Signatures of positive selection and local adaptation to urbanization in white-footed mice (*Peromyscus leucopus*)

Stephen E. Harris[1], Jason Munshi-South[2*]

[1]The Graduate Center, City University of New York (CUNY), New York, NY 10016 USA

[2]Louis Calder Center—Biological Field Station, Fordham University, Armonk, NY 10504 USA

**Corresponding author:* Jason Munshi-South

E-mail: jmunshisouth@fordham.edu

ABSTRACT

Urbanization significantly alters natural ecosystems and has accelerated globally as humans move into dense urban centers. Urban wildlife populations are often highly fragmented by an inhospitable matrix of human infrastructure. Isolated populations may adapt in response to novel urban pressures, but few studies have found evidence of selection in urban environments. We used multiple approaches to examine signatures of selection in transcriptomes from white-footed mice (*Peromyscus leucopus*) in New York City. We analyzed transcriptomes from 48 *P. leucopus* individuals from three urban and three rural populations for evidence of rapid local adaption in isolated urban habitats. We generated a dataset of 154,770 SNPs and analyzed patterns of genetic differentiation between urban and rural sites. We also used genome scans and genotype-by-environment (GEA) analyses to identify signatures of selection in a large subset of genes. Neutral demographic processes may create allele frequency patterns that are indistinguishable from positive selection. Thus, we accounted for demography by simulating a neutral SNP dataset under the inferred demographic history for the sampled *P. leucopus* populations to serve as a null model for outlier analysis. We then annotated outlier genes and further validated them by associating allele frequency differences with two urbanization variables: percent impervious surface and human population density. Many candidate genes were involved in metabolic functions, especially dietary specialization. A subset of these genes have well-established roles in metabolizing lipids and carbohydrates, including transport of cholesterol and desaturation of fatty acids. Our results reveal clear genetic differentiation between rural and urban sites that resulted from rapid local adaptation and drift in urbanizing habitats. The specific outlier loci that we identified suggest that populations of *P. leucopus* are using novel food resources in urban habitats and selection pressures are acting to change metabolic pathways. Our findings support the idea that cities represent novel ecosystems with a unique set of selective pressures.

Keywords: transcriptome, *Peromyscus leucopus*, genotype-environment association, genome scans, positive selection, demographic null-model, urbanization

INTRODUCTION

Traits are adaptive when they increase an organism's fitness in a specific environment (Barrett & Hoekstra 2011). The identification of genotypes underlying adaptive traits is a major goal in evolutionary biology. Many studies have identified the genetic basis underlying adaptation, but they often focus on a small number of well-known, conspicuous traits (Nachman *et al.* 2003; Pool & Aquadro 2007; Linnen *et al.* 2009; Storz *et al.* 2009). With costs of high-throughput DNA sequencing continuing to drop by orders of magnitude (De Wit *et al.* 2015), generating genomic datasets for natural populations of non-model organisms is feasible. These datasets facilitate reverse-genomics approaches where candidate genes behind ecologically relevant, but non-conspicuous, phenotypes are identified based on patterns of variation and signatures of selection in protein-coding sequences (Li *et al.* 2008). Here we examined local adaptation in isolated urban populations of white-footed mice, *Peromyscus leucopus*, in New York City (NYC). We identified regions of *P. leucopus* transcriptomes with divergent allele frequencies suggestive of positive selection. We incorporated a neutral SNP dataset from an inferred demographic history (Harris *et al.* 2016) directly into our null model for identifying outliers. We then examined statistical associations between allele frequencies and environmental measures of urbanization.

Adaptive processes leave a predictable pattern of genetic variation and differentiation along environmental gradients (Savolainen *et al.* 2013). Examining genetic variation using the site frequency spectrum (SFS), the distribution of allele frequencies across the genome, can be an efficient method of detecting these adaptive processes (Merila & Hendry 2014). Our goal was to identify local instances of adaptation from specific patterns in the SFS. Local adaptation, while difficult to identify from genetic signals, is a common pattern in nature (Stinchcombe &

Hoekstra 2008; Bonin 2008; Linnen *et al.* 2009; Hohenlohe *et al.* 2010a; Turner *et al.* 2010; Ellison *et al.* 2011; De Wit & Palumbi 2013), and uncovering the genetic basis of local adaptation has provided insight into a variety of evolutionary processes including speciation, maintenance of genetic diversity, range expansion, and species responses to changing environments (Savolainen *et al.* 2013; Tiffin & Ross-Ibarra 2014). Urban habitats are one of the fastest growing and most rapidly changing environments around the world and may be ideal environments for local adaptation. Urbanization leads to habitat loss and fragmentation, changes in resource availability, novel species interactions, altered community composition, and increased exposure to pollutants (McKinney 2002; Chace & Walsh 2004; Shochat *et al.* 2006; Sih *et al.* 2011). These ecological changes may exert strong selective pressure, and there is mounting evidence that rapid adaptation occurs in many urban organisms. Another cause of rapidly changing environments is global climate change, where increasing temperatures and altered precipitation patterns strongly influence the life history traits of many species (Franks & Hoffmann 2011; Franks *et al.* 2014). These two processes, urbanization and climate change, are not mutually exclusive. Understanding local adaptation in urban habitats may lead to general insights about local adaptation to anthropogenic climate change, such as what traits are involved or how quickly populations respond and adapt to changing environments.

Peromyscus leucopus is one of the most abundant small mammals in North America, preferring the typical oak-hickory forest commonly found in the eastern USA (Wang *et al.* 2008). They are generalists that burrow in a variety of habitats (Metzger 1971; Vessey & Vessey 2007), and feed on a wide-range of invertebrates, nuts, fruit, vegetation, and fungus (Wolff *et al.* 1985; Ostfeld *et al.* 1996). They are especially reliant on oak mast cycles and an important predator of gypsy moths (Ostfeld *et al.* 1996). There is also evidence that *Peromyscus* spp. can adapt to environmental change (Storz *et al.* 2007, 2009, 2010; Mullen & Hoekstra 2008; Linnen *et al.* 2009; Weber *et al.* 2013; Natarajan *et al.* 2013; Munshi-south & Richardson 2016), making them good candidates for the study of local adaptation. White-footed mice are one of the few native mammals that thrive in extremely small, fragmented urban forests in North America (Pergams & Lacy 2007; Rogic *et al.* 2013; Munshi-South & Nagy 2014). *P. leucopus* tend to be found at higher densities in urban patches due to a thick understory and fewer predators and competitors (Rytwinski & Fahrig 2007). Increased density may also be due to limited *P. leucopus* dispersal between urban sites. Munshi-South (2012) found barriers to dispersal between isolated NYC parks, with migrants only moving through significantly vegetated corridors throughout the city. There is also substantial genetic structure between NYC parks as measured by microsatellites (Munshi-South & Kharchenko 2010), genome-wide SNPs (Munshi-South *et al.* 2016) and demographic modeling (Harris *et al.* 2016). We have also previously identified signatures of selection in urban populations of NYC white-footed mice (Harris *et al.* 2013), though we used smaller datasets and more limited approaches than presented here. This study builds on our previous work by employing a larger dataset and more comprehensive statistical analyses to identify signatures of selection in *P. leucopus* populations while explicitly using the inferred demographic history as a null model. We further confirm outlier genes by associating allele frequencies with environmental metrics of urbanization and perform enrichment analyses to predict functional relevance of outlier genes.

Urbanization and global climate change are relatively recent disturbances that rapidly change native ecosystems. Over short timescales, adaptive evolution tends to act on standing genetic variation as opposed to de novo mutations (Barrett & Schluter 2008; Stapley *et al.* 2010). As these pre-existing mutations spread to fixation they produce a detectable signal in the form of 'hard' or 'soft' selective sweeps (Hermisson & Pennings 2005; Messer & Petrov 2013). Additionally, ecologically important traits involved in local adaptation are often quantitative traits with many genes of small effect involved in producing the desired phenotype (Orr 2005; Rockman 2012). To distinguish between these more subtle signatures of selection, we used multiple tests that provide greater statistical power and higher resolution at identifying types and age of selection when used together (Grossman *et al.* 2010; Hohenlohe *et al.* 2011).

We analyzed transcriptomes sequenced from urban and rural populations of *P. leucopus* to produce estimates of nucleotide diversity (π, Tajima 1983), Tajima's *D* (Tajima 1989), and F_{ST} (Wright 1951) and make inferences about the evolutionary processes at work in these populations. We also used a variety of tests to identify outlier genes subject to selection, and took extra steps to account for the potentially confounding effects of demography. Specifically, neutral demographic processes, like population bottlenecks, can produce signatures of variation similar to those produced by selection (Oleksyk *et al.* 2010; Li *et al.* 2012). For example, both selection and a population bottleneck followed by an expansion may produce genomic regions with low genetic diversity, but recent literature discusses how to deal with these overlapping signals (Excoffier *et al.* 2009; Li *et al.* 2012; Vitti *et al.* 2013; Lotterhos & Whitlock 2015). The prevailing approach assumes selection acts on one or a few loci while demographic processes act across the genome. Outlier tests for loci under selection typically generate a null distribution based on an island model of population differentiation (Excoffier *et al.* 2009), and then identify candidate genes with genetic differentiation that exceeds this simulated null distribution. The true demographic history of most organisms is much more complex, and computational approaches have been developed to robustly infer demographic parameters (Gutenkunst *et al.* 2009; Excoffier *et al.* 2013). This inferred demographic history can then be used to construct a more realistic null model, reducing the rate of false positives in tests for selection (Excoffier *et al.* 2009; Yoder *et al.* 2014).

