bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Population genomics of the Anthropocene: urbanization reduces the evolutionary potential of small mammal populations

Jason Munshi-South[1,*], Christine P. Zolnik[1], & Stephen E. Harris[2]

[1]Department of Biological Sciences and the Louis Calder Center—Biological Field Station, Fordham University, 31 Whippoorwill Road, Armonk, NY 10504

[2]Ph.D. Program in Ecology, Evolutionary Biology, and Behavior, the Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016

*corresponding author: jmunshisouth@fordham.edu

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Urbanization results in pervasive habitat fragmentation and reduces standing genetic variation through genetic drift. Loss of genome-wide variation may ultimately reduce the evolutionary potential of animal populations experiencing rapidly changing conditions. In this study, we examined genome-wide variation among 23 white-footed mouse (*Peromyscus leucopus*) populations sampled along an urbanization gradient in the New York City metropolitan area. Genome-wide variation was estimated as a proxy for evolutionary potential using more than 10,000 SNP markers generated by ddRAD-Seq. We found that genome-wide variation is inversely related to urbanization as measured by percent impervious surface cover, and to a lesser extent, human population density. We also report that urbanization results in enhanced genome-wide differentiation between populations in cities. There was no pattern of isolation by distance among these populations, but an isolation by resistance model based on impervious surface significantly explained patterns of genetic differentiation. Isolation by environment modeling also indicated that urban populations deviate much more strongly from global allele frequencies than suburban or rural populations. This study is the first to examine evolutionary potential along an urban-to-rural gradient and quantify urbanization as a driver of population genomics patterns.

Keywords

landscape genetics, Peromyscus, population genomics, ddRADSeq, rapid evolution, SNP, urbanization

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Humans exert an outsized influence on global ecosystem processes (Vitousek et al. 1997) in the Anthropocene, an era that began sometime between the late Pleistocene (Ellis et al. 2013; Ruddiman et al. 2015) and industrialization in the last few centuries (Steffen et al. 2007). Rapid urbanization is a key characteristic of the contemporary Anthropocene. The proportion of the human population in cities increased from 16% to 50% in the last century, and is projected to reach 70% by 2050 (Heilig 2011). Urban land conversion may occur at an even faster rate than population growth, thus resulting in accelerating encroachment of cities on reservoirs of biodiversity around the world (Seto et al. 2012). Urban areas become ecologically homogeneous (Groffman et al. 2014) due to loss of biodiversity that in turn enhances the probability of abrupt state shifts (Barnosky et al. 2012).

Most species are 'urban avoiders' that do not persist after urbanization, but 'urban adapters' and 'exploiters' are facultative or obligate users, respectively, of human resource subsidies (Blair 2001; McKinney 2002). Although Blair's (2001) conception of urban adapters did not explicitly include adaptive evolution after urbanization, local adaptation may have enhanced the ability of many urban adapters to exploit human subsidies. Urban habitats are 'novel ecosystems' composed of unique ecological communities and processes (Hobbs et al. 2006; Hobbs et al. 2009), and thus many species likely faced new, strong selection pressures after urbanization. Humans have undoubtedly altered the evolutionary trajectory of crop, pest, and disease species (Palumbi 2001), but conclusive cases of human-driven evolution in organisms that are not human commensals or pathogens have been difficult to identify (Merilä and Hendry 2014) outside of a few recent examples (Donihue and Lambert 2014).

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Urbanization results in severe habitat fragmentation (Zipperer et al. 2012), a process that is well documented to reduce genetic variation among animal populations (Keyghobadi 2007; Rivera-Ortíz et al. 2014). Loss of genetic variation may in turn reduce the evolutionary potential of populations experiencing rapidly changing ecological conditions (Etterson and Shaw 2001; Hoffmann and Sgrò 2011; Oakley 2013). Definitive evidence of adaptive evolution from standing variation in wild populations is still relatively scarce (Colosimo et al. 2005; Renaut et al. 2011; Domingues et al. 2012), but new user-friendly approaches to generating large population genomic datasets have improved prospects for quantifying and analyzing genetic variation in the wild (Narum et al. 2013). Predicting the loci involved in adaptive responses to urbanization will be difficult in most cases, but estimating genome-wide variation as a general proxy for evolutionary potential is a good substitute (Harrisson et al. 2014).

Evolution can proceed at the nucleotide level through new mutations or changes in frequency of standing genetic variants (Orr 2005), but it is likely that standing variation is more important in cases of rapid evolution (Barrett and Schluter 2008). Results from laboratory experiments (Teotónio et al. 2009; Burke et al. 2010), ancestral variation in model organisms (Rockman 2008; Scarcelli and Kover 2009), humans (J. K. Pritchard et al. 2010), artificial selection in crops (Gibson and Dworkin 2004), viral pathogens (Pennings 2012), and invasive species (Prentis et al. 2008) all support this association.

In this study, we test the following interrelated predictions about the population genomics of white-footed mice (*Peromyscus leucopus* Rafinesque) in the New York City (NYC) metropolitan area:

1) Evolutionary potential as measured by genome-wide variation within populations is inversely related to urbanization of the surrounding landscape.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2) Increasing urbanization of the landscape results in greater genomic structure and differentiation between populations.
3) "Ecological distances" resulting from urbanization are better predictors of genetic differentiation than geographic distances between populations.

Several studies have reported that urbanization results in genetic differentiation, reduced gene flow, and/or loss of genetic variation among fragmented populations (Delaney et al. 2010; Noël and Lapointe 2010; Bjorklund et al. 2010; Chiappero et al. 2011; Shalene Jha and Kremen 2013; Gortat et al. 2014). These studies primarily used microsatellites to estimate population structure and gene flow. However, microsatellites exhibit complex mutational patterns (Selkoe and Toonen 2006) and may not accurately measure genome-wide variation (Chapman and Sheldon 2011). Mitochondrial DNA (mtDNA) sequences are also often used to examine the population genetics of fragmentation, but pervasive selection on the mitochondrial genome make mtDNA markers even less suitable for estimating genome-wide variation (Bazin et al. 2006; Nabholz et al. 2008). Restriction-site associated DNA sequencing (RADseq) has emerged as a powerful alternative to microsatellites or mtDNA because RADseq can generate thousands of single nucleotide polymorphism (SNP) markers relatively quickly and cheaply, even for nonmodel organisms without reference genomes (Davey et al. 2011). In this study, we use a double-digest RADseq protocol (Peterson et al. 2012) to generate over 10,000 SNPs for analyzing the population genomics of white-footed mice.

Our previous microsatellite-based analyses showed substantial genetic structure between white-footed mouse populations in NYC's forest fragments (Munshi-South and Kharchenko 2010). Park area, age, or the extent of habitat within parks did not explain levels of genetic

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variation among these populations (Munshi-South and Nagy 2014), but these studies focused exclusively on forest fragments within NYC that were highly isolated by surrounding urbanization. In a separate analysis we identified SNPs from transcriptomes sequenced from urban and rural populations, and found that population structure was greater among the urban than the rural populations (Harris et al. 2015). This analysis was limited to only six sampling sites, however, and SNPs from coding regions may often deviate from neutral expectations. Here we expand these investigations using large, neutral SNP datasets, and include populations sampled along an urban-to-rural gradient spanning 142 km from the urban core to extensive, rural protected areas.

Land use transformation and anthropogenic barriers may ultimately reduce gene flow between populations (Epps et al. 2007; Balkenhol and Waits 2009; S. Jha 2015), leading to greater genetic structuring in urbanized areas. “Isolation-by-environment” (IBE) processes, or ecological distances, influence gene flow among populations in varied landscapes (Sexton et al. 2014). IBE is empirically detected as positive correlations between ecological and genetic distances that are independent of the effects of geographic distance (Wang and Bradburd 2014). Simulation studies predict that landscape genetic analyses will be most successful when applied to simple landscapes comprised of elements that differ strongly in their ability to impede migration (Jaquiéry et al. 2011). Urbanization produces landscapes with these characteristics, resulting in substantial IBE between otherwise proximal populations due to the relative difficulty of dispersing through the urban matrix.

We previously reported that gene flow between populations of white-footed mice could be largely explained by patterns of urban vegetation cover in NYC (Munshi□South 2012). Here we investigate the association between urbanization and connectivity in a much broader range of

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

130 landscape conditions varying from urban to suburban to relatively undisturbed areas. In our
131 earlier study we used partial Mantel tests and a causal modeling approach to identify ecological
132 distances between populations (based on vegetation cover) that best explained gene flow after
133 factoring out isolation-by-distance, IBD (Cushman and Landguth 2010). While this approach can
134 be successful for landscapes composed of high-contrast cover types (such as NYC), several
135 authors have argued that partial Mantel tests have low statistical power and are prone to false
136 positives (Legendre and Fortin 2010; Graves et al. 2013). Here we use a partial Mantel test as
137 well as a new statistical modeling approach to quantify the relative contributions of urbanization
138 and IBD to genetic differentiation between white-footed mouse populations (Bradburd et al.
139 2013). Specifically, we model the strength of covariance in allele frequencies between
140 populations as a function of IBD and ecological distances due to urbanization of the landscape.
141 This study is the first to examine genome-wide variation along an urban-to-rural gradient and
142 robustly quantify urbanization as a driver of population genomic patterns.
143
144
145 **METHODS**
146 **Study species and sampling sites**
147 White-footed mice are one of the most widespread and abundant small mammals in eastern
148 North America, and occupy a broad range of forest, meadow, and secondary growth habitats.
149 They have served as model systems for population ecology for decades (Vessey and Vessey
150 2007; Brunner et al. 2013) because of their ubiquity and easy trappability. *Peromyscus* spp. more
151 broadly have emerged as model systems for the genomics of adaptation (Bedford and Hoekstra

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2015), and the first reference genomes for the genus are currently being assembled (Kenney-Hunt et al. 2014).

