1; electronic supplementary material, figures S7, S11, S12). Within these broad geographical regions, many subclusters were identified by ADMIXTURE that likely resulted from either intense founder effects, isolation resulting in genetic drift, the inclusion of second- and third-order relatives in the dataset, or a combination of these factors. In the global analysis, four clusters were nested within W Euro (the island of Haida Gwaii, Canada; Vancouver, Canada; Kano, Nigeria and Sonoma County in the western USA) and two within N Euro (Bergen, Norway; Malmo, Sweden). We identified additional well-supported subclusters within the non-Asian cluster at $K = 12$, 15, and 17 that represented individual cities (electronic supplementary material, figure S12).

Our analysis using FINESTRUCTURE identified 101 clusters (figure 2). Of the 39 cities where more than one individual was sampled, 19 cities supported multiple clusters indicating genetic differentiation within cities. As GPS coordinates were not collected, we cannot hypothesize whether these clusters represent distinct populations or were artefacts of sampling relatives, despite removal of individuals with relatedness coefficients greater than 0.20, although the FINESTRUCTURE algorithm should be robust to relatedness when identifying clusters. The Asian and N Euro sampling sites individually had higher co-ancestry coefficients between locations (figure 2) which supported the hierarchical clustering observed using ADMIXTURE.

(ii) Mitochondrial genome

We identified 10 clades within a network-based analysis of 104 mitochondrial haplotypes (figure 3 and electronic supplementary material, tables S3 and S4). Many of the clades had spatial structure concordant with the nuclear genome results (figure 3*a*). We observed clade 1 in China, Russia and western North America. Additionally, clades 6 and 9 contained a single haplotype only observed in China. We

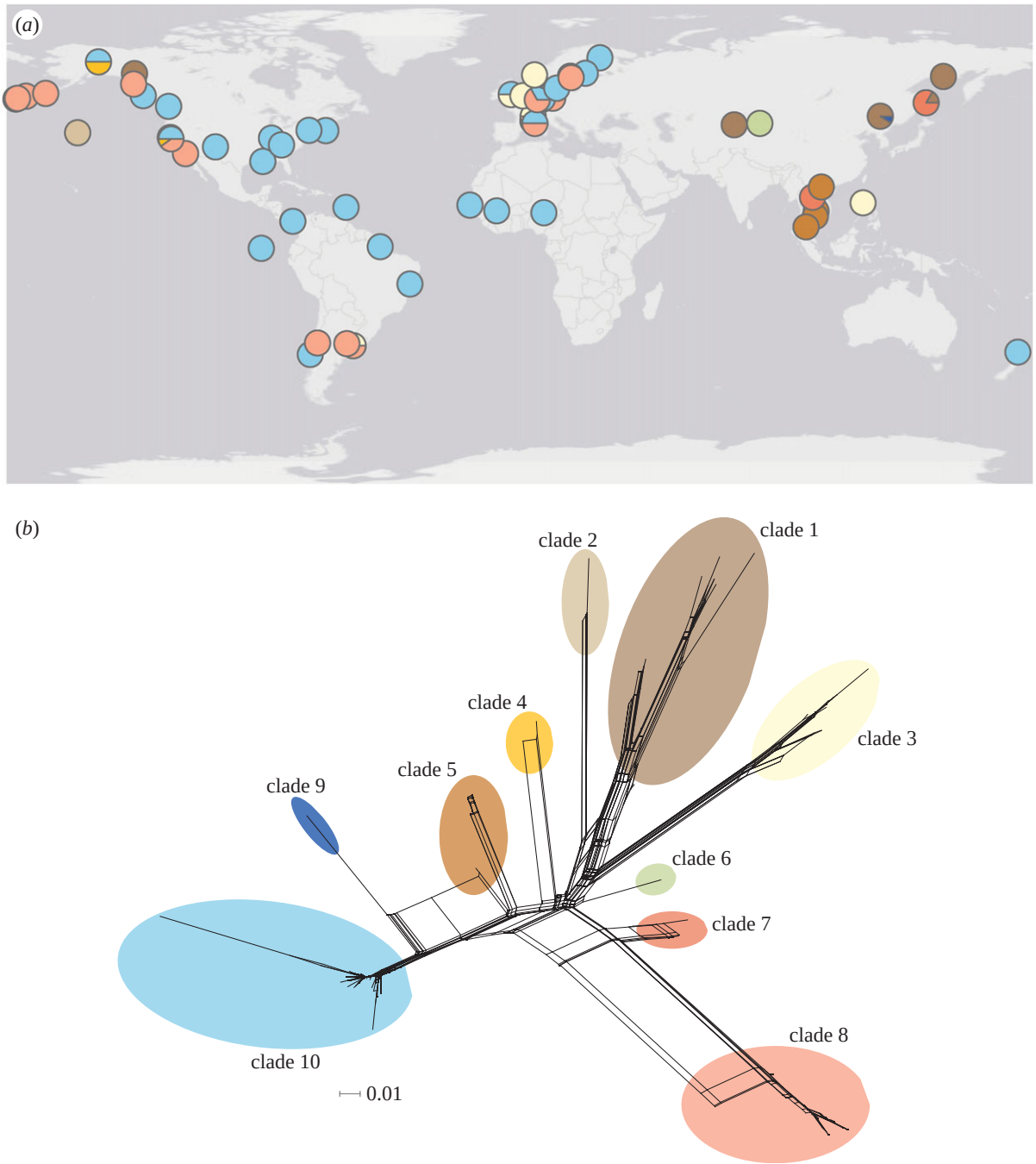

Figure 3. (*a*) Map of the proportion of mitochondrial clades at each sampling site for 144 individuals and (*b*) SNP haplotype network with 104 haplotypes in 10 clades (clade 1: brown; 2: beige; 3: pale yellow; 4: gold; 5: light brown; 6: pale green; 7: pink; 8: light pink; 9: dark blue; 10: light blue).

interpret the diversity of clades within northern China as representative of geographical structure in the ancestral range prior to movement of rats by humans (figure 3 and electronic supplementary material, table S4). In SE Asia, we observed clades 2 (aboard the Bangun Perkasa), 3 (Philippines) and 5 (Cambodia, Thailand and Vietnam). Clade 4 was found in western North America. European samples comprised three divergent clades (3, 8 and 10). Clade 8 was observed across Europe, western North America and South America; this clade shared ancestry with clade 7 that was observed in Russia and Thailand (figure 3).

(b) Range expansion

We thinned our dataset to the sampling site with the largest sample size within each of the 13 clusters supported by ADMIXTURE and analysed the data using TREEMIX. We observed divergence within Asia first, followed by the two independent expansions into western North America. Drift along the backbone of the non-Asian cluster was limited, indicating rapid expansion of rats into Africa, Europe and the Americas (electronic supplementary material, figure S13). Both the population tree topology and PCA (electronic supplementary material, figures S2, S6 and S13) indicated that range expansion occurred in three directions, where one southward and two eastward expansions comprised Asian ancestry, and the westward expansion produced the non-Asian cluster.

(i) Ancestral range

In eastern China, the nuclear genome assigned strongly to a single cluster while mitochondrial diversity encompassed two divergent clades, where samples from western China assigned to both the Chinese and SE Asian clusters and

represented a third mitochondrial clade. This result suggests substructure within the ancestral range, although the samples from northeastern China may not be representative of the ancestral range but instead of an isolated, divergent population that retained high genetic diversity (electronic supplementary material, tables S4 and S5).

(ii) Southern expansion into Southeast Asia

A southward range expansion into SE Asia was supported by the population tree topology, higher heterozygosity, low nuclear F_{ST} with China and elevated co-ancestry coefficients between populations in SE Asia, China and Russia (figure 3; electronic supplementary material, tables S5 and S6). Given evidence for an early southward expansion (electronic supplementary material, figure S10), we hypothesize that the founding of SE Asia was accompanied by a weak bottleneck resulting in relatively low loss of genetic diversity. However, following founding regional diversification occurred as we observed substructure in both the nuclear and mitochondrial genomes (figures 1 and 3; electronic supplementary material, figure S10).

(iii) Two independent eastward expansions

We observed population divergence along the first eastward expansion from eastern Russia into the Aleutian Archipelago based on PCA (electronic supplementary material, figure S9). Both the population tree topology and PCA indicate that a second eastward expansion progressed from Asia to western North America (electronic supplementary material, figures S9 and S13). While the Western North America cluster was observed in both northern and southern Pacific coast localities (electronic supplementary material, figure S8A), we cannot extrapolate that this cluster represents the entirety of the coastline. Specifically, Sitka, Ketchikan, Vancouver and the Bay Area are all located between the Alaskan cities and San Diego County that comprise the Western North America cluster. Further, the timing of these expansions is an open question. While the population tree indicated divergence of these two expansions prior to divergence of the non-Asian cluster, the historical record attributes brown rats in the Aleutian Archipelago to Russian fur traders in the 1780s [25], which is not consistent with rats entering Europe in the 1500s [13]. Thus, evidence of early divergence may be a consequence of unsampled Asian populations sharing ancestry with the Aleutian Archipelago and Western North America clusters.

(iv) Westward range expansion into Europe

The low drift along the backbone of the population tree for the non-Asian cluster was indicative of rapid westward expansion (electronic supplementary material, figure S13). Limited inferences could be drawn about western Asia and the Middle East because of sampling constraints, yet we hypothesize that the region was colonized by the range expansion of the non-Asian cluster. We observed three mitochondrial clades in Europe, where clade 3 shared ancestry with SE Asia and clade 8 shared ancestry with eastern Russia, whereas clade 10 is a European derived clade (figure 3 and electronic supplementary material, table S6). Thus, Europe may have been independently colonized three times, although the routes remain an open question. We hypothesize that clade 10 arrived overland around the Mediterranean Sea, similar to black rats [26]. We hypothesize that following the independent colonizations, the genetic backgrounds admixed prior to divergence between the N Euro and W Euro clusters given the low nuclear F_{ST} (electronic supplementary material, table S5).

Notably, we detected genetic differentiation of Bergen, Norway and Malmo, Sweden within the N Euro cluster (figure 1). This pattern suggests drift following either a strong founder effect or population isolation and limited gene flow. Isolation is likely driving the pattern observed in Bergen, which is separated from eastern Norway by mountains that are thought to limit movement of commensal rodents [27].

