Scans for positive selection reveal candidate genes and local adaptation of *Peromyscus leucopus* populations to urbanization

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 its rate is only expected to increase globally as more humans move into urban centers. Urbanized landscapes are often highly fragmented. Isolated populations within these fragments may adapt in response to novel urban ecosystems, but few studies have found strong evidence of evolutionary responses in urban environments. We used multiple genome scan and genotype-environment association (GEA) approaches to examine signatures of selection in transcriptomes from urban white-footed mice (*Peromyscus leucopus*) in New York City. We scanned transcriptomes from 48 *P. leucopus* individuals from six environmentally heterogeneous locations (three urban and three rural) for evidence of rapid local adaption in isolated urban habitats. We analyzed 154,770 SNPs and identified patterns of genetic differentiation between urban and rural sites and signatures of selection in a large subset of genes. Neutral demographic processes can create allele frequency patterns that are indistinguishable from positive selection. We accounted for this by simulating a neutral SNP dataset under the inferred demographic history for the sampled *P. leucopus* populations to serve as a null model when choosing outliers. We annotated the resulting outlier genes and further validated them by associating allele frequency differences with environmental measures of urbanization, percent impervious surface and human population density. The majority of 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 likely resulted from rapid local adaptation in urbanizing habitats. The specific candidate loci that we identified suggest that populations of *P. leucopus* are using novel food resources in urban habitats or locally adapting

47 through changes in their metabolism. Our data support the idea that cities represent novel
48 ecosystems with a unique set of selective pressures.
49
50 *Keywords*: transcriptome, *Peromyscus leucopus,* genotype-environment association, genome
51 scans, positive selection, demographic null-model
52
53
54
55

INTRODUCTION

Traits are adaptive when they increase an organism's fitness in a specific environment (Barrett & Hoekstra 2011). The identification of specific 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). In the current era of high-throughput DNA sequencing, where costs continue to drop by orders of magnitude (De Wit *et al.* 2015), it is now feasible to generate genomic datasets for natural populations of non-model organisms. Researchers can use a reverse-ecology approach 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 NYC. We scanned *P. leucopus* transcriptomes and identified regions and genes with divergent and skewed allele frequencies indicative of positive selection. We incorporated a neutral SNP dataset from an inferred demographic history directly into our null model. We then examined the statistical association between allele frequencies and environmental measures of urbanization.

Traditional approaches for identifying local adaptation involve reciprocal transplant or common garden experiments (Merila & Hendry 2014), but local adaptation also leaves a predictable pattern of genetic variation and differentiation along environmental gradients across the genome (Savolainen *et al.* 2013). Measuring changes in the site frequency spectrum (SFS), the distribution of allele frequencies across sites, from genomic data can be an efficient method of detecting past selection (Merila & Hendry 2014). Positive directional

79 selection increases interspecific variation at selected loci compared to the genomic background
80 (Beaumont 2005), decreases nucleotide diversity around the selected locus through genetic
81 hitchhiking (Hermisson 2009), and skews the SFS towards excess low and high frequency
82 variants (Nielsen 2005). Balancing selection leaves a generally opposite pattern with decreased
83 intraspecific genetic diversity (Nielsen 2005), low genetic differentiation between sites (Foll &
84 Gaggiotti 2008), and an excess of intermediate frequency alleles (Nielsen 2005). Negative, or
85 purifying selection reduces genetic diversity and differentiation, and only low frequency variants
86 increase in the SFS (Nielsen 2005).

87 Local adaptation has increasingly been shown to occur across multiple taxa
88 (Stinchcombe & Hoekstra 2008; Bonin 2008; Linnen *et al.* 2009; Hohenlohe *et al.* 2010a; Turner
89 *et al.* 2010; Ellison *et al.* 2011; De Wit & Palumbi 2013). Uncovering the genetic basis of local
90 adaptation has provided insight into a variety of evolutionary processes including speciation,
91 maintenance of genetic diversity, range expansion, and species response to changing
92 environments (Savolainen *et al.* 2013; Tiffin & Ross-Ibarra 2014). Cities represent one of the
93 fastest growing and most rapidly changing environments around the world. Urbanization leads
94 to habitat loss and fragmentation, changes in resource availability, novel species interactions,
95 altered community composition, and increased exposure to pollutants (McKinney 2002; Chace &
96 Walsh 2004; Shochat *et al.* 2006; Sih *et al.* 2011). Each of these ecological consequences may
97 exert strong selective pressure, and there is mounting evidence that rapid adaptation occurs in
98 many urban organisms. Another cause of rapidly changing environments is global climate
99 change, where increasing temperatures and altered precipitation patterns strongly influence the
100 life history traits of many species (Franks & Hoffmann 2011). These two processes,
101 urbanization and climate change, are not mutually exclusive, however. Understanding local

adaptation in urban habitats may lead to general insights about local adaptation to future climate change threats, both of which represent cases of general rapid evolution in changing environments. What traits are most likely involved in local adaptation? How quickly do populations respond to selective pressures and adapt locally? What environmental variables have the largest impact on populations and drive local adaptation? Are the same genes and alleles involved in local adaptation also involved in similarly changing environments, i.e. is there evidence of convergent local adaptation?

White-footed mice are good candidates for local adaptation because they are widespread and are one of the few native mammals that thrive in extremely small, fragmented urban forests (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 along 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 found evidence of divergence and selection in urban populations of NYC white-footed mice (Harris *et al.* 2013), though we used much smaller datasets and less sophisticated approaches than presented here. Collectively, strong selective pressures from urbanization, lack of gene flow between NYC parks, genetic structure found between geographically close urban sites, and evidence of urbanization driving neutral allele frequency patterns in urban populations (Munshi-

South *et al.* 2016) makes it likely that populations of urban white-footed mice are adapting to strong selective pressures in spite of the influence of genetic drift.

Urbanization and global climate change are relatively recent disturbances that rapidly change native ecosystems. Over short timescales, standing genetic variation, as opposed to novel mutations in organisms, often underlies adaptation (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). In order to distinguish 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 used 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 made inferences about the evolutionary processes at work in these populations. Several studies have used this suite of population genetic statistics to detect candidate genes that are the target of selection (Stajich & Hahn 2005; Hohenlohe *et al.* 2010a; Tennessen *et al.* 2010; Nadeau *et al.* 2012). Major challenges in solely using π or Tajima's *D* are distinguishing between types of selection, and then disentangling demographic processes from selection (Biswas & Akey 2006). The difficulty arises because neutral demographic processes, like population bottlenecks, produce signatures of variation in the genome similar to those produced by selection (Oleksyk *et al.* 2010; Li *et al.* 2012). For example, a population bottleneck followed by an expansion will create genomic regions with low genetic diversity that

147 resembles signatures from selection. Alleles present in the few breeding individuals during the
148 bottleneck will become widespread during the expansion (Pavlidis *et al.* 2010). There has been
149 much discussion on how to deal with the confounding effects of demographic history on
150 identifying selection (Excoffier *et al.* 2009; Li *et al.* 2012; Vitti *et al.* 2013; Lotterhos &
151 Whitlock 2015). The prevailing approach is to produce genome-wide data and assume selection
152 acts on one or a few loci while demographic processes act across the genome. Outlier tests for
153 loci under selection generate a null distribution, usually based on an island model of population
154 differentiation (Excoffier *et al.* 2009), and then identify candidate genes with genetic
155 differentiation beyond the null model's limits. The true demographic history of most organisms
156 is much more complex, and computational approaches have been developed to robustly infer
157 demographic parameters (Gutenkunst *et al.* 2009; Excoffier *et al.* 2013). The inferred
158 demographic history can then be used to construct a more realistic null model, reducing the rate
159 of false positives in outlier based tests of selection (Excoffier *et al.* 2009; Yoder *et al.* 2014).
160 We used the inferred demographic history of urban populations of *P. leucopus* (Harris *et*
161 *al.* 2016) to simulate comparable SNP datasets to our observed sequence data. We then used two
162 genome scan tests that identify outlier loci based on population differentiation and the SFS,
163 respectively. Bayescan uses a Bayesian approach to identify SNPs that show extreme allele
164 frequency divergence between populations (Foll & Gaggiotti 2008). SweeD is a likelihood
165 based test that finds evidence of selective sweeps by looking for regions with a SFS that deviates
166 from neutral expectations (Pavlidis *et al.* 2013). We also used an emerging approach for
167 identifying loci underlying local adaptation by examining associations between allele frequencies
168 and environmental variables. Several tests have been developed based on the relationship
169 between genotypes and environmental variables, falling under the general category of genotype-

environment association (GEA) tests (Joost *et al.* 2007; Coop *et al.* 2010; Frichot *et al.* 2013; Lotterhos & Whitlock 2015). GEA tests perform better than genome scan based 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 loci identification (De Villemereuil *et al.* 2014). GEA tests also identify 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.

