# Genetic Adaptation in New York City Rats

Arbel Harpak [ID][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

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

[3]Department of Biology, Stanford University

[4]Department of Biological Sciences, Fordham University

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

[6]Department of Biology, San Francisco State University

†These authors contributed equally to this work.

*Corresponding authors: E-mails: arbel.arbel@gmail.com; jmunshisouth@fordham.edu.

Accepted: 17 November 2020

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 1) high-frequency, extended haplotypes that could indicate selective sweeps and 2) 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 suggest that many sweeps began after the split from the ancestral population. Together, our results suggest several hypotheses on adaptation in rats living in proximity to humans.

Key words: adaptation, population genetics, rodents, selective sweeps, urban evolution.

Significance

Rats are arguably the poster child of evolutionary success in urban environments. Did genetic adaptation play a role in this success story? We scanned the genomes of New York City brown rats and found intriguing genetic signatures of adaptation near genes associated with metabolism, diet, the nervous system, and locomotory behavior. Our results suggest several hypotheses on adaptation in urban rats.

Introduction

Urbanization has the potential to drive dramatic ecological and evolutionary consequences for wildlife (Seto et al. 2012; Rivkin et al. 2019). The most commonly examined evolutionary responses to urbanization are changes in gene flow and in the intensity of genetic drift (Beninde et al. 2016; Munshi-South et al. 2016; Johnson and Munshi-South 2017; Miles et al. 2018). A small but growing number of studies have also examined adaptive evolution in urban environments, including morphological adaptations to urban infrastructure (Brown and Brown 2013; Winchell et al. 2016), adaptive life history changes, thermal tolerance to warming cities (Brans et al. 2018; Diamond et al. 2018), and behavioral

changes (Mueller et al. 2013). The recent adoption of genomic scans to identify loci under positive selection can help generate hypotheses about adaptive phenotypes in cities (Ravinet et al. 2018; Theodorou et al. 2018).

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 are a notorious threat to urban quality of life (Pimentel et al. 2000; Himsworth et al. 2013). Recent analyses have revealed some of the relationships between invasive urban rodent populations that spread around the world with humans (Aplin et al. 2011; Jones et al. 2013; Puckett et al. 2016; Puckett and Munshi-South 2019) and the influence of heterogeneous urban environments on gene flow (Combs, Byers, et al. 2018; Combs, Puckett, et al. 2018). 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 with house mice, possibly due to a population bottleneck ~20,000 years ago (Deinum et al. 2015). More recently, they have experienced major range expansions, presumably due to their association with agrarian, and later urban, human societies (Puckett and Munshi-South 2019). 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 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 (NYC) over a 120-year period (Puckett et al. 2020).

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 (Rost et al. 2009), though resistance to recent second-generation anticoagulants is less widespread and may not be monogenic (Heiberg 2009). 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 (Zeng et al. 2018). 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 (Puckett et al. 2020). 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 overrepresented among putatively selected regions (Harris et al. 2013; Harris and Munshi-South 2017; Abueg 2018). This same *Peromyscus* population exhibited shorter toothrows consistent with a change in diet (Yu et al. 2017). 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 (Sabeti et al. 2002; Ferrer-Admetlla et al. 2014; Garud et al. 2015)—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 (Puckett et al. 2016; Puckett and Munshi-South 2019). We find evidence for recent selective sweeps near genes associated with the nervous system, metabolism of endogenous compounds, diet, apoptotic processes, organ morphogenesis, and locomotory behavior.

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) (Combs, Puckett, et al. 2018). The mean coverage was $\geq 15\times$ 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 Material online) (Deinum et al. 2015).