For this study, we sampled white-footed mice from 23 sites in the NYC metropolitan area (Table 1). The 12 sites within NYC limits were the same as in previous studies, although here we combined samples from Pelham Bay and Hunters Island in the Bronx, and Alley Pond and Cunningham Parks in Queens, because evolutionary clustering results showed that these pairs of sites were not strongly differentiated from each other (Munshi-South and Kharchenko 2010). These urban sites contained secondary forest typically dominated by oaks, hickories, maples, and/or tulip trees, with very thick understories composed primarily of invasive plants. Suburban and rural sites contained similar canopies but the understories were largely cleared of thick vegetation by rampant deer herbivory.

At each site, we trapped white-footed mice over a period of one to three nights using two or more 7x7 grids Sherman live traps (9”x3”x3”). Traps within grids were placed 15 m apart, and grids were located several hundred meters apart to avoid trapping close relatives. We collected ear punches or 1cm tail clippings from up to 25 mice at each site, and stored the tissue in 80% EtOH before transfer to a -20 C freezer in the laboratory. Animal handling procedures were approved by Fordham University’s Institutional Animal Care and Use Committee (Protocol No. JMS-13-03).

We chose sites that qualitatively represented typical urban (N = 12 sites), suburban (N = 4), and rural (N = 7) sites in the NYC metropolitan area based on levels of development (Figure 1). However, suburban counties adjacent to NYC have higher human population densities than major cities in other parts of North America. To facilitate comparisons with other urban areas,

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

174 we calculated the size of the human population and percent impervious surface in geographic
175 buffers around each site as proxies for relative urbanization (Figure 1b).

176 We first incorporated GPS coordinates for our study sites into base layers (2010
177 TIGER/Line® Shapefiles—County or States and equivalent) available from the US Census
178 Bureau. Buffers were created in ArcGIS v10.1 (ESRI, Redlands, CA) around each site's GPS
179 coordinates with radii of 500 m, 1,000 m, 1,500 m and 2,000 m. These buffers were chosen
180 because they are in excess of the typical lifetime dispersal distance of white-footed mice (Vessey
181 and Vessey 2007). To determine human population size inside each buffer, we used U.S. Census
182 Blocks as these provide the smallest geographic unit with 100% census data. We first calculated
183 the area (km^2) of each census block and then intersected the census block and buffer layers. We
184 interpolated human population size within each buffer based on the percentage of area for each
185 census block that intercepted each buffer ([area of census block within buffer / area of census
186 block] x population of census block). We then summed all of the interpolated population sizes
187 that fell within each buffer.

188 To determine percent impervious surface within each study site buffer, we used the 2011
189 Percent Developed Imperviousness layer from USGS National Land Cover Data (Xian et al.
190 2011). We then calculated zonal statistics to measure the geometry of the raster file and
191 summarize the cell values of the raster that fall within each buffer. From these data we calculated
192 the average percent impervious surface within each landscape buffer for each site.
193

194 **ddRADSeq, SNP genotyping, and population genomic statistics**

195 We generated SNP genotypes for 233 individuals using a double-digest RADseq protocol
196 (Peterson et al. 2012). We aimed to obtain genotypes from 10 individuals from each site, but

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sample dropout due to poor DNA yield during library preparation resulted in variability in the sample size for each site (Table 1). In brief, we extracted DNA from tail snips or ear punches using Qiagen DNEasy tissue kits (Qiagen, Valencia, CA) with an RNAse treatment, and then digested 500-1,000 ng of DNA using the enzyme combination of SphI-HF and MluCI for one hour following the manufacturer's instructions (New England Biolabs, Ipswich, MA). A Qubit 2.0 fluorometer (Life Technologies, Norwalk, CT) was used to quantify DNA concentration at several steps in the library preparation. Next, digested DNA was cleaned with 1.5X AMPure XP magnetic beads (Beckman-Coulter, Brea, CA) and custom in-line 'flex' barcodes and P1/P2 adapters were ligated to 200-400 ng of digested DNA for each sample (Peterson et al. 2012). Up to 48 individual libraries with unique barcodes were then pooled in equimolar amounts and cleaned with 1.5X AMPure XP beads. Next we selected DNA fragments of known sizes (376-412 bp) from the pooled libraries using a Pippin Prep (Sage Science, Beverly, MA). We then conducted multiple PCR amplifications using 20 ng of size-selected DNA and Phusion High-fidelity PCR reagents and manufacturers' PCR conditions (New England Biolabs, Ipswich, MA). This PCR step added a second, unique index sequence and Illumina sequencing primers to the pooled libraries so each individual sample contained a unique combination of the in-line barcode and index. We then pooled the PCRs for each size-selected library, cleaned the pools using 1.5X AMPure XP beads, and checked the libraries using an Agilent BioAnalyzer (Agilent Technologies, Santa Clara, CA) for DNA concentration and the correct distribution of fragment size. The libraries were sequenced using three lanes of Illumina HiSeq 2000 2x100 bp paired-end sequencing at the New York University Center for Genomics and Systems Biology (New York, NY, USA).

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As an initial check on the quality of our Illumina sequence data, we analyzed the raw reads in FastQC (Andrews 2010). Subsequent demultiplexing, quality filtering and *de novo* SNP calling was conducted using the Stacks 1.21 pipeline (Catchen et al. 2013). First, we used the *process_radtags* script to filter out low-quality reads and demultiplex the remaining reads according to their unique combination of in-line barcode and index. Based on FastQC results, we trimmed all reads to 96 bp to remove poor-quality base calls at the ends of reads. Next, we concatenated the single- and paired-end reads for each individual into one fastq file because the two paired reads did not overlap and were not aligned to a reference genome. To identify RAD loci and call SNPs, we used the *denovo_map.pl* script in Stacks with default settings except for the minimum number of identical reads required to create a 'stack' (m = 7), number of mismatches allowed between RAD loci for a single individual (M = 3), and the number of mismatches allowed when building the catalog (n = 2). After building the initial catalog of loci, we used the *populations* script in Stacks to filter loci for those that occurred in at least 22 / 23 sampling sites (p = 22) and in at least 50% of individuals at each site (r = 0.5). This script also produced genotype output in multiple formats (i.e. Genepop, Structure) with one randomly-selected SNP from each locus (--write_random_SNP), and generates summary statistics such as observed heterozygosity, nucleotide diversity, and pairwise F_{ST} between all populations (Table 2).

We loaded the RAD loci and individual data from Stacks into a MySQL database and visualized the output using the Stacks webserver. Based on the results of the first pipeline run, we removed 26 individuals because their small number of reads resulted in very small SNP datasets and excessive missing genotypes compared to other samples. Highly related individuals in our dataset could also bias downstream analyses. We avoided relatives at our urban sites using

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

242 relatedness values from previous microsatellite studies of the same individuals (Munshi-South
243 and Kharchenko 2010). For the SNP dataset, we identified highly related individuals by
244 calculating kinship coefficients in the software package KING (Manichaikul et al. 2010).
245 Kinship analysis identified 16 pairs of individuals that were related at the half-sib level or
246 greater. We removed one of the pair members from the dataset, resulting in a final dataset
247 comprised of 191 of the original 233 individuals. We then re-ran the Stacks pipeline on the
248 screened dataset of 191 individuals.
249
250 **Modeling genome-wide variation within populations**
251 We modeled four measures of genome-wide variation within populations against our two
252 urbanization proxies to test the hypothesis that urbanization is associated with reduced genetic
253 variation. Summary statistics describing genome-wide variation were calculated for each
254 population using Stacks, including observed heterozygosity (H_O), nucleotide diversity (π),
255 number of private loci, and the percent of polymorphic loci. We examined genetic (dependent)
256 variables using 13 candidate general linear regression models (GLM) consisting of combinations
257 of human population size and percent impervious surface estimated using different geographic
258 buffers as described above (500 m, 1,000 m, 1,500 m, 2,000 m): an intercept-only model, eight
259 univariate models with one variable measured at one buffer size, and four bivariate models
260 including both human population and impervious surface estimated for the same buffer size.
261 Human population size and percent impervious surface were ln-transformed prior to analysis to
262 improve normality. If a model performed substantially better than the intercept-only model, then
263 we interpreted that result as evidence of a statistical effect of the urbanization variable(s) on
264 genetic diversity within white-footed mouse populations. We calculated maximum likelihood