In this study, we examined transcriptomes generated from RNAseq for 48 *Peromyscus 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 find loci with patterns of genetic variation that deviate from neutral expectations. Next, we used several tests of selection that use our transcriptome-wide SNP datasets to determine whether these deviations are due to recent selection in urban populations of white-footed mice. 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 the candidate genes

193 experiencing selection from ecological pressures in urban habitats. We used overlapping results
194 from multiple tests and environmental associations in order to generate a reliable list of candidate
195 genes involved in the local adaptation of *P. leucopus* populations to the urban environment. This
196 study is the first to use transcriptome-wide patterns of genetic variation for analyses of local
197 adaptation in cities. Evidence of local adaptation in urban populations reveals how urbanization
198 acts as an evolutionary force, gives insights into important traits for local adaptation, and
199 provides an example of the speed of evolution in rapidly changing environments.
200
201 **MATERIALS AND METHODS**
202 **Sampling, library preparation, and transcriptome assembly**
203 We sampled white-footed mice from 2009 - 2013. We randomly chose eight individual
204 white-footed mice (equal numbers of males and females) from six sampling locations
205 representative of urban and rural habitats (Fig. 1) (Harris *et al*. 2013, 2015). Three sampling sites
206 occurred within NYC parks: Central Park in Manhattan (CP), New York Botanical Gardens in
207 the Bronx (NYBG), and Flushing Meadow—Willow Lake in Queens (FM). These sites
208 represented urban habitats surrounded by high-volume roads and dense human infrastructure.
209 The remaining three sites occurred ~100 km outside of NYC in rural, undisturbed habitat
210 representative of natural environments for *Peromyscus leucopus*. High Point State Park is in the
211 Kittatinny Mountains in New Jersey (HIP), Clarence Fahnestock State Park is located in the
212 Hudson Highlands in New York (CFP), and Brookhaven and Wilde Wood State Parks and
213 neighboring sites occur on the northeastern end of Long Island, New York (BHwwp). We
214 sacrificed mice on site and liver, gonad, and brain tissue were harvested in the field for
215 immediate storage in RNAlater (Ambion). In the lab, we extracted total RNA and then removed

ribosomal RNA 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 the Genome Analysis Toolkit pipeline using a Bayesian genotype likelihood model (GATK version 2.8, DePristo *et al.* 2011). See Harris *et al.* 2013 and Harris *et al.* 2015 for full transcriptome sequencing, assembly and SNP calling details.

Summary statistics

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

Sans for positive selection based on population differentiation

Population structure analyses for protein coding sequences show that the three urban sites and three rural sites comprise two distinct groups, but there was also hierarchical structure within each indicating urban sites represent unique evolutionary clusters (Harris *et al.* 2015). 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 markers that show divergence patterns between groups that are 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 is under the effect of selection. To generate more realistic allele frequency distributions, I used Bayescan to analyze coalescent simulations of SNP datasets based on the neutral demographic history inferred specifically for *P. leucopus* populations in Harris *et al.* 2016. We generated 100 sets of 100,000 SNPs each 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 an urban population 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. Within 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 calculated F_{ST} and posterior odds probability were higher than for any value calculated from the simulated dataset.

Analysis for selective sweeps

We also scanned the transcriptome to look for contigs where the observed SFS showed an excess of low frequency and high frequency minor alleles, a signal indicative of a recent selective sweep in the region. 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). I calculated the CLR along sliding windows across the transcriptome using the software program SweeD (Pavlidis *et al.* 2013). SweeD is an extension of the popular Sweepfinder (Nielsen *et al.* 2005) and is optimized for large next generation sequencing (NGS) datasets. SweeD was run separately for each population and on individual contigs directly from vcf files 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 chosen 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, but rather, the CLR statistic is computed using a selective sweep model on the observed dataset and needs to be 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 regions and chose the associated contigs as candidates if their CLR statistic 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. Choosing outliers within the top 0.01% of the distribution is a conservative cutoff value. When looking for regions with genetic patterns of a

selective sweep, Wilches (2014) filtered regions within the top 5% of the distribution. Selective sweeps from artificial selection in rice, *Oryza glaberrima,* were identified with a cutoff value of 0.5% (Chen *et al.* 2014) and regions within the Gorilla genome were identified as significant if CLR scores were in the top 0.5% (McManus *et al.* 2014). We chose an even more stringent filter of 0.01% because we lacked a reference genome and analyses were restricted to relatively short individual contigs.

Genotype-environment association tests for environmental selection

We used LFMM (Frichot *et al.* 2013), a software program that is one of the recently emerging genotype-environment association (GEA) approaches for identifying selection (Hedrick *et al.* 1976; Joost *et al.* 2007; Coop *et al.* 2010; Frichot *et al.* 2013; Lotterhos & Whitlock 2015), to associate outlier SNPs and candidate loci identified above with potential environmental selection pressures. Latent Fixed Mixed Modeling (LFMM) tests for correlations 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, the percent impervious surface within a two-kilometer buffer around each sampling site, human density within a two-kilometer buffer around each sampling site, and simply designating each site urban or rural. We tested all individuals and only the outlier SNPs detected in Bayescan and SweeD. An important first step in using the LFMM algorithm is to define the number of latent factors, K, that can be used to define population structure in the genetic background. To identify the appropriate number of K latent factors in our dataset, we used default parameters and performed a PCA followed by a recommended Tracy-Widom test to find the number of eigenvalues with significant p values $\leq$

0.01 (Patterson *et al.* 2006; Frichot & François 2015). Results suggested the use of six latent factors. Thus, I ran LFMM with default parameters except for a K = 6, an increased number of MCMC cycles = 100,000, and a 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 $q \leq 0.1$.

Functional annotation of candidate gene

The contigs containing outlier SNPs identified using the tests for selection above were obtained from the *P. leucopus* transcriptome. The gene annotation pipeline implemented in Blast2GO (Conesa *et al.* 2005; Götz *et al.* 2008) was used to find homologous sequences from the NCBI non-redundant protein database using BLASTX, and associated gene ontology (GO) terms were retrieved. Gene ontology (GO) terms are a standardized method of ascribing functions to genes. Blast2GO retrieves GO terms associated with BLASTX hits and also uses the KEGG database to describe biochemical pathways linking different enzymes (Ogata *et al.* 1999; Kanehisa *et al.* 2014).

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}. There were differences in genetic diversity between urban and rural populations greater than one standard deviation. 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, with rural populations only slightly more positive than urban populations, though not significantly different. 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 2). These F_{ST} calculations were very similar to calculations made for neutral genome-wide SNP datasets from the same *P. leucopus* populations (Munshi-South *et al.* 2016), and supported findings that these populations lack an isolation-by-distance pattern. Comparisons between rural populations had the lowest F_{ST} values, urban to rural populations had the second lowest, and urban to urban population comparisons had the highest overall F_{ST} values despite being less than 5 km apart (Table 2).

Outlier detection

The test for positive or balancing selection implemented in Bayescan for the global analysis revealed 309 (0.19%) SNPs potentially under the influence of divergent selection. To investigate divergent selection due to urbanization, sampling sites were grouped and classified as urban or rural, and genome scans using Bayescan on this dataset uncovered 40 (0.025%) SNPs with signatures of positive selection (Fig. 2A, Table 3). Eight of these SNPs were found in the global analysis. Individual urban to rural population comparisons did not find any outlier SNPs, and zero SNPs were revealed to be under balancing selection. F_{ST} for outlier SNPs ranged from 0.21 - 0.33, much higher than the population average. When Bayescan was run on the simulated

354 neutral dataset, which included bottlenecks during urban population divergence, there were zero
355 identified outlier SNPs. I did, however, only include outlier SNPs from the observed dataset
356 with FDR and posterior odds values that were smaller and larger, respectively, than the most
357 extreme values for the simulated data (FDR $\leq$ 0.6 and $\log_{10}$(PO) $\geq$-0.196).

