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

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

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

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

**Corresponding author:* Jason Munshi-South

E-mail: jmunshisouth@fordham.edu

ABSTRACT

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

metabolic pathways. Our findings support the idea that cities represent novel ecosystems with a unique set of selective pressures.

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

48 **INTRODUCTION**

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

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

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

Peromyscus leucopus is one of the most abundant small mammals in North America, preferring the typical oak-hickory forest commonly found in the eastern USA (Wang *et al.* 2008). They are generalists that burrow in a variety of habitats (Metzger 1971; Vessey & Vessey 2007), and feed on a wide-range of invertebrates, nuts, fruit, vegetation, and fungus (Wolff *et al.* 1985; Ostfeld *et al.* 1996). They are especially reliant on oak mast cycles and an important predator of gypsy moths (Ostfeld *et al.* 1996). There is also evidence that *Peromyscus* spp. can

94 adapt to environmental change (Storz *et al.* 2007, 2009, 2010; Mullen & Hoekstra 2008; Linnen
95 *et al.* 2009; Weber *et al.* 2013; Natarajan *et al.* 2013; Munshi-south & Richardson 2016), making
96 them good candidates for the study of local adaptation. White-footed mice are one of the few
97 native mammals that thrive in extremely small, fragmented urban forests in North America
98 (Pergams & Lacy 2007; Rogic *et al.* 2013; Munshi-South & Nagy 2014). *P. leucopus* tend to be
99 found at higher densities in urban patches due to a thick understory and fewer predators and
100 competitors (Rytwinski & Fahrig 2007). Increased density may also be due to limited *P.*
101 *leucopus* dispersal between urban sites. Munshi-South (2012) found barriers to dispersal
102 between isolated NYC parks, with migrants only moving through significantly vegetated
103 corridors throughout the city. There is also substantial genetic structure between NYC parks as
104 measured by microsatellites (Munshi-South & Kharchenko 2010), genome-wide SNPs (Munshi-
105 South *et al.* 2016) and demographic modeling (Harris *et al.* 2016). We have also previously
106 identified signatures of selection in urban populations of NYC white-footed mice (Harris *et al.*
107 2013), though we used smaller datasets and more limited approaches than presented here. This
108 study builds on our previous work by employing a larger dataset and more comprehensive
109 statistical analyses to identify signatures of selection in *P. leucopus* populations while explicitly
110 using the inferred demographic history as a null model. We further confirm outlier genes by
111 associating allele frequencies with environmental metrics of urbanization and perform
112 enrichment analyses to predict functional relevance of outlier genes.
113 Urbanization and global climate change are relatively recent disturbances that rapidly
114 change native ecosystems. Over short timescales, adaptive evolution tends to act on standing
115 genetic variation as opposed to de novo mutations (Barrett & Schluter 2008; Stapley *et al.* 2010).
116 As these pre-existing mutations spread to fixation they produce a detectable signal in the form of

117 ‘hard’ or ‘soft’ selective sweeps (Hermisson & Pennings 2005; Messer & Petrov 2013).
118 Additionally, ecologically important traits involved in local adaptation are often quantitative
119 traits with many genes of small effect involved in producing the desired phenotype (Orr 2005;
120 Rockman 2012). To distinguish between these more subtle signatures of selection, we used
121 multiple tests that provide greater statistical power and higher resolution at identifying types and
122 age of selection when used together (Grossman *et al.* 2010; Hohenlohe *et al.* 2011).
123 We analyzed transcriptomes sequenced from urban and rural populations of *P. leucopus*
124 to produce estimates of nucleotide diversity (π, Tajima 1983), Tajima’s *D* (Tajima 1989), and
125 F_{ST} (Wright 1951) and make inferences about the evolutionary processes at work in these
126 populations. We also used a variety of tests to identify outlier genes subject to selection, and took
127 extra steps to account for the potentially confounding effects of demography. Specifically,
128 neutral demographic processes, like population bottlenecks, can produce signatures of variation
129 similar to those produced by selection (Oleksyk *et al.* 2010; Li *et al.* 2012). For example, both
130 selection and a population bottleneck followed by an expansion may produce genomic regions
131 with low genetic diversity, but recent literature discusses how to deal with these overlapping
132 signals (Excoffier *et al.* 2009; Li *et al.* 2012; Vitti *et al.* 2013; Lotterhos & Whitlock 2015). The
133 prevailing approach assumes selection acts on one or a few loci while demographic processes act
134 across the genome. Outlier tests for loci under selection typically generate a null distribution
135 based on an island model of population differentiation (Excoffier *et al.* 2009), and then identify
136 candidate genes with genetic differentiation that exceeds this simulated null distribution. The
137 true demographic history of most organisms is much more complex, and computational
138 approaches have been developed to robustly infer demographic parameters (Gutenkunst *et al.*
139 2009; Excoffier *et al.* 2013). This inferred demographic history can then be used to construct a

140 more realistic null model, reducing the rate of false positives in tests for selection (Excoffier *et*
141 *al.* 2009; Yoder *et al.* 2014).
142 We used the inferred demographic history of urban populations of *P. leucopus* (Harris *et*
143 *al.* 2016) to simulate comparable SNP datasets to our observed sequence data. We then used
144 multiple approaches that identify outlier loci based on population differentiation, the SFS, or
145 associations between allele frequencies and environmental variables. Bayescan uses a Bayesian
146 approach to identify SNPs that exhibit extreme allele frequency divergence between populations
147 (Foll & Gaggiotti 2008). SweeD is a likelihood based test that identifies selective sweeps based
148 on SFS that deviate from neutral expectations. We examined associations between allele
149 frequencies and environmental variation using a genotype-environment association (GEA) test.
150 GEA tests have been shown to perform better than outlier tests under complex demographic
151 scenarios (Lotterhos & Whitlock 2015) but can suffer from a high rate of false positives.
152 Analyses suggest that using genome scan-based outlier tests in conjunction with GEA tests leads
153 to reliable outlier identification (De Villemereuil *et al.* 2014). GEA also identifies local
154 adaptation in polygenic phenotypes where each polymorphism has a relatively weak effect
155 (Frichot *et al.* 2013), because correlations between alleles and environmental variables do not
156 rely on the strength of genetic differentiation or SFS skew between populations. (Pavlidis *et al.*
157 2013). Using multiple analyses with alternative statistical approaches is preferred for genome
158 scans, and provides more power and confidence in results when markers are repeatedly found as
159 outliers (Grossman *et al.* 2010). BayPass also uses a Bayesian approach to identify divergent
160 adaptive processes (Gautier 2015), but explicitly incorporates population demographic history
161 including hierarchical population structure. Conveniently, it also uses a population covariance
162 matrix to associate SNPs with population-specific environmental covariables. This feature

allowed us to use BayPass to identify congruence across outliers identified in Bayescan and LFMM.