(c) Range expansion of rats by Europeans

We detected a fifth range expansion that can be attributed to transport by western European imperial powers (1600s–1800s) to former colonial territories (figures 1 and 2; electronic supplementary material, figures S7 and S12). For example, we observed high proportions of W Euro ancestry in samples from the North and South Islands of New Zealand, which was consistent with the introduction of brown rats by British colonists, as has also been inferred for black rats [26] and the house mouse [28]. We observed admixture on both islands (figure 2) although nuclear ancestry proportions differed between the islands with higher proportions of N Euro and Vancouver ancestry on the North Island. The South Island had higher SE Asia and Western North American ancestry (figures 1 and 2; electronic supplementary material, figure S7); these ancestry components may be attributed to the seal skin trade with southern China by sealers from the USA [29].

The samples from Nigeria and Mali formed a sister clade in FINESTRUCTURE, which likely reflects a shared history as French colonies, although Senegal fell outside of the clade (figure 2). Mali had elevated W Euro ancestry compared with Nigeria which may be a consequence of multiple introductions from European sources. South American countries exhibited a paraphyletic FINESTRUCTURE topology that was suggestive of colonization from multiple locations. This result was also supported by the presence of all three mitochondrial clades found in Europe (figure 3*a*). Further sampling from Portugal and Spain would better resolve the origins of Brazilian populations and clarify relationships of former colonies elsewhere in the world.

The complex distribution of clusters in North America was suggestive of a dynamic colonization history, including independent introductions on both the Atlantic and Pacific coasts (figure 1). We detected mtDNA haplotypes of European ancestry in eastern and central USA, whereas the Pacific seaboard harbours high mtDNA haplotype diversity from European and Asian clades (figure 3). These results were consistent with prior observations of four high-frequency mtDNA haplotypes across Alaska and continental USA, of which three were observed in east Asia and one in Europe [30]. Along the Pacific coast, cities with both Asian and non-Asian nuclear ancestry were observed (figure 1), which parallels the pattern observed in black rats [26]. Given the bicoastal introductions, it is unsurprising to observe admixture in North American cities such as the San Francisco Bay Area and Albuquerque, where each has elevated co-ancestry coefficients with Asian and non-Asian clusters (figure 2). We also observed limited eastward dispersal of Asian genotypes, although other work has found evidence of greater inland penetration [30].

Rats from Haida Gwaii off the coast of British Columbia, Canada, were consistently recovered as a separate cluster in

ADMIXTURE, and had high co-ancestry coefficients and F_{ST} with other populations (figure 2 and electronic supplementary material, table S5), indicating substantial genetic drift following colonization. Rats were introduced to Haida Gwaii in the late 1700s via Spanish and/or British mariners, and have been subject to recent, intensive eradication efforts that may have heightened genetic drift [31].

(d) Intra-urban population structure of brown rats

Brown rats exhibit population structure over a remarkably fine-grained spatial scale (figure 2); specifically, rat population structure exists at the scale of both cities and neighbourhoods. We found evidence of heterogeneity among cities as some appear to support one population, whereas others support multiple populations. For example, we detected a single population across multiple neighbourhoods in Manhattan (NYC, USA), whereas four genetic clusters (figure 2) were observed in a neighbourhood in Salvador, Brazil, a result that confirmed previous microsatellite-based analyses [32]. Although denser sampling will be needed to confirm whether these groups represent distinct populations or reflect oversampling of intracity pockets of highly related individuals, intracity clustering likely represents substructure considering the global design of our SNP dataset. Observations of highly variable intracity structure suggest the following three scenarios: first, effective population size rapidly increases after invasion, possibly driven by high urban resource levels and thus genetic drift may have a relatively weak effect on population differentiation. Second, new immigrants that arrive after initial invasion and establishment of rats in a city may be limited in their capacity to either establish new colonies or join existing colonies [33], thereby limiting ongoing gene flow from other areas owing to competitive exclusion [34]. Gene flow into colonies may also be sex-biased as females were recruited more readily than males in a 2 year behavioural study of brown rats [33]. We did observe gene flow in our dataset, including an individual matching coastal Alaska into the Bay Area and an individual with high Sonoma Valley ancestry in Thailand (figure 1*b*), thus migration owing to contemporary human-assisted movement is possible and ongoing. However, given increasing connectivity owing to trade and continual movement of invasive species [35], we expected greater variability in ancestry proportions within cities than observed (electronic supplementary material, figure S7). Third, cityscapes vary in their connectivity where some cities contain strong physical and/or environmental barriers facilitating differentiation and others do not. Identifying commonalities and differences among cityscapes with one or multiple rat populations should be a goal for understanding how rats interact with their environment, particularly in relation to the effect of landscape connectivity for pest and disease control efforts.

(e) Significance

(i) Comparative phylogeography

Commensalism has given rise to complex demographic and evolutionary histories in globally distributed rodents. One commonality between rodent phylogeographies, including this study, was the detection of geographical structure preceding the evolution of commensalism [5,26,36]. This deep structure accentuated observations of regional human-mediated range expansions, including: *M. m. domesticus* around the western Mediterranean and into central Europe [4]; the eastward expansion of *M. m. castaneus* from the Indian subcontinent into SE Asia and Japan [5] and westward movement of *R. rattus* lineage I from the Indian subcontinent into western Asia and Europe [26]. Brown and black rats have both been transported longer distances, including overland transport from Asia into Europe [26]. The European matrilines were then transported overseas to the Americas, Oceania and Africa [26,37]. Mitochondrial clades representative of both European and Asian diversity were observed in North America for both species indicative of multiple invasion routes [26,30]. However, our nuclear data from western USA and New Zealand identified admixture between Asian and non-Asian evolutionary clusters in brown rats although mitochondrial haplotype diversity was largely from Europe. The extent of evolutionary clustering within the nuclear genome and admixture in black rats remains an open question.

(ii) Understanding the spread of zoonotic pathogens

Understanding the global population structure of brown rats offers novel perspectives on the forces driving the spread of zoonotic disease. Our inference that competitive exclusion may limit entry into established populations, also observed in the house mouse [38], helps explain why zoonotic pathogens do not always exhibit the same spatial distribution as rat hosts as well as the patchy distribution of presumably ubiquitous pathogens within and between cities [39]. While within-colony transmission of disease and natal dispersal between colonies are important factors related to the prevalence of zoonotic disease, our results also suggest that contemporary human-aided transport of infected rats does not contribute to the global spread of pathogens, as we would expect higher variability of ancestry proportions within cities if rats were successfully migrating between cities. Additionally, our results indicate that rats with different genomic backgrounds may have variable susceptibilities to pathogens, though differential susceptibility likely depends on concordance between the geographical origins of pathogens and rats. While this idea needs pathogen-specific testing, it could have substantial implications for global disease transmission.

(iii) Rat eradication programmes for species conservation

Eradication of invasive *Rattus* species on islands and in ecosystems with high biodiversity is a priority for conservation of at-risk species, as rats outcompete or kill native fauna. It remains challenging to gauge the success of eradication programmes, because it is difficult to distinguish between post-intervention survival and reproduction as opposed to recolonization by new immigrants [40]. Understanding fine-scale population genetic structure using dense nuclear marker sets [41], as in this study, would allow managers to more clearly assess outcomes and next steps following an eradication campaign. For example, genomic analyses could illustrate that an area has been recolonized by immigration from specific source populations, thereby allowing managers to shift efforts towards biosecurity to reduce the likelihood of establishment by limiting the influx of potential immigrants.

Data accessibility. Illumina reads: NCBI SRA accession PRJNA344413. Nuclear and mitochondrial SNPs used in this study are available in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.jb3tc [42].

Authors' contributions. E.E.P. and J.M.-S. designed the research, analysed data and wrote the manuscript; J.P. conducted the laboratory work;

M.C., M.J.B., J.E.B., A.C., F.C., E.E.D., A.E., C.G.H., P.D.K., A.K., A.L., L.M.M., S.M., J.R., J.R., T.M.S., O.S., L.Y. and J.M.-S. contributed samples; J.M.-S. obtained funding; all authors edited the draft and final versions of the manuscript.

Competing interests. We have no competing interests.

Funding. This work was supported by the National Science Foundation (grants nos DEB 1457523 and DBI 1531639), and a Fordham University faculty research grant to J.M.-S.

Acknowledgements. We thank two anonymous reviewers for comments. We thank Kaitlin Abrams and Ian Hays for assisting with laboratory work. We thank Annette Backhans, Francois Catzeflis, Gauthier Dobigny, Carol Esson, Tim Giles, Gregory Glass, Sabra Klein, Mare Lõhmus, Patrick McClure, Frank van de Goot, Jordan Reed and his Mongrol Hoard, Richard Reynolds and the Ryders Alley Trencher-fed Society (RATS), Thomas Persson Vinnersten and colleagues at Anticimex, and partners of the Network Rodent-Borne Pathogens for collecting and providing rat samples. The mammal collections at the University of Alaska Museum of the North, Angelo State University, Berkeley Museum of Comparative Zoology, the Burke Museum at University of Washington, the Museum of Southwestern Biology and the Museum of Texas Tech University also graciously provided tissue samples.

