Genetic Adaptation in New York City Rats

Arbel Harpak[1,+], Nandita Garud[2], Noah A. Rosenberg[3], Dmitri A. Petrov[3], Matthew Combs[4,5], Pleuni S. Pennings[6,*] and Jason Munshi-South[4,+,*]

[1] Department of Biological Sciences, Columbia University, New York, NY

[2] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA

[3] Department of Biology, Stanford University, Stanford, CA

[4] Department of Biological Sciences, Fordham University, Armonk, NY

[5] Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY

[6] Department of Biology, San Francisco State University, San Francisco, CA

[+] Correspondence should be addressed to A.H. (arbel.arbel@gmail.com) or J.M.-S. (jmunshisouth@fordham.edu)

[*] P.S.P. and J.M.-S. jointly supervised this work

Abstract

Brown rats (*Rattus norvegicus*) thrive in urban environments by navigating the anthropocentric environment and taking advantage of human resources and by-products. From the human perspective, rats are a chronic problem that causes billions of dollars in damage to agriculture, health and infrastructure. Did genetic adaptation play a role in the spread of rats in cities? To approach this question, we collected whole-genome sequences from 29 brown rats from New York City (NYC) and scanned for genetic signatures of adaptation. We tested for (i) high-frequency, extended haplotypes that could indicate selective sweeps and (ii) loci of extreme genetic differentiation between the NYC sample and a sample from the presumed ancestral range of brown rats in northeast China. We found candidate selective sweeps near or inside genes associated with metabolism, diet, the nervous system and locomotory behavior. Patterns of differentiation between NYC and Chinese rats at putative sweep loci suggests that many sweeps began after the split from the ancestral population. Together, our results suggest several hypotheses on adaptation in rats living in close proximity to humans.

Background

Urbanization has the potential to drive dramatic ecological and evolutionary consequences for wildlife [1, 2]. The most commonly examined evolutionary responses to urbanization are changes in gene flow and in the intensity of genetic drift [3, 4, 5, 6]. A small but growing number of studies have also examined adaptive evolution in urban environments, including morphological adaptations to urban infrastructure [7, 8], adaptive life history changes, thermal tolerance to warming cities [9, 10], and behavioral changes [11]. The recent adoption of genomic scans to identify loci under positive selection can help generate hypotheses about adaptive phenotypes in cities [12, 13].

Commensal rodents—particularly house mice (*Mus musculus*), black rats (*Rattus rattus*), and brown rats (*Rattus norvegicus*)—are the most widespread urban mammals besides humans and a notorious threat to urban quality of life [14, 15]. Recent analyses have revealed some of the relationships between invasive urban rodent populations that spread around the world with humans [16, 17, 18, 19] and the influence of heterogeneous urban environments on gene flow [20, 21]. Much less is known about the role of natural selection in the success of commensal rodents in cities.

Brown rats have considerably less genetic diversity compared to house mice, possibly due to a population bottleneck ∼20,000 years ago [22]. More recently, they have experienced major range expansions, presumably due to their association with agrarian, and later urban, human societies [19]. After reaching Europe around 500 years ago, brown rats rapidly became a prominent urban pest and then spread throughout Africa, the Americas, and Australia as a side effect of European colonialism in the 18th and 19th centuries. These introductions to coastal cities, followed by rapid industrialization and urbanization, potentially exerted strong selection on rat populations. Urban environments changed most dramatically from the late 19th into the 20th century–a period that spans around 500 rat generations. A recent

example is found in evidence for a significant change in rat cranial shape in New York City over a 120 year period [23].

It is likely that selection influenced a number of traits in these expanding populations. For example, multiple rat populations exhibit resistance to first-generation anticoagulant rodenticides (such as Warfarin) associated with nonsynonymous substitutions in the VKORC1 gene [24], though resistance to recent second-generation anticoagulants is less widespread and may not be monogenic [25]. Comparisons of Asian and European rats have suggested that many immune-response genes are highly differentiated between these regions, potentially due to selection related to disease pressures [26]. Rats in NYC evolved longer noses—which have been interpreted as adaptations to cold, and shorter upper tooth rows—which were interpreted as adaptations to higher quality, softer diets [23]. Although no study to date has examined the genomic signatures of selection in urban rats, studies of another rodent in NYC, the white-footed mouse (*Peromyscus leucopus*), provide additional hypotheses on the traits subject to selection in urban rodents. Namely, immune response, detoxification of exogenous compounds, spermatogenesis, and metabolism genes were over-represented among putatively-selected regions [27, 28, 29]. This same *Peromyscus* population exhibited shorter toothrows consistent with a change in diet [30]. These findings suggest that some genetic adaptations in urban rodents have arisen in response to increased disease pressure in dense urban settings, increased exposure to pollutants and shifts to novel diets.

Here, we perform a genomic scan for adaptation in an urban rat population. We search for signatures of recent selective sweeps—long haplotypes at a high frequency [31, 32, 33]—in a sample of rats from NYC. We also search for adaptive changes by identifying regions of extreme differentiation between our NYC sample and a second sample from northeast China—the presumed ancestral range of the species [16, 19]. We find evidence for recent selective sweeps near genes associated with metabolism of endogenous compounds, diet, apoptotic processes, animal organ morphogenesis and locomotory behavior.

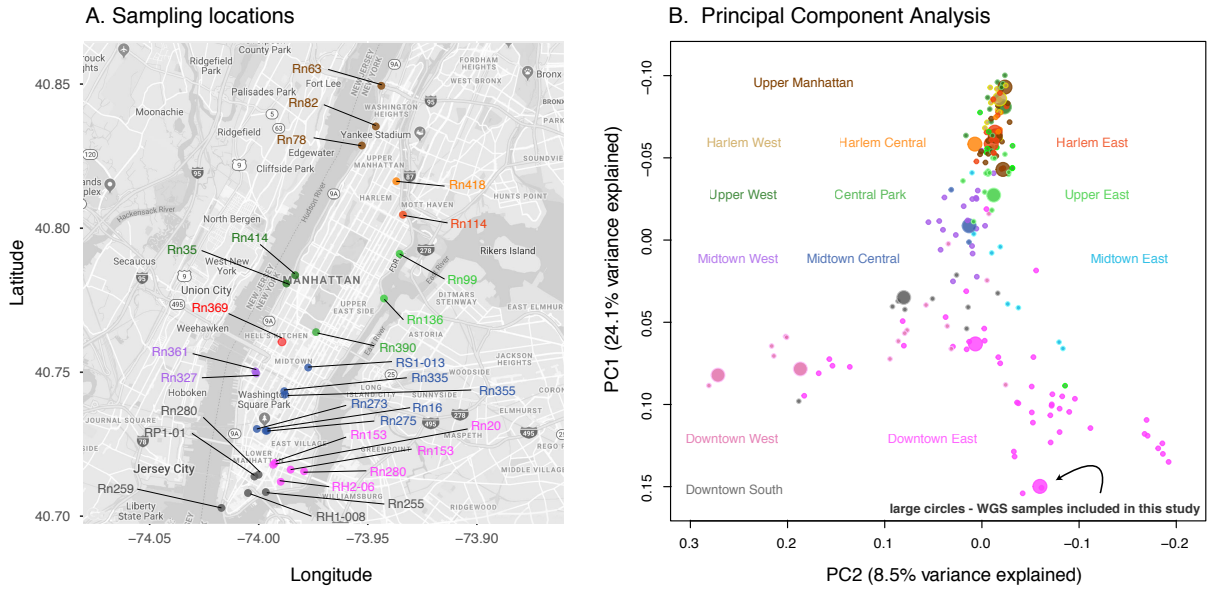

Figure 1: Sampling locations and genetic population structure of NYC brown rats. **(A)** Sampling sites of the 29 rat whole-genome samples used in our selection scans. Color coding corresponds to NYC neighborhoods as in panel (B). **(B)** Principal component analysis of 198 rats reanalyzed from Combs et al. [20], with the 15 rats included in both the Combs et al. [20] study and this study represented by large circles. In both panels, color coding corresponds to NYC neighborhoods.