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

265 estimates of model parameters for each model, and then ranked models using values of *AICc*, the
266 corrected Akaike's information criterion (Burnham and Anderson 2002). The relative quality of
267 models was further assessed based on $\Delta AICc$, and the relative weight (w_i) of each model. For the
268 best models, we compared the null and deviance residuals from the GLM and fitted a model line
269 to the points in a scatterplot (Figure 2) to confirm that the model explained variation in the
270 genetic parameter of interest. We conducted all statistical analyses in R 3.1.2 (R Development
271 Core Team 2008), and used GLM regression in the AICcmodavg package for model selection
272 (Mazerolle 2015).
273
274 **Genetic structure and population differentiation**
275 To investigate whether urbanization results in greater genetic differentiation between
276 populations, we used Discriminant Analysis of Principal Components (DAPC) in the R package
277 *adegenet* (Jombart and Ahmed 2011) to identify evolutionary clusters among 191 white-footed
278 mice. DAPC first reduces total genetic variation (i.e. variance in allele frequencies) into principal
279 components, and then identifies discriminant functions that maximize differences between
280 clusters while minimizing variation within clusters. We used cross-validation in *adegenet* to
281 identify the optimal number of principal components. This procedure uses a randomly-generated
282 training set and validation set of individuals to identify the optimal number of principal
283 components that accurately predict group membership without overfitting. To visualize clusters,
284 we used DAPC scatterplots and barplots (compoplot command in *adegenet*) to visualize
285 membership of individuals in different clusters. Genome-wide SNP datasets often have power to
286 discriminate between all groups, and results from analyses such as DAPC may reflect

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

hierarchical structure. Thus, we ran DAPC on the full dataset as well as subsets of sampling sites to investigate potential hierarchical structure.

We also used the model-based evolutionary clustering approaches in fastSTRUCTURE (Raj et al. 2014) and ADMIXTURE (Alexander et al. 2009). fastSTRUCTURE uses approximate inference of the Bayesian model in the original STRUCTURE (J. Pritchard et al. 2000) whereas ADMIXTURE computes maximum likelihood estimates of parameters to estimate the most likely number of evolutionary clusters, K. We ran fastSTRUCTURE on our data for each value of K from 1-23 using the standard model with a simple prior. The fastSTRUCTURE scripts (*chooseK.py*) also calculate heuristic scores for detecting the range of the most likely values of K. After identifying likely values of K, we re-ran fastSTRUCTURE with the more computationally demanding logistic prior model on a smaller subset of K values. The most likely values of K were determined using the cross validation procedure in ADMIXTURE (Alexander and Lange 2011). To compare clustering results from fastSTRUCTURE and ADMIXTURE, we used the CLUMPAK (Cluster Markov Packager Across K) web server (Kopelman et al. 2015) to align and visualize bar plots for both programs at multiple values of K.

Landscape genomics

To investigate IBD and IBE between white-footed mouse populations, we first tested for IBD using a Mantel test in the *ecodist* package in R. The data were bootstrapped 10,000 times to generate 95% confidence intervals for the Mantel *P* value. We then used the BEDASSLE package in R to estimate the relative contributions of IBD and IBE to genetic differentiation between the sampling sites. We computed allele counts and sample sizes for each population using the --counts function in VCFtools 0.1.12b (Danecek et al. 2011). We examined the

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

following IBE models using BEDASSLE: 1) a simple binary matrix indicating whether the population was located in NYC or outside the city; and 2) a pairwise matrix of "resistance distances" calculated between populations using the Isolation-by-Resistance (IBR) approach in Circuitscape 4.0 (McRae and Beier 2007).

Circuitscape calculates a pairwise matrix of resistance distances between populations based on the ability of a simulated electrical current to flow between adjacent landscape cells connected by resistors with user-defined resistance values. We used the 2011 Percent Impervious Surface layer from USGS National Landcover Data (Xian et al. 2011), and set the resistance value for each 30 m cell as equivalent to its percent impervious surface unless the cell exceeded 70% impervious surface. For cells exceeding 70%, we set the resistance level to 100; in other words, any cell with greater than 70% impervious surface was assumed to be 100X more resistant to migration than a cell with 1% impervious surface. The 70% cutoff for relatively high, quasi-barrier resistance was based on results of an earlier Circuitscape analysis we conducted in NYC (Munshi□South 2012). To calculate resistance distances between all populations, we ran Circuitscape in pairwise mode with raster cells connected to all eight neighboring cells. The analysis also produced a cumulative current map to visualize hypothesized migration between all populations. As an initial check on the success of the resistance distances at explaining variation in pairwise F_{ST}, we conducted a partial Mantel test with 10,000 bootstraps in *ecodist* that factored out the effects of Euclidean geographic distance. We then used these resistance distances for the IBE model in BEDASSLE.

BEDASSLE uses a Markov Chain Monte Carlo (MCMC) approach, and includes several graphing functions for evaluating success of the MCMC posterior parameter estimation. To confirm adequate mixing and convergence of chains, we examined traces and marginal

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

distributions for all parameters. We also examined acceptance rates for MCMC parameter estimation, and when these rates were too high or low we adjusted the tuning parameters and re-ran the analysis for 5-10 million steps. For the final runs, we calculated the median and 95% credible intervals for the α_E : α_D ratio after discarding the first 20% as burn-in. This ratio represents the effect size of ecological distance relative to geographic distance.

The simple BEDASSLE model assumes identical variance of allele frequencies about the global mean allele frequency. However, populations can deviate from the global mean for a number of demographic reasons (i.e. bottlenecks, inbreeding), and outlier populations can have a strong influence on the posterior distributions. To account for this variation, BEDASSLE includes a beta-binomial model that estimates an additional parameter, Φ_K, that measures the strength of drift and lack of fit of each population to the model. To address the impacts of population history, we also ran the beta-binomial model for the two IBE scenarios above. For these models, we examined the posterior distributions of the Φ_K values (recalculated as $F_K = 1 / 1+ \Phi_K$) to identify outlier populations.

To test the relative fit of the models to our data, we generated 1,000 posterior predictive samples for each model and compared the simulated data to our observed data in BEDASSLE. BEDASSLE includes a function that randomly draws parameter values from the posterior distributions and simulates new datasets. These simulated datasets are then used to calculate F_{ST} between all pairs of populations. Simulated F_{ST} values can then be compared to the observed F_{ST} values to examine the models' ability to describe patterns in the real data.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RESULTS

ddRADSeq dataset and summary population genomic statistics

Three lanes of Illumina HiSeq 2x100 PE sequencing produced a total of 1.16 billion reads, of which 936.64 million reads (80.9%) passed initial quality filters. Reads were filtered out due to ambiguous barcode sequences (56.3%), low quality scores (40.9%), and ambiguous RAD tags (2.8%). The catalog generated by STACKS included 880,898 RAD loci with at least 7X coverage, but this number was reduced to 14,930 after requiring loci to be present in 22 / 23 populations and at least 50% of individuals within each population. All populations were well-represented in the final dataset, with averages of 4 to 10 individuals genotyped per locus across the 23 populations (median = 6.6; Table 2).

For all loci that were polymorphic in at least one population, the major allele frequency (P) ranged from 0.894 to 0.957, and the average observed heterozygosity from 0.054 to 0.181 (Table 2). If all invariant positions are included, then the major allele frequency exceeded 0.999 in all populations and the heterozygosity ranged only from 0.001 to 0.002 (Supplementary Table 1). Nearly all of the populations located within NYC exhibited lower genetic variation than suburban or rural populations as measured by the percentage of polymorphic sites, numbers of private alleles, major allele frequency, observed heterozygosity, and nucleotide diversity (π). The only exceptions to these trends were relatively high diversity values for the Pelham Bay (PB) population in NYC, and low values for suburban Saxon Woods (SW), the rural Cary Institute (CIE) and Minnewaska Reserve (MR; Table 2). Excluding those four outliers, heterozygosity ranged from 0.054 to 0.117 in the urban populations and 0.158 to 0.181 in the suburban and rural populations. Nucleotide diversity similarly ranged from 0.066 to 0.116 in NYC and 0.138 to 0.15 in the suburban and rural populations.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

379

380 **Modeling urbanization and genome-wide variation within populations**

381 Sampling sites could be distinguished by their combinations of human population size and
382 percent impervious surface (Figure 1b). The rural sites all clustered at very low values for both
383 urbanization variables. Urban and rural sites exhibited little overlap in the scatterplot, with the
384 exception of Fort Tilden (FT), Jamaica Bay (JB), and Pelham Bay in NYC. Figure 1b shows the
385 relationship between the two urbanization variables measured for a 2 km buffer around sampling
386 sites, but results were qualitatively similar for the smaller buffers.

