Title: Brown rat demography reveals pre-commensal structure in eastern Asia prior to expansion into Southeast Asia during the Song dynasty

Authors: Emily E. Puckett[1,2] and Jason Munshi-South[1]

[1]- Louis Calder Center, Fordham University, Armonk, NY 10504

[2]- Current Address: Integrative Biology, Michigan State University, East Lansing, MI 48824

Corresponding Authors:

Emily E. Puckett

288 Farm Lane, Natural Sciences Building Room 203, Michigan State University

East Lansing, MI 48824

Email: Emily.E.Puckett@gmail.com

and

Jason Munshi-South

31 Whipporwill Road, Louis Calder Center- Biological Field Station, Fordham University

Armonk, NY 10504

Phone: 914-273-3078 ext 20

Email: jmunshisouth@fordham.edu (JM-S)

Keywords: demography, phylogeography, global invasive

ABSTRACT

Fossil evidence indicates that the globally-distributed brown rat (*Rattus norvegicus*) originated in northern China and Mongolia. Historical records report the human-mediated invasion of rats into Europe in the 1500s, followed by global spread due to European imperialist activity during the1600s-1800s. We analyzed 12 genomes representing seven previously identified evolutionary clusters and tested alternative demographic models to infer patterns of range expansion, divergence times, and changes in effective population (N_e) size. We observed three range expansions from the ancestral population that produced the eastern China (diverged ~326kya), Pacific (~42.8kya), and Southeast Asia (~0.68kya) lineages. Our model shows a rapid range expansion from Southeast Asia into Europe 535 years ago (1480AD). We observed declining N_e

within all brown rat lineages from 100-2kya, reflecting population contractions during glacial cycles. N_e increased since 1kya in Asian and European, but not Pacific, evolutionary clusters. Our results support the hypothesis that northern Asia was the ancestral range for brown rats. We suggest that southward human migration across China between 800-1550s AD, resulted in the introduction of rats to Southeast Asia, from which they rapidly expanded into the Middle East then Europe via existing maritime trade routes. Finally, we discovered that North America was colonized separately on both the Atlantic and Pacific seaboards, yet with evolutionary clusters of vastly different ages and genomic diversity levels. Our hypotheses should stimulate discussions among historians and zooarcheologists regarding the relationship between humans and rats.

INTRODUCTION

The genus *Rattus* originated and diversified in eastern and central Asia, and fossil evidence (1) suggests northern China and Mongolia as the likely ancestral range of the cold-hardy brown rat (*Rattus norvegicus*). Yet, their contemporary distribution includes every continent except Antarctica. As a human commensal, brown rats occupy urban and agricultural areas using food, water, and shelter provided by humans. Rats are one of the most destructive invasive mammals, as they spread zoonotic diseases to humans (2), damage food supplies and infrastructure (3), and contribute to the extinction of native wildlife (4). As an invasive species, brown rats outcompete native species for resources and are a primary target of eradication efforts (5). Brown rats have been domesticated as models for biomedical research with inbreeding leading to disease phenotypes similar to humans (6). Finally, they are a nascent model to study evolution within urban landscapes, as they likely experience multiple selection pressures given their global distribution in a range of conditions (7).

The historical record indicates colonization of Europe occurred in the early 1500s, eastern North America by the 1750s, and the Aleutian Archipelago by the 1780s (8, 9). Yet, the timing of historical records may be inconsistent with population divergence times estimated from genomic data for two reasons. First, genetic population divergence is not an instantaneous process, thus sufficient time must pass for allele frequencies to change enough that groups of randomly mating individuals are recognized as unique populations within demographic models. Second, the historical record may not immediately record a species within a new geographic area. We would expect that: population sizes of new species must increase to a noticeable level, species are identifiable (i.e. brown rats can be mistaken for black rats [*R. rattus*]), and records survived through time. Thus, historical records provide minimum dates for introductions.

Research into the global expansion of brown rats has focused on both the routes and timings of different invasions; questions of specific interest include the location of the ancestral range, and when they arrived in Europe, particularly since black rats reached southern Europe by 6kya (10) and Great Britain by the 300s AD (11). Previous phylogeographic studies with mitochondria identified China as the ancestral range based on private haplotypes and ancestral state reconstructions, with multiple expansions into Southeast Asia, Europe, and North America (12-

14). Inference from mitochondria has been limited due to high haplotype diversity observed from locally intense but globally diffuse sampling strategies. Thus, key geographic regions, especially around the Indian Ocean basin and the Middle East, are missing data; sampling these areas would allow us to distinguish clinal versus long-distance expansions, where multiple introductions occurred, and mito-nuclear discordance. A phylogeographic analysis using nuclear genomic data inferred hierarchical clustering along five range expansion routes (15). From the putative ancestral range, brown rats expanded southward into Southeast Asia and eastward into China and Russia (15). The eastward expansion extended to North America with two independent colonizations of the Aleutian Archipelago and sites along the Pacific coast of western North America. From Southeast Asia rats expanded into Europe (15, 16). Given historical records of brown rats in Europe in the 1500s, the likely route was aboard ships conducting maritime trade across the Indian Ocean into the Red Sea and Persian Gulf before moving goods onto land. Although these routes were established by the 200s BC, they intensified in the 1400-1500s AD (17). The fifth range expansion moved rats to eastern North America, the Caribbean, South America, western Africa, and Australasia during the age of European imperialism of the 1600-1800s (15). Ultimately, we inferred the following seven genomic clusters: *Eastern China*, *SE Asia*, *Aleutian*, *Western North America*, *Northern Europe*, *Western Europe*, and *(Western Europe) Expansion*.

Demographic histories of domesticated and commensal species are uniquely compelling as they infer both the species history and how humans interacted with each species over time. Demographies can infer domestication history (18, 19), population declines and recoveries (20, 21), human mediated range expansions (22), and factors influencing extinction (23). Although the demographic history of humans has been extensively researched, inference from domestics and commensal species can elucidate patterns of human expansion and exploration (24). Thus, we tested alternative demographic models to identify patterns of population divergence, infer the timing of range expansions, and estimate effective population size (N_e) changes through time. From the geographic and temporal patterns, we were able to distinguish natural versus human-mediated movement of brown rats.