We searched for selective sweeps using the methods G12 (Garud et al. 2015; Garud and Rosenberg 2015; Harris et al. 2018) and H-scan (Messer Lab website 2014). G12 and H-scan both measure the homogeneity of haplotypes in the sample around a focal single-nucleotide polymorphism (SNP; Materials and 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 *multisite 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 (Harris et al. 2018; Kern and Schrider 2018). 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
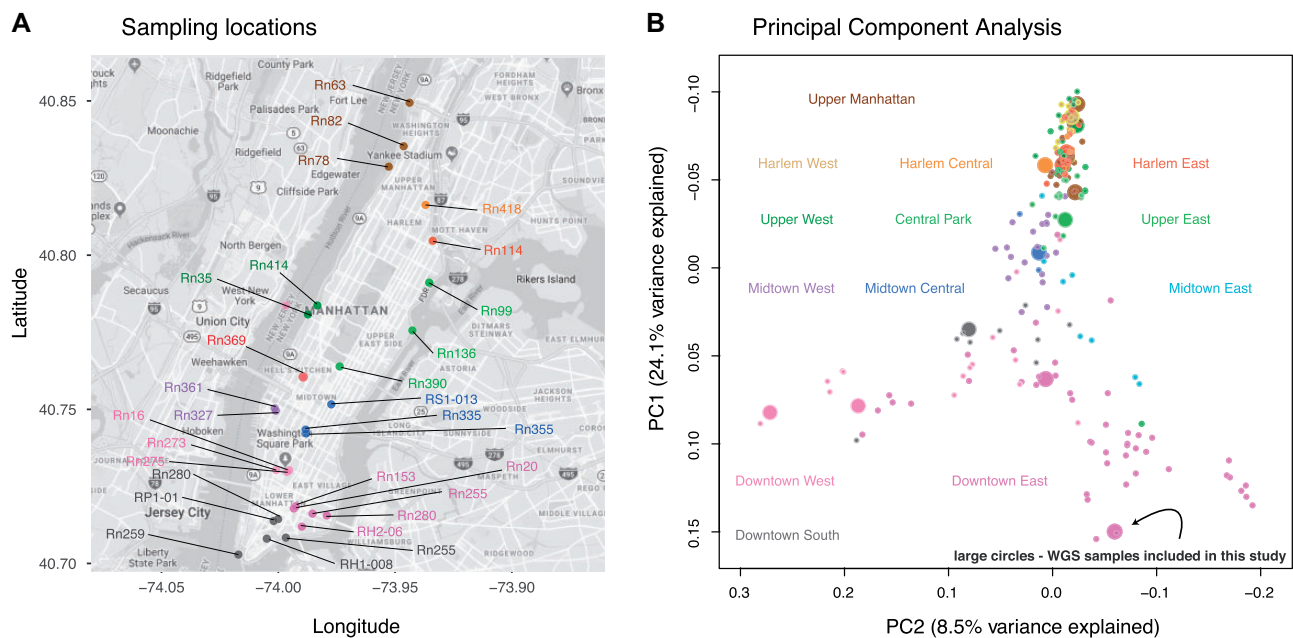
FIG. 1.—Sampling locations and genetic population structure of NYC brown rats. (*A*) Sampling sites of the 29 rat whole-genome samples used in our selection scans. Color coding corresponds to NYC neighborhoods as in panel (*B*). (*B*) Principal component analysis of 198 rats reanalyzed from Combs, Puckett, et al. (2018), with the 15 rats included in both Combs, Puckett, et al. (2018) and this study represented by large circles. In both panels, color coding corresponds to NYC neighborhoods.

similar signals of frequent, long segregating haplotypes (fig. 2 and supplementary fig. S1*A*, Supplementary Material online).

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 (supplementary fig. S2, Supplementary Material online).

We therefore took an alternative approach of focusing on the most extreme empirical values of the selection statistics as putative targets of adaptation (supplementary fig. S7, Supplementary Material online; we note that we still report the neutral simulation results [supplementary table S1, Supplementary Material online] and *P* values computed using the simulation-based null distribution [supplementary table S6, Supplementary Material online]). We calculated G12 and H-scan values across the whole genome (supplementary files S3 and S4, Supplementary Material online). 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 Material online). Multisite genotypes are visibly more homogeneous in top-scoring loci (fig. 3*B* and *C*) than in random loci (fig. 3*A*). In some of the top-scoring loci, we observe a single common haplotype—consistent with the expectation under a hard

selective sweep (Maynard Smith and Haigh 1974; Kaplan et al. 1989; Barton 1998; Kim and Stephan 2002) (fig. 3*B*). In many others, more than two common haplotypes are segregating in the population—consistent with the expectation under a soft selective sweep (Hermisson and Pennings 2005; Pennings and Hermisson 2006) (fig. 3*C* and supplementary file S2, Supplementary Material online).

To gauge the timing of putative selective sweeps, we compared the multisite genotypes of our sample with those of nine rats from Harbin, Heilongjiang Province, China—the presumed ancestral range of the species (Deinum et al. 2015). The Chinese genotypes are often heterogeneous at the loci most homogeneous in the NYC sample (fig. 3*D*). Only 5/17 of the H-scan candidates show significantly elevated H-scores in the Chinese sample (supplementary fig. S4 and file S6, Supplementary Material online). 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 (multisite genotype visualizations for all G12 and H-scan candidate loci can be found in supplementary file S2, Supplementary Material online).