We used the inferred demographic history of urban populations of *P. leucopus* (Harris *et al.* 2016) to simulate comparable SNP datasets to our observed sequence data. We then used multiple approaches that identify outlier loci based on population differentiation, the SFS, or associations between allele frequencies and environmental variables. Bayescan uses a Bayesian approach to identify SNPs that exhibit extreme allele frequency divergence between populations (Foll & Gaggiotti 2008). SweeD is a likelihood based test that identifies selective sweeps based on SFS that deviate from neutral expectations. We examined associations between allele frequencies and environmental variation using a genotype-environment association (GEA) test. GEA tests have been shown to perform better than outlier tests under complex demographic scenarios (Lotterhos & Whitlock 2015) but can suffer from a high rate of false positives. Analyses suggest that using genome scan-based outlier tests in conjunction with GEA tests leads to reliable outlier identification (De Villemereuil *et al.* 2014). GEA also identifies local adaptation in polygenic phenotypes where each polymorphism has a relatively weak effect (Frichot *et al.* 2013), because correlations between alleles and environmental variables do not rely on the strength of genetic differentiation or SFS skew between populations. (Pavlidis *et al.* 2013). Using multiple analyses with alternative statistical approaches is preferred for genome scans, and provides more power and confidence in results when markers are repeatedly found as outliers (Grossman *et al.* 2010). BayPass also uses a Bayesian approach to identify divergent adaptive processes (Gautier 2015), but explicitly incorporates population demographic history including hierarchical population structure. Conveniently, it also uses a population covariance matrix to associate SNPs with population-specific environmental covariables. This feature allowed us to use BayPass to identify congruence across outliers identified in Bayescan and LFMM.

In this study, we examined transcriptomes generated from RNAseq for 48 *P. leucopus* individuals from three urban sites in NYC and three rural sites from the surrounding area. Including population pairs that are near each other and genetically similar, but occur in different environments (urban versus rural), increases the power to identify candidate genes under selection (Lotterhos & Whitlock 2015). We used traditional population genetic summary statistics to generate per-site estimates and identify loci that deviate from neutral expectations. Next, we used several tests of selection to determine whether these deviations are due to recent selection. To increase power, reduce false positives, identify more subtle signals of selection from standing genetic variation, and find candidate genes involved in polygenic phenotypic traits, we simulated a null background model from the inferred demographic history for NYC populations of *P. leucopus.* We examined the association between quantitative metrics of urbanization (percent impervious surface and human population density) and polymorphisms between rural and urban populations to identify candidate genes experiencing selection in NYC. We used overlapping results from multiple tests and environmental associations to generate a robust list of candidate genes involved in local adaptation of *P. leucopus* to the urban environment. Evidence of local adaptation in urban populations reveals how urbanization acts as an evolutionary force, gives insights into important traits for local adaptation, and provides evidence of rapid evolution in novel, human-dominated environments.

MATERIALS AND METHODS

Sampling, library preparation, and transcriptome assembly

We trapped and collected white-footed mice from 2009 - 2013. For full details on sampling and transcriptome sequencing, see Harris et al. (2015). In brief, we randomly chose eight individual white-footed mice (equal numbers of males and females) from six sampling locations representative of urban and rural habitats and with minimal within-site genetic structure (Fig. 1) (Harris *et al.* 2013, 2015). Three sampling sites occurred within NYC parks: Central Park in Manhattan (CP), New York Botanical Gardens in the Bronx (NYBG), and Flushing Meadows—Willow Lake in Queens (FM). These sites represented urban habitats surrounded by high levels of impervious surface cover and high human population density, as previously quantified in Munshi-South et al. (2016). The remaining three sites occurred ~100 km outside of NYC in rural, undisturbed habitat representative of natural environments for *Peromyscus leucopus*. High Point State Park is in the Kittatinny Mountains in New Jersey (HIP), Clarence Fahnestock State Park is located in the Hudson Highlands in New York (CFP), and Brookhaven and Wilde Wood State Parks and neighboring sites occur on the northeastern end of Long Island, New York (BHWWP). We sacrificed mice on site and liver, gonad, and brain tissue were harvested in the field for immediate storage in RNAlater (Ambion). In the lab, we extracted total RNA, removed ribosomal RNA, barcoded each tissue type, and then pooled samples during library preparation. The reverse transcribed cDNA was sequenced using the 454 GS FLX+ and SOLiD 5500 xl systems using standard RNAseq protocols. We called SNPs with the Genome Analysis Toolkit pipeline using a Bayesian genotype likelihood model (GATK version 2.8, DePristo *et al.* 2011) and removed related individuals. See Harris *et al.* 2013, 2015 for full transcriptome sequencing, assembly and SNP calling details, but in short, for SNP calling we required coverage >5X, nucleotide quality >30, no strand bias (FS >35), and SNPs from a uniquely mapped read. We also removed SNPs where every individual was heterozygous, overall depth >10, overall depth <350 and minor allele frequency (MAF) >0.025. The VCF file of SNP genotypes used for demographic inference is on the Dryad digital repository at http://dx.doi.org/10.5061/dryad.d48f9, raw sequencing files for the transcriptome are deposited in the GenBank Sequence Read Archive (SRA Accession no. SRP020005), and transcriptome contigs are available in the Dryad digital repository, doi: 10.5061/dryad.6hc0f.

Summary statistics

SNP information was stored in a VCF (variant call format) file and summary statistics were calculated using vcftools 0.1.12b (Danecek *et al.* 2011). We calculated per-site nucleotide diversity (π), Tajima's *D*, and F_{ST} for each site. We also calculated the statistics for each contig (per-site statistic summed across all SNPs per contig divided by total sites) and calculated the average estimate for each population, including all pairwise population comparisons for $F_{ST.}$

Scans for positive selection based on population differentiation

We used information from multiple previous studies on *P. leucopus* in order to choose our final subset of urban and rural sites for this study. White-footed mice respond surprisingly well to habitat fragmentation (Pergams & Lacy 2007; Rogic *et al.* 2013), including forested urban fragments, which are often densely populated with mice (Munshi-South & Nagy 2014). Previous work suggests that migration is relatively low, only occurring along vegetated pathways between urban parks (Munshi-South 2012). This isolation leads to genetic differentiation between populations in different NYC parks, which was confirmed using microsatellite loci

(Munshi-South & Kharchenko 2010), genome-wide neutral SNPs (Munshi-South *et al.* 2016), and protein coding sequences (Harris *et al.* 2013, 2015). Previous analysis of the demographic history of populations occupying contemporary forest fragments in NYC and the surrounding area estimated that population divergence occurred within the time frame of urbanization (Harris *et al.* 2016). The three urban and three rural sites chosen to investigate patterns of selection in fragmented urban parks in this study represent sampling sites with the strongest evidence of being independent evolutionary clusters. We used the F_{ST} based analysis implemented in Bayescan v. 2.1 (Foll & Gaggiotti 2008) to compare all six population-specific allele frequencies with global averages and identify outlier SNPs. Bayescan identifies loci that exhibit divergence between groups that is stronger than would be expected under neutral genetic processes. Based on a set of neutral allele frequencies under a Dirichlet distribution, Bayescan uses a Bayesian model to estimate the probability that a given locus has been subject to selection. To generate more realistic allele frequency distributions, we used Bayescan for independent coalescent simulations of SNP datasets based on the neutral demographic history inferred specifically for each *P. leucopus* population in (Harris *et al.* 2016). We generated 100 sets of 100,000 SNPs for each population in this study from a three population isolation-with-migration model using the previously inferred parameter estimates for divergence time, effective population size, migration rate, and population size change in the coalescent-based software program, fastsimcoal2 (Excoffier *et al.* 2013). In short, the model represented a deep split between an ancestral population into Long Island, NY and the mainland (including Manhattan) 29,440 generations before present (GBP). Migration was asymmetrical from the mainland into Long Island and a third population (representing the sampling sites in this study) later became isolated 746 GBP. Urban populations were also modeled to include a bottleneck event at the time of divergence.

Finally, we allowed migration to occur between all three populations (Harris *et al.* 2016). Bayescan was run independently on each simulated dataset using default parameters. Using the observed SNP dataset, we performed a global analysis, one Bayescan run where all individuals were partitioned into urban and rural groups, and finally analyses on all individual pairwise population comparisons. Outlier SNPs were retained if they had a false discovery rate (FDR) value ≤ 0.1 and if the posterior odds probability from Bayescan was higher than for any value calculated from the simulated dataset. Outlier SNPs with a FDR ≤ 0.1 were considered significant, implying that diversifying selection better explains allele frequency differences between urban and rural populations (urban vs. rural) and sub-populations (pairwise population comparison) than a neutral null model. A relatively high FDR was chosen for all analyses to ensure inclusion of all putative outlier SNPs.

We reduced the risk of including false positives by also using the software program BayPass (Gautier 2015) to identify putative SNPs showing evidence of divergent selection between populations. We filtered our final outlier SNP list to only include those identified in both Bayescan and BayPass. BayPass incorporates population demographic history when identifying outlier SNPs (Gautier 2015) based on associations between allele frequencies and environmental variables. We ran BayPass using default parameters under the AUX model (Table S2). BayPass uses the XtX differentiation measure to identify differentiated SNPs. We created an empirical distribution of XtX values for each locus by analyzing pseudo-observed data sets (PODs) and chose a 5% threshold value for XtX to use as the cutoff value to differentiate between selection and neutrality (Gautier 2015). PODs were also used to determine a 5% threshold value for Bayes Factors used for associating environmental covariables with allele frequencies.