RESULTS

We sequenced two genomes each from *SE Asia*, *Northern Europe*, *Western Europe*, and the *Western Europe-Expansion* (hereafter- *Expansion*) evolutionary clusters, and one genome each from the *Aleutian* and *Western North America* clusters (NCBI SRA PRJNA344413; Table S1). Average sequencing depth was 28. 2X (range 24-38). We estimated heterozygosity for each individual on the 20 autosomes separately. Samples from *Eastern China* had the highest average chromosomal heterozygosity (0.244) where the *Aleutians* and *Western North America* had the lowest heterozygosity (0.143 and 0.148, respectively; Figure 1B).

Ancestral Gene Flow

We used the ABBA-BABA test to examine historic gene flow between evolutionary clusters and verify the global tree topology by comparing all combinations (n=106) of population tree topologies *D*(H_1, H_2, H_3, *R. rattus*). We observed 97 significant *D*-scores with absolute Z-scores greater than 3 (Table S3), where most patterns were congruent with our inferred population tree (see below, Figure 2A). However, both the *Aleutian* and *Western North America* clusters shared more derived alleles with the *Expansion* cluster than either Asian or other European clusters (Table S3), suggestive of movement of rats across North America. Further, population trees of the form *D*(Asian or European Population, Western North America, Aleutian, *R. rattus*) consistently observed increased derived alleles within the Asian or European clusters than the Pacific clusters, thus a sister relationship between *Western North America* and *Aleutian* was not observed (Table S3). This result was unexpected, as we demonstrate here and in a previous analysis (15) that the two Pacific clusters share a population ancestor; unsampled diversity may have influenced these results.

WGS Demographic model

We conducted model selection by first identifying the model that best represented divergence patterns in Asia and Europe separately, then combined those tree topologies into a global model for parameter estimation. The best Asian model had an ancestral unsampled population with independent divergence events for *Eastern China*, *SE Asia*, and the Pacific cluster that diverged into the *Aleutians* and *Western North America* (Figure S1). For the European model, the best supported model included an unsampled population (likely in the Middle East) diverging from

SE Asia, where *Western Europe* diverged from the unsampled population, and *Northern Europe* and the *Expansion* independently diverged from *Western Europe* (Figure S2).

Using WGS data from 12 genomes representing seven evolutionary clusters, we modeled the nine-population topology inferred from the sub-models (Figure 2, Table S4). We estimated that *Eastern China* diverged from the ancestral population 326kya (90% highest density probably [HPD]: 308-327kya). The Pacific cluster diverged from the ancestral population 42.8kya (HPD: 30.0-41.3kya), then the *Aleutians* and *Western North America* diverged 20.2kya (HPD: 10.7-20.9kya). The divergence that led to the global expansion of rats occurred rapidly, where rats first expanded into *SE Asia* 682 years ago (1333AD; HPD: 997-1406AD), followed by the introduction of rats into the unsampled population 681 generations (HPD: 1001-1409AD). We estimated rapid divergence of rats into Europe including the *Western Europe* divergence 535years ago (1480 AD; HPD: 1462-1511AD), and *Northern Europe* divergence 534 years ago (1481 AD; HPD: 1464-1512AD). Finally, we estimated the *Expansion* cluster diverged 527 years ago (1488 AD; HPD: 1487-1524AD).

We ran the cross-coalescence analysis within MSMC2 to estimate the rate of divergence between the seven clusters. We observed that divergence was complete between both the *Aleutians* or *Western North America* and all other populations (Figure S3). The European clusters showed similar patterns of divergence with *Eastern China* with approximately 60% divergence complete (Figure S3A). Cross-coalescence between the *Aleutians* and *Western North America* increased approximately 200 generations ago before decreasing to 50% (Figure S3C). The four clusters making up the most recent expansions (*SE Asia*, *Northern Europe*, *Western Europe*, and *Expansion*) had signatures of increasing cross-coalescence over the past 1,000 generations (Figure S3 B, D-F). We believe this result was due to a population version of incomplete lineage sorting, where the rapid global range expansion resulted in high rates of coalescence events between instead of within evolutionary clusters.

ddRAD Demographic Models

We used a ddRAD-Seq dataset (Table S2) to further investigate the within-cluster population tree topologies. Given the hypothesis that the ancestral populations for western North America

were in eastern Asia, we first compared these clusters and observed that eastern China and eastern Russia were sister populations (Figure S5). Further, our modeling supported an admixture pulse of 30% 215 generations ago from eastern Russia into rats on Adak Island within the Aleutian Archipelago (Figure S5). Next, we modeled relationships between four populations within Southeast Asia and observed that rats first expanded into central Southeast Asia. There were two independent expansions eastward: one into Cambodia and Vietnam which formed sister populations, and a second to the Philippines (Figure S5).

We split European populations between the *Western* and *Northern* evolutionary clusters. The best supported topology for *Northern Europe* had the continental Netherlands diverging from a sister group including Norway and Sweden (Figure S6). In *Western Europe*, we observed a split between continental Europe (populations in France and Spain) and the British Isles (Figure S7). The Pacific coast of North America contains populations with *Aleutian*, *Western North America*, and *Expansion* genomic signatures (15). To better understand admixed rat populations we modeled two North American populations (Vancouver, Canada and Berkeley, USA). We observed that both populations were best described by admixture between the *Western North America* and *Expansion* clusters (Figure S8).

Effective population size through time

We inferred the change in N_e over time using the multiple sequentially Markovian coalescent (MSMC) model (25) and scaled the estimates to years and N_e using the estimated mutation rate (μ) from the coalescent modeling analysis of 9.90×10^{-8} and 3 generations per year. As Deinum *et al*. (26) estimated μ of 2.96×10^{-9} and the precise generation time for rats is unknown, we present alternative estimates of the MSMC model in Figure S9. We observed two distinct patterns; first, the *Aleutian* and *Western North America* clusters declined sharply in N_e approximately 50kya (Figure 1A); this was consistent with the estimated divergence of the Pacific clusters at 42.8kya. Following divergence, N_e steadily declined until 500 years ago when it stabilized to approximately 1,400 and 1,500 effective individuals in *Aleutian* and *Western North America*, respectively (Figure 1A).