References

1. Jones EP, Eager HM, Gabriel SI, Jóhannesdóttir F, Searle JB. 2013 Genetic tracking of mice and other bioproxies to infer human history. *Trends Genet*. **29**, 298–308. (doi:10.1016/j.tig.2012.11.011)
2. Searle JB *et al.* 2009 Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proc. R. Soc. B* **276**, 201–207. (doi:10.1098/rspb.2008.0958)
3. Gabriel SI, Mathias ML, Searle JB. 2015 Of mice and the 'age of discovery': the complex history of colonization of the Azorean archipelago by the house mouse (*Mus musculus*) as revealed by mitochondrial DNA variation. *J. Evol. Biol*. **28**, 130–145. (doi:10.1111/jeb.12550)
4. Cucchi T, Vigne J-D, Auffray J-C. 2005 First occurrence of the house mouse (*Mus musculus* domesticus Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. *Biol. J. Linn. Soc*. **84**, 429–445. (doi:10.1111/j.1095-8312.2005.00445.x)
5. Suzuki H *et al.* 2013 Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity* **111**, 375–390. (doi:10.1038/hdy.2013.60)
6. Matisoo-Smith E, Robins JH. 2004 Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat. *Proc. Natl Acad. Sci. USA* **101**, 9167–9172. (doi:10.1073/pnas.0403120101)
7. Hulme-Beaman A, Dobney K, Cucchi T, Searle JB. 2016 An ecological and evolutionary framework for commensalism in anthropogenic environments. *Trends Ecol. Evol*. **31**, 633–645. (doi:10.1016/j.tree.2016.05.001)
8. Long JL. 2003 *Introduced mammals of the world: their history, distribution and influence*. Collingwood, Australia, CSIRO Publishing.
9. Pimentel D, Lach L, Zuniga R, Morrison D. 2000 Environmental and economic costs of nonindigenous species in the United States. *Bioscience* **50**, 53–65. (doi:10.1641/0006-3568(2000)050[0053:eaecon]2.3.co;2)
10. Lack JB, Greene DU, Conroy CJ, Hamilton MJ, Braun JK, Mares MA, Van Den Bussche RA. 2012 Invasion facilitates hybridization with introgression in the *Rattus rattus* species complex. *Mol. Ecol*. **21**, 3545–3561. (doi:10.1111/j.1365-294X.2012.05620.x)
11. Song Y, Lan Z, Kohn MH. 2014 Mitochondrial DNA phylogeography of the Norway rat. *PLoS ONE* **9**, e88425. (doi:10.1371/journal.pone.0088425)
12. Smith AT, Xie Y. 2008 *A guide to the mammals of China*. Princeton, NJ: Princeton University Press.
13. Armitage P. 1993 Commensal rats in the New World, 1492–1992. *Biologist* **40**, 174–178.
14. Harper GA, Bunbury N. 2015 Invasive rats on tropical islands: their population biology and impacts on native species. *Glob. Ecol. Conserv*. **3**, 607–627. (doi:10.1016/j.gecco.2015.02.010)
15. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012 Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135. (doi:10.1371/journal.pone.0037135)
16. Alexander DH, Novembre J, Lange K. 2009 Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. **19**, 1655–1664. (doi:10.1101/gr.094052.109)
17. 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**, 1065–1093. (doi:10.1534/genetics.112.145037)
18. Lawson DJ, Hellenthal G, Myers S, Falush D. 2012 Inference of population structure using dense haplotype data. *PloS Genet*. **8**, e1002453. (doi:10.1371/journal.pgen.1002453)
19. Scheet P, Stephens M. 2006 A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet*. **78**, 629–644. (doi:10.1086/502802)
20. Pickrell JK, Pritchard JK. 2012 Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. **8**, e1002967. (doi:10.1371/journal.pgen.1002967)
21. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
22. Excoffier L, Lischer HEL. 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour*. **10**, 564–567. (doi:10.1111/j.1755-0998.2010.02847.x)
23. Danecek P *et al.* 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. (doi:10.1093/bioinformatics/btr330)
24. Weir BS, Cockerham CC. 1984 Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370. (doi:10.2307/2408641)
25. Black L. 1983 Record of maritime disasters in Russian America, part one: 1741–1799. In *Proc. Alaska Maritime Archaeology Workshop, 17–19 May, Sitka, Alaska*. Sea Grant Report 83-9, Sept. 1983, pp. 43–58. Fairbanks, AK: University of Alaska, Alaska Sea Grant College Program.
26. Aplin KP *et al.* 2011 Multiple geographic origins of commensalism and complex dispersal history of black rats. *PLoS ONE* **6**, e26357. (doi:10.1371/journal.pone.0026357)
27. Jones EP, Van Der Kooij J, Solheim R, Searle JB. 2010 Norwegian house mice (*Mus musculus musculus/domesticus*): distributions, routes of colonization and patterns of hybridization. *Mol. Ecol*. **19**, 5252–5264. (doi:10.1111/j.1365-294X.2010.04874.x)
28. Searle JB, Jamieson PM, Gündüz İ, Stevens MI, Jones EP, Gemmill CEC, King CM. 2009 The diverse origins of New Zealand house mice. *Proc. R. Soc. B* **276**, 209–217. (doi:10.1098/rspb.2008.0959)
29. King CM. 2016 How genetics, history and geography limit potential explanations of invasions by house mice *Mus musculus* in New Zealand. *Biol. Invasions* **18**, 1533–1550. (doi:10.1007/s10530-016-1099-0)
30. Lack J, Hamilton M, Braun J, Mares M, Van Den Bussche R. 2013 Comparative phylogeography of invasive *Rattus rattus* and *Rattus norvegicus* in the US reveals distinct colonization histories and dispersal. *Biol. Invasions* **15**, 1067–1087. (doi:10.1007/s10530-012-0351-5)
31. Hobson KA, Drever MC, Kaiser GW. 1999 Norway rats as predators of burrow-nesting seabirds: insights from stable isotope analyses. *J. Wildl. Manage* **63**, 14–25. (doi:10.2307/3802483)
32. Kajdacsi B *et al.* 2013 Urban population genetics of slum-dwelling rats (*Rattus norvegicus*) in Salvador, Brazil. *Mol. Ecol*. **22**, 5056–5070. (doi:10.1111/mec.12455)
33. Calhoun JB. 1962 *The ecology and sociology of the Norway rat*, 288. Bethesda, MD: US Department of Health E, and Welfare, Public Health Service.
34. Waters JM. 2011 Competitive exclusion: phylogeography's 'elephant in the room'? *Mol. Ecol*.

20, 4388–4394. (doi:10.1111/j.1365-294X.2011.05286.x)

35. Banks NC, Paini DR, Bayliss KL, Hodda M. 2015 The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecol. Lett.* **18**, 188–199. (doi:10.1111/ele.12397)
36. Tollenaere C *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. Biogeogr.* **37**, 398–410. (doi:10.1111/j.1365-2699.2009.02228.x)
37. Bastos AD *et al.* 2011 Genetic monitoring detects an overlooked cryptic species and reveals the diversity and distribution of three invasive *Rattus* congeners in South Africa. *BMC Genet.* **12**, 26. (doi:10.1186/1471-2156-12-26)
38. Hardouin EA, Chapuis JL, Stevens MI, van Vuuren JB, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D. 2010 House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evol. Biol.* **10**, 325. (doi:10.1186/1471-2148-10-325)
39. Himsworth CG, Parsons KL, Jardine C, Patrick DM. 2013 Rats, cities, people, and pathogens: a systematic review and narrative synthesis of literature regarding the ecology of rat-associated zoonoses in urban centers. *Vector borne Zoonotic Dis* (*Larchmont, NY*) **13**, 349–359. (doi:10.1089/vbz.2012.1195)
40. Piertney SB, Black A, Watt L, Christie D, Poncet S, Collins MA. 2016 Resolving patterns of population genetic and phylogeographic structure to inform control and eradication initiatives for brown rats *Rattus norvegicus* on South Georgia. *J. Appl. Ecol.* **53**, 332–339. (doi:10.1111/1365-2664.12589)
41. Robins JH, Miller SD, Russell JC, Harper GA, Fewster RM. 2016 Where did the rats of Big South Cape Island come from? *N.Z. J. Ecol.* **40**, 229–234.
42. Puckett EE, Park J, Combs M, Blum MJ, Bryant JE, Caccone A, Costa F, Deinum EE, Esther A, Himsworth CG, Keightley PD, Ko A, Lundkvist Å, McElhinney LM, Morand S, Robins J, Russell J, Strand TM, Suarez O, Yon L, Munshi-South J. 2016 Data from: Global population divergence and admixture of the brown rat (*Rattus norvegicus*). Dryad Digital Repository. (http://dx.doi.org/10.5061/dryad.jb3tc)

SUPPLEMENTAL METHODS

DNA Extraction, RAD sequencing, and SNP calling

We extracted DNA following the manufacturer's protocols using Qiagen DNeasy kits (Valencia, CA). We prepared double digest restriction-site associated DNA sequencing (ddRAD-Seq) libraries with 500-1000ng of genomic DNA from each sample and one negative control made up of water. Briefly, samples were digested with SphI and MluCI before ligation of unique barcoded adapters. We pooled 48 barcoded samples each in 10 libraries at equimolar concentrations. We then selected fragments from 340-412 bp (target = 376 bp) using a Pippin Prep (Sage Science, Beverly, MA). The size-selected pools were PCR-amplified for 10-12 cycles using Phusion PCR reagents (New England Biolabs, Ipswich, MA) and primers that added an Illumina multiplexing read index. Final libraries were checked for concentration and fragment size on a BioAnalyzer (Agilent Technologies, Santa Clara, CA), then sequenced (2 x 125bp paired-end) at the New York Genome Center across five lanes of an Illumina HiSeq 2500.

We demultiplexed the raw reads using the process_radtags script in STACKS v1.35 (Catchen, et al. 2013), then aligned reads for each individual to the *Rattus norvegicus* reference genome (Rnor_6.0) (Gibbs, et al. 2004) using Bowtie v2.2.6 (Langmead and Salzberg 2012) with default parameters. To assess the number of mismatches allowed between stacks and the minimum depth of coverage for each stack when building RADtags (-n and -m flags respectively) in STACKS, we processed two samples under a number of scenarios and compared the number of RADtags that formed as de novo loci versus those that mapped to the reference *R. norvegicus* genome. We first assessed the M parameter by holding m constant at three while varying M between two and five in the ustacks program. We observed a decrease in the undermerged RADs with increasing values of M; we selected M = 4 for both the final RAD processing and as the constant level when we allowed m to vary between two and five. We selected m = 3 to balance between removing real loci and stacks that erroneously mapped to the reference genome. In the cstacks program we assessed the number of allowed mismatches between tags (n) from zero to two. We observed little difference for this parameter between our test values and decided to use n = 2 as a conservative measure.

We initially built the STACKS catalog with all of the reference-aligned samples (n = 447) using the ref_map pipeline. Following processing, we filtered for the following: biallelic SNPs, a minor allele frequency (MAF) greater than or equal to 0.05, SNPs genotyped in 80% of samples, and only one SNP per RADtag (STACKS flag --write_single_snp); additionally, SNPs that mapped to either the Y chromosome or mitochondrial genome were removed. This dataset had 37,730 SNPs. Following sample collection and genotyping, we were informed that *R. rattus* samples had been collected in Mali; we capitalized on this by confirming the species identification for each sample using principal components analysis (PCA) in EIGENSOFT v5.0.2 (Patterson, et al. 2006; Price, et al. 2006), and ADMIXTURE v1.23 (Alexander, et al. 2009) for two clusters. We identified 33 *R. rattus* and 414 *R. norvegicus* samples (Fig. S1).

We reran ref_map using only the confirmed *R. norvegicus* samples, and filtered similarly as described above plus an additional filter to remove individuals with greater than 60% missing data. To add genotypes from 11 of the *R. norvegicus* samples collected in Harbin, China (European Nucleotide Archive ERP001276) (Deinum, et al. 2015), we mapped reads to the Rnor_6.0 genome using SAMTOOLS v1.2 (Li, et al. 2009) then extracted the SNP dataset using mpileup with a position list. We removed related individuals within, but not between, sampling sites by assessing relatedness in KING v1.4 (Manichaikul, et al. 2010). For each pair of individuals with relatedness estimators greater than 0.2, one individual was removed from the analysis (n = 22). Subsequently, we randomly thinned 14 samples from Vancouver, Canada as preliminary analyses indicated oversampling. Thus the final nuclear *R. norvegicus* dataset contained 32,127 SNPs genotyped in 314 individuals (Table S1).