387 Model selection based on AIC_c confirmed that nearly all GLMs with one or more
388 urbanization variables described variation in genomic diversity better than intercept-only models
389 (Table 3). In general, the highest-ranked models included impervious surface cover rather than
390 human population size. The top three models for observed heterozygosity all included percent
391 impervious surface as the only explanatory variable, although models including both statistics
392 were also included in the 95% confidence set. Heterozygosity was negatively associated with
393 percent impervious surface, with the exception of the outlier populations mentioned above (i.e.
394 rural populations with low genetic diversity; Figure 2a). The best model for nucleotide diversity
395 also included percent impervious surface, and this model was clearly superior to any of the other
396 models ($w_i = 0.69$). Nucleotide diversity was also negatively associated with percent impervious
397 surface, and exhibited the same outlier populations as the heterozygosity scatterplot (Figure 2b).
398 Percent polymorphic loci were negatively associated with percent impervious surface (Figure
399 2c). The number of private alleles was the only genetic variation measure that was best explained
400 by human population size; urban populations exhibited low numbers of private alleles (Figure
401 2d). All four top models had ratios of the null to deviance residuals > 1.0. All urbanization

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

coefficients from the top GLMs were also significant at $P < 0.05$, with the coefficients for nucleotide diversity and number of private alleles significant at $P < 0.001$.

Genetic structure and population differentiation

Pairwise F_{ST} calculated using STACKS ranged from a low of 0.033 between a rural (MH) and suburban (LCC) population, and a high of 0.145 between a suburban (MRG) and urban (RR) population. Most values were between 0.05 – 0.10. Some NYC populations exhibited many F_{ST} > 0.10, particularly RR and JB, as did one suburban population (MRG). A simple Mantel test revealed no significant IBD (Mantel $r = 0.041$, 95% CI = -0.046 – 0.124, $P = 0.32$).

Cross-validation identified $N = 23$ as the optimal number of principal components to retain for DAPC analysis (Supplementary Figure S1). The first two discriminant functions distinguished two isolated NYC sites, Jamaica Bay (JB) and Fort Tilden (FT), from the other populations. Discriminant function three separates out a cluster of the other populations on Long Island (AP, FM, FP, KP, RR, WW), a cluster of NYC populations in the middle (IP, NYBG, PB, VC) with suburban populations just below them (LCC, MRG, SW, CPV), and rural populations (CIE, CFP, HIP, HP, MH, MR) at the bottom (Figure 3a). This clustering largely recapitulates the spatial orientation of these populations along a north-south axis. One exception is the most urban population, Central Park (CP), which does not occur where it would be expected based on geography. This placement indicates that Central Park is one of the most isolated and unique NYC populations, along with Fort Tilden and Jamaica Bay. The DAPC compoplot (i.e. a barplot of membership probability) indicated that nearly all individuals could be assigned to their sampling site with high probability (Supplementary Figure S2).

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To clarify relationships between major clusters of sampling sites, we reran the DAPC analysis on two subsets: all populations on the mainland and Manhattan, and all populations located on Long Island. Twelve principal components were retained for the mainland-Manhattan analysis, and the scatterplot of the first two discriminant functions revealed a major, central cluster that recapitulated the geography of the populations. However, three isolated, urban populations were distinct from this cluster (Figure 3b): Central Park (CP), Inwood Hill Park (IP), and Van Cortland Park (VC). The Long Island analysis retained seven principal components, and confirmed that JB and FT are highly distinct populations, as well as the Ridgewood Reservoir (RR; Supplementary Figure S3).

Both the heuristic analysis in fastSTRUCTURE (using the simple prior) and cross-validation in ADMIXTURE identified K = 2 as the most likely number of evolutionary clusters among the 23 white-footed mouse populations we sampled. CLUMPAK confirmed that individual assignment to the two clusters was highly correlated across the two methods (r = 0.93; Figure 4a,b). One cluster (blue in Figure 4a,b) contained all the NYC populations on Long Island, the two populations on Manhattan (CP and IP), and three other suburban / rural populations with atypically low genetic variation (CIE, MR, and SW). All other populations were assigned to the other cluster (orange in Figure 4a,b), except for WW located in rural Long Island which was an admixture of the two clusters in almost equal proportions. When we reran the fastSTRUCTURE analysis for K=1 – 4 using the more accurate logistic prior, the heuristic analysis identified K = 3 as the upper bound on the likely number of evolutionary clusters. Cross-validation in ADMIXTURE also identified K = 3 – 5 as only slightly worse than K = 2 (Figure S4), so we created barplots for these numbers of evolutionary clusters. For K = 3, the blue cluster of urban Long Island, Manhattan, and outlier rural populations was maintained, but

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

447 the other cluster was split into two largely based on location east or west of a north-south axis
448 (Figure 4c). The fourth cluster in the K = 4 ADMIXTURE analysis included a new cluster
449 (Figure 4d) with the two Manhattan populations, a Bronx population (NYBG), and a few distant
450 populations adjacent to or west of the Hudson River (CFP and HIP). The additional cluster for K
451 = 5 included two suburban populations in relative proximity (CPV and LCC; Figure 4e).
452
453 **Landscape genomics**
454 For the simple BEDASSLE model examining presence or absence in the city, the median $\alpha_E : \alpha_D$
455 ratio was 7.14 and the 95% credible set was 6.90 to 7.36. The interpretation of this result is that
456 being located in NYC has an impact of approximately 7 km of extra pairwise geographic
457 distance on genetic differentiation. Comparison of posterior predictive samples for the simple
458 and beta binomial model confirms that the latter is a much better fit to the data (Figure 5a,b). For
459 the beta binomial model, the median $\alpha_E : \alpha_D$ ratio was 0.006 and the 95% credible set was 0.0002
460 to 0.038. This model indicates that presence in NYC has virtually no additional impact over
461 geographic distance on genetic differentiation. However, values of F_k were elevated for most of
462 the urban populations (range in medians from 0.04 – 0.25; mean = 0.12) relative to the suburban
463 (range = 0.01 – 0.06; mean = 0.03) and rural populations (range = 0.001 – 0.11; mean = 0.03),
464 with the exception of two rural populations that also had high values (CIE and MR; Figure 5d).
465 These F_k values indicate that the urban populations deviated substantially more from global
466 allele frequency estimates than the suburban and rural populations.
467 Pairwise resistance distances from the Circuitscape model were significantly associated
468 with F_{ST}, even after factoring out the effects of geographic distance (partial Mantel $r = 0.521$,
469 95% confidence interval = 0.438-0.621, $P < 0.0001$). The simple BEDASSLE model using

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

resistance distances did not converge after trying many different combinations of prior values. The beta binomial model did converge, producing a median $\alpha_E : \alpha_D$ ratio of 0.0001 and 95% negligible impact on genetic differentiation compared to geographic distances, which contradicts the partial Mantel results. Posterior predictive sampling indicated that this beta-binomial model performed well, but systematically did not predict observed $F_{ST} < 0.05$ (Fig. 5c).

DISCUSSION

This study is the first to our knowledge to examine the influence of urbanization on genome-wide SNP variation in city-dwelling species. We found that populations of white-footed mice along an urban-to-rural gradient exhibit a negative correlation between genomic variation and urbanization in and around their habitat. Populations within NYC also exhibited greater genetic differentiation from one another than pairs of populations in rural areas, and urbanization explained a greater proportion of pairwise population differentiation overall than IBD. NYC populations deviated more strongly from global mean allele frequencies than rural populations, indicating that urbanization has substantially altered the evolutionary trajectories of urban wildlife (Donihue and Lambert 2014).

Many recent studies have documented reduced migration and loss of heterozygosity at microsatellite loci among populations in urbanized landscapes (Gortat et al. 2014; S. Jha 2015; Barr et al. 2015). However, for logistical reasons the most variable microsatellite loci are often chosen for population genetic analysis, and thus may not represent unbiased samples of genome-wide diversity (Väli et al. 2008). Heterozygosity measured using microsatellites and traits related to fitness are also often weakly correlated, even though there is a publication bias towards reporting only high heterozygosity-fitness correlations (Chapman et al. 2009). New approaches

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

493 such as ddRAD-Seq used here produce genome-wide SNP markers that are more appropriate for
494 assessing genome-wide variation in relation to ecological factors such as urbanization.
495 Population genomic variation is key to understanding and predicting evolutionary
496 responses to environmental transformation. While urbanization may not directly cause
497 extinctions of most species, it is likely to decrease the evolutionary potential of populations.
498 Results from native species in cities are few. Urban bobcats maintained variation at immune-
499 linked loci due to balancing selection from disease pressures, even after a population bottleneck
500 and population subdivision by freeways (Serieys et al. 2015). Blackbirds colonizing cities also
501 exhibited polymorphisms in a candidate behavioral gene that was strongly associated with urban
502 habitats (Mueller et al. 2013), indicating that functional variation is important for the
503 evolutionary success of urban wildlife. We previously identified candidate genes that may be
504 under selection in urban white-footed mice (Harris et al. 2013), but our results here indicate that
505 these populations have lost as much as half of their genetic diversity compared to nearby rural
506 populations. While candidate genes of large effect may be relevant to understanding some
507 responses to landscape change, screening genome-wide variation offers many advantages for
508 measuring evolutionary potential. Most adaptive traits are polygenic and influenced by many loci
509 of small effect. Information on genome-wide variation will also capture cryptic variation and
510 quantify the amount of standing variation available for responses to future change (Harrisson et
511 al. 2014). Loss of standing variation will thus make it less likely that urban populations will be
512 able to adapt to local conditions, or to global phenomena such as climate change (Franks et al.
513 2014). Mutation is likely too slow of a process for rapid recovery of evolutionary potential in
514 fragmented urban populations. However, increasing connectivity in cities would immediately