Results

To look for selective sweeps in NYC brown rats, we sequenced whole genomes of 29 rats trapped throughout Manhattan, New York City, USA. The animals were chosen to represent the geographic distribution of Manhattan rats while excluding genetic relatives (**Fig. 1**) [20]. The mean coverage was $\geq$ 15x for each of the 29 rats. The mean nucleotide diversity was $0.168 \pm 0.003\%$ (point estimate $\pm$ standard error), slightly lower than the $0.188 \pm 0.0001\%$ estimated for a population at the presumed ancestral range of brown rats in Harbin, China (**Supplementary Materials**) [22].

We searched for selective sweeps using the methods G12 [34, 35] and H-scan [36]. G12 and H-scan both measure the homogeneity of haplotypes in the sample around a focal SNP (**Methods**).

High haplotype homogeneity is a signature of a recent (or sometimes even ongoing) selective sweep. We chose to use G12 and H-scan because they do not require prior phasing of the genotype data—a potential challenge in a small sample. Instead of relying on an assumption of perfectly phased data, G12 and H-scan measure *multi-site genotype homogeneity*. It has been recently argued that—under a wide range of selection parameters and demographic histories—methods using unphased data are almost as statistically powerful as methods based on perfectly phased data [35, 37]. G12 and H-scan were correlated (Spearman $\rho = 0.52$; $p < 2.2 \times 10^{-16}$), and they tended to peak at similar regions of the genome—suggesting that they detect similar signals of frequent, long segregating haplotypes (**Fig. 2; Fig. S1A**).

A common approach to evaluating significance of evidence for selection is to test if an empirical value of a statistic is a likely sample from a specified null distribution. The null distribution can be estimated using simulations of neutral evolution under an inferred demographic model. We initially took this approach, but found that the simulated null distribution of G12 was far from the empirical one, suggesting that the null model was poorly calibrated (**Fig. S2**; **Supplementary Materials**).

We therefore took an alternative approach of focusing on the most extreme empirical values of the selection statistics as putative targets of adaptation (**Fig. S7**; we note that we still report the neutral simulation results (**Table S1**) and p-values computed using the simulation-based null distribution (**Table S6**)). We calculated G12 and H-scan values across the whole genome (**Files S3,S4**). We identified the 100 top-scoring loci with each method, iteratively masking regions around previously-called loci to avoid correlated signals due to linkage disequilibrium (**Supplementary Materials**). Multi-site genotypes are visibly more homogeneous in top-scoring loci (**Fig. 3B,C**) than in random loci (**Fig. 3A**). In some of the top-scoring loci, we observe a single common haplotype—consistent with the expectation under a hard selective sweep [38, 39, 40, 41] (**Fig. 3B**). In many others, more than two