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

186 **MATERIALS AND METHODS**

187 **Sampling, library preparation, and transcriptome assembly**

188 We trapped and collected white-footed mice from 2009 - 2013. For full details on
189 sampling and transcriptome sequencing, see Harris et al. (2015). In brief, we randomly chose
190 eight individual white-footed mice (equal numbers of males and females) from six sampling
191 locations representative of urban and rural habitats and with minimal within-site genetic structure
192 (Fig. 1) (Harris *et al.* 2013, 2015). Three sampling sites occurred within NYC parks: Central
193 Park in Manhattan (CP), New York Botanical Gardens in the Bronx (NYBG), and Flushing
194 Meadows—Willow Lake in Queens (FM). These sites represented urban habitats surrounded by
195 high levels of impervious surface cover and high human population density, as previously
196 quantified in Munshi-South et al. (2016). The remaining three sites occurred ~100 km outside of
197 NYC in rural, undisturbed habitat representative of natural environments for *Peromyscus*
198 *leucopus*. High Point State Park is in the Kittatinny Mountains in New Jersey (HIP), Clarence
199 Fahnestock State Park is located in the Hudson Highlands in New York (CFP), and Brookhaven
200 and Wilde Wood State Parks and neighboring sites occur on the northeastern end of Long Island,
201 New York (BHWWP). We sacrificed mice on site and liver, gonad, and brain tissue were
202 harvested in the field for immediate storage in RNAlater (Ambion). In the lab, we extracted total
203 RNA, removed ribosomal RNA, barcoded each tissue type, and then pooled samples during
204 library preparation. The reverse transcribed cDNA was sequenced using the 454 GS FLX+ and
205 SOLiD 5500 xl systems using standard RNAseq protocols. We called SNPs with the Genome
206 Analysis Toolkit pipeline using a Bayesian genotype likelihood model (GATK version 2.8,
207 DePristo *et al.* 2011) and removed related individuals. See Harris *et al.* 2013, 2015 for full
208 transcriptome sequencing, assembly and SNP calling details, but in short, for SNP calling we

209 required coverage >5X, nucleotide quality >30, no strand bias (FS >35), and SNPs from a
210 uniquely mapped read. We also removed SNPs where every individual was heterozygous, overall
211 depth >10, overall depth <350 and minor allele frequency (MAF) >0.025. The VCF file of SNP
212 genotypes used for demographic inference is on the Dryad digital repository at
213 http://dx.doi.org/10.5061/dryad.d48f9, raw sequencing files for the transcriptome are deposited
214 in the GenBank Sequence Read Archive (SRA Accession no. SRP020005), and transcriptome
215 contigs are available in the Dryad digital repository, doi: 10.5061/dryad.6hc0f.
216

217 **Summary statistics**

218 SNP information was stored in a VCF (variant call format) file and summary statistics
219 were calculated using vcftools 0.1.12b (Danecek *et al.* 2011). We calculated per-site nucleotide
220 diversity (π), Tajima's D, and F_{ST} for each site. We also calculated the statistics for each contig
221 (per-site statistic summed across all SNPs per contig divided by total sites) and calculated the
222 average estimate for each population, including all pairwise population comparisons for $F_{ST.}$
223

224 **Scans for positive selection based on population differentiation**

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

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

255 Finally, we allowed migration to occur between all three populations (Harris *et al.* 2016).
256 Bayescan was run independently on each simulated dataset using default parameters. Using the
257 observed SNP dataset, we performed a global analysis, one Bayescan run where all individuals
258 were partitioned into urban and rural groups, and finally analyses on all individual pairwise
259 population comparisons. Outlier SNPs were retained if they had a false discovery rate (FDR)
260 value ≤ 0.1 and if the posterior odds probability from Bayescan was higher than for any value
261 calculated from the simulated dataset. Outlier SNPs with a FDR ≤ 0.1 were considered
262 significant, implying that diversifying selection better explains allele frequency differences
263 between urban and rural populations (urban vs. rural) and sub-populations (pairwise population
264 comparison) than a neutral null model. A relatively high FDR was chosen for all analyses to
265 ensure inclusion of all putative outlier SNPs.
266 We reduced the risk of including false positives by also using the software program
267 BayPass (Gautier 2015) to identify putative SNPs showing evidence of divergent selection
268 between populations. We filtered our final outlier SNP list to only include those identified in
269 both Bayescan and BayPass. BayPass incorporates population demographic history when
270 identifying outlier SNPs (Gautier 2015) based on associations between allele frequencies and
271 environmental variables. We ran BayPass using default parameters under the AUX model (Table
272 S2). BayPass uses the XtX differentiation measure to identify differentiated SNPs. We created
273 an empirical distribution of XtX values for each locus by analyzing pseudo-observed data sets
274 (PODs) and chose a 5% threshold value for XtX to use as the cutoff value to differentiate
275 between selection and neutrality (Gautier 2015). PODs were also used to determine a 5%
276 threshold value for Bayes Factors used for associating environmental covariables with allele
277 frequencies.