358 Outlier regions showing signatures of selective sweeps from the SweeD analysis were
359 identified using comparisons to neutral expectations. To generate the null distribution of the CLR
360 statistic I tested the 100 SNP datasets simulated under the inferred demographic history for NYC
361 populations of *P. leucopus.* I found that CLR scores in the top 5% of the distribution were
362 generally 2x - 3x lower than for the top 5% of the observed dataset. I ran SweeD runs on
363 observed SNPs within individual contigs and identified outliers by filtering for a CLR score $\geq$
364 3.53 (the maximum CLR from simulated data). I also chose regions that fell within the top
365 0.01% of the observed distribution (Fig. 2B). SweeD identified regions with SFS patterns that fit
366 a selective sweep model in 55 contigs (40,908 contigs in *P. leucopus* transcriptome, 0.13%)
367 within urban populations (Table 4). Contig 35790-44, which codes for the lipid transporter
368 *Apolipoprotein B100*, had the highest CLR score, CLR = 8.56, and all outliers had CLR scores $\geq$
369 4.97. There was no overlap of outliers between Bayescan and SweeD.
370
371 **Environmental associations**
372 We used LFMM to examine statistical associations of outlier SNPs with environmental
373 measures of urbanization. Thirty of 40 outliers identified from Bayescan could be associated
374 with at least one of the three environmental variables tested, which clearly delineate urban and
375 rural sampling locations (Fig. 3A, Table 3). All 30 of the identified SNPs were associated with
376 whether a site was classified as urban or rural. Only seven of the outlier SNPs were associated

377 with percent impervious surface surrounding the sampling site and five were associated with
378 human density. Twenty-six of the 55 outlier contigs in urban populations containing selective
379 sweep regions as identified in SweeD could be associated with one of the environmental
380 variables (Table 4). Again, all 26 significant associations involved classification of a site as
381 either urban or rural. Fourteen outliers from SweeD were associated with percent impervious
382 surface and eight were associated with human density surrounding the sampling location. Some
383 contigs containing outlier SNPs associated with environmental variables were unique to
384 individual urban populations, possibly indicating local adaptation within parks or selection on a
385 polygenic trait.
386
387 **Functional annotation**
388 The full contig sequences containing the outlier SNPs were obtained from the *P. leucopus*
389 transcriptome (Harris *et al.* 2015) and used to identify functional annotations. Of the 40 contigs
390 identified by Bayescan as divergent between urban and rural populations, 36 could be annotated
391 with gene names and functional information (Table 3). Of these, 29 were also associated with
392 urban environmental variables. For the Bayescan outlier sequences, the ten most frequent gene
393 ontology terms attributed to the DNA sequences involved organismal metabolism (Table S1).
394 Some outliers occurred within well-studied genes with known functions and biochemical
395 pathways. These included a farnesoid-x-receptor (FXR, Contig 25795-154) gene, the protein
396 ABCC8 (Contig 26183-148), a Hermansky-Pudlak syndrome gene (Hps1, Contig 36706-36),
397 KDM8, a histone demethylase (Contig 7750-426), a myosin light chain kinase (MYLK, Contig
398 7975-4180), and the gene SORBS2 (Contig 37967-26). These genes were identified as likely

399 experiencing divergent selection between urban and rural populations and showed environmental
400 associations with urbanization.

401 When we used results from SweeD, we found regions within 55 contigs that showed a
402 signature of a selective sweep (Table 4). Forty-nine could be annotated with gene names and
403 gene ontology terms, and 25 were also associated with urbanization. Overall, sequences were
404 associated with metabolic processes, similar to the outliers found in Bayescan, and many genes
405 were involved with basic metabolic functions such as glycolysis and ATP production (Table S1).
406 A few contigs were annotated with well-studied genes and clearly understood functions. Contig
407 35790-44 was annotated as the gene APOB, an apolipoprotein, and Contig 10636-348 was an
408 aflatoxin reductase gene AKR7A1. There was also the gene FADS1, part of the fatty acid
409 denaturase family (Contig 342-1776), a heat-shock protein (Hsp90, Contig 3964-627), and a
410 hepatocyte growth factor activator gene (Contig 8960-388). Most gene annotations did not have
411 known phenotypic traits related to their function, but KEGG analysis revealed several contigs
412 involved in the same biochemical pathways: galactose metabolism, fructose metabolism, and
413 mannose metabolism (Fig. S1).
414

415 **DISCUSSION**

416 The results of this study provide insight into the genetic basis of local adaptation, which
417 is key for understanding the ecological and evolutionary processes that affect biodiversity and
418 how organisms respond to changing environments. We hypothesized that populations of *P.*
419 *leucopus* in urban habitat fragments within NYC adapt in response to selective pressures from
420 urbanization. Previous work supports this claim. Clear evidence of population structure
421 between urban and rural sampling sites from neutral non-coding (Harris *et al.* 2016) and protein

422 coding datasets (Harris *et al.* 2015) suggests NYC populations of white-footed mice are
423 genetically isolated. Urbanization also impacts genetic diversity across the genome (Munshi-
424 South *et al.* 2012, Harris *et al.* 2015, Harris *et al.* 2016). *P. leucopus* populations along an
425 urban-to-rural gradient in NYC had reduced nucleotide diversity and heterozygosity in urban
426 populations (Munshi-South *et al.* 2016). Additionally, demographic inference indicates that NYC
427 populations became isolated within the timeframe of urban settlement (Harris *et al.* 2016).
428 We previously found evidence for older occurrences of divergent selection in NYC
429 white-footed mice by investigating non-synonymous polymorphisms between pooled
430 transcriptome samples (Harris *et al.* 2013). There was little overlap between previous results and
431 those found here, but that was not surprising, as this data-set was much larger, covered more
432 sampling sites, and looked at recent signatures of selection. Two of the eleven previously
433 identified candidate genes (Harris *et al.* 2013) were direct matches to outliers in this current
434 analysis (Serine protease inhibitor a3c and Solute carrier organic anion transporter 1A5), and
435 three other genes were from the same gene families or involved in the same biological processes
436 as those described here. One gene was an aldo-keto reductase protein, part of the same gene
437 family as our SweeD identified aflatoxin reductase gene (Contig 10636-348). The aldo-keto
438 reductase gene family comprises a large group essential for metabolizing various natural and
439 foreign substances (Hyndman *et al.* 2003). Two others, camello-like 1 and a cytochrome P450
440 (CYPA1A) gene, are involved in metabolism of drugs and lipids. In *Peromyscus* spp., CYPA1A
441 is directly expressed along with Hsp90 (outlier from current SweeD analysis) when exposed to
442 environmental toxins (Settachan 2001). Collectively, these findings suggest that urban
443 populations of *P. leucopus* may be adapting in response to selective pressures from urbanization.

444 In this study, we observed patterns of divergent positive selection between urban and
445 rural populations of *P. leucopus*, and were able to associate outlier SNPs, while annotating the
446 parent contig, with environmental variables representative of urbanization. The majority of
447 candidate genes deal with organismal metabolism, particularly diet-related breakdown of lipids
448 and carbohydrates. We discuss what these finding mean for organisms as they are exposed to
449 novel urban ecosystems, and for understanding the ecological processes and time frame of recent
450 local adaptation in general.

451

452 **The utility of using genome scan methods to test for selection**

453 Over the past decade, genome scan methods have become a feasible and common way
454 for investigating polymorphisms across the genome in order to detect and disentangle neutral
455 (demographic) and adaptive (selection) evolutionary processes (De Villemereuil *et al.* 2014).
456 One of the most popular approaches looks at locus specific allele frequency differentiation
457 between sampling locations as measured by F_{ST} (Lewontin & Krakauer 1973; Weir &
458 Cockerham 1984). Sites with extremely high allele frequency differences may be subjects of
459 positive directional selection. Bayescan (Foll & Gaggiotti 2008) builds on this idea and
460 identifies outliers using a Bayesian approach. Bayescan calculates the posterior probability of a
461 site being under the influence of selection by testing two models, one that includes selection and
462 one that does not. The model that does not invoke selection is based on a theorized neutral
463 distribution of allele frequencies.

464 While Bayescan has been shown to be the most robust differentiation method with
465 respect to confounding demographic processes (Pérez-Figueroa *et al.* 2010; De Villemereuil *et*
466 *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). One way to avoid false positives is to build population structure and a specific demographic history directly into the null distribution of F_{ST}. We dealt with the issue of type I errors by running Bayescan on simulated SNP datasets generated under the neutral inferred demographic history for urban populations of *P. leucopus* in NYC (Harris *et al.* 2016). We only included outliers if their posterior probability was greater than any found from simulations. The outliers captured when comparing urban to rural sites made up 0.025% of the total number of loci analyzed from the transcriptome. This number 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, 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 because we included the known demographic history in our tests, because of the relatively recent isolation of urban populations of *P. leucopus*, or due to the fact that we did not have complete transcriptome sequences for our populations.