Analysis for selective sweeps

We also identified outlier regions when the observed SFS showed an excess of low frequency and high frequency minor alleles, a signal indicative of a recent selective sweep. The composite likelihood ratio (CLR) statistic is used to identify regions where the observed SFS matches the expected SFS generated from a selective sweep (Kim & Stephan 2002; Nielsen *et al.* 2005; Pavlidis *et al.* 2010). We calculated the CLR along sliding windows across the transcriptome using the software program SweeD (Pavlidis *et al.* 2013). SweeD is an extension of Sweepfinder (Nielsen *et al.* 2005) that is optimized for large next generation sequencing (NGS) datasets. SweeD was run separately for each population and on individual contigs using default parameters except for setting a sliding window size of 200 bp and using the folded SFS, as we lacked an outgroup to infer the ancestral state. The window within each contig with the highest CLR score is the likely location of a selective sweep. Similar to the method used for Bayescan analyses, statistical significance was established from a null distribution generated by running SweeD on SNP datasets simulated under the inferred demographic history for *P. leucopus* populations (Harris *et al.* 2016). SweeD does not inherently identify outlier regions. The CLR is computed using a selective sweep model on the observed data and then compared to a neutral model calibrated with the background SFS generated from simulations. As before, we used 100 datasets with 100,000 SNPs each, simulated under the inferred neutral demographic history for urban and rural populations of white-footed mice in NYC. The CLR was calculated using SweeD for all simulated datasets and the resulting distribution was used to set a significance cutoff. For the observed dataset, we lacked a genome to provide clear linkage information so SweeD was run separately on each contig. We identified outlier contigs if their

CLR value was greater than any produced when calculated for neutral simulations. We also required outliers to fall within the top 0.01% of the CLR distribution for the observed SNPs.

Genotype-environment association tests for environmental selection

We used the GEA approach of LFMM: Latent Factor Mixed Models (Frichot *et al.* 2013) to associate outlier SNPs and candidate loci identified above with potential environmental selection pressures. LFMM examines associations between environmental and genetic variation while accounting for the neutral genetic background and structure between populations (Frichot *et al.* 2013). We tested three environmental variables associated with urbanization: 1) percent impervious surface within a 2 km buffer around each sampling site, 2) human density within a two-kilometer buffer around each sampling site, and 3) designating each site as urban or rural. We previously found that variables 1-2 are significantly associated with genome-wide variation in *P. leucopus* populations in the NYC metropolitan area (Munshi-South *et al*. 2016). Our final data set included all individuals but only the subset of outlier SNPs that were detected in Bayescan and SweeD. LFMM requires the user to define the number of latent factors, K, that describe population structure in the dataset. To identify the appropriate number of K latent factors, we performed a genetic PCA followed by a Tracy-Widom test to find the number of eigenvalues with *P* values ≤ 0.01 (Patterson *et al.* 2006; Frichot & François 2015). Based on this approach, we ran LFMM with default parameters except for K = 6, number of MCMC cycles = 100,000, and burn-in = 50,000. Using author recommendations, we combined 10 replicate runs and readjusted the p values to increase the power of the test. LFMM uses |z|- scores to report the probability of a SNP's association with an environmental variable. After correcting for multiple testing, we used a cutoff value of q ≤ 0.1.

Similar to the approach described above, we increased statistical power by repeating the GEA test in a separate analysis. We used the auxiliary variable model in the program BayPass (Gautier 2015) to identify associations between allele frequencies and environmental variables. We filtered our final list of markers to only include those identified in both LFMM and BayPass. PODs were also used to determine a 5% threshold value for Bayes Factors used for associating environmental covariables with allele frequencies.

Functional annotation of candidate gene

We used the gene annotation pipeline in Blast2GO (Conesa *et al.* 2005; Götz *et al.* 2008) to find sequences from the NCBI non-redundant protein database that were homologous to our outlier contigs identified above. We then retrieved associated gene ontology (GO) terms. Blast2GO retrieves GO terms associated with BLASTX hits and uses the KEGG database to describe biochemical pathways linking different enzymes (Ogata *et al.* 1999; Kanehisa *et al.* 2014). For downstream enrichment analyses, we also used the Ensembl gene annotation system (Aken *et al.* 2016) to find homologous *Mus musculus* genes for each *P. leucopus* contig (Table S3). We further interpreted the outlier gene lists using g:Profiler (Reimand *et al.* 2016) to identify gene ontology terms enriched in our outlier gene list compared to the fully annotated *Mus musculus* genome. Poorly updated gene annotation databases can significantly affect results and g:Profiler is one of the most comprehensive and most often updated gene annotation databases available (Wadi *et al.* 2016). We used the g:Profiler webserver and identified enriched terms from the full outlier gene list using default parameters displaying only significant results (Table S3). We visualized and summarized the enriched gene ontology list using the revigo webserver (Supek *et al.* 2011).

RESULTS

Genetic diversity statistics

We retained 154,770 total SNPs for use in looking at patterns of genetic variation and performing tests of selection. For each population we obtained estimates of nucleotide diversity, Tajima's *D*, and pairwise F_{ST}. Urban populations had a two-fold decrease in nucleotide diversity compared to the rural populations (Table 1). The average nucleotide diversity for all three rural populations was 0.224 ± 0.034, while the average for urban populations was only 0.112 ± 0.019. The average Tajima's *D* calculation within populations did not show substantial differences between populations (Table 1). For all populations, Tajima's *D* was slightly positive. Average pairwise F_{ST} calculated using vcftools ranged from a low of 0.018 ± 0.364 between two rural populations (CFP – HIP) to a high of 0.110 ± 0.520 between two urban populations (CP – FM, Table S5). These F_{ST} values were similar to F_{ST} for neutral genome-wide SNP datasets from the same *P. leucopus* populations (Munshi-South *et al.* 2016). Comparisons between rural populations had the lowest F_{ST} values, urban to rural pairs had the second lowest, and urban to urban pairs had the highest overall F_{ST} values despite occurring less than 5 km apart (Table S5).

Outlier detection

The global Bayescan analysis identified 309 SNPs potentially under the influence of divergent selection. After sampling sites were grouped as urban or rural, Bayescan identified 40 SNPs with signatures of positive selection (Fig. 2A, Table 2). Eight of these SNPs were also found in the global analysis. Individual urban to rural population comparisons did not find any outlier SNPs, and zero SNPs exhibited signatures of balancing selection. F_{ST} for outlier SNPs ranged from 0.21 - 0.33, much higher than the population median of 0.059. Bayescan identified zero outlier SNPs in the simulated neutral dataset. However, we only included outlier SNPs from the observed dataset with FDR and posterior odds values that were smaller and larger, respectively, than the most extreme values for the simulated data (FDR $\leq$ 0.6 and $\log_{10}$(PO) $\geq$-0.196).

To generate the null distribution of the CLR statistic for analyses in SweeD, we tested the 100 SNP datasets simulated under the inferred demographic history for NYC populations of *P. leucopus*. We found that CLR scores in the top 5% of the simulated distribution were generally 2-3X lower than values in the top 5% of the observed dataset. We ran SweeD on observed SNPs within individual contigs and identified outliers by filtering for a CLR score $\geq$ 3.53 (the maximum CLR from simulated data). We also chose regions that fell within the top 0.01% of the observed distribution (Fig. 2B). SweeD identified regions with SFS patterns that fit a selective sweep model in 55 contigs within urban populations (Table 3). Contig 35790-44, annotated as the lipid transporter *Apolipoprotein B100*, had the highest CLR (8.56). All outliers had CLR scores $\geq$ 4.97. Bayescan and SweeD did not identify any of the same outliers.

The BayPass analysis identified 59 SNPs that showed evidence divergent selection. We used PODs to estimate a null distribution and identified SNPs with XtX values $\geq$ 8.35 (top 5% of the null distribution). BayPass also identified 33 of the 40 outliers (82.5 %) from the Bayescan analysis, and 26 of the 55 outliers (47.3 %) from the SweeD analysis.

Environmental associations

Thirty of the 40 (75%) outliers identified using Bayescan were significantly associated with at least one of the three environmental variables tested using LFMM (Fig. 3A, Table 2). All

30 of the identified SNPs were associated with the binary classification of urban vs. rural. Only seven of the outlier SNPs were associated with percent impervious surface and five were associated with human population density. Twenty-six of the 55 outlier contigs identified using SweeD were associated with one of the environmental variables (Table 3). Again, all 26 regions were associated with the urban vs. rural site classification. Fourteen outliers from SweeD were associated with percent impervious surface and eight were associated with human population density. Some contigs associated with environmental variables were outliers in only one urban population, possibly indicating local adaptation within parks, selection on a polygenic trait, or genetic drift.

The only environmental variable significantly associated with SNPs in the BayPass analysis was urban or rural classification. Percent human density and percent impervious surface cover did not show significant associations. All outliers identified in Bayescan, BayPass, SweeD, and LFMM showed associations with urban versus rural classification (5% threshold value, BF ≥ 17.8, Table S2).

Functional annotation

The full contig sequences containing outlier SNPs were obtained from the *P. leucopus* transcriptome (Harris *et al.* 2015) and used to identify functional annotations. Of the 40 contigs identified by Bayescan as divergent between urban and rural populations, 36 were annotated with gene names and functional information (Table 2). Of these, 29 were also associated with urban environmental variables. The ten most frequent GO terms among the Bayescan outliers involved organismal metabolism (Table S1). Some outliers occurred within sequences homologous with genes of known functions and biochemical pathways. These outliers included a farnesoid-x-receptor (FXR, Contig 25795-154), a myosin light chain kinase (MYLK, Contig 7975-418), and the gene SORBS2 (Contig 37967-26).