278

279 **Analysis for selective sweeps**

280 We also identified outlier regions when the observed SFS showed an excess of low
281 frequency and high frequency minor alleles, a signal indicative of a recent selective sweep. The
282 composite likelihood ratio (CLR) statistic is used to identify regions where the observed SFS
283 matches the expected SFS generated from a selective sweep (Kim & Stephan 2002; Nielsen *et al.*
284 2005; Pavlidis *et al.* 2010). We calculated the CLR along sliding windows across the
285 transcriptome using the software program SweeD (Pavlidis *et al.* 2013). SweeD is an extension
286 of Sweepfinder (Nielsen *et al.* 2005) that is optimized for large next generation sequencing
287 (NGS) datasets. SweeD was run separately for each population and on individual contigs using
288 default parameters except for setting a sliding window size of 200 bp and using the folded SFS,
289 as we lacked an outgroup to infer the ancestral state. The window within each contig with the
290 highest CLR score is the likely location of a selective sweep. Similar to the method used for
291 Bayescan analyses, statistical significance was established from a null distribution generated by
292 running SweeD on SNP datasets simulated under the inferred demographic history for *P.*
293 *leucopus* populations (Harris *et al.* 2016). SweeD does not inherently identify outlier regions.
294 The CLR is computed using a selective sweep model on the observed data and then compared to
295 a neutral model calibrated with the background SFS generated from simulations. As before, we
296 used 100 datasets with 100,000 SNPs each, simulated under the inferred neutral demographic
297 history for urban and rural populations of white-footed mice in NYC. The CLR was calculated
298 using SweeD for all simulated datasets and the resulting distribution was used to set a
299 significance cutoff. For the observed dataset, we lacked a genome to provide clear linkage
300 information so SweeD was run separately on each contig. We identified outlier contigs if their

301 CLR value was greater than any produced when calculated for neutral simulations. We also
302 required outliers to fall within the top 0.01% of the CLR distribution for the observed SNPs.
303
304 **Genotype-environment association tests for environmental selection**
305 We used the GEA approach of LFMM: Latent Factor Mixed Models (Frichot *et al.* 2013)
306 to associate outlier SNPs and candidate loci identified above with potential environmental
307 selection pressures. LFMM examines associations between environmental and genetic variation
308 while accounting for the neutral genetic background and structure between populations (Frichot
309 *et al.* 2013). We tested three environmental variables associated with urbanization: 1) percent
310 impervious surface within a 2 km buffer around each sampling site, 2) human density within a
311 two-kilometer buffer around each sampling site, and 3) designating each site as urban or rural.
312 We previously found that variables 1-2 are significantly associated with genome-wide variation
313 in *P. leucopus* populations in the NYC metropolitan area (Munshi-South *et al*. 2016). Our final
314 data set included all individuals but only the subset of outlier SNPs that were detected in
315 Bayescan and SweeD. LFMM requires the user to define the number of latent factors, K, that
316 describe population structure in the dataset. To identify the appropriate number of K latent
317 factors, we performed a genetic PCA followed by a Tracy-Widom test to find the number of
318 eigenvalues with *P* values ≤ 0.01 (Patterson *et al.* 2006; Frichot & François 2015). Based on this
319 approach, we ran LFMM with default parameters except for K = 6, number of MCMC cycles =
320 100,000, and burn-in = 50,000. Using author recommendations, we combined 10 replicate runs
321 and readjusted the p values to increase the power of the test. LFMM uses |z|- scores to report the
322 probability of a SNP's association with an environmental variable. After correcting for multiple
323 testing, we used a cutoff value of $q \leq 0.1$.

324 Similar to the approach described above, we increased statistical power by repeating the
325 GEA test in a separate analysis. We used the auxiliary variable model in the program BayPass
326 (Gautier 2015) to identify associations between allele frequencies and environmental variables.
327 We filtered our final list of markers to only include those identified in both LFMM and BayPass.
328 PODs were also used to determine a 5% threshold value for Bayes Factors used for associating
329 environmental covariables with allele frequencies.
330
331 **Functional annotation of candidate gene**
332 We used the gene annotation pipeline in Blast2GO (Conesa *et al.* 2005; Götz *et al.* 2008)
333 to find sequences from the NCBI non-redundant protein database that were homologous to our
334 outlier contigs identified above. We then retrieved associated gene ontology (GO) terms.
335 Blast2GO retrieves GO terms associated with BLASTX hits and uses the KEGG database to
336 describe biochemical pathways linking different enzymes (Ogata *et al.* 1999; Kanehisa *et al.*
337 2014). For downstream enrichment analyses, we also used the Ensembl gene annotation system
338 (Aken *et al.* 2016) to find homologous *Mus musculus* genes for each *P. leucopus* contig (Table
339 S3). We further interpreted the outlier gene lists using g:Profiler (Reimand *et al.* 2016) to
340 identify gene ontology terms enriched in our outlier gene list compared to the fully annotated
341 *Mus musculus* genome. Poorly updated gene annotation databases can significantly affect results
342 and g:Profiler is one of the most comprehensive and most often updated gene annotation
343 databases available (Wadi *et al.* 2016). We used the g:Profiler webserver and identified enriched
344 terms from the full outlier gene list using default parameters displaying only significant results
345 (Table S3). We visualized and summarized the enriched gene ontology list using the revigo
346 webserver (Supek *et al.* 2011).

347

348 **RESULTS**

349 **Genetic diversity statistics**

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

363

364 **Outlier detection**

365 The global Bayescan analysis identified 309 SNPs potentially under the influence of
366 divergent selection. After sampling sites were grouped as urban or rural, Bayescan identified 40
367 SNPs with signatures of positive selection (Fig. 2A, Table 2). Eight of these SNPs were also
368 found in the global analysis. Individual urban to rural population comparisons did not find any
369 outlier SNPs, and zero SNPs exhibited signatures of balancing selection. F_{ST} for outlier SNPs

370 ranged from 0.21 - 0.33, much higher than the population median of 0.059. Bayescan identified
371 zero outlier SNPs in the simulated neutral dataset. However, we only included outlier SNPs
372 from the observed dataset with FDR and posterior odds values that were smaller and larger,
373 respectively, than the most extreme values for the simulated data (FDR ≤ 0.6 and $\log_{10}$(PO) ≥-
374 0.196).