196 The second pattern was concordant between the *Eastern China*, *SE Asia*, *Northern Europe*,
197 *Western Europe*, and *Expansion* clusters. Similar to the Pacific clusters, N_e steadily declined
198 from approximately 100 – 1kya before increasing in the most recent time periods (Figure 1A).
199 Approximately 200 years ago (the first reliable time step), N_e was: 12,000 in *Eastern China*,
200 36,000 in *SE Asia*, 45,000 in *Northern Europe*, 28,000 in *Western Europe*, and 25,000 in
201 *Expansion* (Figure 1A).
202
203 **DISCUSSION**
204 Our demographic model showed that from an unsampled ancestral population, brown rats
205 diverged into three genomic clusters that we sampled in eastern China, two locations in North
206 America, and Southeast Asia. These results were congruent with a hypothesized ancestral range
207 in northern China and Mongolia where brown rat fossils were identified (1). Divergence of
208 brown rats from their sister species *R. nitidus* was estimated at 633kya (27). We estimated
209 divergence of the *Eastern China* cluster at 326kya (Figure 2A, Table S4), which would be
210 consistent with the establishment of ancient population structure within glacial refugia near the
211 end of the interglacial period of the Middle Pleistocene (28). This estimate contrasts with the
212 observed mitogenome diversity, where the five clades that occur within the samples genotyped
213 for the *Eastern China* nuclear cluster diverged since 96kya (HPD: 70-128kya) (13). We
214 previously hypothesized that the incongruence between a single nuclear cluster and multiple
215 deeply diverged mitochondrial clades represented ancestral population structure and later
216 admixture. The discrepancies in the divergence timing estimation between the nuclear (326kya;
217 307-327kya) and mitochondrial (96kya; 70-128kya) genomes could be do either to extirpation or
218 incomplete sampling of mitogenome diversity. Alternatively, our coalescence modeling may
219 have overestimated divergence age between *Eastern China* and the ancestral range, which was
220 suggested by the upper limit of the MSMC analysis (Figure 1A) at 100kya.
221
222 We observed that the Pacific cluster diverged from the ancestral population 42.8kya (Figure 2A,
223 Table S4), where this result was supported by our MSMC analysis (Figure 1A). Divergence of
224 the *Aleutian* and *Western North America* clusters occurred 20.2kya, although MSMC did not
225 identify this signal. The similarity in N_e over time between these two clusters may have been due
226 to either true recent divergence as suggested by cross-coalescence (Figure S4C) or similar

climatic conditions through the Pleistocene when population size decreased for many vertebrates. We emphasize that the divergence of these clusters does not identify the timing of the introduction to the Aleutian Archipelago or the Pacific coast of North America where samples were collected. The historic record indicates rats were moved to the Aleutian Archipelago by Russian fur traders in the 1780s (9). Our regional population model suggested a scenario with gene flow from a population in eastern Russia into Adak Island (Figure S4). We modeled this scenario as an admixture pulse from eastern Russia around 1943, suggesting two introductions of rats onto Adak. Our results suggest that rats from Japan and sites around the Bering Sea in Russia and the USA may be a dynamic region for rat population genomics. Specifically by incorporating both more sites and ancient samples, demographic models could estimate divergence times to differentiate between models of contemporary or historic movement of rats, particularly as human divergence into the region occurred approximately 36kya (29).

We estimated that the *SE Asia* cluster diverged from the ancestral population 0.68kya (1333 AD, Figure 2A, Table S4). The timing of this divergence immediately raises the question of why rats did not expand sooner as overland trade between China and Southeast Asia was established by the 500s AD (30); and maritime trade between these regions and the Indian Ocean basin was established before the 900s AD (31). A partial explanation may be due to the intersection of climate and human demography across eastern Asia. The Medieval Climate Anomaly (850-1250 AD) aided agricultural expansion and human demographic growth in China, specifically expanding urban centers outwards at a time of human movement from northern arid lands to more agriculturally productive lands in the south (see references within 30). However, the end of this climatic period resulted in drought, famine, political instability, and ultimately human demographic contractions in both China and Southeast Asia; fortunes reversed in the late 1400s to mid-1500s as the climate improved and populations expanded again (30). We can view rats' southward expansion from this historical context. As we lack samples between the ancestral range and Southeast Asia, we were unable to determine population divergence along this axis and intermediate time points, although Zeng *et al*. (16) observed a clinal pattern of ancestry in southern China. Thus, the founding of new agrarian communities and increasing inter-connectedness with urban centers would serve as stepping-stones moving rats from northern China to Southeast Asia during the two periods of human demographic expansion.

Our results stand in marked contrast to a similar analysis that identified brown rats in Southeast Asia as the ancestral population with a northward expansion to colonize eastern Asia (16). Both analyses used similar modeling approaches but with different datasets. We suggest that our results differ due to the following two factors: first, we included unsampled populations that allowed for coalescence within that group instead of other contemporary lineages. Inclusion of unsampled populations influences parameter estimates of N_e and migration rates and has been shown to improve or at least not harm parameter estimation within the full model (32). Adding unsampled populations to our model was important given the limited number of chromosomes genotyped, as large sample sizes decrease the effect of unsampled populations on parameter estimates (33). Second, Zeng *et al*. included admixed samples (that were geographically located in southern China) between the SE Asia and northern Asia (our *Eastern China*) genomic clusters where the admixed samples were exclusively grouped with the northern Asia genomic cluster. By grouping the mixed ancestry samples preferentially with the northern Asia cluster, the coalescent model was better supported when the SE Asia population was ancestral. This was likely do to better model fit given that a small proportion of SE Asian alleles were within the northern Asia cluster. We believe this was not due to the true history of the populations but instead to artefacts introduced by inappropriate sample clustering. The cline in ancestry proportions from the two divergent groups may be from a southern expansion from the north and/or a northern expansion from SE Asia, thereby creating a regional admixture zone in southern China where samples were collected (16).