From the initial processing in STACKS, we extracted the SNPs that mapped to the mitochondrial genome to produce a second dataset with 115 SNPs (see Table S5 for base pair positions within the *R. norvegicus* reference mitochondrial genome, GenBank accession AY172581.1). We extracted the same positions from the mitochondrial genomes of samples from Harbin, China. We allowed up to 35% missing data per individual and identified 103 haplotypes using COLLAPSE v1.2 (Posada 2004) in 144 individuals. We built a haplotype network using SPLITSTREE v4.13.1 (Huson and Bryant 2006) and identified the haplotypes grouped into 10 clades (Table S6).

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655-1664.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. STACKS: an analysis tool set for population genomics. Molecular Ecology 22:3124-3140.

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. Molecular Biology and Evolution 32:2547-2558.

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493-521.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23:254-267.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Meth 9:357-359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867-2873.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. Plos Genetics 2:e190.

Posada D. 2004. Collapse v1.2.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904-909.

SUPPLEMENTAL FIGURES

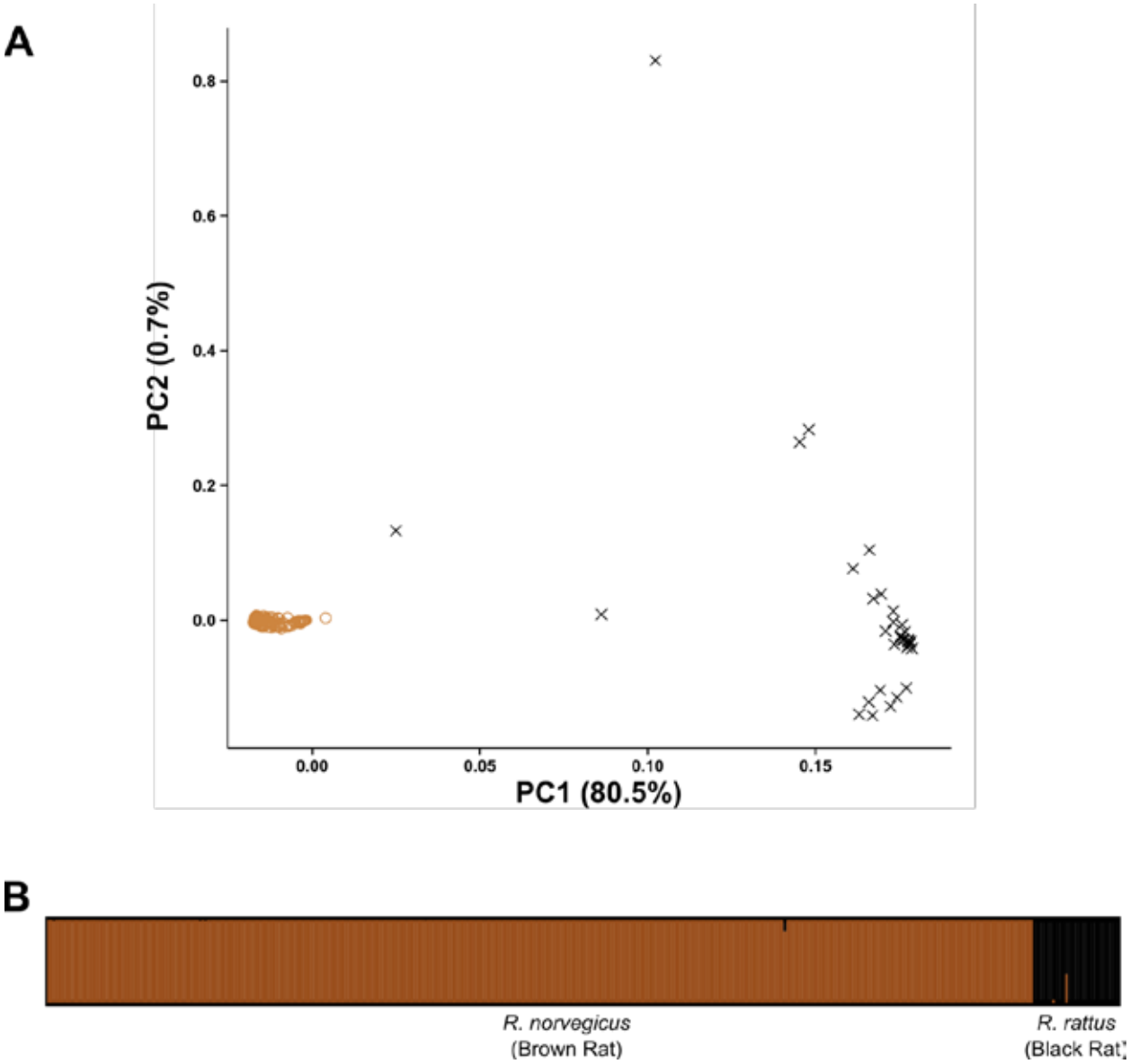

Figure S1- Identification of species as either *Rattus norvegicus* or *R. rattus* using 32k nuclear SNPs in both (A) principal components analysis (PCA; *R. norvegicus*: brown circles, *R. rattus* black Xs), and (B) ancestry proportions from ADMIXTURE.

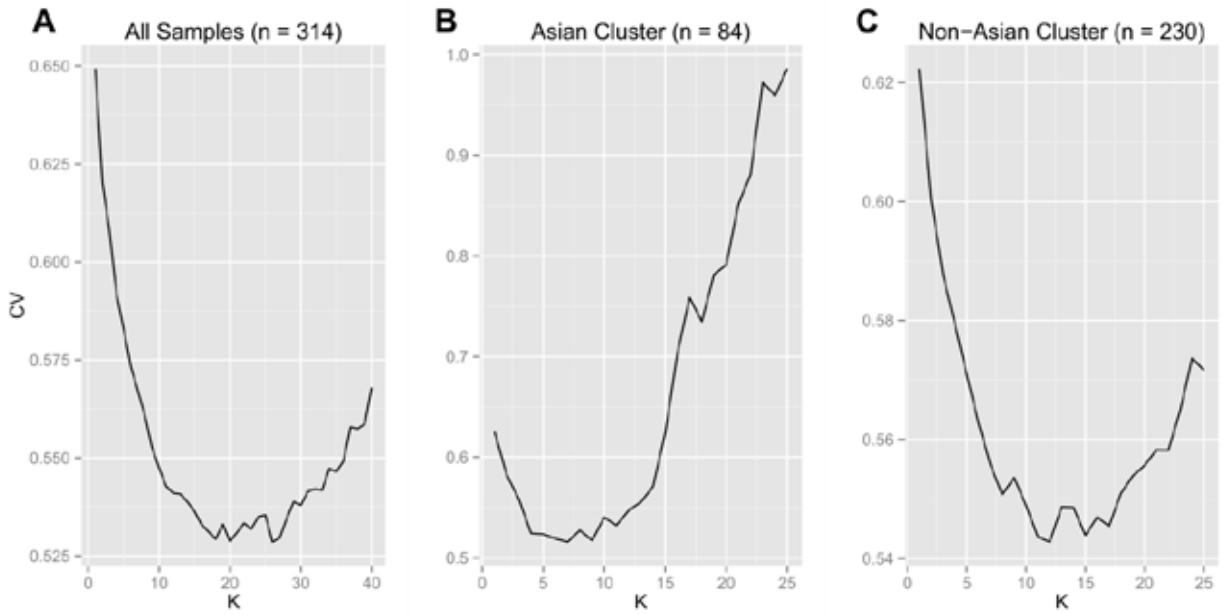

Figure S2- Cross validation error (CV) plots from ADMIXTURE for (A) all *Rattus norvegicus* samples (n = 314), (B) samples identified within the Asian cluster (n = 84), and (C) samples within the non-Asian cluster (n = 230) using datasets with 32k nuclear SNPs. Clustering from the full dataset was run from 1 to 40 clusters (K), whereas subsets of data were run 1 to 25 clusters. At each cluster, 20 repetitions of the program were run to estimate the cross validation error.

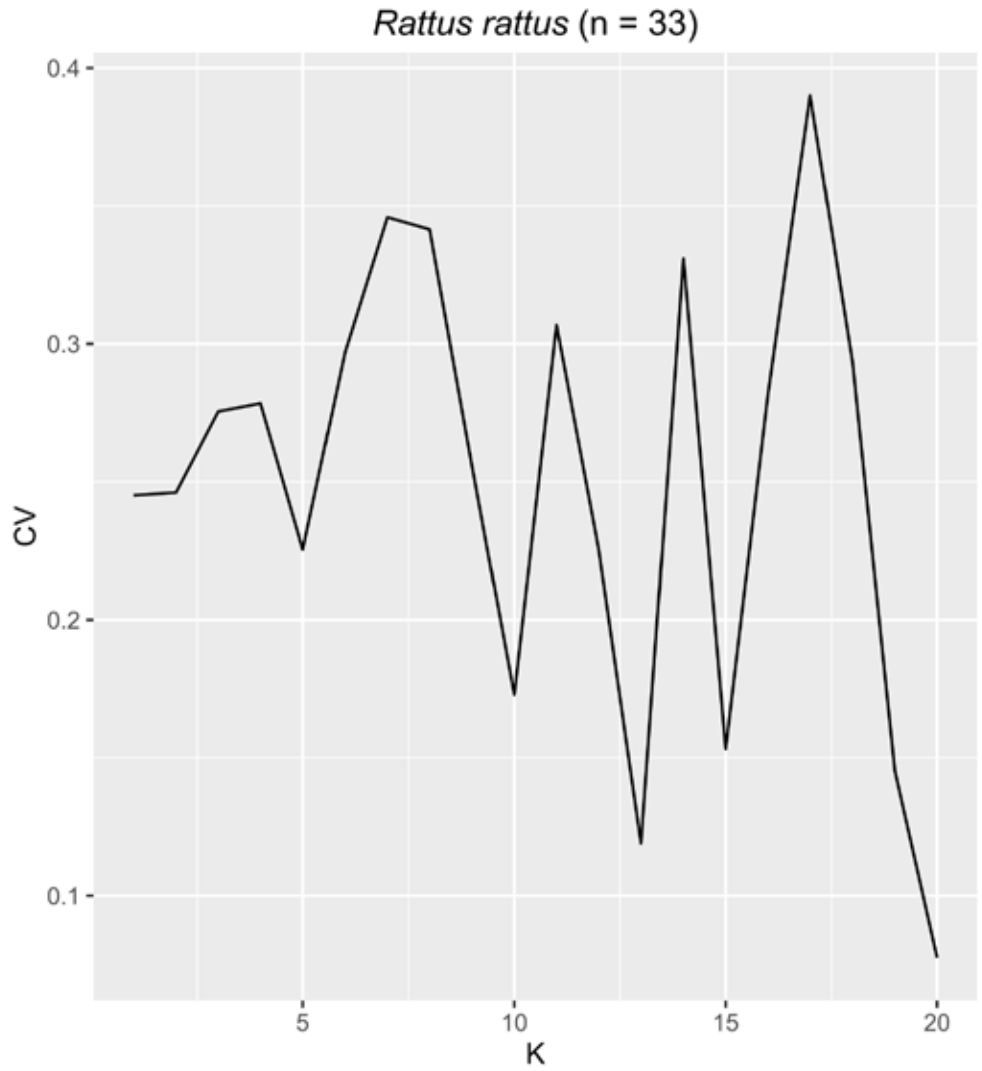

Figure S3- Cross validation error (CV) plot from ADMIXTURE for *Rattus rattus* samples (n = 33) using 32k SNPs from the nuclear genome, where 20 iterations of the program were run for cluster (K) values from 1 to 20.