SweeD, another genome scan approach, looks at patterns in the SFS within a population as opposed to allele differentiation between populations. The statistics developed around the SFS are used to look at genetic hitchhiking around a selected locus that produces a pattern characteristic of a selective sweep (Schlötterer 2003; Pavlidis *et al.* 2008). The main footprint that selective sweeps leave on the SFS is an excess of rare low frequency 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 composite likelihood ratio test based

490 on the ratio between the likelihood of a null (neutral evolution model) and the alternative
491 (selective sweep) hypothesis. Like differentiation based methods, the weakness of hitchhiking
492 methods is the confounding effect certain demographic processes have on the SFS. A strong
493 population bottleneck can lead to variances in the genealogical history so that some loci have
494 decreased genetic diversity and an excess of low frequency variants (Hermisson 2009). Again,
495 however, building the known demographic history into the null model readily reduces false
496 positive rates (Pavlidis *et al.* 2013).

497 We included the *P. leucopus* demographic history into our analysis, and found 0.04% of
498 the transcriptome to contain regions with SFS patterns indicative of selective sweeps. This rate
499 is in line with other studies that found 0.5% of regions in domesticated rice to show evidence of
500 selective sweeps, though this might be unusually high due to artificial selection (Wang *et al.*
501 2014), 0.02% of loci in black cottonwood experiencing selective sweeps across geographic
502 regions (Zhou *et al.* 2014), and 0.02% of regions across the entire Gorilla genome to show
503 hitchhiking patterns (McManus *et al.* 2014).

504 Individual genome scan approaches look at different aspects of genomic structure and by
505 themselves can miss true outliers, type II errors, or identify false positives, type I errors. Several
506 studies have shown that a general principle to follow in order to avoid these errors is to perform
507 multiple tests looking at various aspects of the genome (Nielsen 2005; Grossman *et al.* 2010;
508 Hohenlohe *et al.* 2010b). We used Bayescan and SweeD to identify outliers experiencing
509 positive selection, but did not find any overlapping candidate genes between them. This finding
510 is not necessarily unexpected as the two tests look at different selection scenarios, divergent local
511 selection versus population-wide positive selection in the form of selective sweeps (Hermisson
512 2009). F_{ST} based methods can pick up on divergence between alleles relatively quickly, while

513 models for selective sweeps typically require nearly-fixed derived alleles (Hohenlohe *et al.*
514 2010b). Given the recent time frame of urbanization in NYC, not enough generations may have
515 passed since white-footed mice have become isolated to find complete selective sweeps in loci
516 that overlap with outliers from Bayescan. In the case of NYC populations of *P. leucopus*, it is
517 likely that adaptation is occurring from standing genetic variation in the form of soft sweeps
518 (Hermisson & Pennings 2005), which are not readily identified by programs like SweeD (De
519 Villemereuil *et al.* 2014). To give further support to this idea, we found several outliers across
520 the various tests we ran that are unique to specific urban populations, which is characteristic of
521 soft sweeps, as they and polygenic traits can lead to outlier SNPs unique to populations (Messer
522 & Petrov 2013). Despite the lack of overlapping outlier SNPs between the two tests, further
523 evidence that positive selection is acting in urban populations of *P. leucopus* was found with an
524 additional approach. Independent confirmation of candidate genes came from correlating
525 genotypes and environmental variables, a method that may be more powerful than the genome
526 scans above for identifying SNPs under selection (Savolainen *et al.* 2013).
527
528 **Environmental associations strengthen evidence of local adaptation to urbanization**
529 Genotype-environment association tests are a growing class of methods that provide fine
530 scale detail about the ecological processes driving selection by identifying loci with allele
531 frequencies that are correlated with environmental factors. Several have recently been developed
532 (Joost *et al.* 2007; Coop *et al.* 2010; Frichot *et al.* 2013), and here we used LFMM (Frichot *et al.*
533 2013) to associate outlier SNPs with environmental measurements that capture the effects of
534 urbanization. LFMM is uniquely suited for our dataset as it has been found to perform better
535 than other methods in the presence of hierarchical structure and when polygenic selection is

536 acting on many loci with small effect (De Villemereuil *et al.* 2014). In our dataset, there are
537 many layers of structure including urban and rural differentiation (Harris *et al.* 2015; Harris *et al.*
538 2016), patterns of geographic structure between mainland mice and Long Island, NY (Harris *et*
539 *al.* 2016), and population structure between individual urban parks (Munshi-South &
540 Kharchenko 2010). It also has more power when the sampling size is less than 10 individuals
541 per populations, there is no evidence of IBD, and sampling design of the experiment involves
542 pairs in environmentally heterogeneous habitats (Lotterhos & Whitlock 2015). We sampled
543 eight white-footed mice per population, found no evidence of IBD (Munshi-South *et al.* 2016),
544 and sampled environmentally heterogeneous rural and urban locations.

545 Using LFMM, we found that 75 % and 47 % of outliers from Bayescan and SweeD,
546 respectively, could also be associated with one or more environmental variables. These results
547 complement our findings that positive selection is acting on urban populations of white-footed
548 mice. We acknowledge that impervious surface, human density, or classification as urban may be
549 correlated with a different environmental selection force, but our results ultimately support an
550 evolutionary scenario where isolated urban populations are experiencing divergent positive
551 selection that is strongly affected by one or more environmental variables, likely associated with
552 urbanization. These results are also consistent with other studies combining genome scan
553 methods and GEA tests. Limborg *et al.* (2012) found 62.5 % of the outliers identified in
554 Bayescan to be correlated with temperature or salinity changes in Atlantic herring, and 26.3 % of
555 genome scan outliers could be associated with temperature or latitude in the tree species, *A.*
556 *glutinosa* (De Kort *et al.* 2014).

557 The percent impervious surface and human density around a park, or the classification of
558 sites as urban or rural, are efficient metrics for determining whether a sampling location has been

559 affected by urbanization (Munshi-South *et al.* 2016). We can make several predications about
560 how ecological processes are changing within parks influenced by urbanization. One of the most
561 obvious consequences of human altered environments is habitat loss and fragmentation
562 (McKinney 2002; Sih *et al.* 2011). The act of fragmentation and the building of infrastructure
563 invariably changes the net primary productivity due to increasing percentages of impervious
564 surface or artificial landscapes, parks and yards (Shochat *et al.* 2006). Additionally, species
565 interactions change as organisms are forced into smaller areas or separated by infrastructure
566 (Shochat *et al.* 2006). This includes impediments to migration across the urbanized landscape.
567 Humans often introduce invasive species into habitats (Sih *et al.* 2011) leading to increased
568 competition or novel predator-prey interactions. Urbanization also changes the types and
569 availability of resources available in the altered habitat (McKinney 2002; Sih *et al.* 2011).
570 Pollution is also a major consequence of urbanization (Donihue & Lambert 2014), and can
571 include chemical, noise, or light pollution (Sih *et al.* 2011).
572 Given the rapid alteration of environments during urbanization, behavioral flexibility and
573 phenotypic plasticity are thought to play an important role in a species' response to novel urban
574 ecosystems (Sih *et al.* 2011). Climate change, another form of human-induced rapid
575 environmental change, is often used as a model for understanding plastic and evolutionary
576 responses in organisms. Franks *et al.* (2014), in a comprehensive review of phenotypic changes
577 in plants in response to climate change, reported that the majority of studies showed evidence of
578 plastic responses. They also found many studies showed evidence of adaptation, though not
579 always conclusively. Looking at animal responses to climate, Boutin & Lane (2014) found
580 similar findings but even less conclusive evidence of adaptation versus plasticity, possibly due to
581 the motility of animals and difficulty in establishing common garden or reciprocal transplant

582 experiments. While it is likely that *P. leucopus* in NYC are displaying some plastic phenotypic
583 responses in urban ecosystems, our results provide evidence of heritable evolutionary responses
584 as well.

585 Between divergent allele frequencies, a skewed SFS, and environmental associations, we
586 find several overlapping lines of evidence that support rapid divergent positive selection in
587 white-footed mice. Urban ecologists are increasingly finding evidence of selection acting in
588 urban environments (Donihue & Lambert 2014), and our results are in line with other studies that
589 have found rapid local adaptation to ecological pressures from urbanization. Yeh (2004) found
590 sexually selected tail coloration in juncos was rapidly evolving in urban populations compared to
591 rural ones. European blackbirds show reduced migratory behavior in cities, and there is also
592 evidence of selection on genes underlying anxiety behavior across multiple urban areas (Partecke
593 *et al.* 2006; Mueller *et al.* 2013). Cheptou *et al.* (2008) found weeds in urban vegetation plots
594 surrounded by paved surfaces had a higher percentage of non-dispersing seeds and that this trait
595 was genetically based. In marine species living in the polluted waters around urban areas, rapid
596 adaptation for PCB resistance occurred in both killifish and tomcod (Whitehead *et al.* 2010;
597 Wirgin *et al.* 2011). The realization that a diverse range of taxa may adapt to human induced
598 landscape change suggests rapid adaptation to anthropogenic driven environmental change may
599 be pervasive in nature.
600
601 **Functional roles and ecological relevance of candidate genes**
602 The model rodent species *Mus musculus*, *Rattus norvegicus*, and *Cricetulus griseus*, all
603 have deeply sequenced, assembled and annotated reference genomes. These resources allowed
604 us to annotate 89.5 % of contigs containing outlier SNPs and genomic regions with high quality

gene information. These annotations provided us with information about the traits affected by candidate genes. Urban *P. leucopus* specifically exhibited genetic patterns that suggest positive selection in genes from the mitochondria, a potentially significant finding considering mitochondrial genes are often used for demographic inference (Munshi-South & Nagy 2014). Tests for selection also identified genes that protect cellular health in stressful environments, modulate melanism throughout the body, genes that are involved in epigenetic control of gene expression, or involved in digestion and metabolism of lipids and carbohydrates.