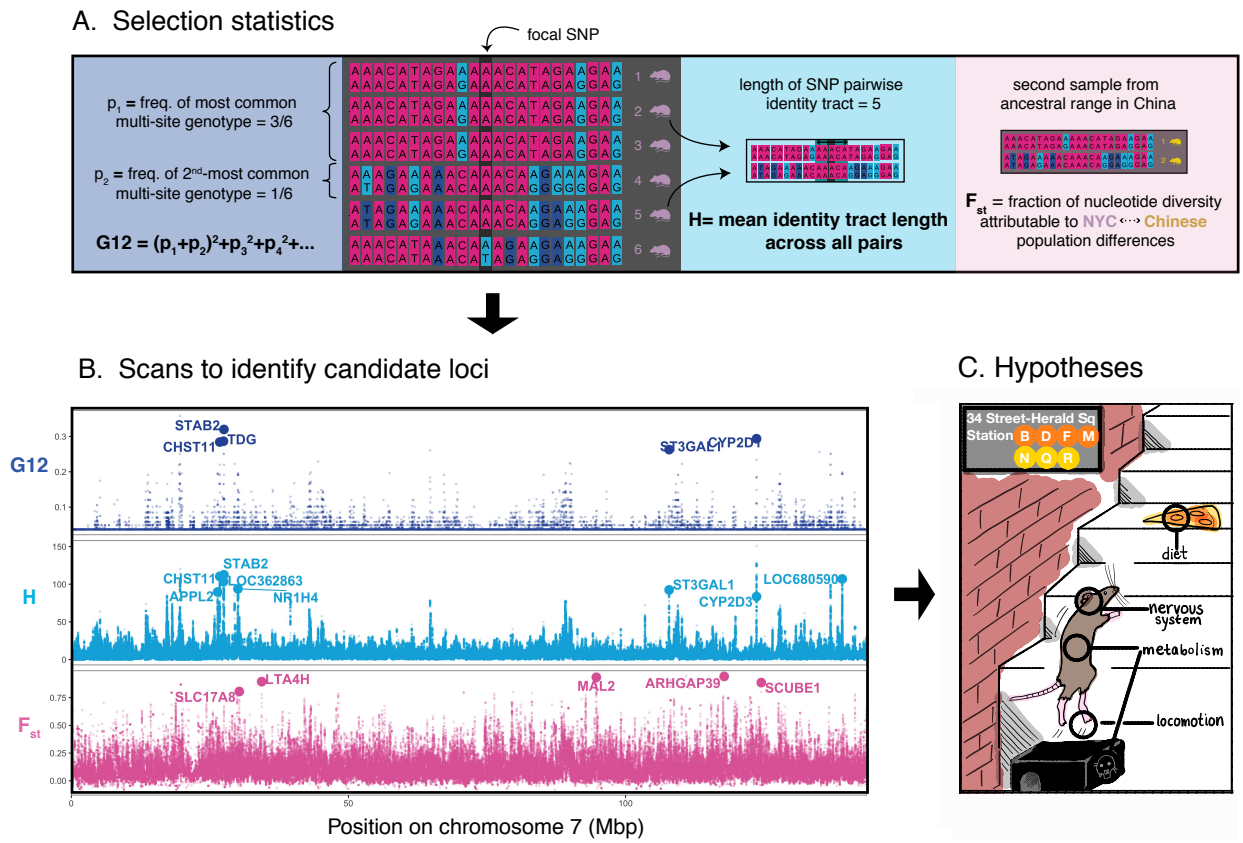

Figure 2: Scanning for signals of adaptation. **(A)** Three selection statistics were used to identify signals of selection. Two of the statistics, G12 and H-scan, detect loci with low haplotype diversity (high homogeneity)—a signature of recent or ongoing selective sweeps. G12 is defined similarly to the homozygosity of multi-site genotypes (in a window of a fixed number of SNPs)—a sum of squared frequencies of multi-site genotypes—except that the two most frequent genotypes are grouped (summed and squared) together. In the illustrated example, the most common multi-site genotype is shared by rats 1,2 and 3. The second-most common genotype is only carried by one rat. H is the mean length of the maximal pairwise identity tract containing the focal SNP—where the mean is taken across all pairs in the sample. In the illustrated example, the tract is 5 SNPs-long for rats 2 and 5. Finally, F_{st} was used to measure genetic differentiation between the NYC sample and a sample from the presumed ancestral range of brown rats in northeast China. Extreme differentiation at a locus may also be indicative of selection since the populations' split. **(B)** Values of the three statistics along chromosome 7. G12 and H-scan peak at similar regions of the genome, suggesting that both statistics pick up on similar signals of frequent, long segregating haplotypes. Large circles denote the locations of candidates: top-scoring loci that are also less than 20kb away from protein-coding genes (names of the corresponding genes are noted). Corresponding plots for all other autosomes can be found in Supplementary File S1. **(C)** Functions associated with candidate genes suggest traits that may have been subject to selection in NYC rats, including diet, the nervous system, metabolism and locomotion.

common haplotypes are segregating in the population—consistent with the expectation under a soft selective sweep [42, 43] (**Fig. 3C; File S2**).

To gauge the timing of putative selective sweeps, we compared the multi-site genotypes of our sample to those of 9 rats from Harbin, Heilongjiang Province, China—the presumed ancestral range of the species [22]. The Chinese genotypes are often heterogeneous at the loci most homogeneous in the NYC sample (**Fig. 3D**). Only 5/17 of the H-scan candidates show significantly elevated H-scores in the Chinese sample (**Fig. S4; File S6**). These patterns are consistent with most of these putative selective sweeps occurring in ancestors of the NYC sample after the split from the ancestral population in China (Multi-site genotype visualizations for all G12 and H-scan candidate loci can be found in **File S2**).

We next took a partially complementary approach to find candidate targets of adaptation: searching for regions of extreme genetic differentiation between our NYC sample and the Chinese sample. We measured differentiation using mean per-SNP F_{st} [44] across SNPs in 10kb sliding windows (**File S5**), and again focused on the 100 top-scoring loci genome-wide (**Table S3**). In the **Supplementary Materials** we discuss the relationship between top-scoring loci of one statistic and scores in the other. In short, if a genotype homogeneity peak is due to a recent selective sweep that has occurred after the split from the ancestral population, then we expect high F_{st} values in the same region. Indeed, F_{st} values are slightly elevated (80th percentile for H-scan, 76th percentile for G12) in genotype homogeneity candidates (**Fig. S1A**). This pattern is consistent with a selective sweep elevating local differentiation from the Chinese population, although we note that even if no selective sweeps have occurred, then F_{st} is expected to be somewhat elevated from the genome-wide baseline at loci ascertained for high homogeneity in the NYC sample [45].

We next asked if we could associate biological functions to these top-scoring loci. For each candidate locus, we identified a single closest gene. The top-scoring G12 and H-scan loci are farther from protein-coding genes than is expected under a permutation-based null (Wilcoxon $p < 0.004$ for both, but not for F_{st}; **Fig. 2C**; **Supplementary Materials**). This result may suggest that genic sweeps are rare—at least among loci identified with our

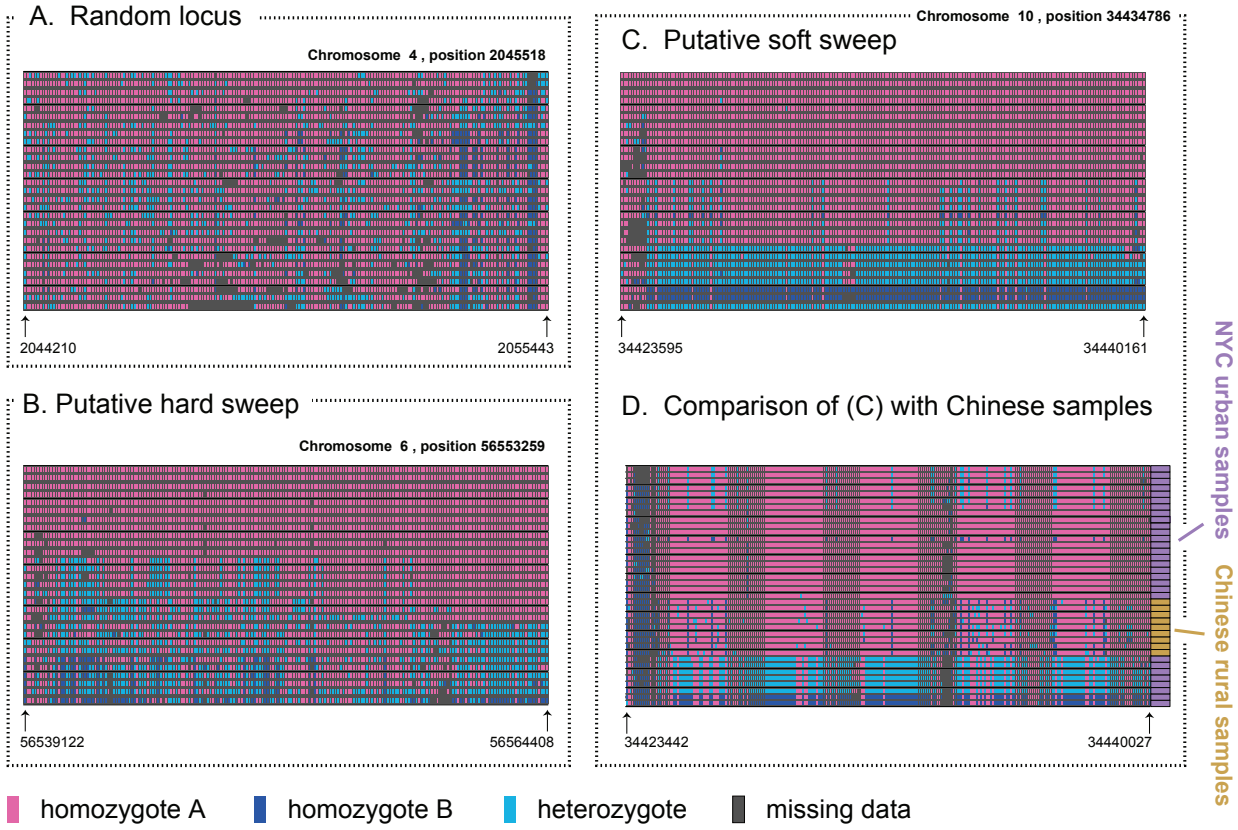

Figure 3: Multi-site genotype structure. Each row confers to a multi-site genotype of one rat. **(A)** At a randomly-chosen locus, there is no clear clustering of multi-site genotypes. **(B)** An example of a locus that was detected as a putative target of selection with H-scan, in which there is one multi-site genotype frequent (11/29) in the NYC sample. This pattern is consistent with a hard selective sweep. **(C)** A candidate locus identified with G12 in which there are more than two frequent multi-site genotypes—consistent with a soft selective sweep. **(D)** A comparison of the NYC samples to a sample of 9 rats from the presumed ancestral range of the species in China, around the same focal site as in panel (C). Multi-site genotypes are relatively heterogeneous in the Chinese sample where homogeneity is observed in the NYC sample. This observation is consistent with a recent onset of selection after the split from the ancestral population in China.

methods.