Finally, we show that from Southeast Asia there was a rapid range expansion into Europe approximately 1480 AD (Table S4); here our phylogeographic pattern but not divergence time results were congruent with those of Zeng and colleagues (16). While our coalescent models support the population topology of divergence into *Western Europe* then into *Northern Europe* (Figure S3), it is important to note that parameter estimates of divergence between those clusters were one generation apart. Thus, the range expansion into Europe occurred so rapidly that precise topology patterns and divergence times may be difficult to recover. Further, potential stepping-stone areas in India, the Middle East, and coastal eastern Africa have been sparsely sampled; although samples from Iran cluster with *Western Europe* with small ancestry

proportions from *Northern Europe* and *SE Asia* (16). This pattern was consistent with the rapid range expansion into Europe we observed, thus not allowing sufficient time for drift to establish unique allele frequency patterns. We estimated that the *Expansion* cluster diverged from *Western Europe* approximately 1488 AD (HPD: 1464-1512 AD; Table S4). While European seafaring to North America occurred during this time, this particular estimate may be too young to represent the actual introduction into New York, USA and instead may represent the timing of divergence of the ancestral population likely within Great Britain.

Ancestral population size

We observed that each evolutionary cluster declined in N_e since 100kya (Figure 1A), consistent with many studies that observed population declines during the Pleistocene. The decline lasted longer for the two Pacific clusters, consistent with the lower heterozygosity estimates (Figure 1B). We observed an increasing trend in the most recent time segments for the two Asian and three European clusters. This result was consistent with our demographic models showing a rapid and recent global range expansion.

Range expansion via human-mediated movements

Our results identified that the global range expansion of rats occurred recently (1480s AD) and rapidly from Southeast Asia into Europe. This stands in marked contrast to previous assumptions that brown rats were transported westward along the Silk Road through central Asia into Europe. This is counterintuitive as overland trade routes from central China to Persia were established 2.1kya (105 BC) and goods reached Rome by 46 BC (17). Brown rats are native to the geographic region through which part of the Silk Road connected, unlike black rats which speciated on the Indian subcontinent (34). Assuming that rats evolved their commensal relationship with humans prior to their global range expansion, as observed with house mouse (35), the availability of cities, road networks, and a flow of merchants naturally suggests a way to expand westward. The Silk Road may have not been the route for expansion due to the limited distance that merchants traveled along the route, as goods went further than caravans (17). Further, high aridity and lack of water sources may have limited rat movement through this region. Yet this does not preclude the idea that brown rats may have expanded westward via Silk Road cities and were then extirpated due to the collapse of those cities during changing geo-

politics and shifts towards maritime trade (17). We instead suggest that pulses of southward human demographic expansion from northern China during favorable climatic conditions enabled the expansion of rats into Southeast Asia from which they expanded into Europe. We present this historical narrative as a hypothesis supported by our demographic model, but also to stimulate interest in further study by historians and zooarchaeologists to examine the historical expansion of this globally important invader.

MATERIALS and METHODS

Whole genome sequencing and datasets

We selected ten individuals for whole genome sequencing: two each representing evolutionary clusters within *SE Asia* (Philippines and Cambodia), *Northern Europe* (Sweden and Netherlands), *Western Europe* (England and France), and *Expansion* (New York, USA), and one sample each from the *Aleutian Islands* and *Western North America* (Table S1). We generated paired-end reads for each sample (4ng RNase treated genomic DNA) by sequencing on an Illumina HiSeq 2500 at the New York Genome Center. We also downloaded whole genomes from 11 brown rats and one black rat (*R. rattus*) collected in Harbin, China (ENA ERP001276), although to not bias estimates with unequal sample sizes we ran analyses using only Rnor13 and Rnor14 (26) which were randomly selected. All genomes were mapped to the Rnor_5.0.75 reference (36) using BWA-MEM v0.7.8 (37). We marked duplicates using Picard Tools v1.122 and realigned indels using GATK v3.4.0 (38). Data for 10 genomes are available on the NCBI SRA BioProject PRJNA344413 (13).

Instead of calling variants, we used the genotype likelihoods implemented in ANGSD v0. 915 (39), using a minimum base quality score of 20, minimum mapQ score of 30, then identified sites genotyped in all 12 genomes on the 20 autosomes. We estimated genotype likelihoods using the function implemented in SAMTOOLS (-GL 1) (40). The *R. rattus* individual from China served as the outgroup allowing for identification of ancestral and derived alleles.

Chromosomal Diversity

We estimated heterozygosity on each chromosome for each individual. We exported the genotype likelihoods from ANGSD into PLINK v1.9 (41, 42) and estimated heterozygosity (--het) on each chromosome.

Historic gene flow

The ABBA-BABA test (43) measures the proportion of "ABBA" and "BABA" patterns of ancestral and derived alleles in a four population tree, where an excess of one pattern indicates gene flow from the third population. To test for ancient admixture, we ran all possible ABBA-BABA tests (n = 106) in ANGSD using the multipop version of the analysis that allowed individuals from the same evolutionary cluster to be analyzed together. We used the black rat as the outgroup for the test $D(H_1, H_2, H_3, \textit{R. rattus})$. We calculated *D*-statistics in program R (44) using the estAvgError.R file provided with ANGSD.

WGS Demographic Modeling

We inferred the demographic history of rats by modeling alternative scenarios that compared the observed and expected site frequency spectra (SFS) for multiple individuals from different global locations. We estimated the SFS for each evolutionary cluster, then output a joint SFS for each pair of clusters.