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

515 boost genetic variation if migration between genetically differentiated populations could be
516 reestablished.
517
518 **Modeling urbanization and genome-wide diversity**
519 We found that percent impervious surface cover, and to a lesser extent human population size,
520 were highly correlated with levels of genome-wide diversity. Gradient studies have
521 predominated in urban ecology, but terms such as “urban” and “suburban” have been used in
522 many different contexts without standardization (Magle et al. 2012). Study sites are often defined
523 based on subjective criteria, or chosen simply to represent a linear geographical gradient
524 regardless of the actual pattern of urbanization (Ramalho and Hobbs 2012). We advocate that
525 future landscape genetics studies report percent impervious surface cover in and around study
526 sites to facilitate comparisons between landscapes and species. Impervious surface and human
527 population size are highly correlated, but humans may be present in large numbers outside of
528 cities for recreation in protected areas (Monz et al. 2013). Impervious surface can be readily used
529 to track urbanization over time, and has relevance to both terrestrial and aquatic systems (Walsh
530 et al. 2005). Extensive impervious surface in the form of roads, parking lots, and buildings
531 characterizes all cities (Nowak and Greenfield 2012). Roads in particular have well-characterized
532 negative impacts on the connectivity of wildlife populations (Balkenhol and Waits 2009;
533 Benítez-López et al. 2010). Nearly 70% of global forest cover is now fragmented, and areas
534 subject to urbanization and high-intensity agriculture are most severely affected (Haddad et al.
535 2015). Reporting measures of impervious surface cover will provide much needed
536 standardization.
537

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

538 **Population structure and differentiation**

539 Urban and rural populations were differentiated from each other, with many $F_{ST} > 0.10$. Some
540 populations within NYC had pairwise F_{ST} as high as urban-rural pairs that were much more
541 distant, indicating that isolation within the city is quite high. Higher F_{ST} values reported here are
542 similar to those reported for endangered beach mice (Austin et al. 2015) and Channel Island deer
543 mice (Ozer et al. 2011) that recently experienced strong genetic drift from extirpations and
544 translocations. The lack of IBD reported here, but moderate to strong genetic population
545 differentiation, suggests that dispersal is limited by barriers and high landscape resistance rather
546 than geographic distance alone. Thus, evolutionary clustering and IBE (Wang and Bradburd
547 2014; Sexton et al. 2014) are more appropriate models for understanding patterns of genome-
548 wide diversity among these populations.

549 Discriminant analysis of principal components, and two model-based clustering analyses,
550 sorted individuals into 23 clusters that largely recapitulated geographic patterns. We identified
551 one major split between Long Island and mainland populations, which historical demographic
552 modeling indicated is likely related to glacial retreat and ecological succession in the region
553 (Harris et al, unpublished manuscript). The other striking pattern was that several urban
554 populations were outliers in their genetic divergence from the major clusters of sampling sites.
555 Previous clustering analyses using microsatellites also identified most NYC sites as distinct
556 populations (Munshi-South and Kharchenko 2010). SNPs evolve much more slowly than
557 microsatellites, and there is likely still considerable ancestral variation in isolated urban
558 populations. These NYC populations may also not have reached linkage and Hardy-Weinberg
559 equilibrium at many SNP loci because not enough generations have elapsed since isolation, or
560 the populations have not reached mutation-migration-drift equilibrium.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

561

562 **Landscape genomics**

563 Our isolation-by-environment (IBE) analyses produced mixed results. Overall, IBE modeling
564 indicated that urbanization drives genetic differentiation to a greater degree than geographic
565 distance alone. Although the relative influence of IBE to IBD was modest, urban populations
566 deviated to a much greater degree from global allele frequencies than suburban or rural
567 populations. The most likely explanation for this deviation is substantial drift due to inbreeding
568 or bottlenecks in isolated urban habitats. This scenario generally conforms to the strong structure
569 observed among some of the urban populations. However, other factors that caused these
570 populations to deviate from the BEDASSLE model cannot be ruled out, such as unsampled
571 environmental variables (Bradburd et al. 2013).

572 IBR modeling in Circuitscape confirmed that variation in percent impervious surface is
573 highly associated with variation in pairwise F_{ST}, even after factoring out IBD. Partial Mantel
574 tests have several known issues with false positives (Legendre and Fortin 2010), but we
575 previously found that vegetation cover (typically the inverse of impervious surface in NYC)
576 successfully described migration between populations using similar approaches (Munshi□South
577 2012). The BEDASSLE models did not identify the same clear association between impervious
578 surface and genetic differentiation. As above, the strong deviation of urban populations from the
579 global mean allele frequencies indicated that the BEDASSLE model may have underperformed.
580 Resistance distances from Circuitscape are integrated over all possible paths between points on
581 the landscape, and thus BEDASSLE also counts geographic distance twice with unknown
582 consequences when using resistance distances (G. Bradburd, personal communication).

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

583 IBE is an active area of inquiry that holds great promise for understanding the processes
584 that generate genetic variation. However, currently available approaches are relatively new, and
585 several caveats apply to their application (Wang and Bradburd 2014). This study is the first use
586 of BEDASSLE to model the relationship between genetic differentiation and human
587 modification of the environment. In addition to the model adequacy issues raised above, some of
588 the populations analyzed here may not have reached migration-drift equilibrium, or were
589 dominated by idiosyncratic environmental processes. Previous BEDASSLE analyses that
590 detected stronger IBE patterns were conducted over much larger geographic scales among much
591 more strongly differentiated populations, such as sky island birds (Manthey and Moyle 2015),
592 lizards occupying a SE Asian archipelago separated by deep ocean trenches (Barley et al. 2015),
593 and a widespread bird occupying much of the Amazon basin (Harvey and Brumfield 2015).
594 These results suggest that the IBE model in BEDASSLE may currently be best suited to
595 populations that are deeply diverged in both time and space.
596
597 **Conclusions & Future Research**
598 The results presented here demonstrate for the first time that urbanization is associated with a
599 pervasive reduction in genome-wide variation among animal populations. *Peromyscus* spp.,
600 including white-footed mice, are among the most widely-distributed small mammals in North
601 America and increasingly important models for investigating natural variation (Bedford and
602 Hoekstra 2015). Given unchecked urbanization, particularly in the eastern United States, it is
603 likely that many white-footed mouse populations in metropolitan areas have experienced similar
604 declines in standing genetic variation. Preliminary evidence indicates that some loci have
605 experienced selective sweeps in urban white-footed mice (Harris et al. 2013), but the relative

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

606 impacts of urbanization on historical demography and natural selection have not been fully
607 disentangled for these populations. Our models of historical demography using genome-wide
608 SNPs have identified bottlenecks and population divergence on the timeframe of urbanization in
609 NYC. Using an using an expanded transcriptome dataset (Harris et al. 2015), we recently
610 identified signatures of selection in NYC populations using an approach that accounted for these
611 historical demographic patterns (Harris & Munshi-South, unpublished manuscript). We
612 identified dozens of candidate loci under selection that are associated with metabolic and
613 immune processes. These patterns may reflect changes in diet and biotic pressures (i.e. disease
614 and inflammation) in cities. Thus, the loss of genome-wide diversity documented here does not
615 necessarily preclude local adaptation to highly altered, stressful urban environments.
616 Evolutionary potential of urban populations could be improved by restoring connectivity
617 between urban forest patches (Zipperer et al. 2012). Enhanced gene flow between urban forests,
618 as well as between urban and suburban areas, would increase overall genome-wide diversity
619 (although may simultaneously break up local adaptation). Habitat networks that promote gene
620 flow in cities can be constructed from even small gardens and green spaces not explicitly
621 dedicated to biodiversity (Goddard et al. 2010; Vergnes et al. 2012). Microevolutionary
622 processes are rarely integrated into urban conservation due to their perceived complex nature and
623 variation between taxa, but could be useful metrics for assessing landscape connectivity
624 (Stockwell et al. 2003; Kinnison et al. 2007). Converting “gray” infrastructure into “green”
625 networks could also simultaneously address conservation goals while delivering co-benefits to
626 humans in the form of biodiversity experiences (Tanner et al. 2014). A rich literature indicates
627 that humans benefit in myriad ways from access to nature (Fuller et al. 2007), but integrative

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

628 approaches to urban wildlife conservation are a major area for future growth (Shwartz et al.
629 2014).
630
631 **Data Archiving Statement**
632 The DNA sequencing reads will be deposited in NCBI's Short-read Archive (SRA). Other data
633 files will be deposited on Dryad.
634
635 **ACKNOWLEDGMENTS**
636 We thank Matthew Combs and Nicole Fusco for helpful comments on the manuscript, and Mike
637 Hickerson for access to the Hicker-cabin and its surrounding white-footed mouse habitat.
638 Miranda Gonzalez provided substantial assistance in the laboratory, and Gideon Bradburd
639 offered useful advice on landscape genetic analyses. The New York State Department of
640 Environmental Conservation, New York City Department of Parks and Recreation, New York
641 Botanical Garden, and the Connecticut Department of Energy and Environmental Protection
642 granted permission to collect white-footed mouse samples. Research reported in this publication
643 was supported by the National Institute of General Medical Sciences of the National Institutes of
644 Health under award number R15GM099055 to JM-S. The content is solely the responsibility of
645 the authors and does not represent the official views of the National Institutes of Health.
646

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

LITERATURE CITED

Alexander, D.H., and K. Lange. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics **12**:246.