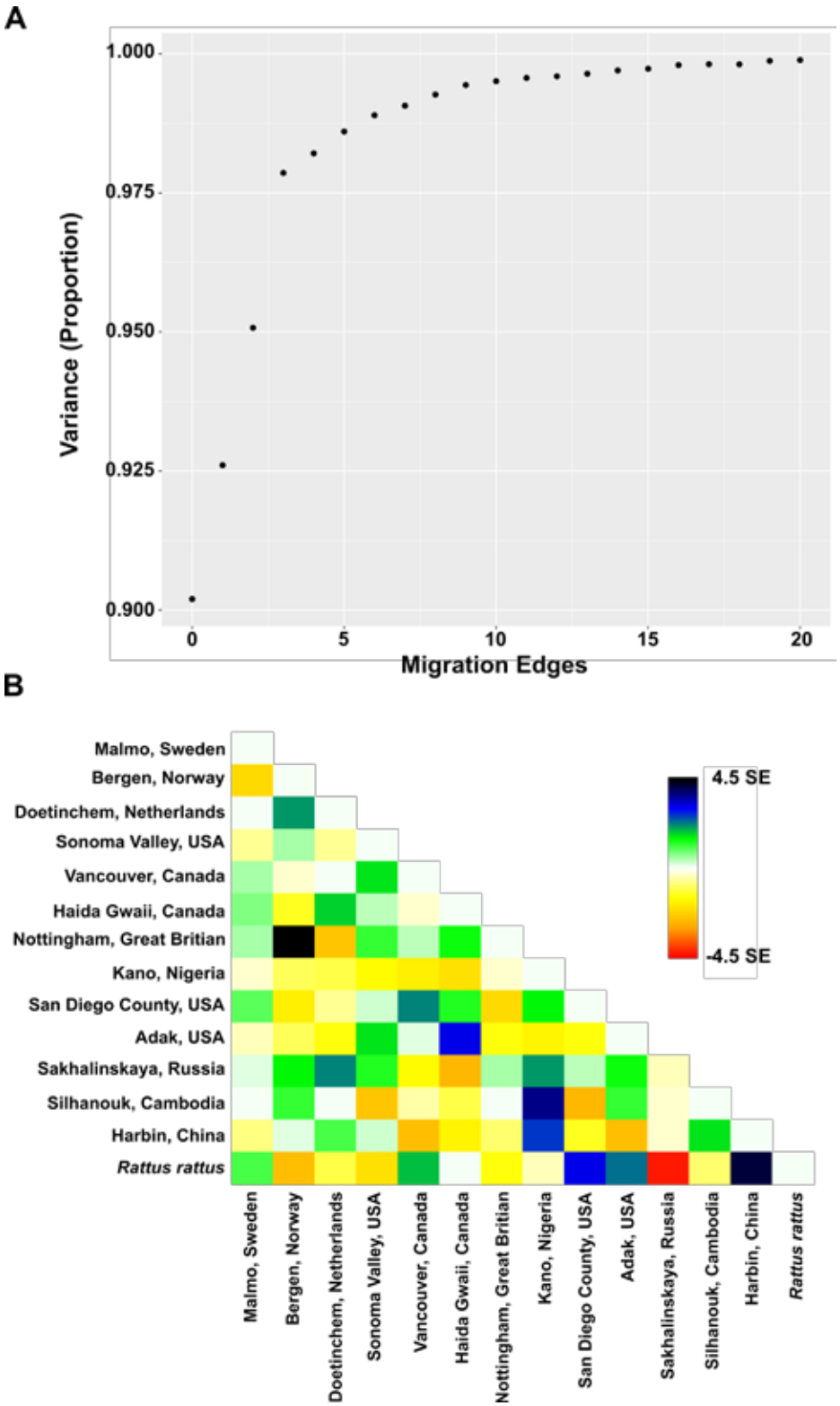

Figure S4- (A) The proportion of variance explained with increasing numbers of migration edges of a population tree with 13 sampling areas from TREEMIX. (B) The plot of pairwise residuals for the selected tree with three migration edges (Figure S3).

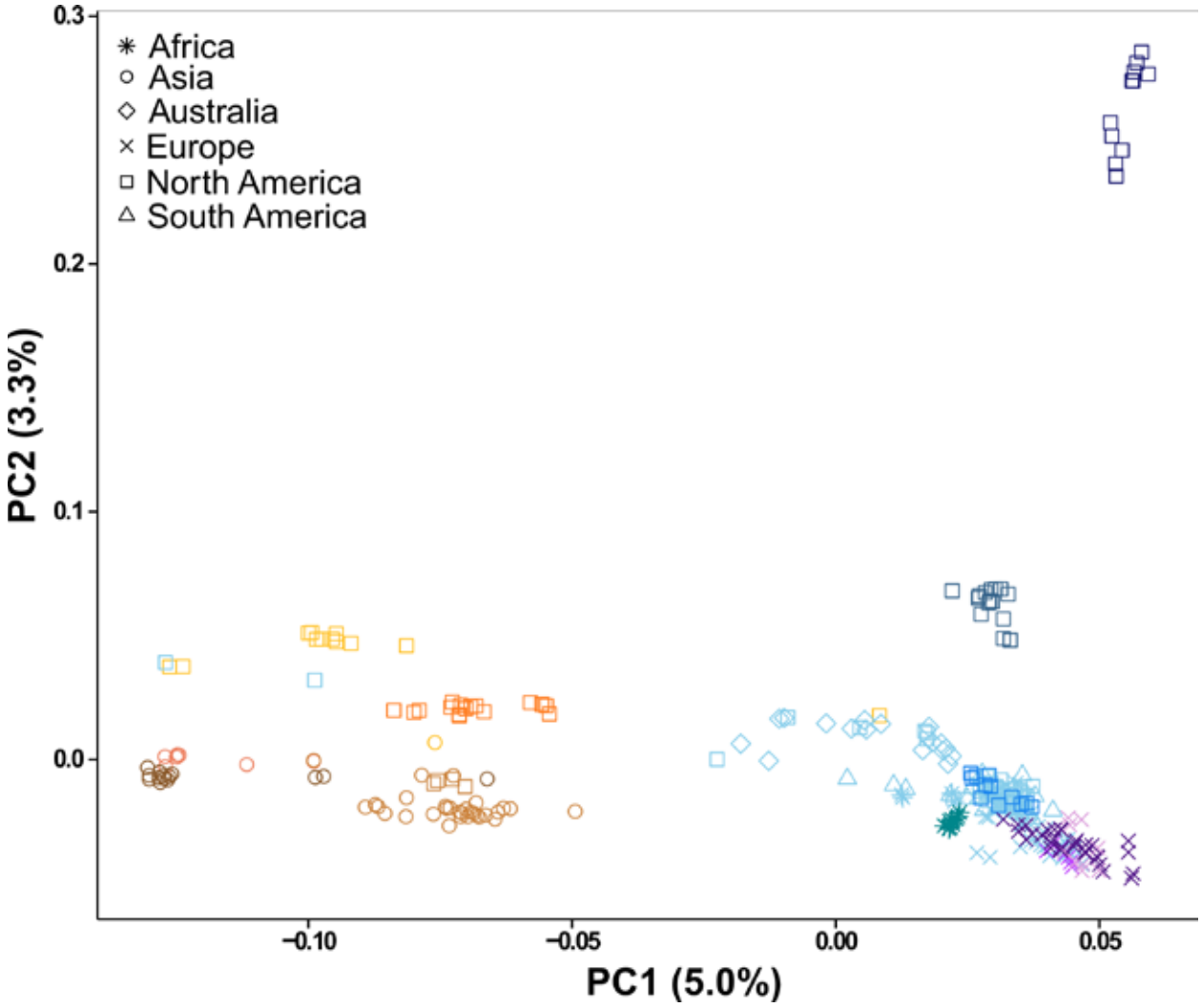

Figure S5- Principal components analysis using 32k nuclear SNPs for worldwide *Rattus norvegicus* samples for the first two principal components. Continents are designated by shape (Asia: circles; Europe: X; Africa: star; North America: square; South America: triangle; New Zealand: diamond) with substructured populations designated by color for the 13 clusters inferred using model-based ancestry analyses (Figs. 2, S2).