Of the 55 contigs with signatures of selection identified by SweeD, forty-nine were annotated with gene names and gene ontology terms, and 25 were significantly associated with urbanization variables. Many of these sequences were homologous with genes involved with basic metabolic functions such as glycolysis and ATP production (Table S1). Contig 35790-44 was homologous to the gene APOB, an apolipoprotein, and Contig 10636-348 to an aflatoxin reductase gene AKR7A1. Other outliers were identified as the gene FADS1, part of the fatty acid denaturase family (Contig 342-1776), and a heat-shock protein (Hsp90, Contig 3964-627). Most gene annotations did not have known phenotypic traits related to their function, but KEGG analysis revealed several contigs involved in the same biochemical pathways: galactose metabolism, fructose metabolism, and mannose metabolism (Fig. S1).

The results from g:Profiler and Revigo show that the identified outlier genes have functions primarily related to metabolic processes. There were 101 GO terms that were significantly overrepresented in the list of outlier genes compared to the curated *Mus musculus* gene list from g:Profiler (Table S3). The top 5 GO terms that occurred with the highest frequency across the outlier genes were metabolic process, cellular process, organic substance metabolic process, cellular metabolic process, and primary metabolic process, respectively (Table S4). Metabolic processes comprised 82% of the overrepresented GO terms. There were also several unique clusters with multiple GO terms dealing with proteolysis, organic substance transport, and nitrogen utilization. The largest cluster of individual GO terms dealt with lipid metabolism and response to lipids (Table S4).

DISCUSSION

The results of this study provide insight into the genetic basis of local adaptation when populations evolve in response to rapidly changing environments. We previously found evidence for older occurrences of divergent selection in NYC white-footed mice by investigating non-synonymous polymorphisms in pooled transcriptome samples (Harris *et al.* 2013). There was little overlap between previous results and those found here, but this dataset was much larger, included more sampling sites, and used analyses that identify more recent signatures of selection. However, two of the eleven previously identified candidate genes (Harris *et al.* 2013) were direct matches to outliers in this current analysis (Serine protease inhibitor a3c and Solute carrier organic anion transporter 1A5), and three other genes were from the same gene families or involved in the same biological processes. One gene, an aldo-keto-reductase protein, is part of the same gene family as the aflatoxin reductase gene (Contig 10636-348) identified in this study. The aldo-keto reductase gene family comprises a large group of essential enzymes for metabolizing various natural and foreign substances (Hyndman *et al.* 2003). Two others, camello-like 1 and a cytochrome P450 (CYPA1A) gene, are involved in metabolism of drugs and lipids. In *Peromyscus* spp., CYPA1A is directly expressed along with Hsp90 (outlier from current SweeD analysis) when exposed to environmental toxins (Settachan 2001).

In this study, we observed patterns of divergent positive selection between urban and rural populations of *P. leucopus*, and were able to associate outlier SNPs with environmental variables relevant to urbanization. The majority of candidate loci were annotated with GO terms that are significantly associated with organismal metabolism, particularly breakdown of lipids and carbohydrates. We discuss what these findings mean for organisms inhabiting novel urban ecosystems, and more generally for understanding the ecological processes and time frame of local adaptation in changing environments.

The utility of using genome scan methods to test for selection

Over the past decade, genome scans have become feasible methods to detect and disentangle neutral and adaptive evolutionary processes (De Villemereuil *et al.* 2014). One of the most popular approaches looks at locus-specific allele frequency differentiation between sampling locations as measured by F_{ST} (Lewontin & Krakauer 1973; Weir & Cockerham 1984). Sites with extremely high allele frequency differences may be subjects of positive directional selection. Bayescan (Foll & Gaggiotti 2008) calculates the posterior probability that a site is under the influence of selection by testing models with and without selection. The model that does not invoke selection is based on a theorized neutral distribution of allele frequencies.

While Bayescan has been shown to be relatively robust to confounding demographic processes (Pérez-Figueroa *et al.* 2010; De Villemereuil *et al.* 2014), population bottlenecks, hierarchical structure, recent migration, or variable times to most-recent-common-ancestor (MRCA) between populations can artificially inflate F_{ST} values (Hermisson 2009; Lotterhos & Whitlock 2014). We minimized false positives by incorporating population structure and a specific demographic history for *P. leucopus* in NYC directly into the null distribution of F_{ST}. (Harris *et al.* 2016). We only included outliers if their posterior probability was greater than probabilities calculated from simulations. The outliers comprised 0.024% of the total number of loci analyzed from the transcriptome. This percentage is in line with candidates uncovered from a similar study (0.05%) that looked at high and low altitude populations of the plant *S. chrysanthemifolius* (Chapman *et al.* 2013). Many studies find higher percentages of outlier loci using Bayescan; for example, 4.5% in the American pika across its range in British Colombia (Henry & Russello 2013), and 5.7% in Atlantic herring across their range (Limborg *et al.* 2012). Our lower overall percentage of outliers may be due to the use of the inferred demographic history to establish outlier cutoffs and reduce false positives, or because of the relatively recent isolation or strength of selection in urban populations.

SweeD, another genome scan approach, examines patterns within a population's SFS rather than allelic differentiation between populations. The main footprint that selective sweeps leave on the SFS is an excess of rare low-and high-frequency variants (Nielsen 2005). The SweepFinder method (Nielsen *et al.* 2005), recently upgraded to the NGS compatible SweeD (Pavlidis *et al.* 2013), uses a CLR test based on the ratio between the likelihood of a neutral and selective sweep hypothesis. As above, the weakness of hitchhiking methods is the confounding influence certain demographic processes have on the SFS (Hermisson 2009). However, building a robustly inferred demographic history into the null model substantially reduces false positive rates (Pavlidis *et al.* 2013).

We included the *P. leucopus* demographic history into our analysis, and found 0.019% of the transcriptome to contain SFS patterns indicative of selective sweeps. This rate is in line with other studies that reported that 0.5% of regions in domesticated rice (Wang *et al.* 2014), 0.02% of loci in black cottonwood (Zhou *et al.* 2014), and 0.02% of the gorilla genome (McManus *et al.* 2014) show evidence of selective sweeps or hitchhiking.

Several studies have shown that performing multiple tests that employ diverse theoretical approaches is the best way to avoid Type I and II errors in genome outlier analyses (Nielsen 2005; Grossman *et al.* 2010; Hohenlohe *et al.* 2010b). We used Bayescan and SweeD to identify signatures of positive selection, and confirmed outliers using BayPass to identify divergent selection while incorporating genetic structure. While BayPass confirmed the majority of outliers identified using other methods (Table S2), there was no overlap between Bayescan and SweeD outliers. This discrepancy is likely due to the different selection scenarios underlying each test, i.e. divergent local selection versus population-wide positive selection in the form of selective sweeps (Hermisson 2009). F_{ST} based methods can respond to allelic divergence relatively quickly, while models for selective sweeps typically require nearly-fixed derived alleles (Hohenlohe *et al.* 2010b). Given the recency of urbanization in NYC, many selective sweeps may be ongoing or otherwise incomplete. Selection may also be acting on standing genetic variation in the form of soft sweeps (Hermisson & Pennings 2005) that are not readily identified by SweeD (De Villemereuil *et al.* 2014). We identified several outliers that were unique to specific urban populations, which is characteristic of soft sweeps and polygenic traits (Messer & Petrov 2013). Despite the lack of overlapping outliers between the two tests, further confirmation of outlier genes experiencing positive selection was provided by genotype-environment association tests. These methods may often be more powerful than the genome scans above (Savolainen *et al.* 2013).

Environmental associations strengthen evidence of local adaptation to urbanization

GEA tests are a growing class of methods that identify loci with allele frequencies that are associated with environmental factors (Joost *et al.* 2007; Coop *et al.* 2010; Frichot *et al.* 2013). Here we used LFMM (Frichot *et al.* 2013) to associate outlier SNPs with environmental metrics of urbanization. LFMM performs better than other methods in the presence of hierarchical structure and when polygenic selection is acting on many loci with small effect (De Villemereuil *et al.* 2014). Hierarchical structure in our dataset includes urban and rural differentiation (Harris *et al.* 2015; Harris *et al.* 2016), patterns of geographic structure between mainland mice and Long Island, NY (Harris *et al.* 2016), and population structure between individual urban parks (Munshi-South & Kharchenko 2010). Simulations also suggest that LFMM is superior when sample size is less than 10 individuals per population, there is no pattern of IBD, and the study compares environmentally divergent habitats (Lotterhos & Whitlock 2015). We sampled eight white-footed mice per population, found no evidence of IBD (Munshi-South *et al.* 2016), and sampled environmentally divergent rural and urban locations.

Using LFMM, we found that 75 % and 47 % of outliers from Bayescan and SweeD, respectively, were significantly associated with one or more urbanization variables. BayPass also confirmed associations between all outlier SNPs and urbanization variables, though only with the binary classification of a site as urban or rural. These results are consistent with other studies combining genome scan methods and GEA tests. Limborg *et al.* (2012) found 62.5% of the outliers identified in Bayescan were correlated with temperature or salinity in Atlantic herring, and 26.3% of genome scan outliers were associated with temperature or latitude in a tree species (De Kort *et al.* 2014). We acknowledge that percent impervious surface, human population density, or binary classification as urban versus rural may not capture the specific, causative selection pressures acting on white-footed mouse populations. We used these metrics as general proxies for ecological processes that in urbanized habitats. The percent of impervious surface around a park is likely representative of habitat fragmentation, as urban infrastructure changes the net primary productivity due to increasing percentages of impervious surface or artificial landscapes, parks and yards (Shochat *et al.* 2006). This fragmentation then leads to changing species interactions as migration is impeded or organisms are forced into smaller areas (Shochat *et al.* 2006). The percent human density surrounding an urban park can serve as a proxy for the multitude of ecological changes humans impose on their surrounding environment. Humans often introduce invasive species into cities (Sih *et al.* 2011), leading to increased competition or novel predator-prey interactions. Urbanization and increasing human density also change the types and availability of resources in the altered habitat (McKinney 2002; Sih *et al.* 2011). Finally, classifying our sites as urban or rural can generally capture the main differences in urban and natural sites. For example, pollution is a major consequence of urbanization (Donihue & Lambert 2014), and urban areas often include increased chemical, noise, or light pollution (Sih *et al.* 2011).