Alexander, D.H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research **19**:1655–64.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Austin, J.D., J.A. Gore, D.U. Greene, and C. Gotteland. 2015. Conspicuous genetic structure belies recent dispersal in an endangered beach mouse (Peromyscus polionotus trissyllepsis). Conservation Genetics **16**:915–28.

Balkenhol, N., and L. Waits. 2009. Molecular road ecology: exploring the potential of genetics for investigating transportation impacts on wildlife. Molecular Ecology **18**:4151–64.

Barley, A.J., P.J. Monnahan, R.C. Thomson, L.L. Grismer, and R.M. Brown. 2015. Sun skink landscape genomics: assessing the roles of microevolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. Molecular Ecology **24**:1696–1712.

Barnosky, A.D., E.A. Hadly, J. Bascompte, E.L. Berlow, J.H. Brown, M. Fortelius, W.M. Getz, et al. 2012. Approaching a state shift in Earth's biosphere. Nature **486**:52–58.

Barrett, R.D.H., and D. Schluter. 2008. Adaptation from standing genetic variation. Trends in Ecology & Evolution **23**:38–44.

Barr, K.R., B.E. Kus, K.L. Preston, S. Howell, E. Perkins, and A.G. Vandergast. 2015. Habitat fragmentation in coastal southern California disrupts genetic connectivity in the cactus wren (Campylorhynchus brunneicapillus). Molecular Ecology **24**:2349–63.

Bazin, E., S. Glémin, and N. Galtier. 2006. Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals. Science **312**:570–72.

Bedford, N.L., and H.E. Hoekstra. 2015. Peromyscus mice as a model for studying natural variation. eLife **4**:e06813.

Benítez-López, A., R. Alkemade, and P.A. Verweij. 2010. The impacts of roads and other infrastructure on mammal and bird populations: A meta-analysis. Biological Conservation **143**:1307–16.

Bjorklund, M., I. Ruiz, and J. Senar. 2010. Genetic differentiation in the urban habitat: the great tits (Parus major) of the parks of Barcelona city. Biological Journal of the Linnean Society **99**:9–19.

Blair, R.B. 2001. Birds and Butterflies Along Urban Gradients in Two Ecoregions of the United States: Is Urbanization Creating a Homogeneous Fauna? In Biotic Homogenization, edited by Julie L. Lockwood and Michael L. McKinney, 33–56. New York, NY: Springer US.

Bradburd, G.S., P.L. Ralph, and G.M. Coop. 2013. Disentangling the Effects of Geographic and Ecological Isolation on Genetic Differentiation. Evolution **67**:3258–73.

Brunner, J.L., S. Duerr, F. Keesing, M. Killilea, H. Vuong, and R.S. Ostfeld. 2013. An Experimental Test of Competition among Mice, Chipmunks, and Squirrels in Deciduous Forest Fragments. PLoS ONE **8**:e66798.

Burke, M.K., J.P. Dunham, P. Shahrestani, K.R. Thornton, M.R. Rose, and A.D. Long. 2010. Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature **467**:587–90.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

692 Burnham, K.P., and D.R. Anderson. 2002. Model selection and multi-model inference: a
693 practical information-theoretic approach. Springer.
694 Catchen, J., P.A. Hohenlohe, S. Bassham, A. Amores, and W.A. Cresko. 2013. Stacks: an
695 analysis tool set for population genomics. Molecular Ecology **22**:3124–40.
696 Chapman, J.R., S. Nakagawa, D.W. Coltman, J. Slate, and B.C. Sheldon. 2009. A quantitative
697 review of heterozygosity–fitness correlations in animal populations. Molecular Ecology
698 **18**:2746–65.
699 Chapman, J.R., and B.C. Sheldon. 2011. Heterozygosity is unrelated to adult fitness measures in
700 a large, noninbred population of great tits (Parus major). Journal of Evolutionary Biology
701 **24**:1715–26.
702 Chiappero, M.B., G.M. Panzetta-Dutari, D. Gómez, E. Castillo, J.J. Polop, and C.N. Gardenal.
703 2011. Contrasting genetic structure of urban and rural populations of the wild rodent
704 Calomys musculinus (Cricetidae, Sigmodontinae). Mammalian Biology - Zeitschrift Für
705 Säugetierkunde **76**:41–50.
706 Colosimo, P.F., K.E. Hosemann, S. Balabhadra, G. Villarreal Jr, M. Dickson, J. Grimwood, J.
707 Schmutz, R.M. Myers, D. Schluter, and D.M. Kingsley. 2005. Widespread parallel
708 evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science **307**:1928–
709 33.
710 Cushman, S.A., and E.L. Landguth. 2010. Spurious correlations and inference in landscape
711 genetics. Molecular Ecology **19**:3592–3602.
712 Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, et
713 al. 2011. The variant call format and VCFtools. Bioinformatics **27**:2156–58.
714 Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, and M.L. Blaxter. 2011.
715 Genome-wide genetic marker discovery and genotyping using next-generation
716 sequencing. Nature Reviews Genetics **12**:499–510.
717 Delaney, K.S., S.P.D. Riley, and R.N. Fisher. 2010. A Rapid, Strong, and Convergent Genetic
718 Response to Urban Habitat Fragmentation in Four Divergent and Widespread
719 Vertebrates. PLoS ONE **5**:e12767.
720 Domingues, V.S., Y.-P. Poh, B.K. Peterson, P.S. Pennings, J.D. Jensen, and H.E. Hoekstra.
721 2012. Evidence of adaptation from ancestral variation in young populations of beach
722 mice. Evolution **66**:3209–23.
723 Donihue, C.M., and M.R. Lambert. 2014. Adaptive evolution in urban ecosystems. AMBIO
724 **44**:194–203.
725 Ellis, E.C., J.O. Kaplan, D.Q. Fuller, S. Vavrus, K.K. Goldewijk, and P.H. Verburg. 2013. Used
726 planet: A global history. Proceedings of the National Academy of Sciences **110**:7978–85.
727 Epps, C.W., J.D. Wehausen, V.C. Bleich, S.G. Torres, and J.S. Brashares. 2007. Optimizing
728 dispersal and corridor models using landscape genetics. Journal of Applied Ecology
729 **44**:714–24.
730 Etterson, J.R., and R.G. Shaw. 2001. Constraint to Adaptive Evolution in Response to Global
731 Warming. Science **294**:151–54.
732 Franks, S.J., J.J. Weber, and S.N. Aitken. 2014. Evolutionary and plastic responses to climate
733 change in terrestrial plant populations. Evolutionary Applications **7**:123–39.
734 Fuller, R.A., K.N. Irvine, P. Devine-Wright, P.H. Warren, and K.J. Gaston. 2007. Psychological
735 benefits of greenspace increase with biodiversity. Biology Letters **3**:390–94.
736 Gibson, G., and I. Dworkin. 2004. Uncovering cryptic genetic variation. Nature Reviews
737 Genetics **5**:681–90.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