Moving forward, we only considered protein-coding genes $\leq 20kb$ away from top-scoring loci as potential targets of adaptation—leaving 19, 17 and 32 candidates for G12, H-scan and F_{st} respectively (**Tables S1-S3**). One G12 candidate lies close to a a cluster of Cytochrome P450 genes (It is specifically closest to CYP2D1, but there are five CYP2D genes $\leq 50kb$ away). CYP-encoded enzymes help metabolize endogenous compounds and detoxify exogenous chemicals [46] and have been associated with Warfarin resistance [47, 48]. This

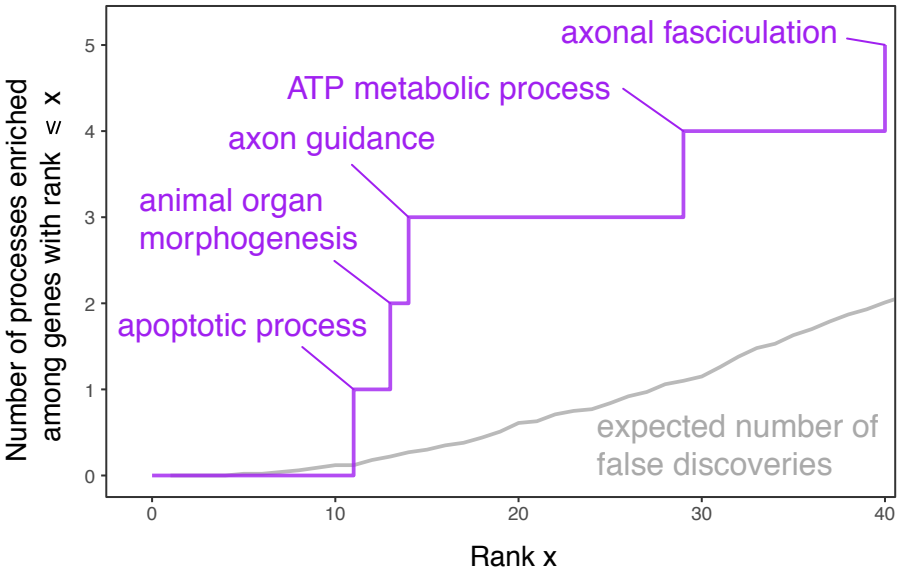

Figure 4: Biological processes enriched among 40 top-scoring genes. This figure focuses on candidate targets of adaptation arising from a comparison of G12 values in genes. Each gene was given the maximum score of all focal SNPs in the gene, and genes were ranked by this maximal score. We consider Gene Ontology (GO) biological processes associated with at least three top-scoring genes as enriched, and the y-axis shows their cumulative count. Blue text labels show GO biological processes associated with exactly three of the (x value) top-scoring genes. The gray line shows the expected number of false (biological process) discoveries among the top X candidates, estimated based on permutations of gene scores.

gene family has expanded substantially in rodents [49].

Several of the genes—3 of 17 for H-scan and 2 of 19 G12 candidates—are olfactory receptor genes. Changes in the olfactory nervous system can alter odor perception and behaviors driven by precise chemical cues—and have been suggested to be targets of adaptation in numerous mammalian populations [50]. However, this is not a significant enrichment compared with a null generated by permuting the location of the candidate loci along the genome (Fisher's Exact Test; **Supplementary Materials**).

We do, however, find Gene Ontology (GO) biological processes [51] that are enriched among top-scoring genes. Among the top fourteen G12 values inside genes (coresponding to 0.27 expected false discoveries, **Methods**), three biological processes are associated with at least three different genes: apoptotic process, animal organ morphogenesis and axon guidance (**Fig. 4**; largely overlapping results using H-scan instead of G12 are shown in **Fig. S3**).

One potential candidate that we had considered before performing our scans was VKORC1, as previous research has pointed to selective sweeps at this locus in rodents following the broad application of Warfarin as a rodenticide [24, 52]. The Chinese and NYC rats were highly differentiated at a 10kb window centered on VKORC1 (98th chromosomal percentile of F_{st}). At the same time, multi-site genotypes around VKORC1 did not appear homogeneous, and G12 and H-scan are close to the chromosomal medians (**Fig. S5**). Zooming in on the VKORC1 coding sequence, we found that no known resistance-conferring alleles appear to have fixed in the population (**Supplementary Material**), and, further, that the data contained no nonsynonymous SNPs. To our knowledge, all of the experimentally-confirmed resistance-conferring variants in VKORC1 are nonsynonymous [24]. We detected 10 intronic SNPs, and two synonymous SNPs (in codon 68 coding for a Histidine and in codon 82 coding for Isoleucine)—both previously observed in NYC rats [53] (**Fig. S6**).

Discussion

We performed scans for adaptation in NYC rats and identified the top-scoring loci as potential targets of recent selective sweeps. Near or within top-scoring loci, we find genes associated with the nervous system, metabolism apoptotic processes and morphogenic traits.

One trait of central interest as a target of adaptation is rodenticide resistance. Some of the genes we identified near candidate loci (e.g. the CYP2D gene cluster and AHR; **Tables S1-S5**) may be associated with rodenticide resistance. In VKORC1—which we had a priori considered as a possible target of such adaptation—we find no evidence for a recent selective sweep (though levels of differentiation between Chinese and NYC rats in the genomic region surrounding VKORC1 were high, see **Fig. S5**). We also detect no nonsynonymous polymorphisms in the gene (**Fig. S6**; but see [53] for evidence of two nonsynonymous variants

associated with resistance segregating at low frequencies in NYC). Warfarin became the rodenticide of choice in New York in the late 1950s [54], but was gradually phased out in the late 1970s and replaced by more effective second-generation anticoagulant rodenticides (SGAR) after high levels of resistance were reported from Europe [55, 56] and the United States [57], and specifically New York [58]. It is possible that sweeps that have occurred during these few decades are too old to be detectable with our methods, although resistance-associated VKORC1 polymorphisms are still segregating in some European populations [24]. Another possibility is that a VKORC1 sweep did not occur in the NYC population.

Natural populations of brown rats have been reported to show resistance to SGAR as well [59, 60]. VKORC1 may play a role in SGAR resistance—but likely a smaller role than it does for first-generation anticoagulants [25]: lab rats homozygous for the resistance mutation Y139C showed lower mortality in response to some SGARs (difenacoum and bromadiolone), but not others (brodifacoum, difethialone and flocoumafen) [61]. At the same time, rats resistant to the SGAR bromadialone showed increased expression of a suite of CYP genes [62]. We may therefore hypothesize that the signals of adaptation we observe near the CYP2D gene cluster could underlie changes in response to SGAR through the clearance of exogenous compounds. However, to our knowledge, levels of anticoagulant (including warfarin) resistance in NYC rats have not been quantified in recent decades.