375 To generate the null distribution of the CLR statistic for analyses in SweeD, we tested the
376 100 SNP datasets simulated under the inferred demographic history for NYC populations of *P.*
377 *leucopus*. We found that CLR scores in the top 5% of the simulated distribution were generally
378 2-3X lower than values in the top 5% of the observed dataset. We ran SweeD on observed SNPs
379 within individual contigs and identified outliers by filtering for a CLR score ≥ 3.53 (the
380 maximum CLR from simulated data). We also chose regions that fell within the top 0.01% of
381 the observed distribution (Fig. 2B). SweeD identified regions with SFS patterns that fit a
382 selective sweep model in 55 contigs within urban populations (Table 3). Contig 35790-44,
383 annotated as the lipid transporter *Apolipoprotein B100*, had the highest CLR (8.56). All outliers
384 had CLR scores ≥ 4.97. Bayescan and SweeD did not identify any of the same outliers.

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

389

390 **Environmental associations**

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

393 30 of the identified SNPs were associated with the binary classification of urban vs. rural. Only
394 seven of the outlier SNPs were associated with percent impervious surface and five were
395 associated with human population density. Twenty-six of the 55 outlier contigs identified using
396 SweeD were associated with one of the environmental variables (Table 3). Again, all 26 regions
397 were associated with the urban vs. rural site classification. Fourteen outliers from SweeD were
398 associated with percent impervious surface and eight were associated with human population
399 density. Some contigs associated with environmental variables were outliers in only one urban
400 population, possibly indicating local adaptation within parks, selection on a polygenic trait, or
401 genetic drift.
402 The only environmental variable significantly associated with SNPs in the BayPass
403 analysis was urban or rural classification. Percent human density and percent impervious surface
404 cover did not show significant associations. All outliers identified in Bayescan, BayPass,
405 SweeD, and LFMM showed associations with urban versus rural classification (5% threshold
406 value, BF ≥ 17.8, Table S2).
407
408
409 **Functional annotation**
410 The full contig sequences containing outlier SNPs were obtained from the *P. leucopus*
411 transcriptome (Harris *et al.* 2015) and used to identify functional annotations. Of the 40 contigs
412 identified by Bayescan as divergent between urban and rural populations, 36 were annotated with
413 gene names and functional information (Table 2). Of these, 29 were also associated with urban
414 environmental variables. The ten most frequent GO terms among the Bayescan outliers involved
415 organismal metabolism (Table S1). Some outliers occurred within sequences homologous with

416 genes of known functions and biochemical pathways. These outliers included a farnesoid-x-
417 receptor (FXR, Contig 25795-154), a myosin light chain kinase (MYLK, Contig 7975-418), and
418 the gene SORBS2 (Contig 37967-26).

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

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

439

440 **DISCUSSION**

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

457 In this study, we observed patterns of divergent positive selection between urban and
458 rural populations of *P. leucopus*, and were able to associate outlier SNPs with environmental
459 variables relevant to urbanization. The majority of candidate loci were annotated with GO terms
460 that are significantly associated with organismal metabolism, particularly breakdown of lipids
461 and carbohydrates. We discuss what these findings mean for organisms inhabiting novel urban

462 ecosystems, and more generally for understanding the ecological processes and time frame of
463 local adaptation in changing environments.

464

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

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

474 While Bayescan has been shown to be relatively robust to confounding demographic
475 processes (Pérez-Figueroa *et al.* 2010; De Villemereuil *et al.* 2014), population bottlenecks,
476 hierarchical structure, recent migration, or variable times to most-recent-common-ancestor
477 (MRCA) between populations can artificially inflate F_{ST} values (Hermisson 2009; Lotterhos &
478 Whitlock 2014). We minimized false positives by incorporating population structure and a
479 specific demographic history for *P. leucopus* in NYC directly into the null distribution of F_{ST}.
480 (Harris *et al.* 2016). We only included outliers if their posterior probability was greater than
481 probabilities calculated from simulations. The outliers comprised 0.024% of the total number of
482 loci analyzed from the transcriptome. This percentage is in line with candidates uncovered from
483 a similar study (0.05%) that looked at high and low altitude populations of the plant *S.*
484 *chrysanthemifolius* (Chapman *et al.* 2013). Many studies find higher percentages of outlier loci

485 using Bayescan; for example, 4.5% in the American pika across its range in British Colombia
486 (Henry & Russello 2013), and 5.7% in Atlantic herring across their range (Limborg *et al.* 2012).
487 Our lower overall percentage of outliers may be due to the use of the inferred demographic
488 history to establish outlier cutoffs and reduce false positives, or because of the relatively recent
489 isolation or strength of selection in urban populations.

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

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

504 Several studies have shown that performing multiple tests that employ diverse theoretical
505 approaches is the best way to avoid Type I and II errors in genome outlier analyses (Nielsen
506 2005; Grossman *et al.* 2010; Hohenlohe *et al.* 2010b). We used Bayescan and SweeD to identify
507 signatures of positive selection, and confirmed outliers using BayPass to identify divergent