Gene ontology vocabulary assigns gene function according to biological process, molecular function, and cellular component. Across all candidate genes and gene ontology terms, involvement with mitochondria was one of the most common assignments (Table S1). Whether genes were involved in energy production through metabolism of food or were actual mitochondrial proteins, it appears evolution in mitochondria and metabolic processes is extremely important for *P. leucopus* living in urban parks. Mitochondrial genes were traditionally used as neutrally evolving markers, but researchers are finding evidence of selection on mitochondrial DNA across taxa (Oliveira *et al.* 2008; Balloux 2010). One example includes mitochondrial haplotypes associated with more efficient non-shivering thermogenisis and higher fitness in over-wintering shrews (Fontanillas *et al.* 2005). In *Peromyscus leucopus*, Pergams & Lacy (2007) found complete mitochondrial haplotype replacement in present-day white-footed mice living in the urban Chicago environment compared to haplotypes found in museum skins collected from before urbanization. The agent of selection is not clear, but independent research found evidence of negative selection acting on the mitochondrial D-loop gene in NYC *P. leucopus* (Munshi-South & Nagy 2014). These findings are not surprising. Many mitochondria-related metabolic functions are affected by the same environmental variables that change in

response to urbanization, like temperature (Urban = heat island effect) (Balloux 2010), population density (Urban = barriers to dispersal around parks) (Lankau & Strauss 2011; Munshi-South 2012), or resource availability (Urban = increased non-native prey) (Burcelin *et al*. 2002). In novel urban ecosystems, *P. leucopus* may be experiencing different energy requirements than rural counterparts.

One example of uniquely urban energy requirements comes from the signature of a selective sweep and a strong correlation with urban site classification found in the heat-shock protein Hsp90. Heat shock proteins are a gene family that have repeatedly been found to play a pivotal role in adaptation to environmental stress (Limborg *et al.* 2012). In a landmark study, cryptic variation in Hsp90 specifically, was found to act as a capacitor for the loss of eyes in cavefish (Rohner *et al*. 2013). Essentially, under normal environmental conditions, Hsp90 masks phenotypic variation in eye size, but under high stress conditions, Hsp90 is effectively inhibited allowing for eye size variation and eventual selection for unmasked phenotypic traits. In *Peromyscus* spp., Hsp90 acts 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). When *P. maniculatus* was exposed to soils inundated with the toxin, 2,3,7,8 TCDD, maintenance of their circadian rhythm was affected and mice became active 3 hours earlier than under normal conditions (Settachan 2001). The aldo-keto reductase gene, aflatoxin aldehyde reductase (AKR7), was also an outlier in our analyses and is also important for metabolizing environmental toxins (Hyndman *et al.* 2003). Aflatoxin is a natural carcinogen often found in cereals and nuts contaminated with the fungus, *A. flavus* and is metabolically activated by cytochrome P450 (Jin & Penning 2007). In experiments on *Rattus norvegicus*, researchers found AKR7 is upregulated in the liver when exposed to various classes of toxins

651 and quickly acts to metabolize them, protecting cellular health (Ellis *et al.* 2003). We found *P.*
652 *leucopus* caught in NYC had more enlarged, scarred, and fatty livers than those from rural
653 populations (personal observation), and this may be directly related to ecological conditions in
654 urban environments that promote environmental toxin accumulation. Due to proximity to human
655 infrastructure, urban soils consistently show increased levels of heavy metal contamination
656 (McDonnell *et al.* 1997). Urban ecosystems also experience the heat island effect with higher
657 temperatures than rural locations (McDonnell *et al.* 1997), leaf litter that quickly decomposes but
658 is of poor quality (Pouyat *et al.* 1997), and NYC in particular experiences high humidity in
659 warmer months (National Oceanic and Atmospheric Administration, NOAA). The combination
660 of constantly decaying vegetation, high temperatures, and high humidity is ideal for healthy
661 communities of the fungus *A. flavus*, the primary producer of aflatoxins. Hsp90, AKR7, and
662 cytochrome P450 may be under selective pressures in NYC to efficiently metabolize higher
663 concentrations of toxins in *P. leucopus* exposed to polluted urban soils or food sources in NYC.
664 Energy requirements may also be different in in urban populations because of dietary
665 shifts. We found a surprising number of candidate genes with functions related to the
666 metabolism and transport of lipids and carbohydrates. These genes were strongly correlated with
667 environmental measures of urbanization, with clearly divergent allele frequencies between urban
668 and rural sites (Fig. 3B). APOB-100 is the primary apolipoprotein that binds and transports
669 lipids, including both forms of cholesterol (HDL and LDL), and *Mus musculus* knock-out
670 models result in hyperglycemia and obesity (Lloyd *et al.* 2008). FADS1, a farnesoid-x-receptor,
671 is a nuclear receptor antagonist that is involved in bile synthesis and modulates high fat diets,
672 with variation in expression affecting rates of obesity in mice (Li *et al.* 2013). Manually curated
673 protein annotations show MYLK and SORBS2 are both directly involved in the gastrointestinal

674 system, involved in smooth muscle contractions and absorption of water and sodium in the
675 intestine, respectively (Magrane & Consortium 2011; Consortium 2014). ABCC8 is an ATP-
676 binding cassette transporter, and knock-out mice models lack insulin secretion in response to
677 glucose (Seghers *et al.* 2000). Finally, KEGG analysis found that two contigs (10636-348 and
678 27546-129) represent proteins that are both directly involved in Galactose, Fructose and
679 Mannose metabolism (Ogata *et al.* 1999).

680 These candidate genes suggest that white-footed mice in isolated urban parks are
681 responding to resource differences between urban and rural habitats. One prediction is urban *P.*
682 *leucopus* consume a diet with higher overall fat content. The typical diet of *P. leucopus* across
683 its range consists of arthropods, fruits, nuts, various green vegetation, and fungus (Wolff *et al.*
684 1985). They are especially reliant on oak mast cycles and an important predator of gypsy moths
685 (Ostfeld *et al.* 1996). They are generalists and opportunistic in the food they eat, and thus many
686 different food resources could drive diet differences in urban versus rural systems. Urbanization
687 in NYC has lead to relatively small green patches that are surrounded by a dense urban matrix.
688 The high percent of impervious surface is detrimental to the persistence of white-tailed deer in
689 urban parks, leading to their exclusion throughout the majority of NYC. An overabundance of
690 deer, like what occurs in our rural sampling sites, leads to the removal of the vegetative
691 understory and inhibits regeneration of many plants (Stewart 2001). In these heavily browsed
692 habitats lacking a thick vegetative understory, there is direct correlation with length of deer
693 browsing in the area and invertebrate species diversity and abundance (Stewart 2001; Allombert
694 *et al.* 2005). As the understory is cleared by deer there are fewer food resources and habitats for
695 woodland invertebrates.

This is not the case for urban parks that often have extremely thick and healthy understories (Leston & Rodewald 2006). Although the understory of urban forest fragments is typically composed of invasive plants, such an understory can produce a number of 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 are likely so successful in urban ecosystems because they take advantage of the new food sources in urban habitats, including seeds and other plant parts from an invasive understory layer, as well as invertebrates that may thrive in urban fragments. There has been much research on adaptation to diet specialization, especially in human populations. One well known case involves mutations in the human lactase gene that lead to lactase persistence, most likely in response to a cattle domestication event (Enattah *et al.* 2008). Another study that looked at more subtle shifts in allele frequencies across human populations found outlier SNPs within genes that more efficiently metabolize proteins found in the root and tuber based diets that humans switched to as they moved into polar ecoregions (Hancock *et al.* 2010). There is also growing evidence of adaptation in native predators in order to consume exotic or toxic prey species (Carlsson *et al.* 2009), for example, larger mouthparts in the Australian soapberry bug to increase foraging on invasive balloon vines (Carroll *et al.* 2005).