These genes could also be under selective pressure from other environmental toxins, inflammatory responses, or novel dietary items. For example, one of our candidate genes, AHR, has been implicated in adaptation to polychlorinated biphenyl (PCB) contamination in killifish and tomcod in the Hudson and other rivers in the area [63, 64]. On a possibly related note, our GO analysis identified the nervous system genes as a possible target (**Fig. 4,S3**). This putative adaptation may also relate to neuronal or hormonal responses to the environment, but may also underlie behavioral changes.

Other candidate genes and biological processes highlighted in **Figs. 4,S3 and Tables**

S1-3 are associated with many phenotypes in humans and lab rodents that are also plausible targets of selection. However, the sheer number of phenotypic associations for each gene makes it difficult to generate clear hypotheses about phenotypes under selection in wild urban rats. In addition, in the absence of a sample from rural counterparts of our urban study population, we cannot discern whether any of the adaptations are driven by the urban environment with certainty. Even with this caveat, there are some potentially promising leads that follow from this study. As one example, the gene CACNA1C (7th highest score among genes for both G12 and H-scan, **Tables S4-5**) has been repeatedly associated with psychiatric disorders in humans, and the effects may be modulated by early-life stress [65]. Thus, rat behavioral phenotypes associated with anxiety (such as anti-predator responses or responses to novel stimuli) are promising areas for future investigation. CACNA1C has also been demonstrated to influence social behavior and communication in rats [66]. CACNA1C and multiple other top-scoring genes like FGF12, EPHA3 and CHAT (**Table S4**) may also affect locomotion in rodents [67, 68]. The gait or other locomotory phenotypes could have also undergone adaptive changes, given that urban rats must move through a highly artificial, constructed environment that differs markedly from naturally vegetated habitats.

Given that urban rats are so closely associated with human city dwellers, future work might explicitly address whether rats respond to the pressures of city living that are also experienced by humans. One striking commonality between urban humans and rats is their diet. Today, the human urban diet contains an increasingly large proportion of highly processed sugars and fats that lead to a number of public health concerns. Some of these health concerns could conceivably apply to rats as well [69, 70, 71]. Indeed, several of the candidate genes we identified, ST6GALNAC1 (G12, **Table S1**), CHST11 and GBTG1 (H-scan, **Table S2**), are linked with oligosaccharide or carbohydrate metabolic processes.

The main approach we used to identify targets of adaptation was to search for long haplotypes segregating at a high frequency. The results of our scans should be viewed only

as preliminary hypotheses for future testing: in the absence of a sensible null model, the specificity of our genotype-homogeneity outlier approach remains unclear. In addition, the power of our approach to detect genetic adaptation as a whole is limited; it is most powerful for ongoing or recent selective sweeps—and the definition of recency here is also a function of the sample size (as larger samples can reveal more recent coalescence events) [35, 72]. Other modes of adaptation, such as polygenic adaptation [73, 74, 75, 76], may be common in NYC rats, but entirely missed in this study. For example, the putative adaptation of cranial morphology identified in NYC rats is likely polygenic [23]. Therefore, there is ample room for detecting targets and quantifying the mode of adaptation in urban populations in future research.

With the increasing urbanization of our planet, the effects of human activity on animals that inhabit cities merit further attention. Genetic adaptation and phenotypic plasticity have both been suggested as crucial drivers of success for rats and other species in urban settings—including great success in their capacity as human pests. Our results suggest that in the case of NYC brown rats, this adaptation—at least in part—could be rooted in rapid changes in the genetic composition of the population through natural selection.

Acknowledgements

This work was funded by a fellowship from the Simons Society of Fellows (#633313) and a fellowship from the Stanford Center for Evolutionary and Human Genomics (CEHG) to A.H., National Science Foundation (NSF) grant DBI-1458059 to N.A.R. and D.A.P. and NSF grants DEB 1457523 and MRI 1531639 to J.M.-S. We thank Emily Puckett for her assistance in implementing demographic models from [19]. We thank Andrés Bendesky, Zach Fuller, Ana Pinharanda, Molly Przeworski and Nasa Sinnott-Armstrong for comments on the manuscript.

Methods

Data Availability

A Variant Calling Format (VCF) file that includes both the NYC sample and the Chinese sample, as well as all Supplementary Files can be found at `https://doi.org/10.5061/dryad.08kprr4zn`.

Data collection

The NYC brown rat samples used in this study were a subset of nearly 400 rats collected previously for population genomic studies. Full details of this sampling effort are available in Combs et al. [20]. In brief, we trapped brown rats across the island of Manhattan in NYC between June 2014 and December 2015. Rats were trapped using lethal snap traps baited with a mixture of peanut butter, oats, and bacon, housed in bait stations (Bell Labs) and set for 24-hr periods. Traps were typically set outside earthen burrow systems or other areas with signs of ongoing rat activity. We then sampled 3-4 cm of tail tissue that was stored in 70% ethanol for downstream genetic analyses.

For comparison with the NYC sample, we used publicly-available whole genome sequences of 11 brown rats sampled from a 500 km^2 area around the city of Harbin, Heilongjiang Province, China (European Nucleotide Archive ERP001276). Two BAM files were corrupt and our analysis is therefore based on 9 of the 11 Chinese samples. Heilongjiang Province is presumed to represent the ancestral range of *R. norvegicus*. This sample was previously used to examine the demographic history of brown rats [22]. Sequencing, mapping, variant calling and filtering were produced using a similar bioinformatic processing pipeline that we used for the NYC whole-genome samples and that is described below.

Sequencing, read alignment, and variant calling

We sent RNAse-treated DNA extracts from 29 rat samples to the New York Genome Center for whole-genome resequencing on an Illumina HiSeq 2500. Rat samples were chosen to represent the diversity of rat colonies sampled across Manhattan (**Fig. 1**). The sequencing was designed to achieve at least 15X coverage for each sample. After sequencing, the NY Genome Center aligned cleaned and filtered reads to the *R. norvegivus* v. Rn 5.0 reference genome using BWA, and created a sorted, aligned BAM file using *Picard toolkit* [77]. Duplicate reads were then removed using *Picard toolkit*, followed by base quality score recalibration and local realignment around indels using *GATK* v3.7 [78].

We also used *GATK* for variant calling in both the NYC and Chinese samples. We made a file containing genotype information available for download from `https://doi.org/10.5061/dryad.08kprr4zn` in Variant Calling Format (VCF) for the *Rattus norvegicus* samples used in the study, including the 29 NYC individuals and the 9 Chinese individuals of the Deinum et al. study [22]. In the **Supplementary Materials**, we describe the bioinformatic procedures to produce this VCF.

Selection statistics and calling candidate loci

For each SNP called in the NYC sample, we computed two statistics: G12 [33, 35], and H-scan [36]. For a sample of n diploids, both methods take as input n multi-site genotypes, where at each SNP, and for each individual, the two nucleotide alleles are replaced by pseudo-alleles corresponding to one of the (two or three) unique genotypes.