Given the large number of evolutionary clusters to model, we first modeled the relationship between *Eastern China*, *SE Asia*, *Aleutian*, and *Western North America* by comparing five four-population models and five five-population models that included an unsampled population (Figure S1). The best supported scenario (Model 6 in Figure S1) had a topology that included an ancestral unsampled ghost population with independent divergence of *Eastern China*, *SE Asia*, and the Pacific clusters. We then modeled four populations of a five-tree topology between an unsampled population, *SE Asia*, *Northern Europe*, *Western Europe*, and the *Expansion*. Our previous work on brown rat phylogeography suggested that rats expanded into Europe from SE Asia (15); thus, we tested the topology between the three European clusters arising from an unsampled population (likely in the Middle East) which was founded from SE Asia. The best supported scenario (Figure S2) had an initial divergence of *Western Europe* from the unsampled population, with *Northern Europe* and the *Expansion* diverging independently. For all testing

models, we did not allow population size to change through time, and we set the mutation rate at 2.5×10^{-8} mutations per generation.

We combined the best supported topologies from both sub-models into a nine-population model. We ran this model both with and without independent growth rates on each cluster, and log likelihoods were higher for the model without changes in growth parameters (although that may have been due to the reduction in parameter space), thus we report that model. Unlike in the sub-models described above, we allowed the mutation rate parameter to be estimated with the model. We used three generations per year to convert parameter estimates; all time calculations were done since 2015.

We ran 50 iterations of the nine-population model in fastsimcoal2, then identified the iteration with the highest estimated likelihood. Using these point estimates, we generated 500 samples of pairwise SFS each containing 30,000 markers that served as pseudo-observed data for estimating parameters ranges under the best supported model. We observed nine outlier iterations based on the difference between the maximum observed and estimated likelihoods in fastsimcoal2; thus, we removed those samples before calculating the 90% highest probability density (HPD) using the *HDInterval* v0.1.3 package (45) in R.

ddRAD-Seq Demographic Modeling

While our WGS had many more loci, there was limited geographic representation, as well as fewer individuals sampled; therefore, we built regional models from double digest restriction enzyme associated DNA sequencing (ddRAD-Seq; NCBI SRA BioProject PRJNA344413)) data to explore additional population tree topologies. We estimated the SFS of each genomic cluster or population of interest in ANGSD using the reference aligned Illumina reads instead of the previously called SNPs.

We built regional models within the evolutionary clusters for eastern Asia/Pacific, SE Asia, Northern Europe, and Western Europe. We used this reductive approach to limit the number of parameters being estimated. Within each region, we compared topologies between populations suggested by previous population structure analyses (15). We used the same fastsimcoal2 run

parameters as described above; however, we did not create pseudo-observed datasets for parameter estimation, unless noted, as our interest was in topology. A secondary reason we did not further explore population parameters within the regions was that we observed these datasets tended to overestimate divergence times, likely due to unsorted variation remaining within populations until coalescence with the unsampled ancestral population. Finally, we investigated population topology and admixture proportions in Vancouver, Canada and Berkeley, USA since each site was identified as admixed in our previous analysis (Table S2).

Effective population size through time

We estimated the change in effective population size over time in each evolutionary cluster using MSMC2 (25). To call variants, we used SAMTOOLS v1.3 (40) mpileup across all samples with a minimum mapping quality of 18 and the coefficient to downgrade mapping qualities for excessive mismatches at 50. We then utilized the variant calling in BCFTOOLS v1.3 with the consensus caller and excluded indels which limited the dataset to bi-allelic SNPs. As there was not a brown rat reference panel, we phased the 12 individuals plus two inbred lines (SS/Jr and WKY/NHsd; NCBI SRA accessions ERR224465 and ERR224470, respectively (6)) for each of the 20 autosomes using fastPHASE v1.4.8 (46).

We estimated change in N_e over time for each of the seven evolutionary clusters using two haplotypes for the *Aleutian* and *Western North American* clusters and four haplotypes for each other cluster. We also estimated the proportion of population divergence over time using the cross-population analysis, and combined results from individual populations with the cross-population analysis using the combineCrossCoal.py script provided.

DATA ACCESSIBILITY

Data for whole genome sequences from 10 brown rats available on NCBI SRA BioProject PRJNA344413.

ACKNOWLEDGEMENTS

This work was funded by National Science Foundation grants DEB 1457523 and MRI 1531639 to JM-S. The mammal collections at the University of Alaska Museum of the North, University

442 of California- Berkeley Museum of Vertebrate Zoology, the Burke Museum at the University of
443 Washington, and the Museum of Texas Tech University also graciously provided tissue samples.
444

FIGURE LEGENDS

Figure 1- (A) Plots of change in N_e over time using MSMC where the x-axis is years before the present. (B) Heterozygosity measured on each chromosome (n=20) for each of 12 brown rat samples. A box and whisker plot of mean chromosomal heterozygosity was overlaid the point estimates where colors indicate genomic cluster assignment. Additional details of sample locations available in Table S1. Each evolutionary cluster was represented by a different color: *eastern China*- dark brown; *SE Asia*- light brown; *Aleutian*- orange; *Western North America*- yellow; *Northern Europe*- purple; *Western Europe*- light blue; and *Expansion*- medium blue.

