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

Emily E. Puckett1, Jane Park1, Matthew Combs1, Michael J. Blum2, Juliet E. Bryant3, Adalgisa Caccone4, Federico Costa5, Eva E. Deinum6,7, Alexandra Esther8, Chelsea G. Himsworth9, Peter D. Keightley6, Albert Ko10, Åke Lundkvist11, Lorraine M. McElhinney12, Serge Morand13, Judith Robins14,15, James Russell15,16, Tanja M. Strand11, Olga Suarez17, Lisa Yon18 and Jason Munshi-South1

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...
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 one to 40; given that a single SNP per RAD-tag was retained, we met the criterion for unlinked data for this analysis. 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 SAMSTOCKS 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 S4a) and the residuals of the population tree (electronic supplementary material, figure S4b) and chose the model with three migration edges. We decided to thin the sampling areas owing to uneven
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 S2r 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 S2b). The predominant four clusters reflected differentiation between: China, Southeast (SE) Asia, the Aleutian Archipelago and Western North America and an extended non-Asian cluster found in Europe, Africa, the Americas and New Zealand.

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.
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 S2c). 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 3e). 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
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 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 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 to 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 3a). 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
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 gene flow in our dataset, including an individual matching 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 gene flow in our dataset, including a 2 year behavioural study of brown rats [33].

(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;
References