G12 examines a symmetric window of a fixed number of SNPs around each focal SNP. Denoting the frequency of the k unique multi-site genotypes at a window as $p_1, p_2, p_3, ..., p_k$,

ranked from most common to most rare (**Fig. 2A**), G_{12} is defined as

$$G_{12} = (p_1 + p_2)^2 + p_3^2 + ... + p_k^2.$$

We used a window size of 201 SNPs to compute G12 values, conferring to a mean length of 29kbp and mean linkage disequilibrium (r^2) of 0.135 between the SNPs at the ends of the window (**Fig. S10**).

H-scan is the mean pairwise identity tract length across all pairs of individuals in the sample. Namely, for a given focal site, let h_{ij} denote the length (in some distance metric, see below) of the maximal tract of pseudo-allele identity between individuals i and j that includes the focal site (and zero if no such tract exists—i.e. the pseudo-allele differs at the focal site). H is defined as the average of h_{ij} across all pairs,

$$H = \frac{1}{\binom{n}{2}} \sum_{i<j} h_{ij}.$$

To run H-scan [36], a distance metric must be defined for the length of an identity-by-state tract for a given pair of individuals (h_{ij}). We set the metric to be the number of SNPs rather than physical distance (base pairs) or genetic distance—an option that requires the input of a genetic map.

Lastly, we computed mean Weir-Cockerham per-SNP F_{st} in windows of 10kb with a 1kb step. In the **Supplementary Materials**, we provide further detail on pre-processing procedures, handling missing data and the method used to define candidate loci around top-scoring SNPs.

Biological process enrichment analysis

In **Fig. 4** and **Fig. S3**, we examine whether top-scoring genes are enriched for any Gene Ontology (GO) biological process categories. We downloaded protein-coding gene annotations for the Rn5 reference genome from the UCSC genome browser [79]. We scored each gene by the maximal value of the statistic (G12 or H-scan) among focal SNPs inside the gene. We then ranked the genes by their score. **Table S4** (H-scan) and **Table S5** (G12) detail the gene scores, ranks, and associations. We considered a biological function category as enriched (considered as a "discovery") if it was associated with at least three genes among the 50 top-ranking genes. In order to avoid spurious enrichment due to autocorrelated genotype homogeneity, we iteratively masked regions around genes in categories identified as enriched. Namely, with each new biological function category that we considered as enriched, we masked regions (i.e., excluded genes in these regions) that are within 30kb upstream and downstream of the three genes yielding the enrichment. This masking threshold is a modification of the masking approach we used for the genome-wide scans (**Supplementary Materials**), as we have found that without this modification, some gene clusters (e.g. the CYP2D cluster), that may also share biological process annotations, were over-represented due to shared G12 or H-scan signals.

We estimated the false discovery rate (FDR) for considering enriched categories among the top r ranked genes as "discoveries" (enriched biological processes). This FDR, appearing in grey in **Fig. 4** and **Fig. S3**, is the number of biological process categories that are expected to be associated with at least 3 of the genes ranked $1, ..., r$ by chance alone, i.e. if ranks are independent of scores. Therefore, we simulated the null distribution of the number of discoveries by permuting scores across genes, reranking and counting the number of categories associated with at least 3 of the genes in the top r ranks. The FDR was then estimated as the average number of categories observed across 100 independent iterations of this permutation

procedure. This permutation approach tests the null of independence of biological function and selection score, while controlling for the co-dependence of biological process categories by maintaining their joint distribution of the number of categories associated with each gene.

References

[1] Karen C Seto, Burak Güneralp, and Lucy R Hutyra. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*, 109(40):16083–16088, 2012.

[2] L Ruth Rivkin, James S Santangelo, Marina Alberti, Myla FJ Aronson, Charlotte W de Keyzer, Sarah E Diamond, Marie-Josée Fortin, Lauren J Frazee, Amanda J Gorton, Andrew P Hendry, et al. A roadmap for urban evolutionary ecology. *Evolutionary Applications*, 12(3):384–398, 2019.

[3] Marc TJ Johnson and Jason Munshi-South. Evolution of life in urban environments. *Science*, 358(6363):eaam8327, 2017.

[4] Jason Munshi-South, Christine P Zolnik, and Stephen E Harris. Population genomics of the anthropocene: Urbanization is negatively associated with genome-wide variation in white-footed mouse populations. *Evolutionary Applications*, 9(4):546–564, 2016.

[5] Joscha Beninde, Stephan Feldmeier, Maike Werner, Daniel Peroverde, Ulrich Schulte, Axel Hochkirch, and Michael Veith. Cityscape genetics: structural vs. functional connectivity of an urban lizard population. *Molecular Ecology*, 25(20):4984–5000, 2016.

[6] Lindsay S Miles, J Chadwick Johnson, Rodney J Dyer, and Brian C Verrelli. Urbanization as a facilitator of gene flow in a human health pest. *Molecular Ecology*, 27(16):3219–3230, 2018.

[7] Kristin M Winchell, R Graham Reynolds, Sofia R Prado-Irwin, Alberto R Puente-Rolón, and Liam J Revell. Phenotypic shifts in urban areas in the tropical lizard anolis cristatellus. *Evolution*, 70(5):1009–1022, 2016.

[8] Charles R Brown and Mary Bomberger Brown. Where has all the road kill gone? *Current Biology*, 23(6):R233–R234, 2013.

[9] Kristien I Brans, Robby Stoks, and Luc De Meester. Urbanization drives genetic differentiation in physiology and structures the evolution of pace-of-life syndromes in the water flea Daphnia magna. *Proceedings of the Royal Society B: Biological Sciences*, 285(1883):20180169, 2018.

[10] Sarah E Diamond, Lacy D Chick, Abe Perez, Stephanie A Strickler, and Ryan A Martin. Evolution of thermal tolerance and its fitness consequences: Parallel and non-parallel responses to urban heat islands across three cities. *Proceedings of the Royal Society B: Biological Sciences*, 285(1882):20180036, 2018.

[11] Jakob C Müller, Jesko Partecke, Ben J Hatchwell, Kevin J Gaston, and Karl L Evans. Candidate gene polymorphisms for behavioural adaptations during urbanization in blackbirds. *Molecular Ecology*, 22(13):3629–3637, 2013.

[12] Mark Ravinet, Tore Oldeide Elgvin, Cassandra Trier, Mansour Aliabadian, Andrey Gavrilov, and Glenn-Peter Sætre. Signatures of human-commensalism in the house sparrow genome. *Proceedings of the Royal Society B: Biological Sciences*, 285(1884):20181246, 2018.

[13] Panagiotis Theodorou, Rita Radzevičiūtė, Belinda Kahnt, Antonella Soro, Ivo Grosse, and Robert J Paxton. Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (Bombus lapidarius L.). *Proceedings of the Royal Society B: Biological Sciences*, 285(1877):20172806, 2018.

[14] David Pimentel, Lori Lach, Rodolfo Zuniga, and Doug Morrison. Environmental and economic costs of nonindigenous species in the United States. *BioScience*, 50(1):53–66, 2000.

[15] Chelsea G Himsworth, Kirbee L Parsons, Claire Jardine, and David M Patrick. 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*, 13(6):349–359, 2013.