Between divergent allele frequencies, a skewed SFS, environmental associations, and overrepresented GO terms, we find several overlapping lines of evidence that support rapid divergent selection in white-footed mice. Evidence of selection operating in urban environments is accumulating (Donihue & Lambert 2014), and our results are in line with other studies that have found rapid local adaptation to urbanization. Yeh (2004) found sexually-selected tail coloration in juncos was rapidly evolving in urban populations compared to rural ones. European blackbirds show reduced migratory behavior in cities, and there is also evidence of selection on genes underlying anxiety behavior across multiple urban areas (Partecke *et al.* 2006; Mueller *et al.* 2013). Cheptou *et al.* (2008) reported that weeds in urban vegetation plots surrounded by paved surfaces showed heritable changes in seed morphology and disperal. Thompson *et al.* (2016) found parallel adaptive evolution to urbanization in white clover, *T. repens*, by identifying reduced cyanogenesis and freezing tolerance in plants in response to warmer minimum ground temperatures in urban areas relative to rural areas. Rapid adaptation for polychlorinated biphenyl (PCB) resistance occurred in both killifish and tomcod inhabiting urban water bodies (Whitehead *et al.* 2010; Wirgin *et al.* 2011).

Functional roles and ecological relevance of candidate genes

The model rodents *Mus musculus, Rattus norvegicus,* and *Cricetulus griseus* all have deeply sequenced, assembled and annotated reference genomes. These resources allowed us to annotate 89.5% of outlier loci with high quality functional information. Urban *P. leucopus* exhibited signatures of positive selection in genes with GO terms overrepresented for organismal metabolic processes, specifically digestion and metabolism of lipids and carbohydrates.

While not significantly overrepresented, association with mitochondrial processes was another of the most common annotations among our outlier loci (Table S1). While we can only speculate until further physiological studies are conducted, our evidence suggests that the evolution of mitochondrial and metabolic processes has been important to the success of *P. leucopus* living in NYC's urban forests. Mitochondrial genes have often been used to describe neutral population variation, but researchers have found ample evidence of selection acting on the mitochondrial genome (Oliveira *et al.* 2008; Balloux 2010). For example, specific mitochondrial haplotypes are associated with more efficient thermogenesis and higher fitness in over-wintering shrews (Fontanillas *et al.* 2005). Pergams & Lacy (2007) found complete mitochondrial haplotype replacement in contemporary *P. leucopus* in Chicago compared to haplotypes sequenced from museum skins collected before urbanization. The agent of selection is not clear, but Munshi-South and Nagy (2014) also identified signatures of selection in mitochondrial D-loop haplotypes from contemporary *P. leucopus* in NYC. Many mitochondria-related metabolic functions are affected by the same environmental variables that change in response to urbanization, such as temperature (Balloux 2010), reduced migration (Lankau & Strauss 2011; Munshi-South 2012), or resource availability (Burcelin *et al.* 2002).

Urban *P. leucopus* may experience different energy budgets, physiological stressors or diets compared to rural counterparts. The signatures of selection reported for certain genes here support this scenario, such as heat-shock protein Hsp90. Heat shock proteins have repeatedly been found to play a pivotal role in adaptation to environmental stress (Limborg *et al.* 2012). Hsp90 was significantly enriched for 12 GO terms from the g:Profiler analysis with the majority associated with protein metabolism. In *Peromyscus* spp., Hsp90 is a chaperone for many proteins, including a suite of metabolizing receptors activated by dioxin-like industrial toxins often found in polluted soil samples (Settachan 2001). Another outlier from our analyses, aflatoxin aldehyde reductase (AKR7), was also significantly enriched for 8 GO terms primarily involved with single organism metabolism and is important for metabolizing environmental toxins (Hyndman *et al.* 2003).Urban soils are often much more contaminated with toxins than soils in adjacent rural areas (McDonnell *et al.* 1997).

We found a surprising number of candidate genes with functions related to the metabolism and transport of lipids and carbohydrates. These genes were strongly correlated with environmental measures of urbanization, with clearly divergent allele frequencies between urban and rural sites (Fig. 3B). APOB-100 is the primary apolipoprotein that binds and transports lipids, including both forms of cholesterol (HDL and LDL). The outlier gene, APOB-100, was significantly enriched for 9 GO terms with the primary cluster involved in single-organism metabolism, or anabolic / catabolic processes involving one organism and abiotic stimuli. FADS1, a farnesoid-x-receptor, is a nuclear receptor antagonist that is involved in bile synthesis and modulates high fat diets, with variation in expression affecting rates of obesity in mice (Li *et al.* 2013). FADS1 was enriched for 23 GO terms including five for lipid metabolism and regulation of lipid biosynthesis. Manually curated protein annotations show MYLK (10 significantly enriched GO terms; Metabolism) and SORBS2 (2 significantly enriched GO terms; Cellular processes) are both directly involved in the gastrointestinal system, including smooth muscle contractions and absorption of water and sodium in the intestine, respectively (Magrane & Consortium 2011; Consortium 2014). Finally, KEGG analysis identified two contigs (10636-348: 8 enriched GO terms and 27546-129: 22 enriched GO terms) that represent proteins that are both directly involved in galactose (primarily found in dairy products), fructose and mannose (both naturally found in fruits, seeds, and vegetables) metabolism (Ogata *et al.* 1999).

These candidate genes suggest that white-footed mice in isolated urban parks may be evolving in response to resource differences between urban and rural habitats. One prediction is that urban *P. leucopus* consume a diet with a substantially different fat content than diets of rural populations. The typical diet of *P. leucopus* across its range consists of arthropods, fruits, nuts, various green vegetation, and fungus (Wolff *et al.* 1985). Given that white-footed mice are opportunistic generalists, many different food resources could differ between urban and rural habitats. Urbanization in NYC has produced relatively small green patches that are surrounded by a dense urban matrix and largely free of white-tailed deer. The overabundance of deer outside of NYC removes the vegetative understory and inhibits regeneration of many plants (Stewart 2001), decreasing invertebrate species diversity and abundance (Stewart 2001; Allombert *et al.* 2005). In contrast, urban parks often have extremely thick and healthy understories composed of invasive plants (Leston & Rodewald 2006) thatproduce novel seed and fruit resources (McKinney 2008), as well as support a high abundance, if not diversity, of invertebrate prey (McDonnell *et al.* 1997). *P. leucopus* in NYC may successfully take advantage of these new food sources in urban habitats. We hypothesize that urban *P. leucopus* consume significantly different amounts or types of fats than their rural counterparts due to altered abundance of seeds, invertebrates, or direct human subsidies. Local adaptation in urban populations may allow these mice to more efficiently metabolize different types or amounts of lipids and carbohydrates.

ACKNOWLEDGMENTS

We thank Mike Hickerson for his helpful comments and advice on many analyses and for access to lab space for analyses and writing. We thank Diego Alvarado-Serrano, Alexander T. Xue, Tyler Joseph, and Champak Reddy for their invaluable comments and advice concerning bioinformatics and demographic analyses. This research was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R15GM099055 to JM-S and a NSF Graduate Research Fellowship to SEH. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

REFERENCES

Aken BL, Ayling S, Barrell D *et al.* (2016) The Ensembl gene annotation system. *Database*, **2016**, baw093.

Allombert S, Stockton S, Martin JL (2005) A natural experiment on the impact of overabundant deer on forest invertebrates. *Conservation Biology*, **19**, 1917–1929.

Balloux F (2010) The worm in the fruit of the mitochondrial DNA tree. *Heredity*, **104**, 419–420.

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, **12**, 767–780.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology*

& Evolution, **23**, 38–44.

Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular Ecology*, **17**, 3583–4.

Burcelin R, Crivelli V, Dacosta A, Roy-Tirelli A, Thorens B (2002) Heterogeneous metabolic adaptation of C57BL/6J mice to high-fat diet. *American Journal of Physiology. Endocrinology and Metabolism*, **282**, E834–E842.

Chace JF, Walsh JJ (2004) Urban effects on native avifauna: a review. *Landscape and Urban Planning*, **74**, 46–69.

Chapman M a, Hiscock SJ, Filatov D a (2013) Genomic Divergence during Speciation Driven by Adaptation to Altitude. *Molecular Biology and Evolution*, **30**, 2553–67.

Cheptou P-O, Carrue O, Rouifed S, Cantarel A (2008) Rapid evolution of seed dispersal in an urban environment in the weed Crepis sancta. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 3796–9.

Conesa A, Götz S, García-Gómez JM *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–6.

Consortium TU (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, **42**, D191-8.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–23.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–8.

Donihue CM, Lambert MR (2014) Adaptive evolution in urban ecosystems. *Ambio*, 1–10.

Ellison CE, Hall C, Kowbel D *et al.* (2011) Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences*, **108**, 2831–2836.

Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M (2013) Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics*, **9**, e1003905.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–98.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–93.