We hypothesize that urban *P. leucopus* have much higher fat content in their diets due to increased seed or invertebrate abundance or the inclusion of high-fat human food waste, and local adaption is occurring to more efficiently metabolize the increased lipids and carbohydrates. There is strong genetic evidence that divergent positive selection is occurring between urban and rural mice, but in order to confirm hypotheses, it would be worth performing common garden experiments to measure metabolic rates when mice from different habitats are fed a consistent

719 diet, or sequencing these same candidate genes across a broader range of urban and rural sites to
720 look for similar signatures of selection. It might also be worthwhile to associate outlier SNPs
721 with more fine scale ecological measurements like temperature, environmental pollutant level, or
722 vegetative understory cover. Diet analyses between sites can also be undertaken and with the use
723 of a metabarcoding approach using next generation sequencing, the entire diet can easily be
724 identified from *P. leucopus* waste (Pompanon *et al.* 2012; Soininen *et al.* 2013).
725

726 **CONCLUSIONS**

727 Results strongly suggest that populations of *Peromyscus leucopus* within urban parks in
728 NYC are adapting to the effects of urbanization. Focusing on protein-coding regions of the
729 genome, using multiple tests of selection that analyze different parts of genomic structure, and
730 associating outliers with environmental variables that capture the ecological changes imposed by
731 urbanization allowed us to narrow in on specific genes underlying recent adaptation in urban
732 habitats. In line with the definition of an 'urban adapter' (McKinney 2002), the generalist *P.*
733 *leucopus* is successful in urban parks, and our results suggest white-footed mice may be adapting
734 to changing dietary resources in urban ecosystems and potentially metabolizing increased
735 chemical pollutants in their environment. While we find definitive evidence of genetic variation
736 between urban and rural sampling sites, further work needs to be done to look at specific
737 polymorphisms and their impact on translation and protein folding.

738 Next steps should include SNP assays or full sequencing of outlier genes in more
739 individuals from an increased number of sites across the urban - rural gradient. With this further
740 confirmation, ecological based studies of diet can be pursued. Humans are increasingly altering
741 the natural landscape through urbanization and indirectly through global climate change.

742 Despite this, there are few studies with clear evidence of adaptation in novel urban ecosystems.
743 Our study begins to address this issue using the statistical power of genomic datasets and finds
744 that rapid adaptation is possible in recently disturbed ecosystems. By providing further
745 understanding of contemporary evolution in response to urbanization, we have begun to answer
746 important questions about the traits involved in adaptation to human modified landscapes and
747 what environmental variables most likely drive this adaptation. Hopefully, these insights can be
748 used for urban ecosystem management as global biodiversity continues to deal with
749 unprecedented environmental change in the Anthropocene.
750

751 **ACKNOWLEDGMENTS**

752 This research was supported by the National Institute of General Medical Sciences of the
753 National Institutes of Health under award number R15GM099055 to JM-S and a NSF Graduate
754 Research Fellowship to SEH. The content is solely the responsibility of the authors and does not
755 represent the official views of the National Institutes of Health.
756
757

758 **REFERENCES**

759 Allombert S, Stockton S, Martin JL (2005) A natural experiment on the impact of overabundant
760 deer on forest invertebrates. *Conservation Biology*, **19**, 1917–1929.
761 Balloux F (2010) The worm in the fruit of the mitochondrial DNA tree. *Heredity*, **104**, 419–420.
762 Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level.
763 *Nature Reviews Genetics*, **12**, 767–780.

764 Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology*
765 *& Evolution*, **23**, 38–44.
766 Beaumont M a (2005) Adaptation and speciation: what can F(st) tell us? *Trends in Ecology &*
767 *Evolution*, **20**, 435–40.
768 Biswas S, Akey JM (2006) Genomic insights into positive selection. *Trends in Genetics*, **22**,
769 437–46.
770 Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with
771 functional genomics. *Molecular Ecology*, **17**, 3583–4.
772 Boutin S, Lane JE (2014) Climate change and mammals: Evolutionary versus plastic responses.
773 *Evolutionary Applications*, **7**, 29–41.
774 Burcelin R, Crivelli V, Dacosta A, Roy-Tirelli A, Thorens B (2002) Heterogeneous metabolic
775 adaptation of C57BL/6J mice to high-fat diet. *American Journal of Physiology.*
776 *Endocrinology and Metabolism*, **282**, E834–E842.
777 Carlsson NOL, Sarnelle O, Strayer DL (2009) Native predators and exotic prey - An acquired
778 taste? *Frontiers in Ecology and the Environment*, **7**, 525–532.
779 Carroll SP, Loye JE, Dingle H *et al.* (2005) And the beak shall inherit - Evolution in response to
780 invasion. *Ecology Letters*, **8**, 944–951.
781 Chace JF, Walsh JJ (2004) Urban effects on native avifauna: a review. *Landscape and Urban*
782 *Planning*, **74**, 46–69.
783 Chapman M a, Hiscock SJ, Filatov D a (2013) Genomic Divergence during Speciation Driven by
784 Adaptation to Altitude. *Molecular Biology and Evolution*, **30**, 2553–67.

785 Chen S, Zhang G, Shao C *et al.* (2014) Whole-genome sequence of a flatfish provides insights
786 into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics*,
787 **46**.
788 Cheptou P-O, Carrue O, Rouifed S, Cantarel A (2008) Rapid evolution of seed dispersal in an
789 urban environment in the weed Crepis sancta. *Proceedings of the National Academy of*
790 *Sciences of the United States of America*, **105**, 3796–9.
791 Conesa A, Götz S, García-Gómez JM *et al.* (2005) Blast2GO: a universal tool for annotation,
792 visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–6.
793 Consortium TU (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids*
794 *Research*, **42**, D191–8.
795 Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to
796 identify loci underlying local adaptation. *Genetics*, **185**, 1411–23.
797 Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools.
798 *Bioinformatics*, **27**, 2156–2158.
799 DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and
800 genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–8.
801 Donihue CM, Lambert MR (2014) Adaptive evolution in urban ecosystems. *Ambio*, 1–10.
802 Ellis EM, Slattery CM, Hayes JD (2003) Characterization of the rat aflatoxin B1 aldehyde
803 reductase gene, AKR7A1. Structure and chromosomal localization of AKR7A1 as well as
804 identification of antioxidant response elements in the gene promoter. *Carcinogenesis*, **24**,
805 727–737.

806 Ellison CE, Hall C, Kowbel D *et al.* (2011) Population genomics and local adaptation in wild
807 isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences*,
808 **108**, 2831–2836.
809 Enattah NS, Jensen TGK, Nielsen M *et al.* (2008) Independent Introduction of Two Lactase-
810 Persistence Alleles into Human Populations Reflects Different History of Adaptation to
811 Milk Culture. *American Journal of Human Genetics*, **82**, 57–72.
812 Endler J (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton, NJ.
813 Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M (2013) Robust Demographic
814 Inference from Genomic and SNP Data. *PLoS Genetics*, **9**, e1003905.
815 Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured
816 population. *Heredity*, **103**, 285–98.
817 Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both
818 dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–93.
819 Fontanillas P, Dépraz A, Giorgi MS, Perrin N (2005) Nonshivering thermogenesis capacity
820 associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew,
821 Crocidura russula. *Molecular Ecology*, **14**, 661–670.
822 Franks SJ, Hoffmann A a. (2011) Genetics of Climate Change Adaptation. *Annual Review of*
823 *Genetics*, **46**, 185–208.
824 Franks SJ, Weber JJ, Aitken SN (2014) Evolutionary and plastic responses to climate change in
825 terrestrial plant populations. *Evolutionary Applications*, **7**, 123–139.
826 Frichot E, François O (2015) LEA□: An R package for landscape and ecological association
827 studies. *Methods in Ecology and Evolution*, **6**.