738 Goddard, M.A., A.J. Dougill, and T.G. Benton. 2010. Scaling up from gardens: biodiversity
739 conservation in urban environments. Trends in Ecology & Evolution **25**:90–98.
740 Gortat, T., R. Rutkowski, A. Gryczyńska, A. Pieniążek, A. Kozakiewicz, and M. Kozakiewicz.
741 2014. Anthropopressure gradients and the population genetic structure of Apodemus
742 agrarius. Conservation Genetics **16**:649–59.
743 Graves, T.A., P. Beier, and J.A. Royle. 2013. Current approaches using genetic distances
744 produce poor estimates of landscape resistance to interindividual dispersal. Molecular
745 Ecology **22**:3888–3903.
746 Groffman, P.M., J. Cavender-Bares, N.D. Bettez, J.M. Grove, S.J. Hall, J.B. Heffernan, S.E.
747 Hobbie, K.L. Larson, J.L. Morse, and C. Neill. 2014. Ecological homogenization of
748 urban USA. Frontiers in Ecology and the Environment **12**:74–81.
749 Haddad, N.M., L.A. Brudvig, J. Clobert, K.F. Davies, A. Gonzalez, R.D. Holt, T.E. Lovejoy, et
750 al. 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science
751 Advances **1**:e1500052.
752 Harris, S.E., J. Munshi-South, C. Obergfell, and R. O'Neill. 2013. Signatures of Rapid Evolution
753 in Urban and Rural Transcriptomes of White-Footed Mice (Peromyscus leucopus) in the
754 New York Metropolitan Area. PLoS ONE **8**:e74938.
755 Harris, S.E., R.J. O'Neill, and J. Munshi-South. 2015. Transcriptome resources for the white-
756 footed mouse (Peromyscus leucopus): new genomic tools for investigating ecologically
757 divergent urban and rural populations. Molecular Ecology Resources **15**:382–94.
758 Harrisson, K.A., A. Pavlova, M. Telonis-Scott, and P. Sunnucks. 2014. Using genomics to
759 characterize evolutionary potential for conservation of wild populations. Evolutionary
760 Applications **7**:1008–25.
761 Harvey, M.G., and R.T. Brumfield. 2015. Genomic variation in a widespread Neotropical bird
762 (Xenops minutus) reveals divergence, population expansion, and gene flow. Molecular
763 Phylogenetics and Evolution **83**:305–16.
764 Heilig, G.K. 2011. World Urbanization Prospects: The 2011 Revision. New York, NY: United
765 Nations, Department of Economic and Social Affairs (DESA), Population Division,
766 Population Estimates and Projections Section.
767 Hobbs, R.J., S. Arico, J. Aronson, J.S. Baron, P. Bridgewater, V.A. Cramer, P.R. Epstein, J.J.
768 Ewel, C.A. Klink, and A.E. Lugo. 2006. Novel ecosystems: theoretical and management
769 aspects of the new ecological world order. Global Ecology and Biogeography **15**:1–7.
770 Hobbs, R.J., E. Higgs, and J.A. Harris. 2009. Novel ecosystems: implications for conservation
771 and restoration. Trends in Ecology & Evolution **24**:599–605.
772 Hoffmann, A.A., and C.M. Sgrò. 2011. Climate change and evolutionary adaptation. Nature
773 **470**:479–85.
774 Jaquiéry, J., T. Broquet, A.H. Hirzel, J. Yearsley, and N. Perrin. 2011. Inferring landscape
775 effects on dispersal from genetic distances: how far can we go? Molecular Ecology
776 **20**:692–705.
777 Jha, S. 2015. Contemporary human-altered landscapes and oceanic barriers reduce bumble bee
778 gene flow. Molecular Ecology **24**:993–1006.
779 Jha, S., and C. Kremen. 2013. Urban land use limits regional bumble bee gene flow. Molecular
780 Ecology **22**:2483–95.
781 Jombart, T., and I. Ahmed. 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide
782 SNP data. Bioinformatics **27**:3070–71.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

783 Kenney-Hunt, J., A. Lewandowski, T.C. Glenn, J.L. Glenn, O.V. Tsyusko, R.J. O'Neill, J.
784 Brown, et al. 2014. A genetic map of Peromyscus with chromosomal assignment of
785 linkage groups (a Peromyscus genetic map). Mammalian Genome **25**:160–79.
786 Keyghobadi, N. 2007. The genetic implications of habitat fragmentation for animals. Canadian
787 Journal of Zoology **85**:1049–64.
788 Kinnison, M., A. Hendry, and C. Stockwell. 2007. Contemporary evolution meets conservation
789 biology II: Impediments to integration and application. Ecological Research **22**:947–54.
790 Kopelman, N.M., J. Mayzel, M. Jakobsson, N.A. Rosenberg, and I. Mayrose. 2015. Clumpak: a
791 program for identifying clustering modes and packaging population structure inferences
792 across K. Molecular Ecology Resources **15**:1179–91.
793 Legendre, P., and M.-J. Fortin. 2010. Comparison of the Mantel test and alternative approaches
794 for detecting complex multivariate relationships in the spatial analysis of genetic data.
795 Molecular Ecology Resources **10**:831–44.
796 Magle, S.B., V.M. Hunt, M. Vernon, and K.R. Crooks. 2012. Urban wildlife research: Past,
797 present, and future. Biological Conservation **155**:23–32.
798 Manichaikul, A., J.C. Mychaleckyj, S.S. Rich, K. Daly, M. Sale, and W.-M. Chen. 2010. Robust
799 relationship inference in genome-wide association studies. Bioinformatics **26**:2867–73.
800 Manthey, J.D., and R.G. Moyle. 2015. Isolation by environment in White-breasted Nuthatches
801 (Sitta carolinensis) of the Madrean Archipelago sky islands: a landscape genomics
802 approach. Molecular Ecology **24**:3628–38.
803 Mazerolle, M.J. 2015. AICcmodavg: model selection and multimodel inference based on (Q)
804 AIC (c). R Package Version 2.0-3.
805 McKinney, M.L. 2002. Urbanization, biodiversity, and conservation. Bioscience **52**:883–90.
806 McRae, B.H., and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal
807 populations. Proceedings of the National Academy of Sciences USA **104**:19885–90.
808 Merilä, J., and A.P. Hendry. 2014. Climate change, adaptation, and phenotypic plasticity: the
809 problem and the evidence. Evolutionary Applications **7**:1–14.
810 Monz, C.A., C.M. Pickering, and W.L. Hadwen. 2013. Recent advances in recreation ecology
811 and the implications of different relationships between recreation use and ecological
812 impacts. Frontiers in Ecology and the Environment **11**:441–46.
813 Mueller, J.C., J. Partecke, B.J. Hatchwell, K.J. Gaston, and K.L. Evans. 2013. Candidate gene
814 polymorphisms for behavioural adaptations during urbanization in blackbirds. Molecular
815 Ecology **22**:3629–37.
816 Munshi□South, J. 2012. Urban landscape genetics: canopy cover predicts gene flow between
817 white□footed mouse (Peromyscus leucopus) populations in New York City. Molecular
818 Ecology **21**:1360–78.
819 Munshi-South, J., and K. Kharchenko. 2010. Rapid, pervasive genetic differentiation of urban
820 white-footed mouse (Peromyscus leucopus) populations in New York City. Molecular
821 Ecology **19**:4242–54.
822 Munshi-South, J., and C. Nagy. 2014. Urban park characteristics, genetic variation, and historical
823 demography of white-footed mouse (*Peromyscus leucopus*) populations in New York
824 City. PeerJ **2** (March):e310.
825 Nabholz, B., S. Glémin, and N. Galtier. 2008. Strong Variations of Mitochondrial Mutation Rate
826 across Mammals—the Longevity Hypothesis. Molecular Biology and Evolution **25**:120–
827 30.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

828 Narum, S.R., C.A. Buerkle, J.W. Davey, M.R. Miller, and P.A. Hohenlohe. 2013. Genotyping-
829 by-sequencing in ecological and conservation genomics. Molecular Ecology **22**:2841–47.
830 Noël, S., and F.-J. Lapointe. 2010. Urban conservation genetics: Study of a terrestrial salamander
831 in the city. Biological Conservation **143**:2823–31.
832 Nowak, D.J., and E.J. Greenfield. 2012. Tree and impervious cover change in U.S. cities. Urban
833 Forestry & Urban Greening **11**:21–30.
834 Oakley, C.G. 2013. Small effective size limits performance in a novel environment. Evolutionary
835 Applications **6**:823–31.
836 Orr, H.A. 2005. The genetic theory of adaptation: a brief history. Nature Reviews Genetics
837 **6**:119–27.
838 Ozer, F., H. Gellerman, and M.V. Ashley. 2011. Genetic impacts of Anacapa deer mice
839 reintroductions following rat eradication. Molecular Ecology **20**:3525–39.
840 Palumbi, S. 2001. Humans as the World's Greatest Evolutionary Force. Science **293**:1786–90.
841 Pennings, P.S. 2012. Standing genetic variation and the evolution of drug resistance in HIV.
842 PLoS Computational Biology **8**:e1002527.
843 Peterson, B.K., J.N. Weber, E.H. Kay, H.S. Fisher, and H.E. Hoekstra. 2012. Double Digest
844 RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model
845 and Non-Model Species. PLoS ONE **7**:e37135.
846 Prentis, P.J., J.R. Wilson, E.E. Dormontt, D.M. Richardson, and A.J. Lowe. 2008. Adaptive
847 evolution in invasive species. Trends in Plant Science **13**:288–94.
848 Pritchard, J.K., J.K. Pickrell, and G. Coop. 2010. The genetics of human adaptation: hard
849 sweeps, soft sweeps, and polygenic adaptation. Current Biology **20**:R208–15.
850 Pritchard, J., M. Stephens, and P. Donnelly. 2000. Inference of population structure using
851 multilocus genotype data. Genetics **155**:945–59.
852 Raj, A., M. Stephens, and J.K. Pritchard. 2014. FastSTRUCTURE: Variational Inference of
853 Population Structure in Large SNP Data Sets. Genetics **197**:573–89.
854 Ramalho, C.E., and R.J. Hobbs. 2012. Time for a change: dynamic urban ecology. Trends in
855 Ecology & Evolution **27**:179–88.
856 R Development Core Team, R. 2008. R: a language and environment for statistical computing. R
857 Foundation for Statistical Computing. http://www.R-project.org/.
858 Renaut, S., A.W. Nolte, S.M. Rogers, N. Derome, and L. Bernatchez. 2011. SNP signatures of
859 selection on standing genetic variation and their association with adaptive phenotypes
860 along gradients of ecological speciation in lake whitefish species pairs (Coregonus spp.).
861 Molecular Ecology **20**:545–59.
862 Rivera-Ortíz, F.A., R. Aguilar, M.D.C. Arizmendi, M. Quesada, and K. Oyama. 2014. Habitat
863 fragmentation and genetic variability of tetrapod populations. Animal Conservation
864 **18**:249–58.
865 Rockman, M.V. 2008. Reverse engineering the genotype–phenotype map with natural genetic
866 variation. Nature **456**:738–44.
867 Ruddiman, W.F., E.C. Ellis, J.O. Kaplan, and D.Q. Fuller. 2015. Defining the epoch we live in.
868 Science **348**:38–39.
869 Scarcelli, N., and P.X. Kover. 2009. Standing genetic variation in FRIGIDA mediates
870 experimental evolution of flowering time in Arabidopsis. Molecular Ecology **18**:2039–
871 49.
872 Selkoe, K.A., and R.J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using
873 and evaluating microsatellite markers. Ecology Letters **9**:615–29.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