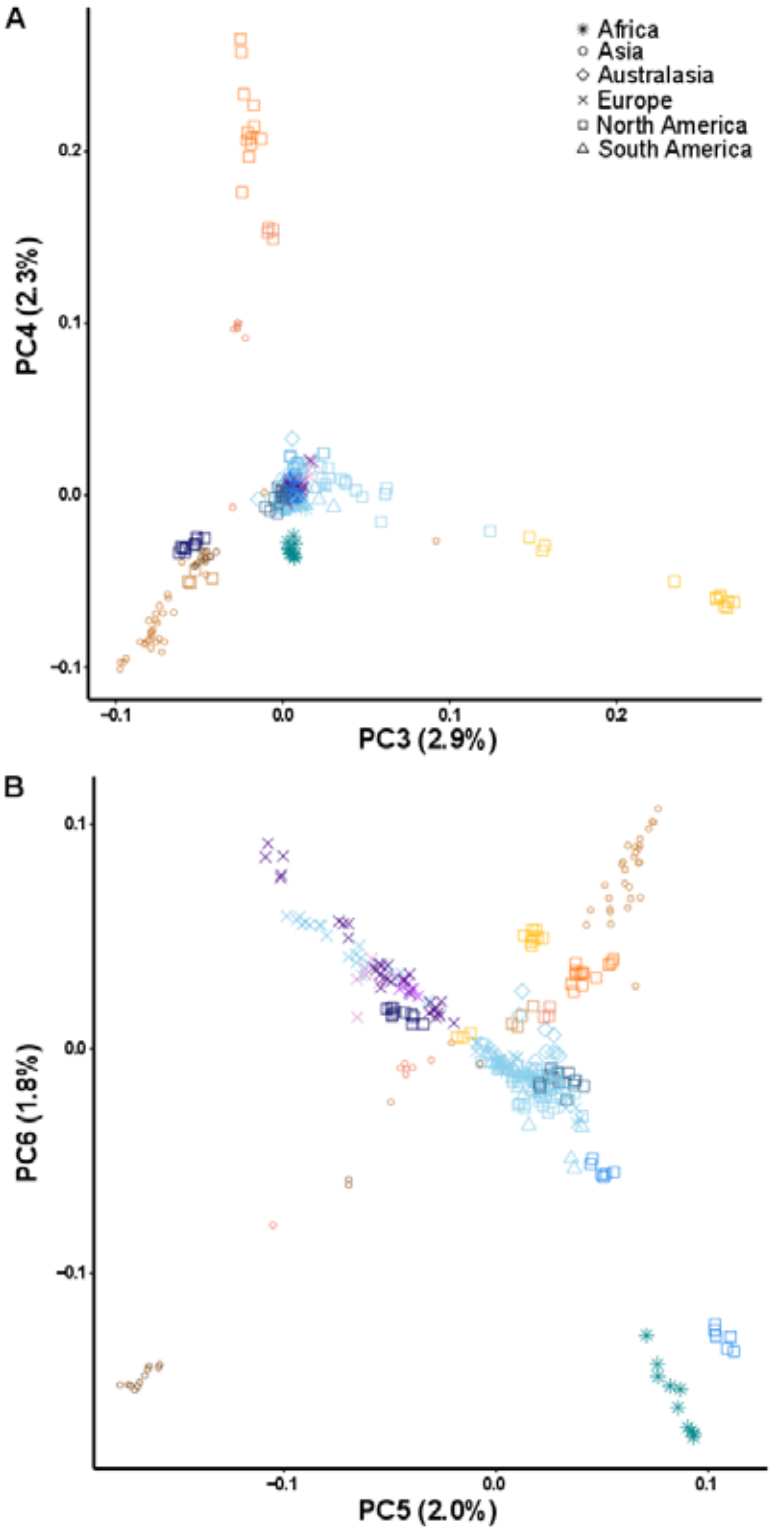

Figure S6- Principal components analysis using 32k nuclear SNPs for worldwide *Rattus norvegicus* samples for (A) the third and fourth PCs, and (B) the fifth and sixth PCs. Continents are designated by shape (Asia: circles; Europe: X; Africa: star; North America: square; South America: triangle; New Zealand: diamond) with substructured populations designated by color for the 13 clusters inferred using model based ancestry analyses (Figures 1, S7).

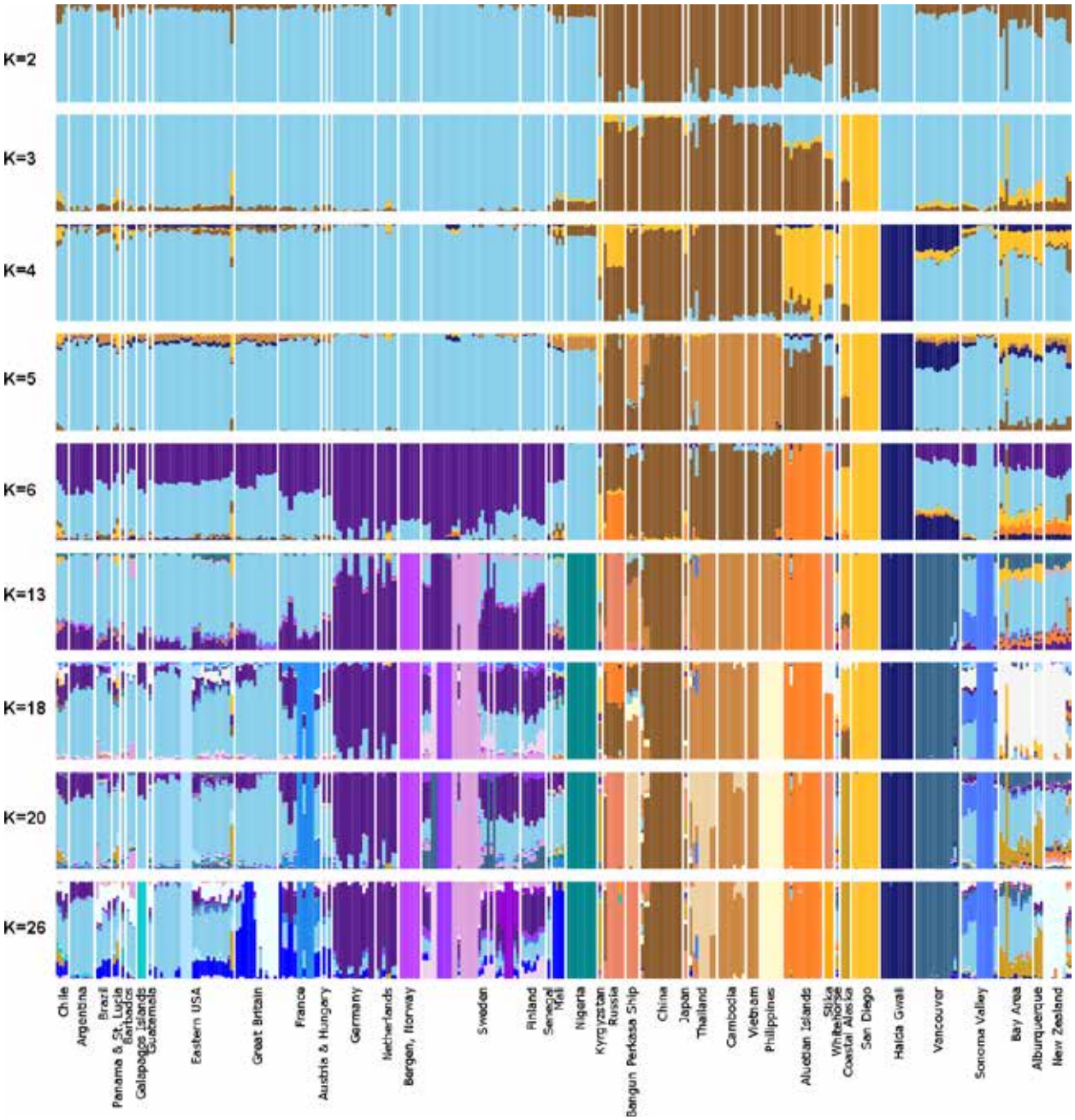

Figure S7- ADMIXTURE plots for all *Rattus norvegicus* samples (n = 314) using 32k SNPs from the nuclear genome with varying cluster (K) values. The higher cluster values were those supported based on cross validation error (see Figure S2A).

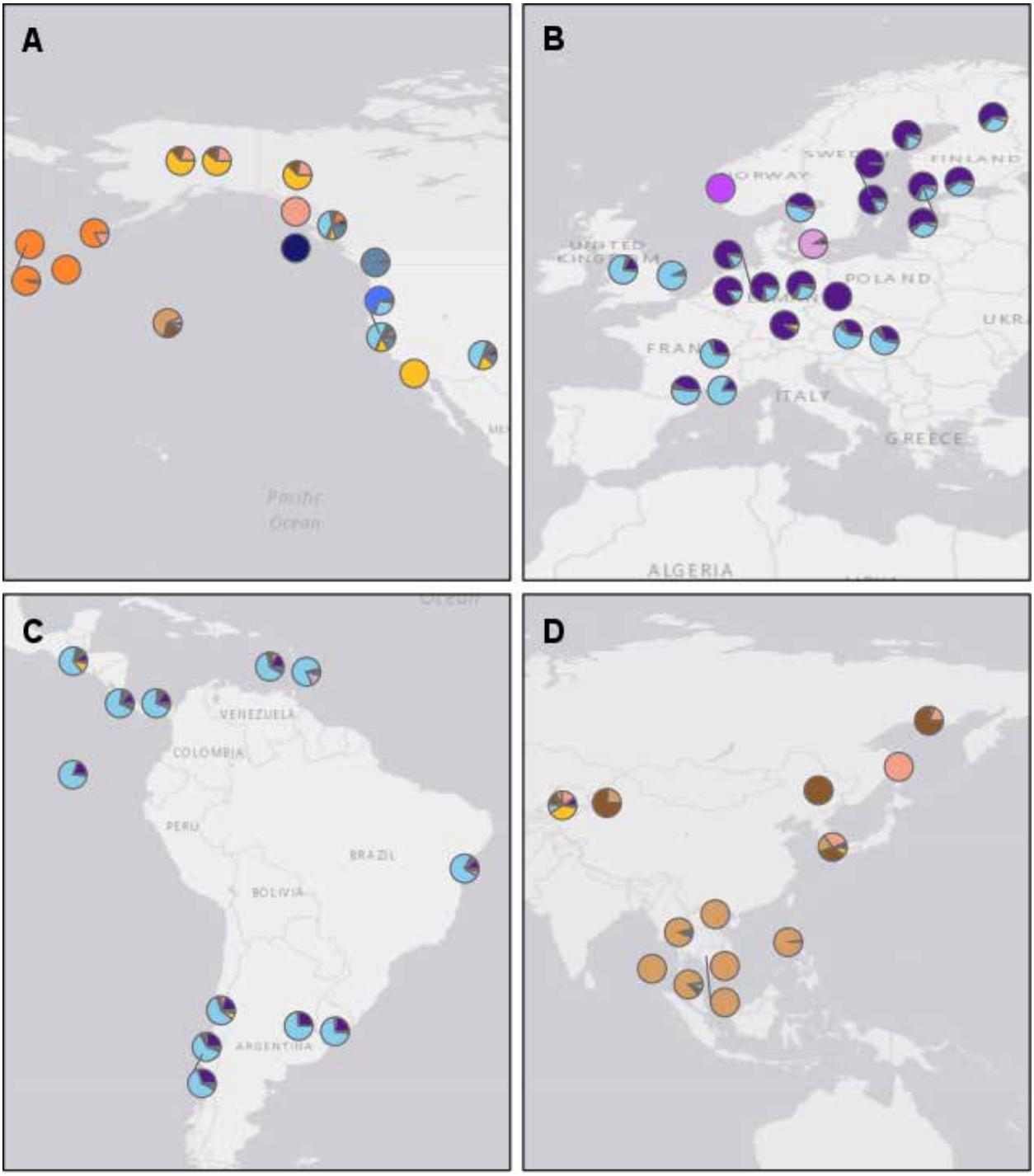

Figure S8- Continent specific maps (A- western North America, B- South America, C- Europe, D- Asia) with average proportion of ancestry inferred from 32k nuclear SNPs and based on estimates from 13 clusters (Figure 1; China: brown; SE Asia: light brown; Aleutian Archipelago: orange; San Diego: gold; Europe: light blue; Scandinavia: purple; Nigeria: turquoise; Sonoma Valley: medium blue; Haida Gwaii: dark blue; Vancouver: cerulean; Bergen: medium purple; Malmo: light purple). Geographic locations were jiggered so all sites could be viewed.