[16] Emily E Puckett, Jane Park, Matthew Combs, Michael J Blum, Juliet E Bryant, Adalgisa Caccone, Federico Costa, Eva E Deinum, Alexandra Esther, Chelsea G Himsworth, et al. Global population divergence and admixture of the brown rat (Rattus norvegicus). *Proceedings of the Royal Society B: Biological Sciences*, 283(1841):20161762, 2016.

[17] Ken P Aplin, Hitoshi Suzuki, Alejandro A Chinen, R Terry Chesser, José Ten Have, Stephen C Donnellan, Jeremy Austin, Angela Frost, Jean Paul Gonzalez, Vincent Herbreteau, et al. Multiple geographic origins of commensalism and complex dispersal history of black rats. *PLoS One*, 6(11):e26357, 2011.

[18] Eleanor P Jones, Heidi M Eager, Sofia I Gabriel, Fría Jóhannesdóttir, and Jeremy B Searle. Genetic tracking of mice and other bioproxies to infer human history. *Trends in Genetics*, 29(5):298–308, 2013.

[19] Emily E Puckett and Jason Munshi-South. Brown rat demography reveals pre-commensal structure in Eastern Asia before expansion into Southeast Asia. *Genome Research*, 29(5):762–770, 2019.

[20] Matthew Combs, Emily E Puckett, Jonathan Richardson, Destiny Mims, and Jason Munshi-South. Spatial population genomics of the brown rat (Rattus norvegicus) in New York City. *Molecular Ecology*, 27(1):83–98, 2018.

[21] Matthew Combs, Kaylee A Byers, Bruno M Ghersi, Michael J Blum, Adalgisa Caccone, Federico Costa, Chelsea G Himsworth, Jonathan L Richardson, and Jason Munshi-South. Urban rat races: spatial population genomics of brown rats (Rattus norvegicus) compared across multiple cities. *Proceedings of the Royal Society B: Biological Sciences*, 285(1880):20180245, 2018.

[22] Eva E Deinum, Daniel L Halligan, Rob W Ness, Yao-Hua Zhang, Lin Cong, Jian-Xu Zhang, and Peter D Keightley. Recent evolution in Rattus norvegicus is shaped by declining effective population size. *Molecular Biology and Evolution*, 32(10):2547–2558, 2015.

[23] Emily E Puckett, Emma Sherratt, Matthew Combs, Elizabeth J Carlen, William Harcourt-Smith, and Jason Munshi-South. Variation in brown rat cranial shape shows directional selection over 120 years in new york city. *Ecology and Evolution*, 2020.

[24] Simone Rost, Hans-Joachim Pelz, Sandra Menzel, Alan D MacNicoll, Vanina León, Ki-Joon Song, Thomas Jäkel, Johannes Oldenburg, and Clemens R Müller. Novel mutations in the VKORC1 gene of wild rats and mice–a response to 50 years of selection pressure by warfarin? *BMC Genetics*, 10(1):4, 2009.

[25] Ann-Charlotte Heiberg. Anticoagulant resistance: a relevant issue in sewer rat (rattus norvegicus) control? *Pest Management Science: Formerly Pesticide Science*, 65(4):444–449, 2009.

[26] Lin Zeng, Chen Ming, Yan Li, Ling-Yan Su, Yan-Hua Su, Newton O Otecko, Ambroise Dalecky, Stephen Donnellan, Ken Aplin, Xiao-Hui Liu, et al. Out of southern east asia of the brown rat revealed by large-scale genome sequencing. *Molecular Biology and Evolution*, 35(1):149–158, 2017.

[27] Stephen E Harris, Jason Munshi-South, Craig Obergfell, and Rachel O'Neill. Signatures of rapid evolution in urban and rural transcriptomes of white-footed mice (peromyscus leucopus) in the New York metropolitan area. *PLoS One*, 8(8):e74938, 2013.

[28] Stephen E Harris and Jason Munshi-South. Signatures of positive selection and local adaptation to urbanization in white-footed mice (Peromyscus leucopus). *Molecular Ecology*, 26(22):6336–6350, 2017.

[29] Linelle Ann Lacson Abueg. Landscape genomics of white-footed mice (Peromyscus leucopus) along an urband-to-rural gradient in the New York City metropolitan area. Master's thesis, Fordham University, New York, 2018.

[30] Aimy Yu, Jason Munshi-South, and Eric J Sargis. Morphological differentiation in white-footed mouse (mammalia: rodentia: cricetidae: Peromyscus leucopus) populations from the new york city metropolitan area. *Bulletin of the Peabody Museum of Natural History*, 58(1):3–16, 2017.

[31] Pardis C Sabeti, David E Reich, John M Higgins, Haninah ZP Levine, Daniel J Richter, Stephen F Schaffner, Stacey B Gabriel, Jill V Platko, Nick J Patterson, Gavin J McDonald, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909):832, 2002.

[32] Anna Ferrer-Admetlla, Mason Liang, Thorfinn Korneliussen, and Rasmus Nielsen. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular biology and evolution*, 31(5):1275–1291, 2014.

[33] Nandita R Garud, Philipp W Messer, Erkan O Buzbas, and Dmitri A Petrov. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. *PLoS Genetics*, 11(2):e1005004, 2015.

[34] Nandita R Garud and Noah A Rosenberg. Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. *Theoretical Population Biology*, 102:94–101, 2015.

[35] Alexandre M Harris, Nandita R Garud, and Michael DeGiorgio. Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. *Genetics*, 210(4):1429–1452, 2018.

[36] Messer lab website, url: https://messerlab.org/resources/, Accessed September 2014.

[37] Andrew D Kern and Daniel R Schrider. diplos/hic: an updated approach to classifying selective sweeps. *G3: Genes, Genomes, Genetics*, 8(6):1959–1970, 2018.

[38] John Maynard Smith and John Haigh. The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1):23–35, 1974.

[39] Norman L Kaplan, Richard R Hudson, and Charles H Langley. The "hitchhiking effect" revisited. *Genetics*, 123(4):887–899, 1989.

[40] Nicholas H Barton. The effect of hitch-hiking on neutral genealogies. *Genetics Research*, 72(2):123–133, 1998.

[41] Yuseob Kim and Wolfgang Stephan. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 160(2):765–777, 2002.

[42] Joachim Hermisson and Pleuni S Pennings. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4):2335–2352, 2005.

[43] Pleuni S Pennings and Joachim Hermisson. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, 23(5):1076–1084, 2006.

[44] Bruce S Weir and C Clark Cockerham. Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6):1358–1370, 1984.

[45] Mattias Jakobsson, Michael D Edge, and Noah A Rosenberg. The relationship between FST and the frequency of the most frequent allele. *Genetics*, 193(2):515–528, 2013.

[46] Ping Feng and Zhijun Liu. Complex gene expansion of the CYP2D gene subfamily. *Ecology and Evolution*, 8(22):11022–11030, 2018.

[47] Ann K Daly and Barry P King. Pharmacogenetics of oral anticoagulants. *Pharmacogenetics and Genomics*, 13(5):247–252, 2003.

[48] Kazuki Takeda, Yoshinori Ikenaka, Tsutomu Tanikawa, Kazuyuki D Tanaka, Shouta MM Nakayama, Hazuki Mizukawa, and Mayumi Ishizuka. Novel revelation of warfarin resistant mechanism in roof rats (rattus rattus) using pharmacokinetic/pharmacodynamic analysis. *Pesticide Biochemistry and Physiology*, 134:1–7, 2016.