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} (Weir and Cockerham 1984) across SNPs in 10-kb sliding windows (supplementary file S5, Supplementary Material online) and again focused on the 100 top-scoring
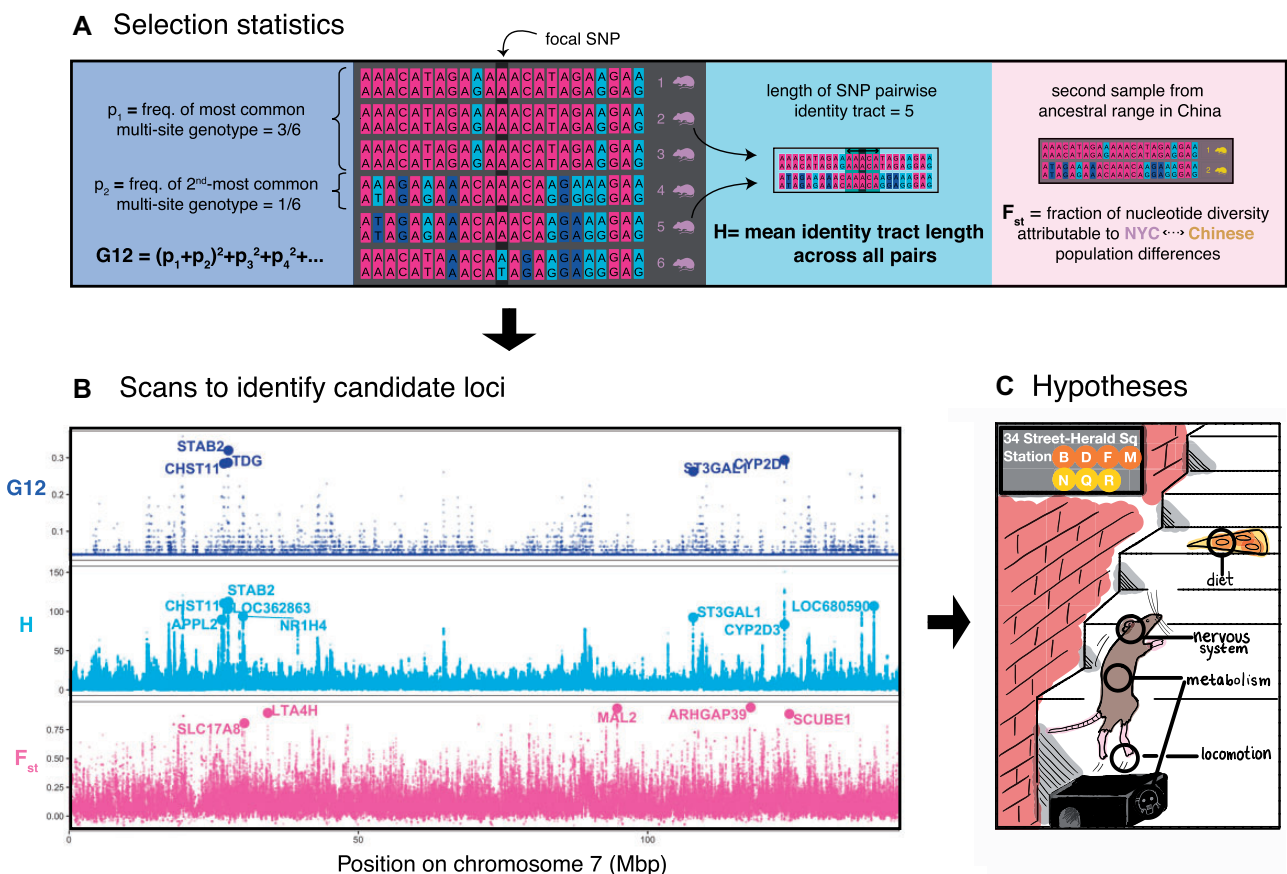
Fig. 2.—Scanning for signals of adaptation. (A) Three selection statistics were used to identify signals of selection. Two of the statistics, G12 and H-scan, detect loci with low haplotype diversity (high homogeneity)—a signature of recent or ongoing selective sweeps. G12 is defined similarly to the homozygosity of multisite genotypes (in a window of a fixed number of SNPs)—a sum of squared frequencies of multisite genotypes—except that the two most frequent genotypes are grouped (summed and squared) together. In the illustrated example, the most common multisite genotype is shared by rats 1–3. 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 five 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 <20 kb 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, Supplementary Material online. (C) Biological functions associated with candidate genes may point to traits that have been subject to selection in NYC rats, including diet, the nervous system, metabolism, and locomotion.

loci genome-wide (supplementary table S3, Supplementary Material online). In the Supplementary Material online, 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 (supplementary fig. S1A, Supplementary Material online). 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 (Jakobsson et al. 2013).

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 and Supplementary Material online). This result may suggest that genic sweeps are rare—at least among loci identified with our methods.