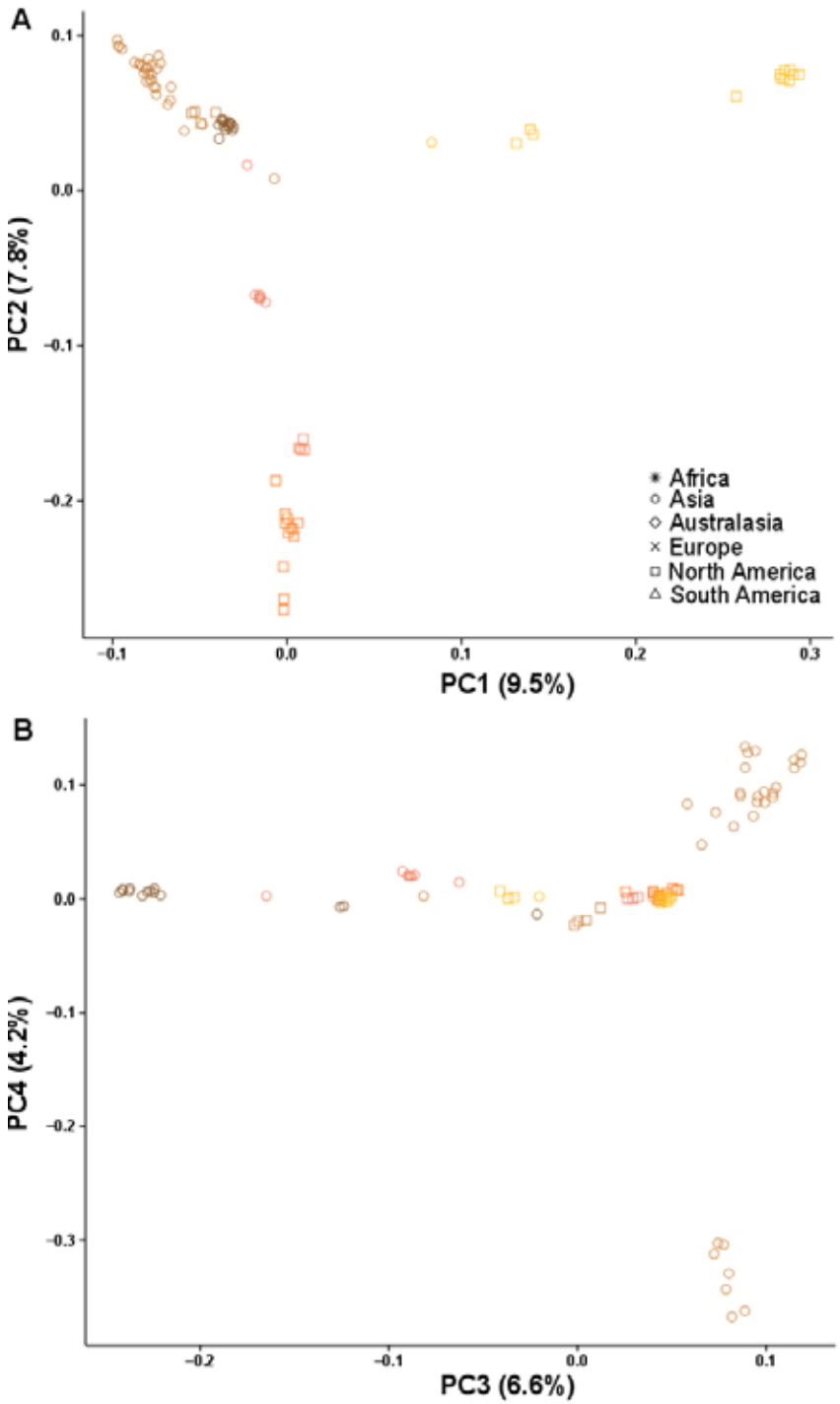

Figure S9- Principal components analysis using 32k nuclear SNPs for the Asian *Rattus norvegicus* samples (n = 84) for (A) the first and second PCs, and (B) the third and fourth PCs. Continents are designated by shape (Asia: circles; North America: square) with substructured populations designated by color for the 13 clusters inferred using model based ancestry analyses (Figures 1, S7; dark brown: China, light brown: SE Asia, pink: Russia, orange: Aleutian Archipelago, gold: San Diego).

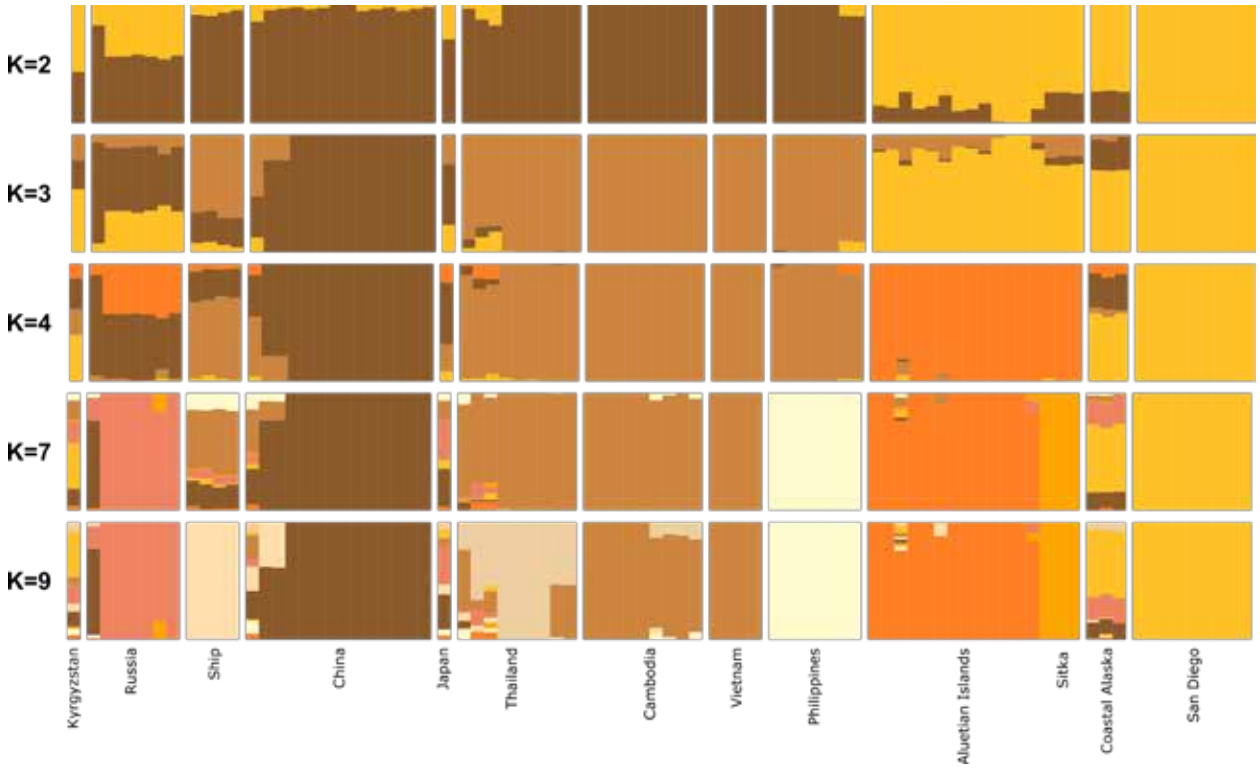

Figure S10- ADMIXTURE plots for the subset of *Rattus norvegicus* samples assigned to the Asian cluster (n = 84) using 32k SNPs from the nuclear genome with varying cluster (K) values based upon support from the cross validation error plots (see Figure S2B).