828 Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci
829 and environmental gradients using latent factor mixed models. *Molecular Biology and*
830 *Evolution*, **30**, 1687–1699.
831 Götz S, García-Gómez JM, Terol J *et al.* (2008) High-throughput functional annotation and data
832 mining with the Blast2GO suite. *Nucleic Acids Research*, **36**, 3420–35.
833 Grossman SR, Shylakhter I, Karlsson EK *et al.* (2010) A composite of multiple signals
834 distinguishes causal variants in regions of positive selection. *Science*, **327**, 883–6.
835 Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint
836 demographic history of multiple populations from multidimensional SNP frequency data.
837 *PLoS Genetics*, **5**, e1000695.
838 Hancock AM, Witonsky DB, Ehler E *et al.* (2010) Colloquium paper: human adaptations to diet,
839 subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proceedings of the*
840 *National Academy of Sciences of the United States of America*, **107**, 8924–8930.
841 Harris SE, Munshi-South J, Obergfell C, O'Neill R (2013) Signatures of Rapid Evolution in
842 Urban and Rural Transcriptomes of White-Footed Mice (Peromyscus leucopus) in the New
843 York Metropolitan Area. *PLoS ONE*, **8**, e74938.
844 Harris SE, O'Neill RJ, Munshi-South J (2015) Transcriptome resources for the white-footed
845 mouse (Peromyscus leucopus): new genomic tools for investigating ecologically divergent
846 urban and rural populations. *Molecular ecology resources*, **15**, 382–394.
847 Harris SE, Xue AT, Alvarado-Serrano D *et al.* (2016) Urbanization shapes the demographic
848 history of a city-dwelling native rodent. *bioRxiv*, doi:10.1101/032979.
849 Hedrick PW, Ginevan ME, Ewing EP (1976) Genetic polymorphism in heterogeneous
850 environments. *Annual Review of Ecology and Systematics*, **7**, 1–32.

851 Henry P, Russello M a. (2013) Adaptive divergence along environmental gradients in a climate-
852 change-sensitive mammal. *Ecology and Evolution*, **3**, 3906–3917.
853 Hermisson J (2009) Who believes in whole-genome scans for selection? *Heredity*, **103**, 283–284.
854 Hermisson J, Pennings PS (2005) Soft sweeps: Molecular population genetics of adaptation from
855 standing genetic variation. *Genetics*, **169**, 2335–2352.
856 Hohenlohe PA, Bassham S, Etter PD *et al.* (2010a) Population genomics of parallel adaptation in
857 threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.
858 Hohenlohe PA, Phillips PC, Cresko WA (2010b) Using Population Genomics To Detect
859 Selection in Natural Populations: Key Concepts and Methodological Considerations.
860 *International Journal of Plant Sciences*, **171**, 1059–1071.
861 Hohenlohe P a., Phillips PC, Cresko W a. (2011) Using population genomics to detect selection
862 in natural populations: Key concepts and methodological considerations. *International*
863 *Journal of Plant Science*, **171**, 1059–1071.
864 Hyndman D, Bauman DR, Heredia V V., Penning TM (2003) The aldo-keto reductase
865 superfamily homepage. *Chemico-Biological Interactions*, **143-144**, 621–631.
866 Jin Y, Penning TM (2007) Aldo-keto reductases and bioactivation/detoxication. *Annual review*
867 *of pharmacology and toxicology*, **47**, 263–292.
868 Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect
869 candidate loci for selection: towards a landscape genomics approach to adaptation.
870 *Molecular Ecology*, **16**, 3955–69.
871 Kanehisa M, Goto S, Sato Y *et al.* (2014) Data, information, knowledge and principle: Back to
872 metabolism in KEGG. *Nucleic Acids Research*, **42**, 199–205.

873 Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–
874 1241.
875 Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a
876 recombining chromosome. *Genetics*, **160**, 765–777.
877 De Kort H, Vandepitte K, Bruun HH *et al.* (2014) Landscape genomics and a common garden
878 trial reveal adaptive differentiation to temperature across Europe in the tree species Alnus
879 glutinosa. *Molecular Ecology*, 4709–4721.
880 Lankau RA, Strauss SY (2011) Newly rare or newly common: evolutionary feedbacks through
881 changes in population density and relative species abundance, and their management
882 implications. *Evolutionary Applications*, **4**, 338–353.
883 Leston LF V, Rodewald AD (2006) Are urban forests ecological traps for understory birds? An
884 examination using Northern cardinals. *Biological Conservation*, **131**, 566–574.
885 Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the
886 selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
887 Li YF, Costello JC, Holloway AK, Hahn MW (2008) "Reverse ecology" and the power of
888 population genomics. *Evolution*, **62**, 2984–2994.
889 Li F, Jiang C, Krausz KW *et al.* (2013) Microbiome remodelling leads to inhibition of intestinal
890 farnesoid X receptor signalling and decreased obesity. *Nature communications*, **4**, 2384.
891 Li J, Li H, Jakobsson M *et al.* (2012) Joint analysis of demography and selection in population
892 genetics: where do we stand and where could we go? *Molecular Ecology*, **28**, 28–44.
893 Limborg MT, Helyar SJ, De Bruyn M *et al.* (2012) Environmental selection on transcriptome-
894 derived SNPs in a high gene flow marine fish, the Atlantic herring (Clupea harengus).
895 *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.

Lloyd DJ, McCormick J, Helmering J *et al.* (2008) Generation and characterization of two novel mouse models exhibiting the phenotypes of the metabolic syndrome: Apob48-/-Lepob/ob mice devoid of ApoE or Ldlr. *American journal of physiology. Endocrinology and metabolism*, **294**, E496–E505.

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.

Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary Applications*, **7**, 1–14.

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

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.

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, 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.

940 Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing
941 Heliconius butterflies identified by large-scale targeted sequencing. *Philosophical*
942 *Transactions of the Royal Society of London. Series B, Biological Sciences*, **367**, 343–53.
943 Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–
944 218.
945 Nielsen R, Williamson S, Kim Y *et al.* (2005) Genomic scans for selective sweeps using SNP
946 data. *Genome research*, **15**, 1566–75.
947 Ogata H, Goto S, Sato K *et al.* (1999) KEGG: Kyoto encyclopedia of genes and genomes.
948 *Nucleic Acids Research*, **27**, 29–34.
949 Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural
950 selection. *Philosophical transactions of the Royal Society of London. Series B, Biological*
951 *sciences*, **365**, 185–205.
952 Oliveira DCSG, Raychoudhury R, Lavrov D V., Werren JH (2008) Rapidly evolving
953 mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp
954 Nasonia (Hymenoptera: Pteromalidae). *Molecular Biology and Evolution*, **25**, 2167–2180.
955 Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**,
956 119–27.
957 Ostfeld R., Jones C, Wolff J (1996) Of mice and mast: ecological connections in eastern
958 deciduous forests. *BioScience*, **46**, 323–330.
959 Partecke J, Schwabl I, Gwinner E (2006) Stress and the city: Urbanization and its effects on the
960 stress physiology in European Blackbirds. *Ecology*, **87**, 1945–1952.
961 Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**,
962 e190.

963 Pavlidis P, Hutter S, Stephan W (2008) A population genomic approach to map recent positive
964 selection in model species. *Molecular Ecology*, **17**, 3585–98.
965 Pavlidis P, Jensen JD, Stephan W (2010) Searching for Footprints of Positive Selection in
966 Whole-genome SNP Data from Non-equilibrium Populations. *Genetics*.
967 Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N (2013) SweeD: likelihood-based detection of
968 selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, **30**, 2224–34.
969 Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010)
970 Comparing three different methods to detect selective loci using dominant markers. *Journal*
971 *of Evolutionary Biology*, **23**, 2267–2276.
972 Pergams ORW, Lacy RC (2007) Rapid morphological and genetic change in Chicago-area
973 Peromyscus. *Molecular Ecology*, **17**, 450–63.
974 Pompanon F, Deagle BE, Symondson WOC *et al.* (2012) Who is eating what: Diet assessment
975 using next generation sequencing. *Molecular Ecology*, **21**, 1931–1950.
976 Pool JE, Aquadro CF (2007) The genetic basis of adaptive pigmentation variation in Drosophila
977 melanogaster. *Molecular Ecology*, **16**, 2844–2851.
978 Pouyat R V, Mcdonnell MJ, Pickett ST a (1997) Litter decomposition and nitrogen
979 mineralization in oak stands along an urban-rural land use gradient. *Urban Ecosystems*, **1**,
980 117–131.
981 Reznick D, Ghalambor C (2001) The population ecology of contemporary adaptations: What
982 empirical studies reveal about the conditions that promote adaptive evolution. *Genetica*,
983 **112-113**, 183–198.
984 Rockman M V (2012) The QTN program and the alleles that matter for evolution: all that's gold
985 does not glitter. *Evolution*, **66**, 1–17.