Fontanillas P, Dépraz A, Giorgi MS, Perrin N (2005) Nonshivering thermogenesis capacity associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew, Crocidura russula. *Molecular Ecology*, **14**, 661–670.

Franks SJ, Hoffmann A a. (2011) Genetics of Climate Change Adaptation. *Annual Review of Genetics*, **46**, 185–208.

Franks SJ, Weber JJ, Aitken SN (2014) Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary Applications*, **7**, 123–139.

Frichot E, François O (2015) LEA : An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–1699.

Gautier M (2015) Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, **201**, 1555–1579.

Götz S, García-Gómez JM, Terol J *et al.* (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, **36**, 3420–35.

Grossman SR, Shylakhter I, Karlsson EK *et al.* (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*, **327**, 883–6.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, e1000695.

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**, e74938.

Harris SE, O'Neill RJ, Munshi-South J (2015) Transcriptome resources for the white-footed mouse (Peromyscus leucopus): new genomic tools for investigating ecologically divergent urban and rural populations. *Molecular ecology resources*, **15**, 382–394.

Harris SE, Xue AT, Alvarado-Serrano D *et al.* (2016) Urbanization shapes the demographic history of a city-dwelling native rodent. *bioRxiv*, doi:10.1101/032979.

Henry P, Russello M a. (2013) Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. *Ecology and Evolution*, **3**, 3906–3917.

Hermisson J (2009) Who believes in whole-genome scans for selection? *Heredity*, **103**, 283–284.

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

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010a) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hohenlohe PA, Phillips PC, Cresko WA (2010b) Using Population Genomics To Detect Selection in Natural Populations: Key Concepts and Methodological Considerations. *International Journal of Plant Sciences*, **171**, 1059–1071.

Hohenlohe P a., Phillips PC, Cresko W a. (2011) Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *International Journal of Plant Science*, **171**, 1059–1071.

Hyndman D, Bauman DR, Heredia V V., Penning TM (2003) The aldo-keto reductase superfamily homepage. *Chemico-Biological Interactions*, **143**–**144**, 621–631.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–69.

Kanehisa M, Goto S, Sato Y *et al.* (2014) Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, **42**, 199–205.

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

De Kort H, Vandepitte K, Bruun HH *et al.* (2014) Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species Alnus glutinosa. *Molecular Ecology*, 4709–4721.

Lankau RA, Strauss SY (2011) Newly rare or newly common: evolutionary feedbacks through changes in population density and relative species abundance, and their management implications. *Evolutionary Applications*, **4**, 338–353.

Leston LF V, Rodewald AD (2006) Are urban forests ecological traps for understory birds? An examination using Northern cardinals. *Biological Conservation*, **131**, 566–574.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Li YF, Costello JC, Holloway AK, Hahn MW (2008) "Reverse ecology" and the power of population genomics. *Evolution*, **62**, 2984–2994.

Li F, Jiang C, Krausz KW *et al.* (2013) Microbiome remodelling leads to inhibition of intestinal farnesoid X receptor signalling and decreased obesity. *Nature communications*, **4**, 2384.

Li J, Li H, Jakobsson M *et al.* (2012) Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular Ecology*, **28**, 28–44.

Limborg MT, Helyar SJ, De Bruyn M *et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (Clupea harengus). *Molecular ecology*, **21**, 3686–703.

Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an adaptive allele in deer mice. *Science*, **325**, 1095–8.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F ST outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, **24**, 1031–1046.

Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database : the journal of biological databases and curation*.

McDonnell M, McDonnell M, Pickett S *et al.* (1997) Ecosystem processes along an urban to rural gradient. *Urban Ecosystems*, **1**, 21–36.

McKinney ML (2002) Urbanization, biodiversity, and conservation. *Bioscience*, **52**, 883–890.

McKinney ML (2008) Effects of urbanization on species richness: A review of plants and animals. *Urban Ecosystems*, **11**, 161–176.

McManus KF, Kelley JL, Song S *et al.* (2014) Inference of Gorilla Demographic and Selective History from Whole-Genome Sequence Data. *Molecular Biology and Evolution*, **32**, 600–612.

Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, 1–11.

Metzger LH (1971) Behavioral Population Regulation in the Woodmouse, Peromyscus leucopus. *American Midland Naturalist*, **86**, 434–448.

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

Mullen LM, Hoekstra HE (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution*, **62**, 1555–70.

Munshi-South J (2012) Urban landscape genetics: canopy cover predicts gene flow between white-footed mouse (Peromyscus leucopus) populations in New York City. *Molecular Ecology*, **21**, 1360–1378.

Munshi-South J, Kharchenko K (2010) Rapid, pervasive genetic differentiation of urban white-footed mouse (Peromyscus leucopus) populations in New York City. *Molecular Ecology*, **19**, 4242–4254.

Munshi-South J, Nagy C (2014) Urban park characteristics, genetic variation, and historical demography of white-footed mouse (Peromyscus leucopus) populations in New York City.

PeerJ, **2**, e310.

Munshi-south J, Richardson JL (2016) Peromyscus transcriptomics: Understanding adaptation and gene expression plasticity within and between species of deer mice. *Semin Cell Dev Biol*, doi:10.1016/j.semcdb.2016.08.011.

Munshi-South J, Zolnik CP, Harris SE (2016) Population genomics of the Anthropocene: urbani zation is negatively associated with genome-wide variation in white -footed mouse populations. *Evolutionary Applications*, doi:10.1111/eva.12357.

Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5268–5273.

Natarajan C, Inoguchi N, Weber RE *et al.* (2013) Epistasis Among Adaptive Mutations in Deer Mouse Hemoglobin. *Science*, **340**, 1324–1327.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Nielsen R, Williamson S, Kim Y *et al.* (2005) Genomic scans for selective sweeps using SNP data. *Genome research*, **15**, 1566–75.

Ogata H, Goto S, Sato K *et al.* (1999) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **27**, 29–34.

Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural selection. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 185–205.

Oliveira DCSG, Raychoudhury R, Lavrov D V., Werren JH (2008) Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp

Nasonia (Hymenoptera: Pteromalidae). *Molecular Biology and Evolution*, **25**, 2167–2180.

Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119–27.

Ostfeld R., Jones C, Wolff J (1996) Of mice and mast: ecological connections in eastern deciduous forests. *BioScience*, **46**, 323–330.

Partecke J, Schwabl I, Gwinner E (2006) Stress and the city: Urbanization and its effects on the stress physiology in European Blackbirds. *Ecology*, **87**, 1945–1952.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, e190.

Pavlidis P, Jensen JD, Stephan W (2010) Searching for Footprints of Positive Selection in Whole-genome SNP Data from Non-equilibrium Populations. *Genetics*.

Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N (2013) SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, **30**, 2224–34.

Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Pergams ORW, Lacy RC (2007) Rapid morphological and genetic change in Chicago-area Peromyscus. *Molecular Ecology*, **17**, 450–63.

Pool JE, Aquadro CF (2007) The genetic basis of adaptive pigmentation variation in Drosophila melanogaster. *Molecular Ecology*, **16**, 2844–2851.

Reimand J, Arak T, Adler P *et al.* (2016) g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research*, **44**, W83–W89.

Rockman M V (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, **66**, 1–17.

Rogic A, Tessier N, Legendre P, Lapointe F-J, Millien V (2013) Genetic structure of the white-footed mouse in the context of the emergence of Lyme disease in southern Québec. *Ecology and Evolution*, **3**, 2075–88.

Rytwinski T, Fahrig L (2007) Effect of road density on abundance of white-footed mice. *Landscape Ecology*, **22**, 1501–1512.

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–20.

Settachan D (2001) Mechanistic and molecular studies into the effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin and similar compounds in the deer mouse, Peromyscus maniculatus. Texas Tech University.

Shochat E, Warren PS, Faeth SH, McIntyre NE, Hope D (2006) From patterns to emerging processes in mechanistic urban ecology. *Trends in Ecology and Evolution*, **21**, 186–91.

Sih A, Ferrari MCO, Harris DJ (2011) Evolution and behavioural responses to human-induced rapid environmental change. *Evolutionary Applications*, **4**, 367–387.

Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation Genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705–712.

Stewart AJA (2001) The impact of deer on lowland woodland invertebrates: A review of the evidence and priorities for future research. *Forestry*, **74**, 259–270.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–70.

Storz JF, Runck AM, Moriyama H, Weber RE, Fago A (2010) Genetic differences in hemoglobin function between highland and lowland deer mice. *The Journal of*

Experimental Biology, **213**, 2565–74.

Storz JF, Runck AM, Sabatino SJ *et al.* (2009) Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 14450–5.

Storz J, Sabatino S, Hoffmann F (2007) The molecular basis of high-altitude adaptation in deer mice. *PLoS Genetics*, **3**.

Supek F, Bosnjak M, Skunca N, Smuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, **6**.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–95.

Thompson K, Renaudin M, Johnson M (2016) Urbanization drives parallel adaptive clines in plant populations 3. *bioarXiv*, **50773**.

Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, **29**, 673–680.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin S V (2010) Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature Genetics*, **42**, 260–3.

Vessey S, Vessey KB (2007) Linking behavior, life history and food supply with the population dynamics of white-footed mice (Peromyscus leucopus). *Integrative Zoology*, **2**, 123–130.

De Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: When and how much should we trust them? *Molecular*

Ecology, **23**, 2006–2019.

Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting natural selection in genomic data. *Annual review of genetics*, **47**, 97–120.

Wadi L, Meyer M, Weiser J, Stein L, Reimand J (2016) Impact of knowledge accumulation on pathway enrichment analysis. *bioarXiv*, **49288**.

Wang G, Wolff JO, Vessey SH *et al.* (2008) Comparative population dynamics of Peromyscus leucopus in North America: influences of climate, food, and density dependence. *Population Ecology*, **51**, 133–142.