Moving forward, we only considered protein-coding genes ≤20 kb away from top-scoring loci as potential targets of adaptation—leaving 19, 17, and 32 candidates for G12, H-scan, and F_{st}, respectively (supplementary tables S1–S3,

A Random locus

Chromosome 4 , position 2045518

2044210 — 2055443

B Putative hard sweep

Chromosome 6 , position 56553259

56539122 — 56564408

C Putative soft sweep

Chromosome 10 , position 34434786

34423595 — 34440161

D Comparison of (C) with Chinese samples

34423442 — 34440027

NYC urban samples

Chinese rural samples

■ homozygote A   ■ homozygote B   ■ heterozygote   ■ missing data
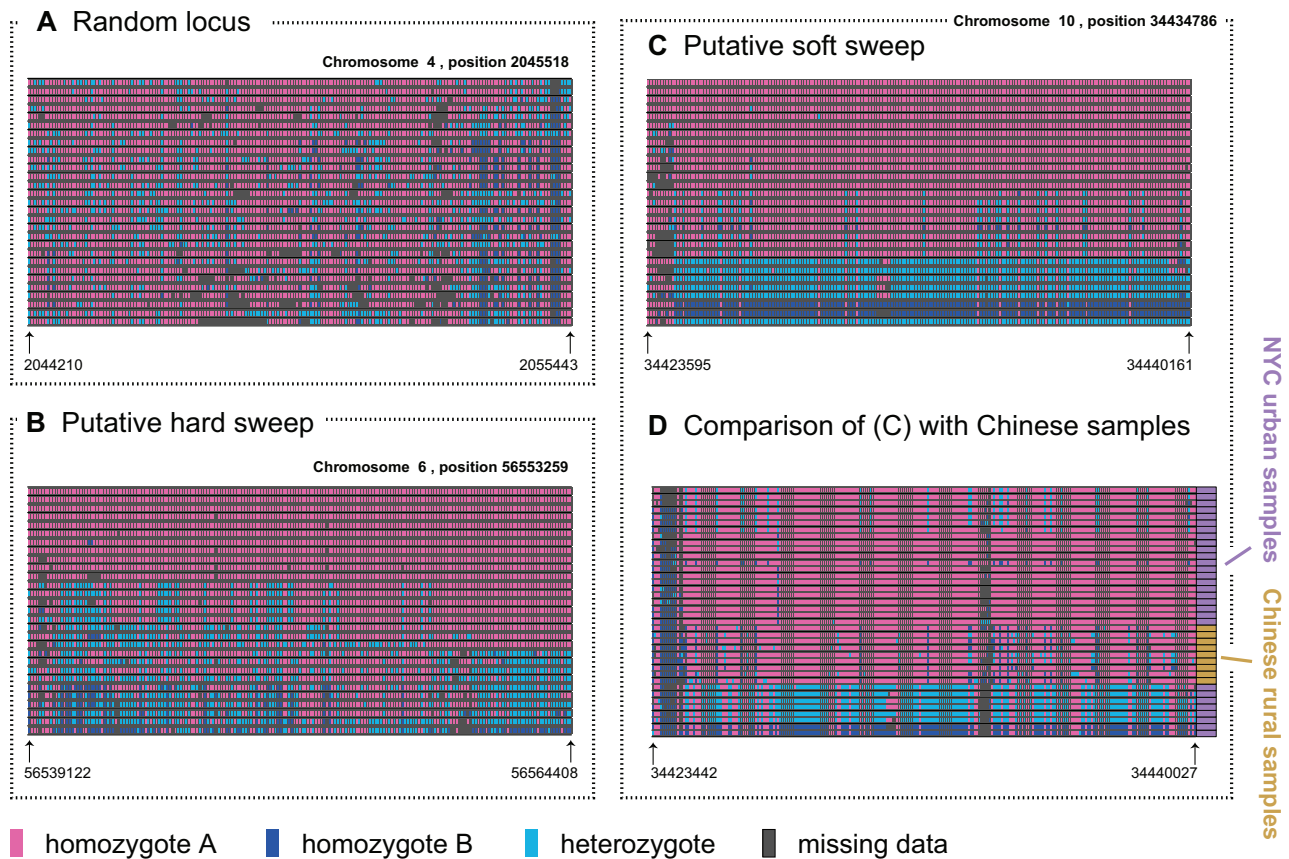
Fig. 3.—Multisite genotype structure. Each row refers to a multisite genotype of one rat. (A) At a randomly chosen locus, there is no clear clustering of multisite genotypes. (B) An example of a locus that was detected as a putative target of selection with H-scan, in which there is one multisite genotype 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 multisite genotypes—consistent with a soft selective sweep. (D) A comparison of the NYC samples to a sample of nine rats from the presumed ancestral range of the species in China, around the same focal site as in panel (C). Multisite 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.

Supplementary Material online). One G12 candidate lies close to a cluster of cytochrome P450 genes (it is specifically closest to CYP2D1, but there are five CYP2D genes ≤50 kb away). CYP-encoded enzymes help metabolize endogenous compounds and detoxify exogenous chemicals (Feng and Liu 2018) and have been associated with Warfarin resistance (Daly and King 2003; Takeda et al. 2016). This gene family has expanded substantially in rodents (Nelson et al. 2004).

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 (Bear et al. 2016). 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 Material online).