[49] David R Nelson, Darryl C Zeldin, Susan MG Hoffman, Lois J Maltais, Hester M Wain, and Daniel W Nebert. Comparison of cytochrome p450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics and Genomics*, 14(1):1–18, 2004.

[50] Daniel M Bear, Jean-Marc Lassance, Hopi E Hoekstra, and Sandeep Robert Datta. The evolving neural and genetic architecture of vertebrate olfaction. *Current Biology*, 26(20):R1039–R1049, 2016.

[51] Mary Shimoyama, Jeff De Pons, G Thomas Hayman, Stanley JF Laulederkind, Weisong Liu, Rajni Nigam, Victoria Petri, Jennifer R Smith, Marek Tutaj, Shur-Jen Wang, et al. The rat genome database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic acids research*, 43(D1):D743–D750, 2014.

[52] Simone Rost, Andreas Fregin, Vytautas Ivaskevicius, Ernst Conzelmann, Konstanze Hörtnagel, Hans-Joachim Pelz, Knut Lappegard, Erhard Seifried, Inge Scharrer, Edward GD Tuddenham, et al. Mutations in vkorc1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature*, 427(6974):537, 2004.

[53] Stephanie Grieb. Novel and recurring mutations in the VKORC1 gene in Norway rats from New York City and Long Island. Master's thesis, Hofstra University, Long Island, New York, 2017.

[54] Karl Paul Link. The discovery of dicumarol and its sequels. *Circulation*, 19(1):97–107, 1959.

[55] C Mary Boyle. Case of apparent resistance of Rattus norvegicus berkenhout to anticoagulant poisons. *Nature*, 188(4749):517, 1960.

[56] WILLIAM B JACKSON. Anticoagulant resistance in Europe. *Pest Control*, 37(3):51, 1969.

[57] WB Jackson, PJ Spear, and CG Wright. Resistance of norway rats to anticoagulant rodenticides confirmed in the United States. *Pest Control*, 1971.

[58] JE Brooks and AM Bowerman. Anticoagulant resistance in wild norway rats in new york. *Epidemiology & Infection*, 71(2):217–222, 1973.

[59] Hans-Joachim Pelz. Spread of resistance to anticoagulant rodenticides in germany. *International Journal of Pest Management*, 53(4):299–302, 2007.

[60] Alan P Buckle, Colin V Prescott, and Keith J Ward. Resistance to the first and second generation anticoagulant rodenticides-a new perspective. 1994.

[61] Agnes Grandemange, Michael Hans Kohn, Romain Lasseur, Christiane Longin-Sauvageon, Philippe Berny, and Etienne Benoit. Consequences of the y139f vkorc1

mutation on resistance to avks: in-vivo investigation in a 7th generation of congenic y139f strain of rats. *Pharmacogenetics and Genomics*, 19(10):742, 2009.

[62] Mette Drude Markussen, Ann-Charlotte Heiberg, Carsten Alsbo, Peter Stein Nielsen, Sakari Kauppinen, and Michael Kristensen. Involvement of hepatic xenobiotic related genes in bromadiolone resistance in wild norway rats, rattus norvegicus (berk.). *Pesticide Biochemistry and Physiology*, 88(3):284–295, 2007.

[63] Zhanpeng Yuan, Simon Courtenay, R Christopher Chambers, and Isaac Wirgin. Evidence of spatially extensive resistance to pcbs in an anadromous fish of the hudson river. *Environmental health perspectives*, 114(1):77–84, 2006.

[64] Isaac Wirgin, Nirmal K Roy, Matthew Loftus, R Christopher Chambers, Diana G Franks, and Mark E Hahn. Mechanistic basis of resistance to pcbs in atlantic tomcod from the hudson river. *Science*, 331(6022):1322–1325, 2011.

[65] Anna L Moon, Niels Haan, Lawrence S Wilkinson, Kerrie L Thomas, and Jeremy Hall. Cacna1c: association with psychiatric disorders, behavior, and neurogenesis. *Schizophrenia Bulletin*, 44(5):958–965, 2018.

[66] Theresa M Kisko, Moria D Braun, Susanne Michels, Stephanie H Witt, Marcella Rietschel, Carsten Culmsee, Rainer KW Schwarting, and Markus Wöhr. Sex-dependent effects of cacna1c haploinsufficiency on juvenile social play behavior and pro-social 50-khz ultrasonic communication in rats. *Genes, Brain and Behavior*, page e12552, 2018.

[67] Rachel Wurzman, Patrick A Forcelli, Christopher J Griffey, and Lawrence F Kromer. Repetitive grooming and sensorimotor abnormalities in an ephrin-a knockout model for autism spectrum disorders. *Behavioural Brain Research*, 278:115–128, 2015.

[68] Yukiko Hanada, Yukiko Nakamura, Yoshiyuki Ozono, Yusuke Ishida, Yasumitsu Takimoto, Manabu Taniguchi, Kazuya Ohata, Yoshihisa Koyama, Takao Imai, Tetsuo Mori-

hana, et al. Fibroblast growth factor 12 is expressed in spiral and vestibular ganglia and necessary for auditory and equilibrium function. *Scientific Reports*, 8(1):11491, 2018.

[69] E Guiry and M Buckley. Urban rats have less variable, higher protein diets. *Proceedings of the Royal Society B*, 285(1889):20181441, 2018.

[70] Albrecht I Schulte-Hostedde, Zvia Mazal, Claire M Jardine, and Jeffrey Gagnon. Enhanced access to anthropogenic food waste is related to hyperglycemia in raccoons (Procyon lotor). *Conservation Physiology*, 6(1):coy026, 2018.

[71] Jeremy Lyons, Gabriela Mastromonaco, Darryl B Edwards, and Albrecht I Schulte-Hostedde. Fat and happy in the city: Eastern chipmunks in urban environments. *Behavioral Ecology*, 28(6):1464–1471, 2017.

[72] Florencia Schlamp, Julian van der Made, Rebecca Stambler, Lewis Chesebrough, Adam R Boyko, and Philipp W Messer. Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Molecular Ecology*, 25(1):342–356, 2016.

[73] Russell Lande. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution*, 34(2):292–305, 1980.

[74] Nicholas H Barton and Peter D Keightley. Multifactorial genetics: understanding quantitative genetic variation. *Nature Reviews Genetics*, 3(1):11, 2002.

[75] Jonathan K Pritchard and Anna Di Rienzo. Adaptation–not by sweeps alone. *Nature Reviews Genetics*, 11(10):665, 2010.

[76] Jonathan K Pritchard, Joseph K Pickrell, and Graham Coop. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20(4):R208–R215, 2010.

[77] Picard toolkit. `http://broadinstitute.github.io/picard/`, 2019.

[78] Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303, 2010.

[79] Maximilian Haeussler, Ann S Zweig, Cath Tyner, Matthew L Speir, Kate R Rosenbloom, Brian J Raney, Christopher M Lee, Brian T Lee, Angie S Hinrichs, Jairo Navarro Gonzalez, et al. The UCSC genome browser database: 2019 update. *Nucleic Acids Research*, 47(D1):D853–D858, 2018.