Wang M, Yu Y, Haberer G *et al.* (2014) The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. *Nature Genetics*, 982–988.

Weber JN, Peterson BK, Hoekstra HE (2013) Discrete genetic modules are responsible for complex burrow evolution in Peromyscus mice. *Nature*, **493**, 402–405.

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

Whitehead A, Triant D, Champlin D, Nacci D (2010) Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population. *Molecular Ecology*, **19**, 5186–5203.

Wirgin I, Roy NK, Loftus M *et al.* (2011) Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson River. *Science (New York, N.Y.)*, **331**, 1322–5.

De Wit P, Palumbi SR (2013) Transcriptome-wide polymorphisms of red abalone (Haliotis rufescens) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, **22**, 2884–97.

De Wit P, Pespeni MH, Palumbi SR (2015) SNP genotyping and population genomics from expressed sequences -current advances and future possibilities. *Molecular Ecology*, **24**.

Wolff JO, Dueser RD, Berry K (1985) Food Habits of Sympatric Peromyscus leucopus and Peromyscus maniculatus. *Journal of Mammalogy*, **66**, 795–798.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, 323–354.

Yeh PJ (2004) Rapid evolution of a sexually selected trait following population establishment in a novel habitat. *Evolution*, **58**, 166–174.

Yoder JB, Stanton-Geddes J, Zhou P *et al.* (2014) Genomic signature of adaptation to climate in Medicago truncatula. *Genetics*, **196**, 1263–1275.

Zhou L, Bawa R, Holliday J a. (2014) Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (Populus trichocarpa). *Molecular Ecology*, **23**, 2486–2499.

FIGURES AND TABLES

Table 1. Summary population genomic statistics (mean ± standard error) for three urban and three rural populations of white-footed mice (*Peromyscus leucopus*) examined in this study.

Population	Nucleotide diversity (π)	Tajima's *D*
Urban		
CP	0.131 ±0.0012	0.318 ±0.005
FM	0.112 ±0.0012	0.301 ±0.006
NYBG	0.094 ±0.0011	0.280 ±0.006
Rural		
BHwwp	0.198 ±0.0012	0.350 ±0.004
CFP	0.211 ±0.0012	0.336 ±0.004
HIP	0.263 ±0.0011	0.349 ±0.004

Table 2. Outlier loci (*N* = 33) in the urban to rural comparison identified using Bayescan and confirmed with BayPass. Last three columns indicate whether the locus was also significantly associated with environmental variables across urban (CP, FM, NYBG) and rural (BHwwp, CFP, HIP) populations in the LFMM analysis. I = percent impervious surface, D = human density, C = Urban or Rural Classification

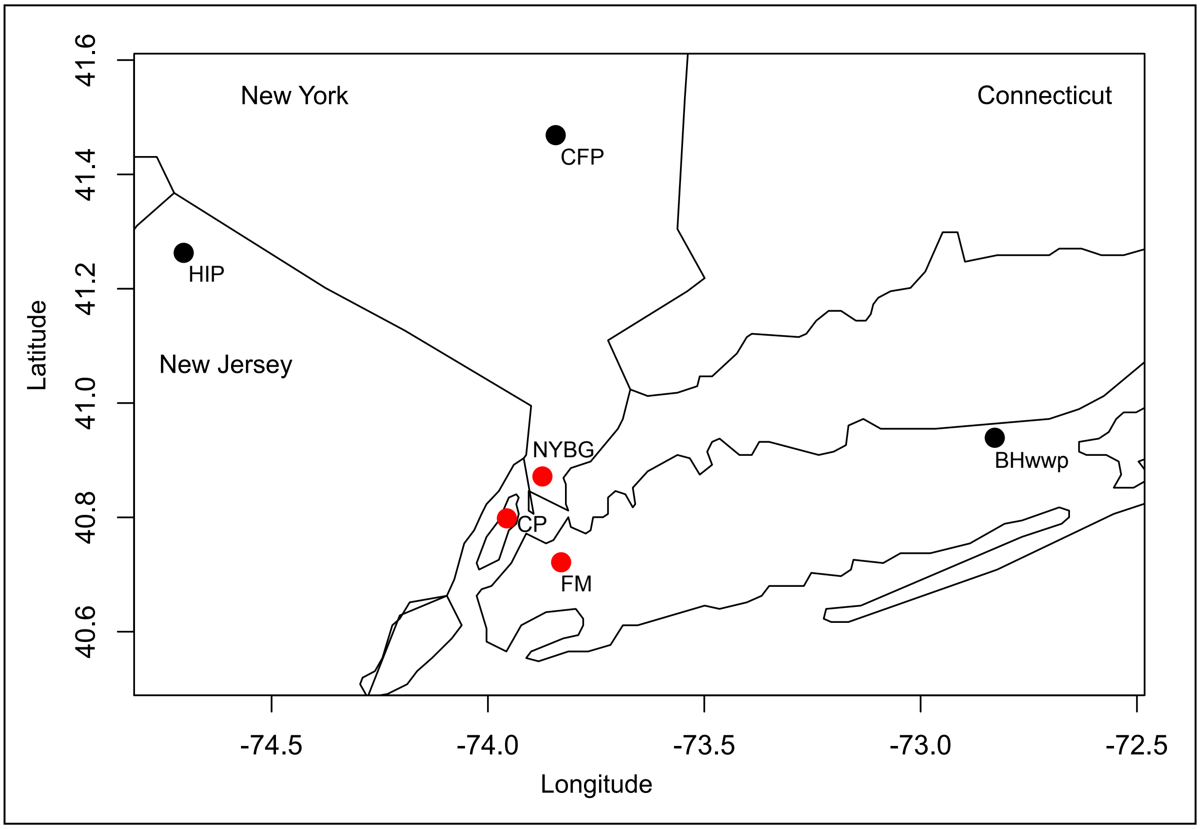
Urban to Rural		LFMM results		
Outliers	Gene	I	D	C
27691-127	retroviral nucleocapsid protein gag containing protein	-	-	+
25795-154	af478441_1farnesoid-x-receptor alpha splice variant 1	-	-	+
37015-34	tubulin folding cofactor e-like isoform x6	-	-	-
902-1236	alkyldihydroxyacetonephosphate peroxisomal	-	-	+
3135-709	transmembrane 9 superfamily member 1 isoform 2	-	-	-
27707-127	autophagy-related protein 2 homolog a isoform x2	-	+	+
38397-23	--	-	-	-
3567-665	gram domain-containing protein 3	-	+	+
2482-790	protein diaphanous homolog 1 isoform x1	-	-	+
37967-26	sorbin and sh3 domain-containing protein 2 isoform x3	-	-	+
17974-242	40s ribosomal protein s15a-like protein	+	+	+
36437-38	jnk sapk-inhibitory isoform cra_a	-	-	+
7975-418	myosin light chain smooth muscle	-	-	+
12107-321	--	+	-	+
5754-511	otu domain-containing protein 3	-	-	-
27887-125	26s proteasome non-atpase regulatory subunit 9	-	-	+
1749-927	utrophin isoform x2	-	-	-
29218-108	n-alpha-acetyltransferase 50 isoform x1	-	-	-
31201-85	transmembrane protein 115	-	-	-
22365-204	transmembrane protein 19 isoform x1	+	+	+
7690-428	casp8-associated protein 2	-	-	+
2260-821	a kinase anchor protein isoform cra_a	-	-	+
1371-1036	signal recognition particle 9 kda protein	-	-	+
19-4220	cytoplasmic dynein 1 heavy chain 1	+	+	+
20787-217	adp-ribosylation factor-like protein 1	-	-	+
36491-37	5-oxoprolinase isoform x1	-	-	+
23896-185	low molecular weight phosphotyrosine protein phosphatase-like	+	-	+
1396-1029	proteasome activator complex subunit 1	-	-	+
11279-335	mitochondrial ribosomal protein l37	-	-	-

26257-147	PREDICTED: uncharacterized protein C1orf167 homolog	-	-	+
31894-78	--	-	-	+
14102-290	succinate dehydrogenase	-	-	+
40819-1	adaptin ear-binding coat-associated protein 1	+	-	+

Table 3. Outlier loci ($N = 26$) identified using SweeD and confirmed with BayPass. Last three columns indicate whether the locus was also significantly associated with environmental variables across urban (CP, FM, NYBG) and rural (BHwwp, CFP, HIP) populations in the LFMM analysis. Columns to the left of the outliers show the population in which the SweeD outlier was identified. I = percent impervious surface, D = human density, C = Urban or Rural Classification

Population				SweeD		LFMM results		
CP	FM	NYBG	Combined	Outliers		I	D	C
-	-	-	+	10099-359	--	-	-	-
+	-	-	-	10636-348	aflatoxin b1 aldehyde reductase member 2	+	-	+
-	-	-	+	113-2629	--	-	-	-
+	-	-	+	124-2491	--	-	-	-
+	-	-	-	12718-311	--	-	-	-
-	-	+	-	1583-971	isoform cra_a	-	-	+
-	-	-	+	17779-244	--	-	-	-
-	+	+	-	17856-243	serine protease inhibitor a3c-like	+	+	+
-	-	-	+	23358-193	--	-	-	+
-	+	-	-	243-1951	solute carrier family member 13	-	-	+
+	+	+	+	25500-158	--	-	-	-
-	-	+	-	2736-755	--	-	-	-
-	-	+	-	27546-129	6- liver type	-	-	+
-	-	-	+	28127-122	sarcosine mitochondrial	-	-	+
-	-	-	+	28528-117	--	-	-	-
-	-	-	+	29117-109	--	-	-	-
-	+	-	-	31034-87	--	-	-	-
+	-	-	-	342-1776	fatty acid desaturase 1	+	-	+
+	+	+	+	35790-44	apolipoprotein b- partial	-	-	-
-	-	-	+	37202-32	PREDICTED: poly	+	-	+
-	-	-	+	37400-30	--	-	-	-
-	-	-	+	39-3749	alpha-aminoadipic semialdehyde mitochondrial	+	+	+
-	-	-	+	3964-627	heat shock protein alpha class a member 1	-	-	+
-	-	-	+	408-1655	disintegrin and metalloproteinase domain-containing protein 9 isoform x1	-	-	+
-	-	-	+	50-3466	--	-	-	-
-	-	-	+	533-1512	fructose- -bisphosphatase 1	-	-	+