We do, however, find Gene Ontology (GO) biological processes (Shimoyama et al. 2015) that are enriched among top-

scoring genes. Among the top fourteen G12 values inside genes (corresponding to 0.27 expected false discoveries, Materials and 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 supplementary fig. S3, Supplementary Material online).

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 (Rost et al. 2004, 2009). The Chinese and NYC rats were highly differentiated at a 10-kb window centered on VKORC1 (98th chromosomal percentile of F_{st}). At the same time, multisite genotypes around VKORC1 did not appear homogeneous, and G12 and H-scan are close to the chromosomal medians (supplementary fig. S5, Supplementary Material online). Zooming in on the VKORC1 coding sequence, we found that no known resistance-conferring alleles

appear to have fixed in the population (Supplementary Material online), and, further, that the data contained no nonsynonymous SNPs. To our knowledge, all of the experimentally confirmed resistance-conferring variants in VKORC1 are nonsynonymous (Rost et al. 2009). We detected ten 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 (Grieb 2017) (supplementary fig. S6, Supplementary Material online).

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; supplementary tables S1–S5, Supplementary Material online) 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 supplementary fig. S5, Supplementary Material online). We also detect no nonsynonymous polymorphisms in the gene (supplementary fig. S6, Supplementary Material online; but see Grieb [2017] 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 (Link 1959) but was gradually phased out in the late 1970s and replaced by more effective second-generation anticoagulant rodenticides (SGARs) after high levels of resistance were reported from Europe (Boyle 1960; Jackson 1969) and the United States (Jackson et al. 1971), and specifically New York (Brooks and Bowerman 1973). 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 (Rost et al. 2009). 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 (Buckle et al. 1994; Pelz 2007). VKORC1 may play a role in SGAR resistance—but likely a smaller role than it does for first-generation anticoagulants (Heiberg 2009): 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) (Grandemange et al. 2009). At the same time, rats resistant to the SGAR bromadialone showed increased expression of a suite of
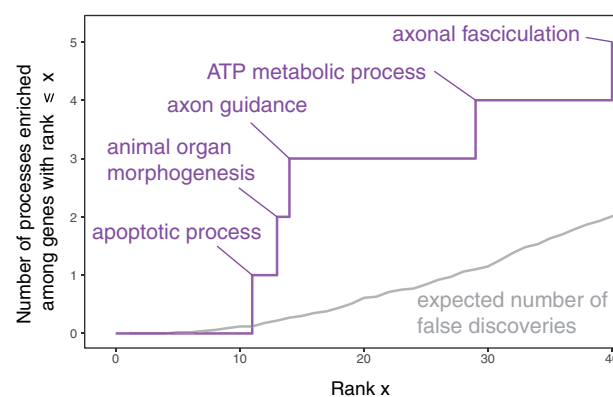

Fig. 4.—Biological processes enriched among 40 top-scoring genes. This figure focuses on candidate targets of adaptation arising from a comparison of G12 values in genes. Each gene was given the maximum score of all focal SNPs in the gene, and genes were ranked by this maximal score. We consider GO biological processes associated with at least three top-scoring genes as enriched, and the y axis shows their cumulative count. Purple 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.

CYP genes (Markussen et al. 2007). 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 contamination in killifish and tomcod in the Hudson and other rivers in the area (Yuan et al. 2006; Wirgin et al. 2011). On a possibly related note, our GO analysis identified nervous system genes as a possible target (fig. 4 and supplementary fig. S3, Supplementary Material online). 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 figure 4 and supplementary figure S3 and tables S1–S3, Supplementary Material online, 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 (seventh highest

score among genes for both G12 and H-scan, supplementary tables S4 and S5, Supplementary Material online) has been repeatedly associated with psychiatric disorders in humans, and the effects may be modulated by early-life stress (Moon et al. 2018). Thus, rat behavioral phenotypes associated with anxiety (such as antipredator 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 (Kisko et al. 2020). CACNA1C and multiple other top-scoring genes like FGF12, EPHA3, and CHAT (supplementary table S4, Supplementary Material online) may also affect locomotion in rodents (Wurzman et al. 2015; Hanada et al. 2018). 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 (Lyons et al. 2017; Guiry and Buckley 2018; Schulte-Hostedde et al. 2018). Indeed, several of the candidate genes we identified, ST6GALNAC1 (G12, supplementary table S1, Supplementary Material online), CHST11, and GBTG1 (H-scan, supplementary table S2, Supplementary Material online), 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) (Schlamp et al. 2016; Harris et al. 2018). Other modes of adaptation, such as polygenic adaptation (Lande 1980; Barton and Keightley 2002; Pritchard and Di Rienzo 2010; Pritchard et al. 2010), 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 (Puckett et al. 2020). 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 several hypotheses on how the success of NYC brown rats may be rooted—at least in part—in rapid changes in the genetic composition of the population through natural selection.