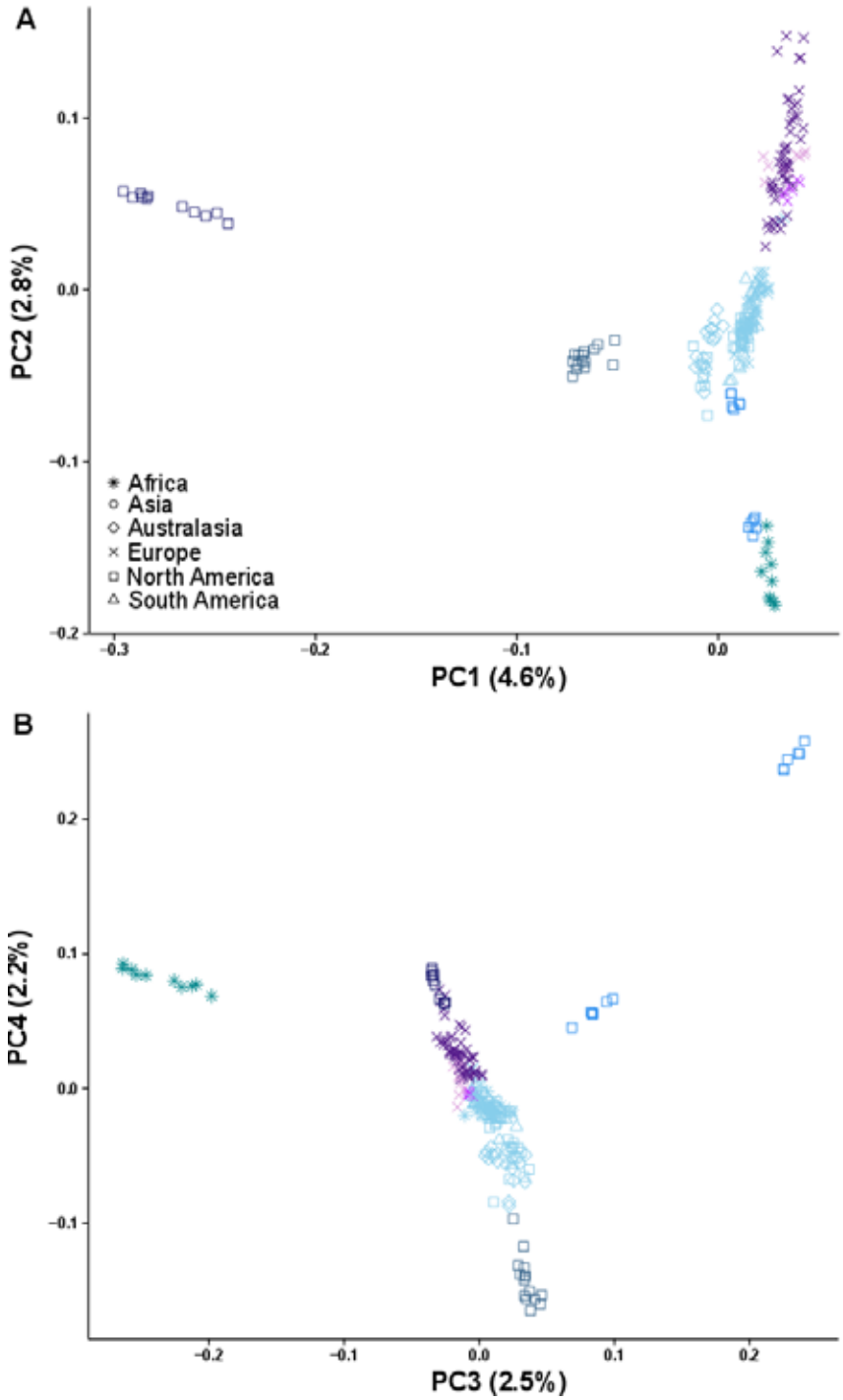

Figure S11- Principal components analysis for the non-Asian *Rattus norvegicus* samples (n = 230) using 32k SNPs from the nuclear genome for (A) the first and second PCs, and (B) the third and fourth PCs. Continents are designated by shape (Europe: X; Africa: star; North America: square; South America: triangle; New Zealand: diamond) with substructured populations designated by color for the 13 clusters inferred using model based ancestry analyses (Figures 1, S7; light blue: Europe, dark purple: Scandinavia, dark blue: Haida Gwaii, Canada, cerulean: Vancouver, Canada, turquoise: Kano, Nigeria, medium blue: Sonoma Valley, USA, medium purple: Bergen, Norway, light purple: Malmo, Sweden).

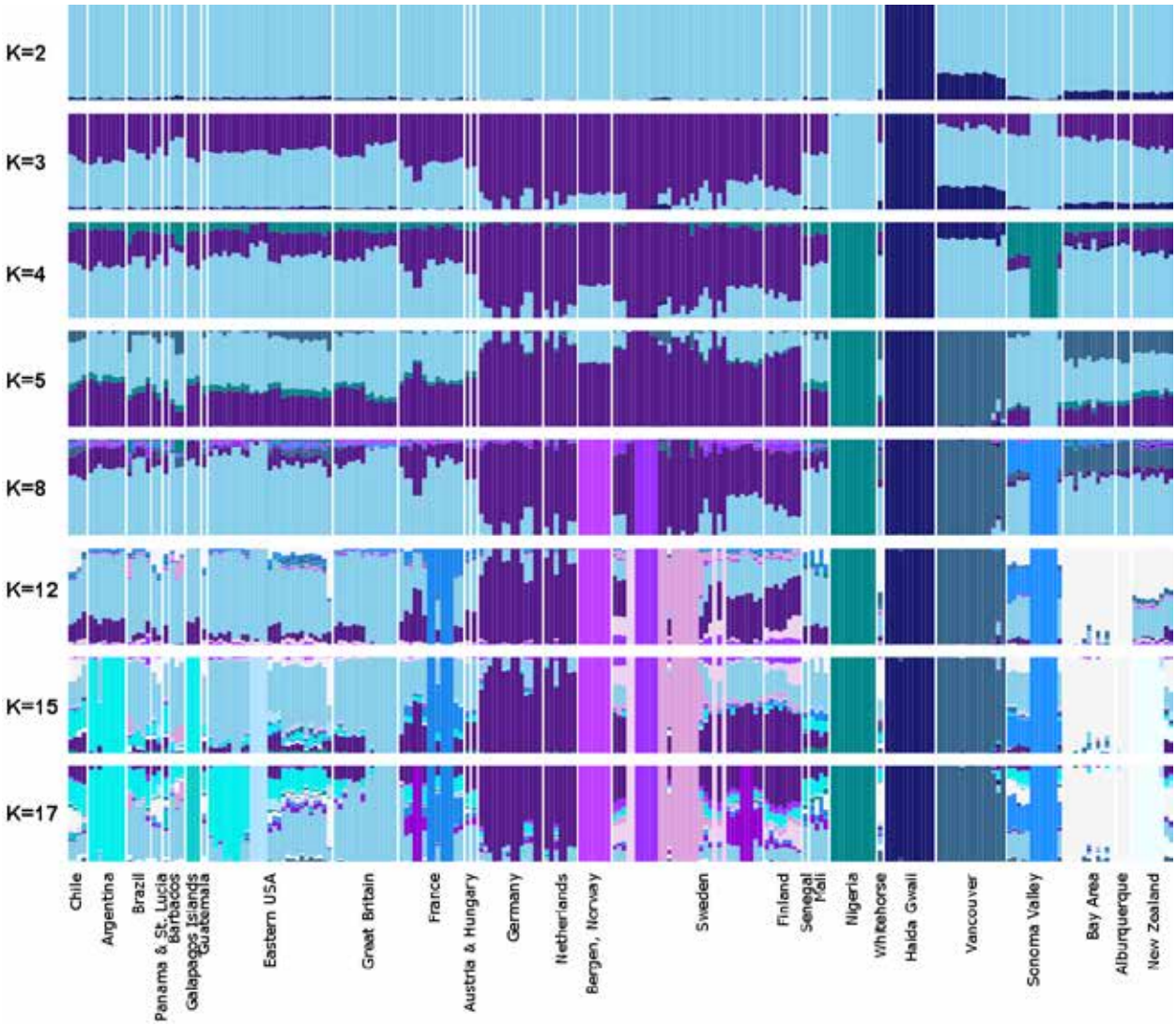

Figure S11- ADMIXTURE plots for the subset of *Rattus norvegicus* samples assigned to the non-Asian cluster (n = 230) using 32k SNPs from the nuclear genome with varying cluster (K) values based upon support from the cross validation error plots (see Figure S2C). Note that the highest cluster values identify signatures not observed in the highest cluster values from the full analysis.

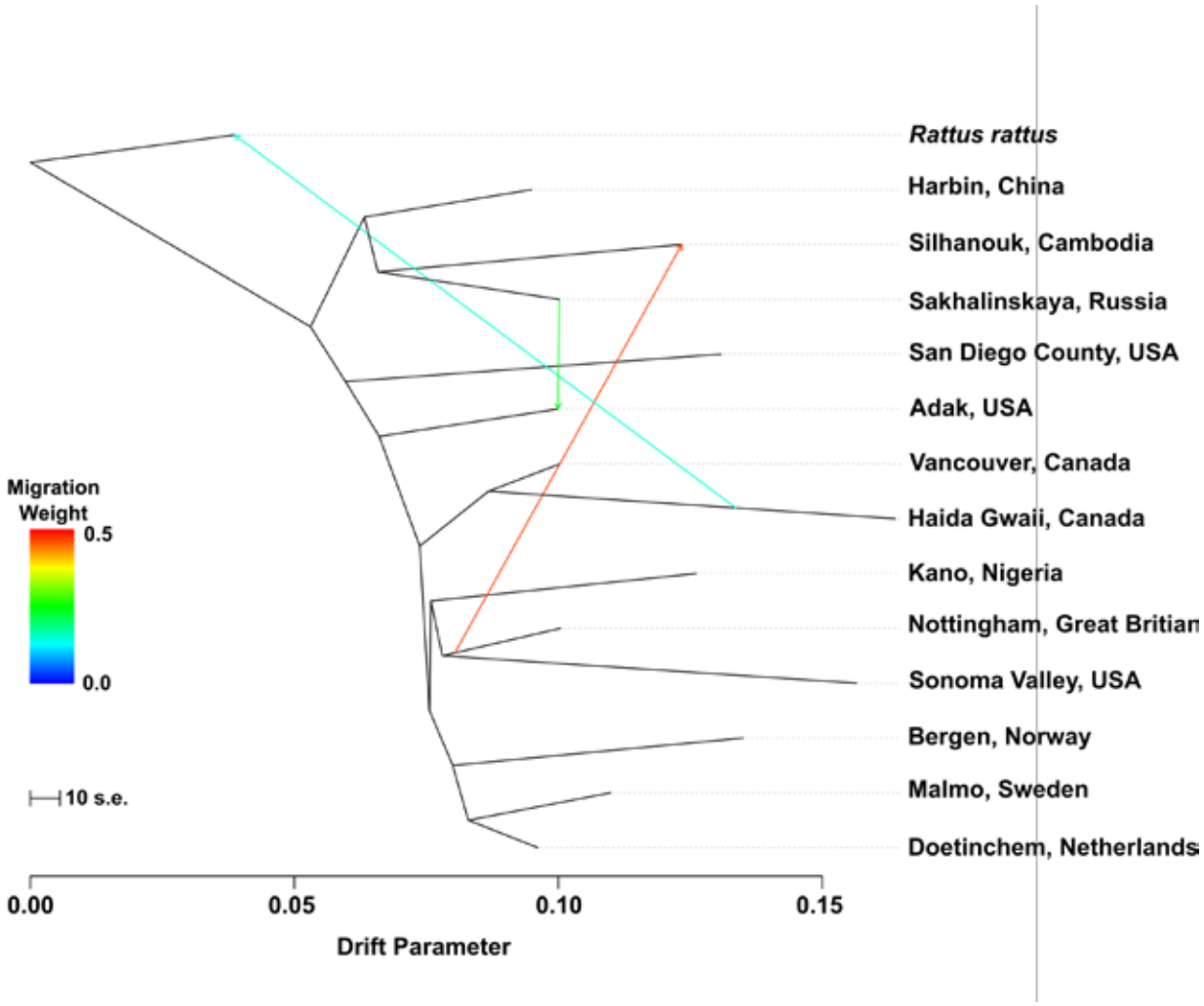

Figure S13- Population tree with 13 sampling areas, one for each of 13 supported evolutionary clusters (see Figures 1, S2A, S7) produced using TREEMIX using 32k SNPs from the nuclear genome with three migration edges added.