986 Rogic A, Tessier N, Legendre P, Lapointe F-J, Millien V (2013) Genetic structure of the white-
987 footed mouse in the context of the emergence of Lyme disease in southern Québec. *Ecology*
988 *and Evolution*, **3**, 2075–88.
989 Rohner N, Jarosz DF, Kowalko JE *et al.* (2013) Cryptic variation in morphological evolution:
990 HSP90 as a capacitor for loss of eyes in cavefish. *Science*, **342**, 1372–5.
991 Rytwinski T, Fahrig L (2007) Effect of road density on abundance of white-footed mice.
992 *Landscape Ecology*, **22**, 1501–1512.
993 Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature*
994 *Reviews Genetics*, **14**, 807–20.
995 Schlötterer C (2003) Hitchhiking mapping--functional genomics from the population genetics
996 perspective. *Trends in Genetics*, **19**, 32–8.
997 Seghers V, Nakazaki M, DeMayo F, Aguilar-Bryan L, Bryan J (2000) Sur1 Knockout Mice
998 study of K ATP -independent pathways that regulate insu-. *The Journal of biological*
999 *chemistry*, **275**, 9270–9277.
1000 Settachan D (2001) Mechanistic and molecular studies into the effects of 2,3,7,8-
1001 tetrachlorodibenzo-p-dioxin and similar compounds in the deer mouse, Peromyscus
1002 maniculatus. Texas Tech University.
1003 Shochat E, Warren PS, Faeth SH, McIntyre NE, Hope D (2006) From patterns to emerging
1004 processes in mechanistic urban ecology. *Trends in Ecology and Evolution*, **21**, 186–91.
1005 Sih A, Ferrari MCO, Harris DJ (2011) Evolution and behavioural responses to human-induced
1006 rapid environmental change. *Evolutionary Applications*, **4**, 367–387.
1007 Soininen EM, Ravolainen VT, Bråthen KA *et al.* (2013) Arctic Small Rodents Have Diverse
1008 Diets and Flexible Food Selection. *PloS one*, **8**, e68128.

1009 Stajich JE, Hahn MW (2005) Disentangling the effects of demography and selection in human
1010 history. *Molecular Biology and Evolution*, **22**, 63–73.
1011 Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation Genomics: the next generation. *Trends*
1012 *in Ecology & Evolution*, **25**, 705–712.
1013 Stewart AJA (2001) The impact of deer on lowland woodland invertebrates: A review of the
1014 evidence and priorities for future research. *Forestry*, **74**, 259–270.
1015 Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics:
1016 finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–70.
1017 Storz JF, Runck AM, Sabatino SJ *et al.* (2009) Evolutionary and functional insights into the
1018 mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proceedings of*
1019 *the National Academy of Sciences of the United States of America*, **106**, 14450–5.
1020 Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*,
1021 **105**, 437–460.
1022 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA
1023 polymorphism. *Genetics*, **123**, 585–95.
1024 Tennessen J a., Madeoy J, Akey JM (2010) Signatures of positive selection apparent in a small
1025 sample of human exomes. *Genome Research*, 1327–1334.
1026 Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand
1027 local adaptation. *Trends in Ecology & Evolution*, **29**, 673–680.
1028 Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin S V (2010) Population resequencing
1029 reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature Genetics*, **42**, 260–
1030 3.

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.

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.

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.

Wilches R (2014) Evolution of genes related to temperature adaptation in drosophila melanogaster as revealed by qtl and population genetics analyses. University of Munchen.

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 statistic averages for six *P. leucopus* populations

Population	Nucleotide Diversity, π (mean ± SD)	Tajima's *D* (mean ± SD)
CP	0.131 ±0.173	0.318 ±0.522
FM	0.112 ±0.166	0.301 ±0.522
NYBG	0.094 ±0.153	0.280 ±0.500
BHwwp	0.198 ±0.186	0.350 ±0.549
CFP	0.211 ±0.184	0.336 ±0.543
HIP	0.263 ±0.182	0.349 ±0.569

Table 2. Average pairwise F_{ST} among six *P. leucopus* populations

	F_{ST} (mean ± SD)					
	BHwwp	HIP	CFP	CP	NYBG	FM
BHwwp	0					
HIP	0.042 ±0.376	0				
CFP	0.034 ±0.400	0.018 ±0.364	0			
CP	0.089 ±0.458	0.060 ±0.417	0.063 ±0.447	0		
NYBG	0.070 ±0.477	0.054 ±0.428	0.043 ±0.462	0.092 ±0.536	0	
FM	0.056 ±0.468	0.061 ±0.420	0.057 ±0.456	0.109 ±0.520	0.085 ±0.535	0

1086 **Table 3.** Results for selection from Bayescan and associations with environmental variables

1087 across urban (CP, FM, NYBG) and rural (BHwwp, CFP, HIP) populations. I = percent

1088 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	-	-	-
35973-42	isoform cra_a	-	-	+
36706-36	hermansky-pudlak syndrome 1 protein homolog	-	-	+
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	+	+	+
36701-36	isoform cra_b	-	-	-
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	+	-	+
38691-21	protein mdm4	-	-	-
1396-1029	proteasome activator complex subunit 1	-	-	+
7750-426	lysine-specific demethylase 8	-	-	+

11279-335	mitochondrial ribosomal protein l37	-	-	-
26257-147	PREDICTED: uncharacterized protein C1orf167 homolog	-	-	+
31894-78	--	-	-	+
26183-148	atp-binding cassette sub-family c member 8-like	+	-	+
14102-290	succinate dehydrogenase	-	-	+
40819-1	adaptin ear-binding coat-associated protein 1	+	-	+

Table 4. Results for selection from SweeD and associations with environmental variables across urban (CP, FM, NYBG) and rural (BHwwp, CFP, HIP) populations. Columns to the left of the outliers show the population where the SweeD identified outlier was found. 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	+	-	+
+	-	-	-	1115-1128	--	-	-	-
-	-	-	+	113-2629	--	-	-	-
-	-	-	+	11470-332	--	-	-	-
+	-	-	-	1156-1114	--	-	-	-
+	-	-	+	124-2491	--	-	-	-
+	-	-	-	12718-311	--	-	-	-
-	+	-	+	13665-297	--	-	-	-
-	-	-	+	14528-283	solute carrier family 22 (organic anion transporter) member 7	+	-	+
-	-	-	+	148-2324	--	-	-	-
-	-	+	-	1583-971	isoform cra_a	-	-	+
-	-	+	+	1596-968	small ubiquitin-related modifier 2 isoform 2	-	-	+
-	-	-	+	17779-244	--	-	-	-
-	-	-	+	1782-919	solute carrier family 39 (zinc transporter) member 1	-	-	+
-	+	+	-	17856-243	serine protease inhibitor a3c-like	+	+	+
-	-	-	+	2034-860	dehydrogenase reductase (sdr family) member 3	-	-	+
-	-	+	-	20378-220	--	-	-	-
-	-	-	+	21270-213	--	-	-	-
-	-	-	+	22908-200	--	-	-	-
-	-	-	+	23358-193	--	-	-	+
-	+	-	-	243-1951	solute carrier family member 13	-	-	+
+	+	+	+	25500-158	--	-	-	-
-	+	-	-	26488-144	pentatricopeptide repeat domain-containing protein mitochondrial isoform x2	+	-	+
-	-	+	-	2736-755	--	-	-	-
-	-	+	-	27546-129	6- liver type	-	-	+
-	-	-	+	280-1900	--	-	-	-
-	-	-	+	28127-122	sarcosine mitochondrial	-	-	+
-	-	-	+	28528-117	--	-	-	-
-	-	-	+	289-1886	--	-	-	-
-	-	-	+	29117-109	--	-	-	-

-	+	-	-	31034-87	--	-	-	-
+	-	-	-	342-1776	fatty acid desaturase 1	+	-	+
+	+	+	+	35790-44	apolipoprotein b- partial	-	-	-
-	-	-	+	37202-32	PREDICTED: poly	+	-	+
-	-	-	+	37400-30	--	-	-	-
-	-	-	+	37830-27	--	-	-	-
-	-	-	+	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	-	-	+
-	-	-	+	42-3615	2-oxoglutarate dehydrogenase	-	-	+
+	-	-	-	4384-592	--	-	-	-
-	-	-	+	4818-563	exportin-t isoform x1	+	+	+
-	-	-	+	50-3466	--	-	-	-
-	-	-	+	533-1512	fructose- -bisphosphatase 1	-	-	+

1110 **Figure 1.** Map of NYC and surrounding area showing included sample localities. Sites in Red
1111 are urban parks within New York City. CP = Central Park; NYBG = New York Botanical
1112 Gardens; FM = Flushing Meadow/Willow Lake; BHwwp = Brookhaven and Wilde Wood State
1113 Park; CFP = Clarence Fahnestock State Park; HIP = High Point State Park

1114 **Figure 2.** (a) Bayescan results; Red line is FDR = 0.1. (b) Sweed results, Red line corresponds to
1115 p value ≤ 0.0001

1116 **Figure 3**. (a) Population values for three environmental variables. (b) Allele frequencies for
1117 three contigs found as outliers in a genome scan and GEA test