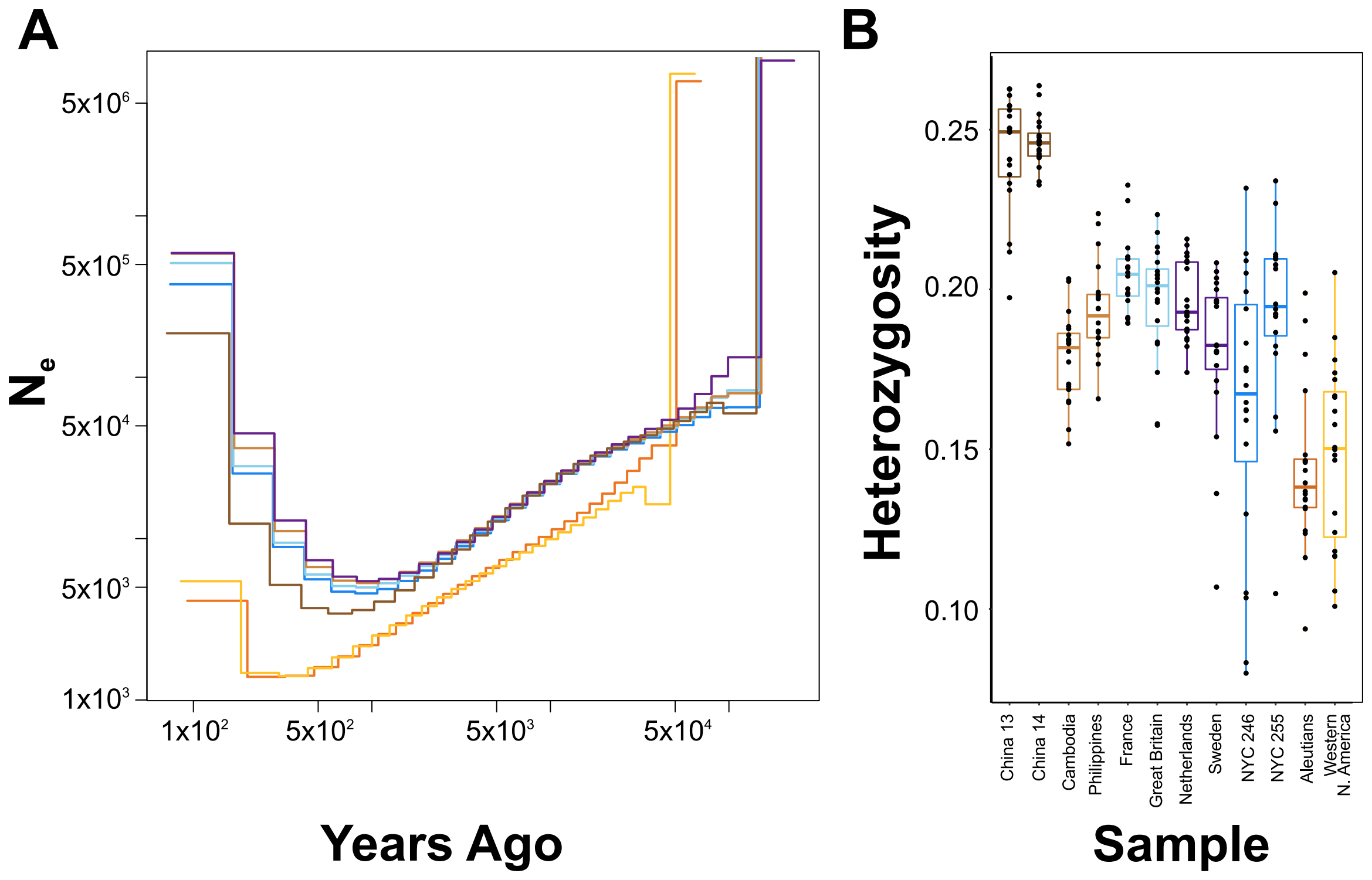
Figure 2- (A) The best supported demographic model contained nine evolutionary clusters inclusive of two unsampled populations. The divergence times in generations and N_e are listed in Table S4. (B) Map of global sampling locations including WGS (plus sign, +) and ddRAD-Seq (circles) samples, where locations utilizing both datasets had a plus within a circle. Evolutionary clusters were represented different colors: *Eastern China*- dark brown; *SE Asia*- light brown; *Aleutian*- orange; *Western North America*- yellow; *Northern Europe*- purple; *Western Europe*- light blue; and *Expansion*- medium blue.