508 selection while incorporating genetic structure. While BayPass confirmed the majority of
509 outliers identified using other methods (Table S2), there was no overlap between Bayescan and
510 SweeD outliers. This discrepancy is likely due to the different selection scenarios underlying
511 each test, i.e. divergent local selection versus population-wide positive selection in the form of
512 selective sweeps (Hermisson 2009). F_{ST} based methods can respond to allelic divergence
513 relatively quickly, while models for selective sweeps typically require nearly-fixed derived
514 alleles (Hohenlohe *et al.* 2010b). Given the recency of urbanization in NYC, many selective
515 sweeps may be ongoing or otherwise incomplete. Selection may also be acting on standing
516 genetic variation in the form of soft sweeps (Hermisson & Pennings 2005) that are not readily
517 identified by SweeD (De Villemereuil *et al.* 2014). We identified several outliers that were
518 unique to specific urban populations, which is characteristic of soft sweeps and polygenic traits
519 (Messer & Petrov 2013). Despite the lack of overlapping outliers between the two tests, further
520 confirmation of outlier genes experiencing positive selection was provided by genotype-
521 environment association tests. These methods may often be more powerful than the genome
522 scans above (Savolainen *et al.* 2013).
523
524 **Environmental associations strengthen evidence of local adaptation to urbanization**
525 GEA tests are a growing class of methods that identify loci with allele frequencies that
526 are associated with environmental factors (Joost *et al.* 2007; Coop *et al.* 2010; Frichot *et al.*
527 2013). Here we used LFMM (Frichot *et al.* 2013) to associate outlier SNPs with environmental
528 metrics of urbanization. LFMM performs better than other methods in the presence of
529 hierarchical structure and when polygenic selection is acting on many loci with small effect (De
530 Villemereuil *et al.* 2014). Hierarchical structure in our dataset includes urban and rural

531 differentiation (Harris *et al.* 2015; Harris *et al.* 2016), patterns of geographic structure between
532 mainland mice and Long Island, NY (Harris *et al.* 2016), and population structure between
533 individual urban parks (Munshi-South & Kharchenko 2010). Simulations also suggest that
534 LFMM is superior when sample size is less than 10 individuals per population, there is no
535 pattern of IBD, and the study compares environmentally divergent habitats (Lotterhos &
536 Whitlock 2015). We sampled eight white-footed mice per population, found no evidence of IBD
537 (Munshi-South *et al.* 2016), and sampled environmentally divergent rural and urban locations.
538 Using LFMM, we found that 75 % and 47 % of outliers from Bayescan and SweeD,
539 respectively, were significantly associated with one or more urbanization variables. BayPass also
540 confirmed associations between all outlier SNPs and urbanization variables, though only with the
541 binary classification of a site as urban or rural. These results are consistent with other studies
542 combining genome scan methods and GEA tests. Limborg *et al.* (2012) found 62.5% of the
543 outliers identified in Bayescan were correlated with temperature or salinity in Atlantic herring,
544 and 26.3% of genome scan outliers were associated with temperature or latitude in a tree species
545 (De Kort *et al.* 2014). We acknowledge that percent impervious surface, human population
546 density, or binary classification as urban versus rural may not capture the specific, causative
547 selection pressures acting on white-footed mouse populations. We used these metrics as general
548 proxies for ecological processes that in urbanized habitats. The percent of impervious surface
549 around a park is likely representative of habitat fragmentation, as urban infrastructure changes
550 the net primary productivity due to increasing percentages of impervious surface or artificial
551 landscapes, parks and yards (Shochat *et al.* 2006). This fragmentation then leads to changing
552 species interactions as migration is impeded or organisms are forced into smaller areas (Shochat
553 *et al.* 2006). The percent human density surrounding an urban park can serve as a proxy for the

554 multitude of ecological changes humans impose on their surrounding environment. Humans
555 often introduce invasive species into cities (Sih *et al.* 2011), leading to increased competition or
556 novel predator-prey interactions. Urbanization and increasing human density also change the
557 types and availability of resources in the altered habitat (McKinney 2002; Sih *et al.* 2011).
558 Finally, classifying our sites as urban or rural can generally capture the main differences in urban
559 and natural sites. For example, pollution is a major consequence of urbanization (Donihue &
560 Lambert 2014), and urban areas often include increased chemical, noise, or light pollution (Sih *et*
561 *al.* 2011).

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

577

578 **Functional roles and ecological relevance of candidate genes**

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

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

600 Urban *P. leucopus* may experience different energy budgets, physiological stressors or
601 diets compared to rural counterparts. The signatures of selection reported for certain genes here
602 support this scenario, such as heat-shock protein Hsp90. Heat shock proteins have repeatedly
603 been found to play a pivotal role in adaptation to environmental stress (Limborg *et al.* 2012).
604 Hsp90 was significantly enriched for 12 GO terms from the g:Profiler analysis with the majority
605 associated with protein metabolism. In *Peromyscus* spp., Hsp90 is a chaperone for many
606 proteins, including a suite of metabolizing receptors activated by dioxin-like industrial toxins
607 often found in polluted soil samples (Settachan 2001). Another outlier from our analyses,
608 aflatoxin aldehyde reductase (AKR7), was also significantly enriched for 8 GO terms primarily
609 involved with single organism metabolism and is important for metabolizing environmental
610 toxins (Hyndman *et al.* 2003).Urban soils are often much more contaminated with toxins than
611 soils in adjacent rural areas (McDonnell *et al.* 1997).
612 We found a surprising number of candidate genes with functions related to the
613 metabolism and transport of lipids and carbohydrates. These genes were strongly correlated with
614 environmental measures of urbanization, with clearly divergent allele frequencies between urban
615 and rural sites (Fig. 3B). APOB-100 is the primary apolipoprotein that binds and transports
616 lipids, including both forms of cholesterol (HDL and LDL). The outlier gene, APOB-100, was
617 significantly enriched for 9 GO terms with the primary cluster involved in single-organism
618 metabolism, or anabolic / catabolic processes involving one organism and abiotic stimuli.
619 FADS1, a farnesoid-x-receptor, is a nuclear receptor antagonist that is involved in bile synthesis
620 and modulates high fat diets, with variation in expression affecting rates of obesity in mice (Li *et*
621 *al.* 2013). FADS1 was enriched for 23 GO terms including five for lipid metabolism and
622 regulation of lipid biosynthesis. Manually curated protein annotations show MYLK (10

623 significantly enriched GO terms; Metabolism) and SORBS2 (2 significantly enriched GO terms;
624 Cellular processes) are both directly involved in the gastrointestinal system, including smooth
625 muscle contractions and absorption of water and sodium in the intestine, respectively (Magrane
626 & Consortium 2011; Consortium 2014). Finally, KEGG analysis identified two contigs (10636-
627 348: 8 enriched GO terms and 27546-129: 22 enriched GO terms) that represent proteins that are
628 both directly involved in galactose (primarily found in dairy products), fructose and mannose
629 (both naturally found in fruits, seeds, and vegetables) metabolism (Ogata *et al.* 1999).