967



968

969 **Figure 1.** Map of sample localities in the NYC metropolitan area. Sites in red are urban parks
970 within New York City. CP = Central Park; FM = Flushing Meadows—Willow Lake; NYBG =
971 New York Botanical Gardens; BHwwp = Brookhaven and Wildwood State Park; CFP =
972 Clarence Fahnestock State Park; HIP = High Point State Park

973

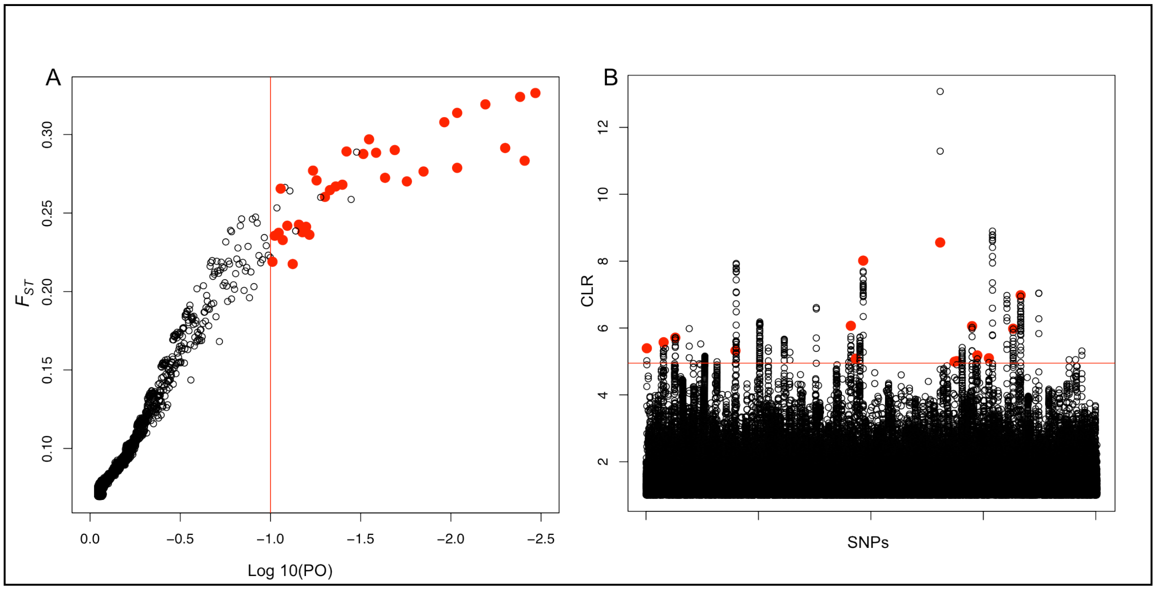


Figure 2. (a) BayeScan 2.1 plot of 154,770 SNPs genome scan analysis between urban and rural populations, including 48 individual white-footed mice from six NYC sampling sites. F_{ST} is on the vertical axis plotted against the $\log_{10}$ of the posterior odds (PO). The vertical red line indicates the cutoff (FDR = 0.1) used for identifying outlier SNPs. The markers on the right side of the vertical line show all outlier SNP candidates and the red circles represent the final accepted outlier SNPs from Table 2. (b) SweeD results with each of the 154,770 SNPs plotted from all 48 individuals. The Composite Likelihood Ratio (CLR) is plotted along the vertical access and each unfilled point represents an individual SNP. The horizontal red line indicates the cutoff used for identifying outlier SNPs at $P \leq 0.0001$. The red circles represent the final accepted outlier SNPs from Table 3.

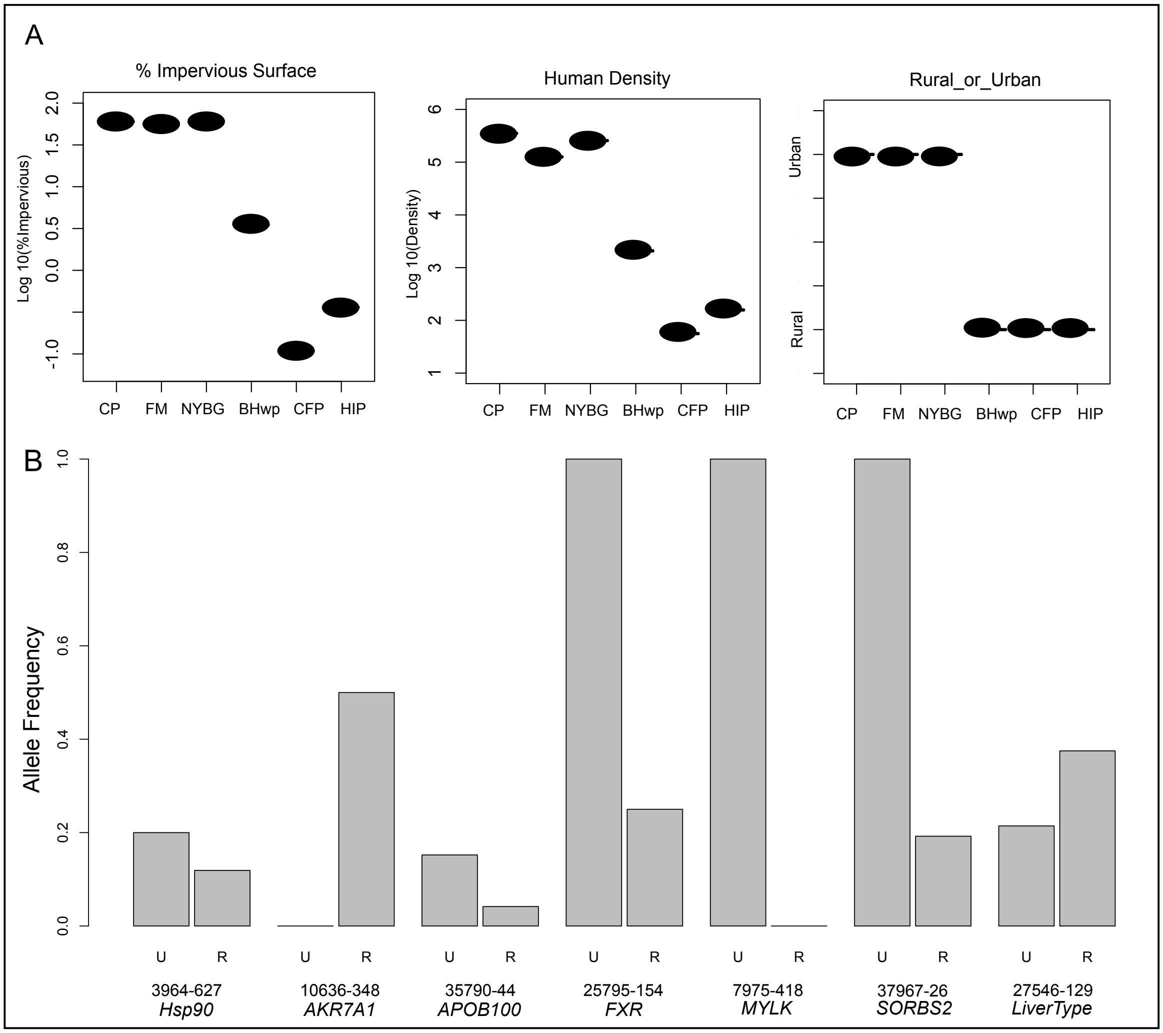


Figure 3. (a) Plot of urbanization metrics for all 6 sampling sites from NYC used in this study. The log10 value of % Impervious Surface and Human Density are plotted along the vertical axis and the oval represents the value for each sampling site. Ovals on the Rural or Urban plot show sample sites designated as either Urban or Rural. (b) Allele frequencies for selected candidate genes found to contain outlier SNPs from both genome scans and GEA tests grouped by urban (U) or rural (R) classification. The frequency of the outlier SNP within each type of population is plotted on the vertical axis.

SUPPORTING INFORMATION

Figure S1. KEGG analysis for biochemical pathways that contain multiple outlier contigs. Colored boxes represent outlier genes.

Table S1. Blast2GO table with BLASTX hits from *M. musculus*, *R. rattus*, and *C griseus* and top three supported Gene Ontology terms for outlier genes from Bayescan and SweeD

Table S2. Excel file containing Bayescan and SweeD outliers and the corresponding BayPass results. Full BayPass results are also included.

Table S3. Excel file containing filtered list of outlier contigs, the homologous *Mus musculus* genes, and the significantly enriched GO terms from g:Profiler.

Table S4. Excel file containing Revigo results. Enriched GO terms from g:Profiler are sorted into largest parent terms and listed based on the frequency of occurrence.

Table S5. Average pairwise F_{ST} among six *P. leucopus* populations.

Table S6. Top Blast hits including NCBI accession numbers for outlier contigs listed in Table 2 and Table 3.