REFERENCES

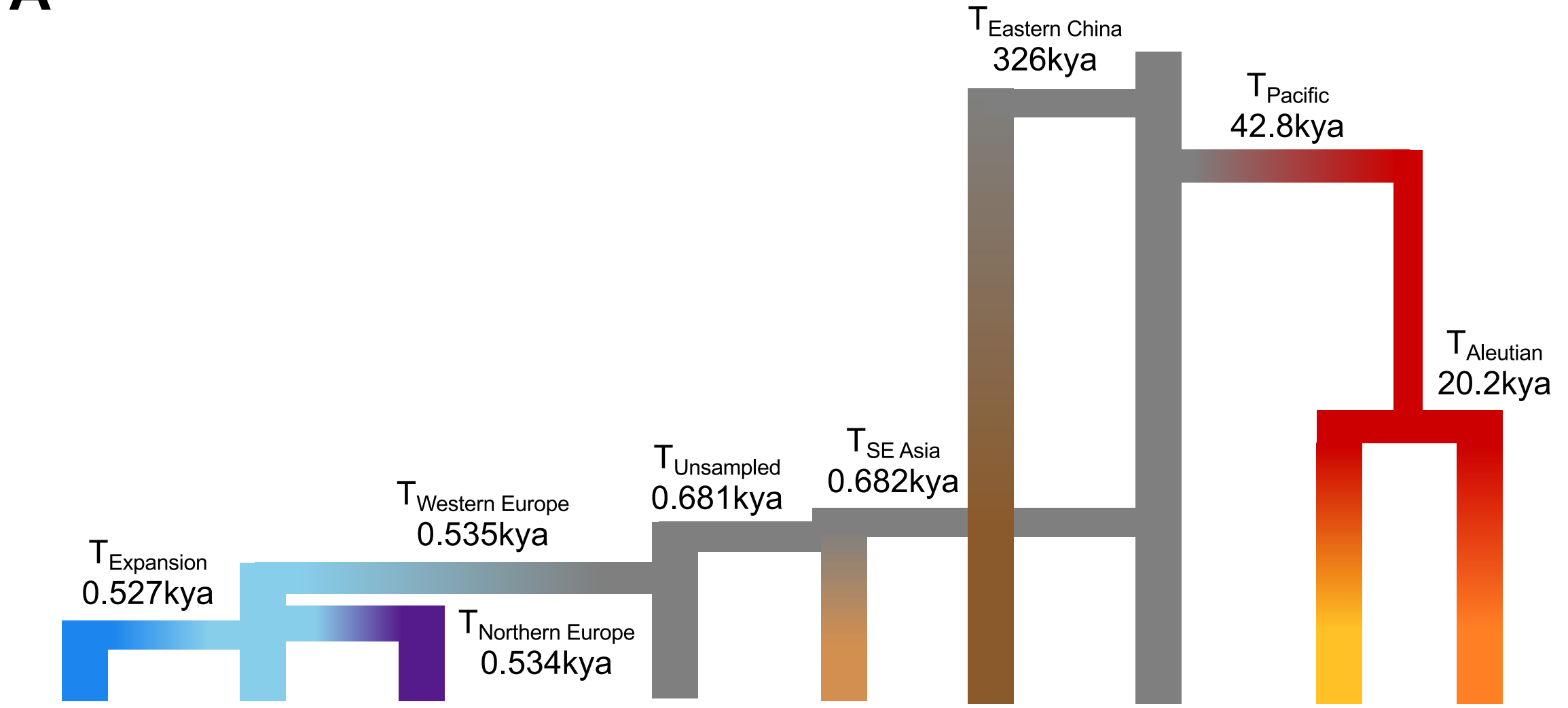
1. Smith AT & Xie Y (2008) *A Guide to the Mammals of China* (Princeton University Press, Princeton, NJ) p 544.
2. Himsworth CG, Parsons KL, Jardine C, & Patrick DM (2013) Rats, cities, people, and pathogens: a systematic review and narrative synthesis of literature regarding the ecology of rat-associated zoonoses in urban centers. *Vector borne and zoonotic diseases (Larchmont, N.Y.)* 13(6):349-359.
3. Pimentel D, Lach L, Zuniga R, & Morrison D (2000) Environmental and Economic Costs of Nonindigenous Species in the United States. *BioScience* 50(1):53-65.
4. Harper GA & Bunbury N (2015) Invasive rats on tropical islands: Their population biology and impacts on native species. *Global Ecology and Conservation* 3(Supplement C):607-627.
5. Jones HP, *et al.* (2016) Invasive mammal eradication on islands results in substantial conservation gains. *Proceedings of the National Academy of Sciences* 113(15):4033-4038.
6. Atanur Santosh S, *et al.* (2013) Genome Sequencing Reveals Loci under Artificial Selection that Underlie Disease Phenotypes in the Laboratory Rat. *Cell* 154(3):691-703.
7. Johnson MTJ & Munshi-South J (2017) Evolution of life in urban environments. *Science* 358(6363).
8. Armitage P (1993) Commensal rats in the New World, 1492-1992. *Biologist* 40:174-178.
9. Black L (1983) Record of maritime disasters in Russian America, Part One: 1741-1799. *Proceedings of the Alaska Maritime Archaeology Workshop, May 17-19, 1983*, (University of Alaska- Fairbanks).
10. Ervynck A (2002) Sedentism or urbanism? On the origin of the commensal black rat (*Rattus rattus*). *Bones and the man: Studies in honour of Don Brothwell*, eds Dobney K & O'Connor T (Oxbow Books, Oxford), pp 95-109.
11. Yalden DW (2003) Mammals in Britain - A historical perspective. *British Wildlife* 14(4):243-251.
12. Song Y, Lan Z, & Kohn MH (2014) Mitochondrial DNA phylogeography of the Norway rat. *PLoS ONE* 9(2):e88425.
13. Puckett EE, Micci-Smith O, & Munshi-South J (In Press) Genomic analyses identify multiple Asian origins and deeply diverged mitochondrial clades in inbred brown rats (*Rattus norvegicus*). *Evolutionary Applications*.
14. Lack J, Hamilton M, Braun J, Mares M, & Van Den Bussche R (2013) Comparative phylogeography of invasive *Rattus rattus* and *Rattus norvegicus* in the U.S. reveals distinct colonization histories and dispersal. *Biological Invasions* 15(5):1067-1087.
15. Puckett EE, *et al.* (2016) Global population divergence and admixture of the brown rat (*Rattus norvegicus*). *Proceedings of the Royal Society B: Biological Sciences* 283(1841):1-9.
16. Zeng L, *et al.* (2017) Out of Southern East Asia of the Brown Rat Revealed by Large-Scale Genome Sequencing. *Molecular Biology and Evolution*.
17. Tucker J (2015) *The Silk Road: China and the Karakorum Highway* (I.B.Tauris & Company, New York, USA) p 254.
18. Frantz LAF, *et al.* (2016) Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* 352(6290):1228-1231.

19. Shannon LM, *et al.* (2015) Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences* 112(44):13639-13644.
20. Vernesi C, *et al.* (2003) The genetic impact of demographic decline and reintroduction in the wild boar (Sus scrofa): A microsatellite analysis. *Molecular Ecology* 12(3):585-595.
21. Gautier M, *et al.* (2016) Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences. *Molecular Biology and Evolution* 33(11):2801-2814.
22. Estoup A, Beaumont M, Sennedot F, Moritz C, & Cornuet JM (2004) Genetic analysis of complex demographic scenarios: Spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution* 58(9):2021-2036.
23. Murray GGR, *et al.* (2017) Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science* 358(6365):951.
24. Jones EP & Searle JB (2015) Differing Y chromosome versus mitochondrial DNA ancestry, phylogeography, and introgression in the house mouse. *Biological Journal of the Linnean Society* 115(2):348-361.
25. Schiffels S & Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. *Nature Genetics* 46(8):919-925.
26. Deinum EE, *et al.* (2015) Recent evolution in *Rattus norvegicus* is shaped by declining effective population size. *Molecular Biology and Evolution* 32(10):2547-2558.
27. Teng H, *et al.* (2017) Population Genomics Reveals Speciation and Introgression between Brown Norway Rats and Their Sibling Species. *Molecular Biology and Evolution* 34(9):2214-2228.
28. Zheng B, Xu Q, & Shen Y (2002) The relationship between climate change and Quaternary glacial cycles on the Qinghai–Tibetan Plateau: review and speculation. *Quaternary International* 97-98(Supplement C):93-101.
29. Moreno-Mayar JV, *et al.* (2018) Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* 553:203.
30. Lieberman V (2009) *Strange Parallels: Southeast Asia in Global Context, c. 800-1830* (Cambridge University Press, Cambridge, Great Britain) p 947.
31. Heng D (2009) *Sino-Malay Trade and Diplomacy from the Tenth through the Fourteenth Century* (Ohio University Press, Athens, USA) p 286.
32. Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology* 13(4):827-836.
33. Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology* 14(1):67-73.
34. Aplin KP, *et al.* (2011) Multiple geographic origins of commensalism and complex dispersal history of black rats. *PLoS ONE* 6(11):e26357.
35. 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(5):375-390.
36. Gibbs RA, *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428(6982):493-521.
37. Li H & Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589-595.