Materials and Methods

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, Puckett, et al. (2018). 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-h 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 (Deinum et al. 2015). 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 15× 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* (Picard 2019). Duplicate reads were then removed using *Picard toolkit*, followed by base quality score recalibration and local realignment around indels using *GATK* v3.7 (McKenna et al. 2010).

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 (last accessed 2015) in Variant Calling Format (VCF) for the *R. norvegicus* samples used in the study, including the 29 NYC individuals and the nine Chinese individuals of the study of Deinum et al. (2015). In the Supplementary Material online, 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 (Garud and Rosenberg 2015; Garud et al. 2015; Harris et al. 2018) and H-scan (Messer Lab website 2014). For a sample of n diploids, both methods take as input n multisite 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 nucleotide genotypes.

G12 examines a symmetric window of a fixed number of SNPs around each focal SNP. Denoting the frequency of the k unique multisite genotypes at a window as $p_1, p_2, p_3, \ldots, p_k$, ranked from most common to most rare (fig. 2A), G_{12} is defined as

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

We used a window size of 201 SNPs to compute G12 values, conferring to a mean length of 29 kb and mean linkage disequilibrium (r^2) of 0.135 between the SNPs at the ends of the window (supplementary fig. S10, Supplementary Material online).

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—that is, 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 (Messer Lab 2014), 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 10 kb with a 1-kb step. In the Supplementary Material online, we provide further detail on preprocessing procedures, handling missing data, and the method used to define candidate loci around top-scoring SNPs.

Biological Process Enrichment Analysis

In figure 4 and supplementary figure S3, Supplementary Material online, we examine whether top-scoring genes are enriched for any GO biological process categories. We downloaded protein-coding gene annotations for the Rn5 reference genome from the UCSC Genome Browser (Haeussler et al. 2019). 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. Supplementary tables S4 (H-scan) and S5 (G12), Supplementary Material online, 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 30 kb 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 Material online), as we have found that without this modification, some gene clusters (e.g., the CYP2D cluster), that may also share biological process annotations, were overrepresented 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 gray in figure 4 and supplementary figure S3, Supplementary Material online, is the number of biological process categories that are expected to be associated with at least three of the genes ranked $1, \ldots, r$ by chance alone, that is, 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 three 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 codependence of biological process categories by maintaining their joint distribution of the number of categories associated with each gene.

Ethical Statement

Collection of rats for this study adhered to all local guidelines. Approval in New York City was granted by the Fordham University Institutional Animal Care and Use Committee

(IACUC; Protocol JMS-16-04) and the NYC Department of Parks and Recreation.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

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.

Literature Cited

Abueg LAL. 2018. Landscape genomics of white-footed mice (*Peromyscus leucopus*) along an urban-to-rural gradient in the New York City metropolitan area [master's thesis]. [New York]: Fordham University.

Aplin KP, et al. 2011. Multiple geographic origins of commensalism and complex dispersal history of black rats. PLoS One 6(11):e26357.

Barton NH. 1998. The effect of hitch-hiking on neutral genealogies. Genet Res. 72(2):123–133.

Barton NH, Keightley PD. 2002. Multifactorial genetics: understanding quantitative genetic variation. Nat Rev Genet. 3(1):11–21.

Bear DM, Lassance J-M, Hoekstra HE, Datta SR. 2016. The evolving neural and genetic architecture of vertebrate olfaction. Curr Biol. 26(20):R1039–R1049.

Beninde J, et al. 2016. Cityscape genetics: structural vs. functional connectivity of an urban lizard population. Mol Ecol. 25(20):4984–5000.

Boyle CM. 1960. Case of apparent resistance of *Rattus norvegicus* Berkenhout to anticoagulant poisons. Nature 188(4749):517.

Brans KI, Stoks R, Meester LD. 2018. Urbanization drives genetic differentiation in physiology and structures the evolution of pace-of-life syndromes in the water flea *Daphnia magna*. Proc R Soc B 285(1883):20180169.

Brooks J, Bowerman A. 1973. Anticoagulant resistance in wild Norway rats in New York. J Hyg. 71(2):217–222.

Brown CR, Brown MB. 2013. Where has all the road kill gone? Curr Biol. 23(6):R233–R234.

Buckle AP, Prescott CV, Ward KJ. 1994. Resistance to the first and second generation anticoagulant rodenticides—a new perspective. In: Halverson WS, Marsh RE, editors. Proceedings of the sixteenth vertebrate pest conference. p. 138–144.

Combs M, Byers KA, et al. 2018. Urban rat races: spatial population genomics of brown rats (*Rattus norvegicus*) compared across multiple cities. Proc R Soc B 285(1880):20180245.

Combs M, Puckett EE, Richardson J, Mims D, Munshi-South J. 2018. Spatial population genomics of the brown rat (*Rattus norvegicus*) in New York City. Mol Ecol. 27(1):83–98.