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

646 abundance of seeds, invertebrates, or direct human subsidies. Local adaptation in urban
647 populations may allow these mice to more efficiently metabolize different types or amounts of
648 lipids and carbohydrates.
649

650 **ACKNOWLEDGMENTS**

651 We thank Mike Hickerson for his helpful comments and advice on many analyses and for access
652 to lab space for analyses and writing. We thank Diego Alvarado-Serrano, Alexander T. Xue,
653 Tyler Joseph, and Champak Reddy for their invaluable comments and advice concerning
654 bioinformatics and demographic analyses. This research was supported by the National Institute
655 of General Medical Sciences of the National Institutes of Health under award number
656 R15GM099055 to JM-S and a NSF Graduate Research Fellowship to SEH. The content is solely
657 the responsibility of the authors and does not represent the official views of the National
658 Institutes of Health.
659

660 **REFERENCES**

661 Aken BL, Ayling S, Barrell D *et al.* (2016) The Ensembl gene annotation system. *Database*,
662 **2016**, baw093.
663 Allombert S, Stockton S, Martin JL (2005) A natural experiment on the impact of overabundant
664 deer on forest invertebrates. *Conservation Biology*, **19**, 1917–1929.
665 Balloux F (2010) The worm in the fruit of the mitochondrial DNA tree. *Heredity*, **104**, 419–420.
666 Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level.
667 *Nature Reviews Genetics*, **12**, 767–780.
668 Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology*

669 *& Evolution*, **23**, 38–44.

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

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

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

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

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

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

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

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

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

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

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

693 Ellison CE, Hall C, Kowbel D *et al.* (2011) Population genomics and local adaptation in wild
694 isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences*,
695 **108**, 2831–2836.

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

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

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

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

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

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

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

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

714 Gautier M (2015) Genome-wide scan for adaptive divergence and association with population-

specific covariates. *Genetics*, **201**, 1555–1579.

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

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

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

Harris SE, Munshi-South J, Obergfell C, O'Neill R (2013) Signatures of Rapid Evolution in Urban and Rural Transcriptomes of White-Footed Mice (Peromyscus leucopus) in the New York Metropolitan Area. *PLoS ONE*, **8**, e74938.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

761 Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the
762 selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
763 Li YF, Costello JC, Holloway AK, Hahn MW (2008) "Reverse ecology" and the power of
764 population genomics. *Evolution*, **62**, 2984–2994.
765 Li F, Jiang C, Krausz KW *et al.* (2013) Microbiome remodelling leads to inhibition of intestinal
766 farnesoid X receptor signalling and decreased obesity. *Nature communications*, **4**, 2384.
767 Li J, Li H, Jakobsson M *et al.* (2012) Joint analysis of demography and selection in population
768 genetics: where do we stand and where could we go? *Molecular Ecology*, **28**, 28–44.
769 Limborg MT, Helyar SJ, De Bruyn M *et al.* (2012) Environmental selection on transcriptome-
770 derived SNPs in a high gene flow marine fish, the Atlantic herring (Clupea harengus).
771 *Molecular ecology*, **21**, 3686–703.
772 Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an
773 adaptive allele in deer mice. *Science*, **325**, 1095–8.
774 Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral
775 parameterization on the performance of F ST outlier tests. *Molecular Ecology*, **23**, 2178–
776 2192.
777 Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local
778 adaptation depends on sampling design and statistical method. *Molecular Ecology*, **24**,
779 1031–1046.
780 Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data.
781 *Database : the journal of biological databases and curation*.
782 McDonnell M, McDonnell M, Pickett S *et al.* (1997) Ecosystem processes along an urban to
783 rural gradient. *Urban Ecosystems*, **1**, 21–36.

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

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

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

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

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

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

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

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

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

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

807 *PeerJ*, **2**, e310.

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

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

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

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

819 Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–
820 218.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rockman M V (2012) The QTN program and the alleles that matter for evolution: all that's gold

does not glitter. *Evolution*, **66**, 1–17.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

899 *Ecology*, **23**, 2006–2019.

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

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

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

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

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

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

913 Whitehead A, Triant D, Champlin D, Nacci D (2010) Comparative transcriptomics implicates
914 mechanisms of evolved pollution tolerance in a killifish population. *Molecular Ecology*, **19**,
915 5186–5203.

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

918 De Wit P, Palumbi SR (2013) Transcriptome-wide polymorphisms of red abalone (Haliotis
919 rufescens) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, **22**, 2884–
920 97.

921 De Wit P, Pespeni MH, Palumbi SR (2015) SNP genotyping and population genomics from

922 expressed sequences -current advances and future possibilities. *Molecular Ecology*, **24**.
923 Wolff JO, Dueser RD, Berry K (1985) Food Habits of Sympatric Peromyscus leucopus and
924 Peromyscus maniculatus. *Journal of Mammalogy*, **66**, 795–798.
925 Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, 323–354.
926 Yeh PJ (2004) Rapid evolution of a sexually selected trait following population establishment in
927 a novel habitat. *Evolution*, **58**, 166–174.
928 Yoder JB, Stanton-Geddes J, Zhou P *et al.* (2014) Genomic signature of adaptation to climate in
929 Medicago truncatula. *Genetics*, **196**, 1263–1275.
930 Zhou L, Bawa R, Holliday J a. (2014) Exome resequencing reveals signatures of demographic
931 and adaptive processes across the genome and range of black cottonwood (Populus
932 trichocarpa). *Molecular Ecology*, **23**, 2486–2499.
933
934
935
936