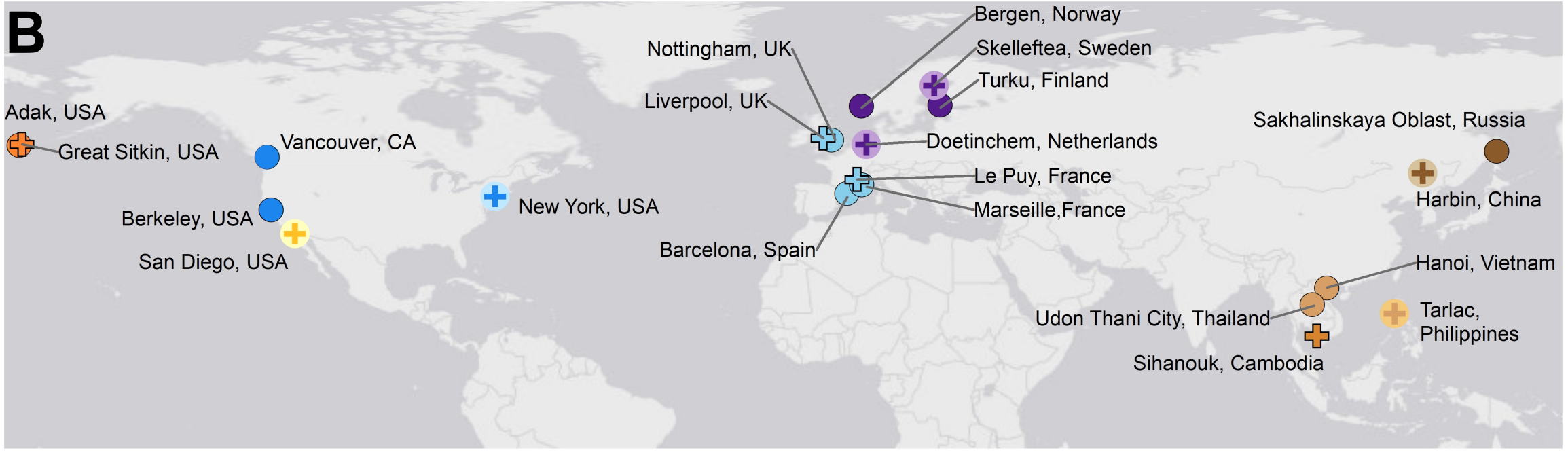
553 38. McKenna A, *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for
554 analyzing next-generation DNA sequencing data. *Genome Research* 20(9):1297-1303.
555 39. Korneliussen TS, Albrechtsen A, & Nielsen R (2014) ANGSD: Analysis of Next
556 Generation Sequencing Data. *BMC Bioinformatics* 15(1):1-13.
557 40. Li H, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
558 25(16):2078-2079.
559 41. Purcell S, *et al.* (2007) PLINK: A tool set for whole-genome association and population-
560 based linkage analyses. *The American Journal of Human Genetics* 81(3):559-575.
561 42. Chang CC, *et al.* (2015) Second-generation PLINK: rising to the challenge of larger and
562 richer datasets. *GigaScience* 4(1):7.
563 43. Green RE, *et al.* (2010) A draft sequence of the Neandertal genome. *Science*
564 328(5979):710-722.
565 44. R Core Team (2013) R: A language and environment for statistical computing (R
566 Foundation for Statistical Computing, Vienna, Austria).
567 45. Meredith M & Kruschke J (2016) Highest (Posterior) Density Intervals (CRAN, CRAN),
568 0.1.3.
569 46. Scheet P & Stephens M (2006) A fast and flexible statistical model for large-scale
570 population genotype data: Applications to inferring missing genotypes and haplotypic
571 phase. *American Journal of Human Genetics* 78(4):629-644.
572


A
B
N_e
$5x10^6$
$5x10^5$
$5x10^4$
$5x10^3$
$1x10^3$
$1x10^2$
$5x10^2$
$5x10^3$
$5x10^4$
Years Ago
Heterozygosity
0.25
0.20
0.15
0.10
China 13
China 14
Cambodia
Philippines
France
Great Britain
Netherlands
Sweden
NYC 246
NYC 255
Aleutians
Western N. America
Sample

A


T Eastern China
326kya
T Pacific
42.8kya
T Aleutian
20.2kya
T Unsampled
0.681kya
T SE Asia
0.682kya
T Western Europe
0.535kya
T Expansion
0.527kya
T Northern Europe
0.534kya
Expansion
Western Europe
Northern Europe
Unsampled
SE Asia
Eastern China
Ancestral
Western North America
Aleutian

B


Bergen, Norway
Nottingham, UK
Skelleftea, Sweden
Turku, Finland
Liverpool, UK
Adak, USA
Great Sitkin, USA
Vancouver, CA
Doetinchem, Netherlands
Sakhalinskaya Oblast, Russia
Le Puy, France
Harbin, China
New York, USA
Berkeley, USA
Marseille,France
Barcelona, Spain
San Diego, USA
Hanoi, Vietnam
Udon Thani City, Thailand
Tarlac, Philippines
Sihanouk, Cambodia