Daly AK, King BP. 2003. Pharmacogenetics of oral anticoagulants. Pharmacogenet Genomics. 13(5):247–252.

Deinum EE, et al. 2015. Recent evolution in *Rattus norvegicus* is shaped by declining effective population size. Mol Biol Evol. 32(10):2547–2558.

Diamond SE, Chick LD, Perez A, Strickler SA, Martin RA. 2018. Evolution of thermal tolerance and its fitness consequences: parallel and non-parallel responses to urban heat islands across three cities. Proc R Soc B 285(1882):20180036.

Feng P, Liu Z. 2018. Complex gene expansion of the CYP2D gene subfamily. Ecol Evol. 8(22):11022–11030.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol Biol Evol. 31(5):1275–1291.

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. PLoS Genet. 11(2):e1005004.

Garud NR, Rosenberg NA. 2015. Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. Theor Popul Biol. 102:94–101.

Grandemange A, et al. 2009. Consequences of the Y139F Vkorc1 mutation on resistance to AVKs: in-vivo investigation in a 7th generation of congenic Y139F strain of rats. Pharmacogenet Genomics. 19(10):742.

Grieb S. 2017. Novel and recurring mutations in the VKORC1 gene in Norway rats from New York City and Long Island [master's thesis]. [Long Island (NY)]: Hofstra University.

Guiry E, Buckley M. 2018. Urban rats have less variable, higher protein diets. Proc R Soc B 285(1889):20181441.

Haeussler M, et al. 2019. The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. 47(D1):D853–D858.

Hanada Y, et al. 2018. Fibroblast growth factor 12 is expressed in spiral and vestibular ganglia and necessary for auditory and equilibrium function. Sci Rep. 8(1):11491.

Harris AM, Garud NR, DeGiorgio M. 2018. Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. Genetics 210(4):1429–1452.

Harris SE, Munshi-South J. 2017. Signatures of positive selection and local adaptation to urbanization in white-footed mice (*Peromyscus leucopus*). Mol Ecol. 26(22):6336–6350.

Harris SE, Munshi-South J, Obergfell C, O'Neill R. 2013. 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.

Heiberg A-C. 2009. Anticoagulant resistance: a relevant issue in sewer rat (*Rattus norvegicus*) control? Pest Manage Sci. 65(4):444–449.

Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169(4):2335–2352.

Himsworth CG, Parsons KL, Jardine C, Patrick DM. 2013. Rats, cities, people, and pathogens: a systematic review and narrative synthesis of literature regarding the ecology of rat-associated zoonoses in urban centers. Vector-Borne Zoonotic Dis. 13(6):349–359.

Jackson W, Spear P, Wright C. 1971. Resistance of Norway rats to anticoagulant rodenticides confirmed in the United States. Pest Control. 39(9):13–14.

Jackson WB. 1969. Anticoagulant resistance in Europe. Pest Control. 37(3):51.

Jakobsson M, Edge MD, Rosenberg NA. 2013. The relationship between FST and the frequency of the most frequent allele. Genetics 193(2):515–528.

Johnson MT, Munshi-South J. 2017. Evolution of life in urban environments. Science 358(6363):eaam8327.

Jones EP, Eager HM, Gabriel SI, Jóhannesdóttir F, Searle JB. 2013. Genetic tracking of mice and other bioproxies to infer human history. Trends Genet. 29(5):298–308.

Kaplan NL, Hudson RR, Langley CH. 1989. The "hitchhiking effect" revisited. Genetics 123(4):887–899.

Kern AD, Schrider DR. 2018. diploS/HIC: an updated approach to classifying selective sweeps. G3 (Bethesda) 8(6):1959–1970.

Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160(2): 765–777.

Kisko TM, et al. 2020. Sex-dependent effects of Cacna1c haploinsufficiency on juvenile social play behavior and pro-social 50-kHz ultrasonic communication in rats. Genes Brain Behav. 19(2):e12552.

Lande R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution 34(2):292–305.

Link KP. 1959. The discovery of dicumarol and its sequels. Circulation 19(1):97–107.

Lyons J, Mastromonaco G, Edwards DB, Schulte-Hostedde AI. 2017. Fat and happy in the city: eastern chipmunks in urban environments. Behav Ecol. 28(6):1464–1471.

Markussen MD, et al. 2007. Involvement of hepatic xenobiotic related genes in bromadiolone resistance in wild Norway rats, Rattus norvegicus (Berk.). Pestic Biochem Physiol. 88(3):284–295.

Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23(1):23–35.

McKenna A, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303.

Messer Lab website. 2014. Messer lab website. Available from: https://messerlab.org/resources/

Miles LS, Johnson JC, Dyer RJ, Verrelli BC. 2018. Urbanization as a facilitator of gene flow in a human health pest. Mol Ecol. 27(16):3219–3230.