937 **FIGURES AND TABLES**

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

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

940
941
942
943
944
945
946
947
948
949
950
951

952 **Table 2.** Outlier loci (*N* = 33) in the urban to rural comparison identified using Bayescan and

953 confirmed with BayPass. Last three columns indicate whether the locus was also significantly

954 associated with environmental variables across urban (CP, FM, NYBG) and rural (BHwwp, CFP,

955 HIP) populations in the LFMM analysis. I = percent impervious surface, D = human density, C

956 = Urban or Rural Classification

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

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

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

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

966

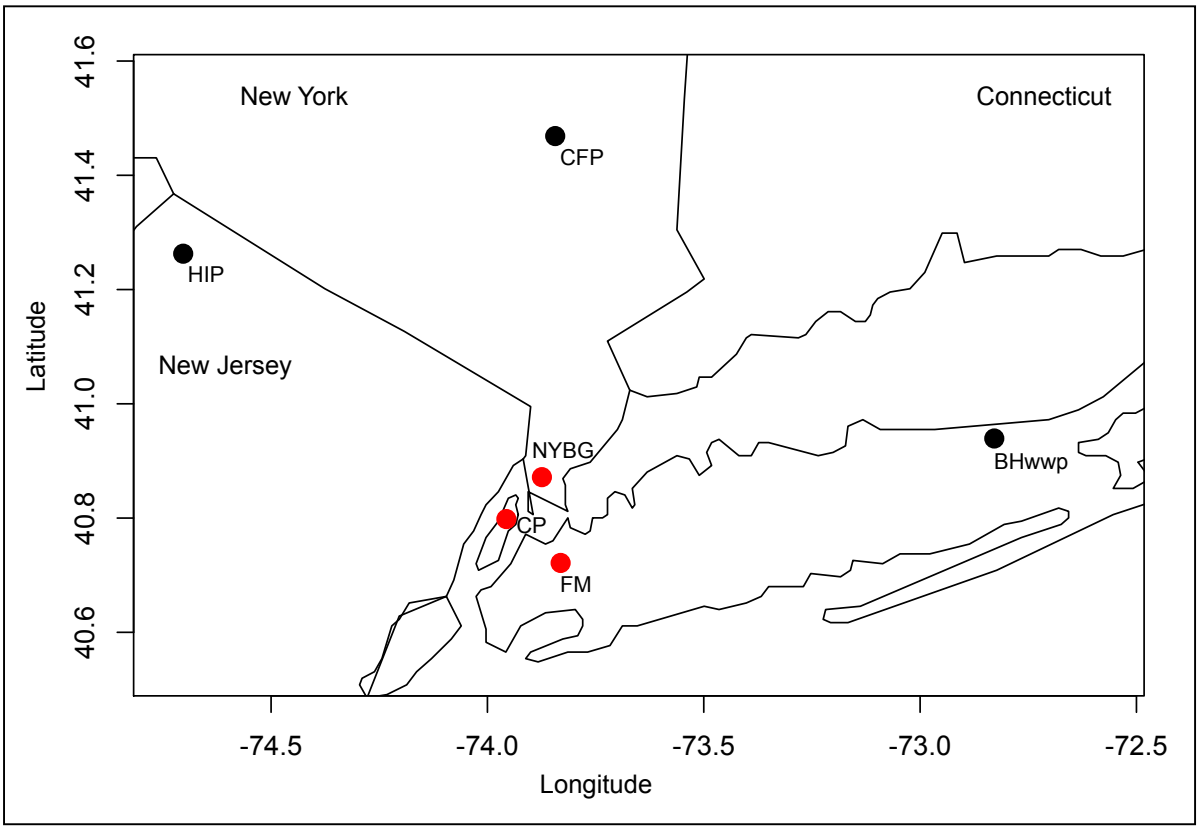

Figure 1. Map of sample localities in the NYC metropolitan area. Sites in red are urban parks within New York City. CP = Central Park; FM = Flushing Meadows—Willow Lake; NYBG = New York Botanical Gardens; BHwwp = Brookhaven and Wildwood State Park; CFP = Clarence Fahnestock State Park; HIP = High Point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

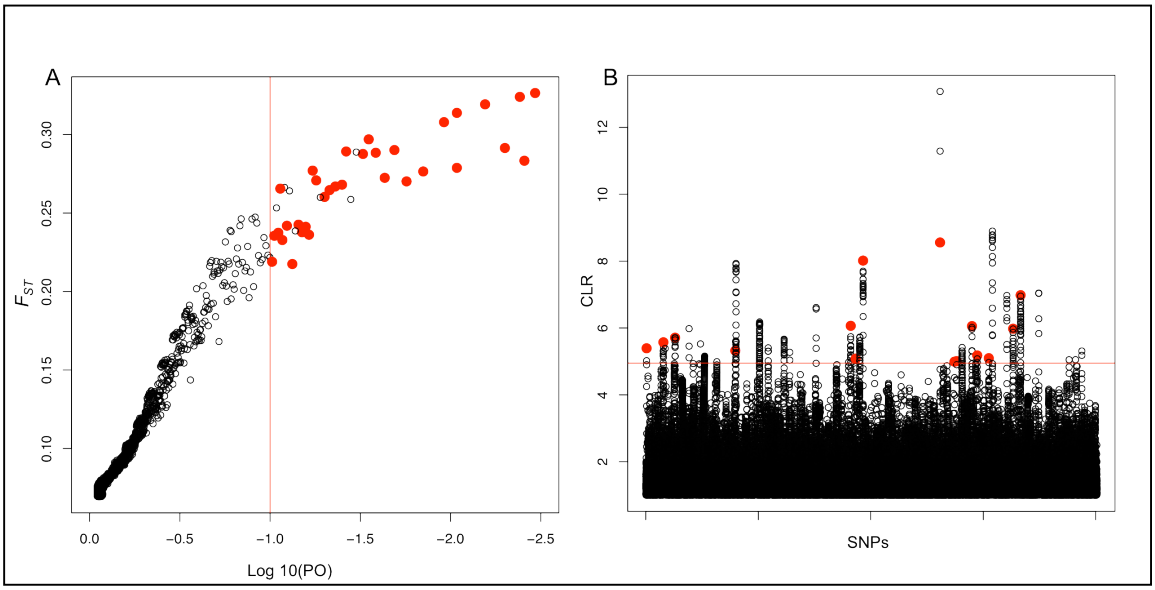

Figure 2. (a) BayeScan 2.1 plot of 154,770 SNPs genome scan analysis between urban and rural populations, including 48 individual white-footed mice from six NYC sampling sites. F_{ST} is on the vertical axis plotted against the $\log_{10}$ of the posterior odds (PO). The vertical red line indicates the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

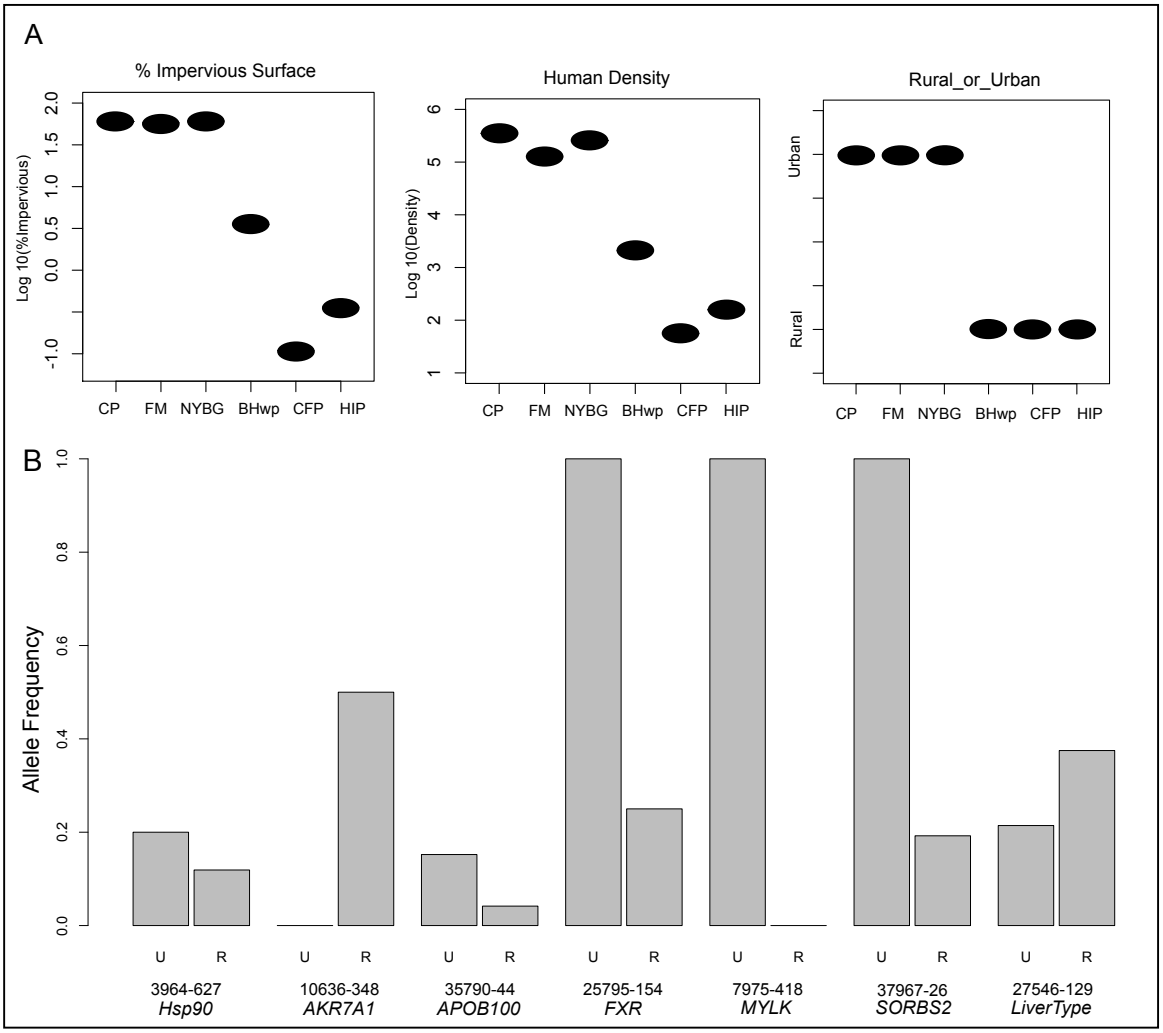

Figure 3. (a) Plot of urbanization metrics for all 6 sampling sites from NYC used in this study. The log10 value of % Impervious Surface and Human Density are plotted along the vertical axis and the oval represents the value for each sampling site. Ovals on the Rural or Urban plot show sample sites designated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

1011

1012 **SUPPORTING INFORMATION**

1013 **Figure S1.** KEGG analysis for biochemical pathways that contain multiple outlier contigs.
1014 Colored boxes represent outlier genes.

1015 **Table S1.** Blast2GO table with BLASTX hits from *M. musculus*, *R. rattus*, and *C griseus* and top
1016 three supported Gene Ontology terms for outlier genes from Bayescan and SweeD

1017 **Table S2.** Excel file containing Bayescan and SweeD outliers and the corresponding BayPass
1018 results. Full BayPass results are also included.

1019 **Table S3.** Excel file containing filtered list of outlier contigs, the homologous *Mus musculus*
1020 genes, and the significantly enriched GO terms from g:Profiler.

1021 **Table S4.** Excel file containing Revigo results. Enriched GO terms from g:Profiler are sorted
1022 into largest parent terms and listed based on the frequency of occurrence.

1023 **Table S5.** Average pairwise F_{ST} among six *P. leucopus* populations.

1024 **Table S6.** Top Blast hits including NCBI accession numbers for outlier contigs listed in Table 2
1025 and Table 3.