Moon AL, Haan N, Wilkinson LS, Thomas KL, Hall J. 2018. CACNA1C: association with psychiatric disorders, behavior, and neurogenesis. Schizophr Bull. 44(5):958–965.

Mueller JC, Partecke J, Hatchwell BJ, Gaston KJ, Evans KL. 2013. Candidate gene polymorphisms for behavioural adaptations during urbanization in blackbirds. Mol Ecol. 22(13):3629–3637.

Munshi-South J, Zolnik CP, Harris SE. 2016. Population genomics of the Anthropocene: urbanization is negatively associated with genome-wide variation in white-footed mouse populations. Evol Appl. 9(4):546–564.

Nelson DR, et al. 2004. Comparison of cytochrome p450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. Pharmacogenet Genomics. 14(1):1–18.

Pelz H-J. 2007. Spread of resistance to anticoagulant rodenticides in Germany. Int J Pest Manage. 53(4):299–302.

Pennings PS, Hermisson J. 2006. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol. 23(5):1076–1084.

Picard. 2019. Picard toolkit. Available from: http://broadinstitute.github.io/picard/

Pimentel D, Lach L, Zuniga R, Morrison D. 2000. Environmental and economic costs of nonindigenous species in the United States. BioScience 50(1):53–66.

Pritchard JK, Di Rienzo A. 2010. Adaptation—not by sweeps alone. Nat Rev Genet. 11(10):665–667.

Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 20(4):R208–R215.

Puckett EE, Munshi-South J. 2019. Brown rat demography reveals pre-commensal structure in Eastern Asia before expansion into Southeast Asia. Genome Res. 29(5):762–770.

Puckett EE, et al. 2016. Global population divergence and admixture of the brown rat (Rattus norvegicus). Proc R Soc B 283(1841):20161762.

Puckett EE, et al. 2020. Variation in brown rat cranial shape shows directional selection over 120 years in New York City. Ecol Evol. 10(11):4739–4748.

Ravinet M, et al. 2018. Signatures of human-commensalism in the house sparrow genome. Proc R Soc B 285(1884):20181246.

Rivkin LR, et al. 2019. A roadmap for urban evolutionary ecology. Evol Appl. 12(3):384–398.

Rost S, et al. 2004. Mutations in vkorc1 cause warfarin resistance and multiple coagulation factor deficiency type 2. Nature 427(6974):537–541.

Rost S, et al. 2009. Novel mutations in the VKORC1 gene of wild rats and mice—a response to 50 years of selection pressure by warfarin? BMC Genet. 10(1):4.

Sabeti PC, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419(6909):832–837.

Schlamp F, et al. 2016. Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. Mol Ecol. 25(1):342–356.

Schulte-Hostedde AI, Mazal Z, Jardine CM, Gagnon J. 2018. Enhanced access to anthropogenic food waste is related to hyperglycemia in raccoons (Procyon lotor). Conserv Physiol. 6(1):coy026.

Seto KC, Güneralp B, Hutyra LR. 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proc Natl Acad Sci U S A. 109(40):16083–16088.

Shimoyama M, et al. 2015. The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. Nucleic Acids Res. 43(D1):D743–D750.

Takeda K, et al. 2016. Novel revelation of warfarin resistant mechanism in roof rats (Rattus rattus) using pharmacokinetic/pharmacodynamic analysis. Pestic Biochem Physiol. 134:1–7.

Theodorou P, et al. 2018. Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (Bombus lapidarius L.). Proc R Soc B 285(1877):20172806.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358–1370.

Winchell KM, Reynolds RG, Prado-Irwin SR, Puente-Rolón AR, Revell LJ. 2016. Phenotypic shifts in urban areas in the tropical lizard Anolis cristatellus. Evolution 70(5):1009–1022.

Wirgin I, et al. 2011. Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson River. Science 331(6022):1322–1325.

Wurzman R, Forcelli PA, Griffey CJ, Kromer LF. 2015. Repetitive grooming and sensorimotor abnormalities in an ephrin—a knockout model for autism spectrum disorders. Behav Brain Res. 278:115–128.

Yu A, Munshi-South J, Sargis EJ. 2017. Morphological differentiation in white-footed mouse (Mammalia: Rodentia: Cricetidae: Peromyscus leucopus) populations from the New York City metropolitan area. Bull Peabody Mus Nat Hist. 58(1):3–16.

Yuan Z, Courtenay S, Chambers RC, Wirgin I. 2006. Evidence of spatially extensive resistance to PCBs in an anadromous fish of the Hudson River. Environ Health Perspect. 114(1):77–84.

Zeng L, et al. 2018. Out of southern east Asia of the brown rat revealed by large-scale genome sequencing. Mol Biol Evol. 35(1):149–158.

Associate editor: Adam Eyre-Walker