874 Serieys, L.E.K., A. Lea, J.P. Pollinger, S.P.D. Riley, and R.K. Wayne. 2015. Disease and
875 freeways drive genetic change in urban bobcat populations. Evolutionary Applications
876 **8**:75–92.
877 Seto, K.C., B. Güneralp, and L.R. Hutyra. 2012. Global forecasts of urban expansion to 2030 and
878 direct impacts on biodiversity and carbon pools. Proceedings of the National Academy of
879 Sciences **109**:16083–88.
880 Sexton, J.P., S.B. Hangartner, and A.A. Hoffmann. 2014. Genetic Isolation by Environment or
881 Distance: Which Pattern of Gene Flow Is Most Common? Evolution **68**:1–15.
882 Shwartz, A., A. Turbé, R. Julliard, L. Simon, and A.-C. Prévot. 2014. Outstanding challenges for
883 urban conservation research and action. Global Environmental Change **28**:39–49.
884 Steffen, W., P.J. Crutzen, and J.R. McNeill. 2007. The Anthropocene: are humans now
885 overwhelming the great forces of nature. AMBIO **36**:614–21.
886 Stockwell, C.A., A.P. Hendry, and M.T. Kinnison. 2003. Contemporary evolution meets
887 conservation biology. Trends in Ecology & Evolution **18**:94–101.
888 Tanner, C.J., F.R. Adler, N.B. Grimm, P.M. Groffman, S.A. Levin, J. Munshi-South, D.E.
889 Pataki, M. Pavao-Zuckerman, and W.G. Wilson. 2014. Urban ecology: advancing science
890 and society. Frontiers in Ecology and the Environment **12**:574–81.
891 Teotónio, H., I.M. Chelo, M. Bradić, M.R. Rose, and A.D. Long. 2009. Experimental evolution
892 reveals natural selection on standing genetic variation. Nature Genetics **41**:251–57.
893 Väli, Ü., A. Einarsson, L. Waits, and H. Ellegren. 2008. To what extent do microsatellite
894 markers reflect genome-wide genetic diversity in natural populations? Molecular Ecology
895 **17**:3808–17.
896 Vergnes, A., I.L. Viol, and P. Clergeau. 2012. Green corridors in urban landscapes affect the
897 arthropod communities of domestic gardens. Biological Conservation **145**:171–78.
898 Vessey, S.H., and K.B. Vessey. 2007. Linking behavior, life history and food supply with the
899 population dynamics of white-footed mice (Peromyscus leucopus). Integrative Zoology
900 **2**:123–30.
901 Vitousek, P., H. Mooney, J. Lubchenco, and J. Melillo. 1997. Human domination of Earth's
902 ecosystems. Science **277**:494–99.
903 Walsh, C.J., A.H. Roy, J.W. Feminella, P.D. Cottingham, P.M. Groffman, and R.P. Morgan II.
904 2005. The urban stream syndrome: current knowledge and the search for a cure. Journal
905 of the North American Benthological Society **24**:706–23.
906 Wang, I.J., and G.S. Bradburd. 2014. Isolation by environment. Molecular Ecology **23**:5649–62.
907 Xian, G., C. Homer, J. Demitz, J. Fry, N. Hossain, and J. Wickham. 2011. Change of impervious
908 surface area between 2001 and 2006 in the conterminous United States. Photogrammetric
909 Engineering and Remote Sensing **77**:758–62.
910 Zipperer, W.C., T.W. Foresman, S.P. Walker, and C.T. Daniel. 2012. Ecological consequences
911 of fragmentation and deforestation in an urban landscape: a case study. Urban
912 Ecosystems **15**:533–44.
913

914 **Table 1**. Characteristics of numbers of white-footed mice sampled at 23 locations along an urban-to-rural gradient. Human population

915 size and percent impervious surface were measured at four buffer sizes around these sites: 0.5 km, 1.0km, 1.5km, and 2.0km.

916

Site	Code	Type	N	Latitude	Longitude	Human Population Size	% Impervious Surface
Alley Pond / Cunningham	AP	City	13	40.747605	-73.742505	1298.3, 11518.3, 32262.4, 59880.1	21.2, 42.7, 48.3, 50.8
Central Park	CP	City	9	40.798231	-73.956198	15908.3, 86289.3, 210342.9, 351698.8	48.6, 59, 64.2, 60.2
Flushing Meadows	FM	City	10	40.721442	-73.830805	1577, 24976.4, 62246.8, 125893.1	32.3, 55.8, 56.9, 58
Forest Park	FP	City	6	40.703356	-73.850889	268.3, 18451.6, 59338.1, 123722	6.8, 40.2, 54.2, 61.3
Fort Tilden	FT	City	4	40.561031	-73.887552	17.2, 557.7, 1144.2, 2357.5	5.5, 11.2, 8.9, 8.5
Inwood Hill Park	IP	City	6	40.873295	-73.925005	1073.6, 31425.1, 65049.1, 121354.2	7.5, 26.2, 29.5, 30
Jamaica Bay	JB	City	5	40.623063	-73.824582	0, 9, 173.5, 1438.4	3.6, 1.9, 2, 3.2
Kissena Park	KP	City	9	40.746547	-73.811192	2039.9, 17360, 47871.4, 107273.1	28.5, 48.1, 58.1, 62.3
NY Botanical Garden	NYBG	City	11	40.871613	-73.874079	6875, 54876.3, 140744.3, 256359.1	44.4, 50.7, 61.2, 60.9
Pelham Bay	PB	City	11	40.879895	-73.804063	1.8, 22.5, 934, 3508.1	0, 7.6, 9.7, 9.5
Ridgewood Reservoir	RR	City	10	40.687347	-73.88711	2725.8, 22195.2, 54859, 143223.9	21.7, 45.4, 57.3, 66.9
Van Cortlandt Park	VC	City	6	40.902086	-73.882341	0, 2106.1, 28583.1, 77541.7	10, 16, 23.3, 27.7
Pleasant Valley	CPV	Suburb	7	41.707149	-73.796406	45.7, 190.1, 406.8, 720	0.5, 0.6, 1, 1
Louis Calder Center	LCC	Suburb	13	41.128624	-73.73042	235.8, 1139, 1945.7, 3118.6	3.1, 5.9, 9.6, 10.6
Mianus River Gorge	MRG	Suburb	6	41.185234	-73.622813	76.4, 369.8, 912.5, 1681.4	0.3, 0.6, 0.7, 0.6
Saxon Woods Park	SW	Suburb	7	40.988344	-73.753852	45.7, 669.5, 3765.2, 8164.7	0.6, 13.6, 17.6, 17.6
C. Fahnestock St. Park	CFP	Rural	9	41.468662	-73.843355	1.9, 9.9, 26.9, 56	0.1, 0.2, 0.1, 0.1
Cary Institute	CIE	Rural	7	41.784468	-73.735476	5.6, 41.3, 122.3, 263.8	0.5, 1, 1.1, 1.3
Highpoint State Park	HIP	Rural	8	41.262754	-74.703121	4.2, 19.7, 56.8, 157.5	0.8, 0.4, 0.3, 0.4
Harriman State Park	HP	Rural	9	41.282946	-74.068317	0, 0, 0.2, 2.8	0.1, 0.1, 0.2, 0.2
Cornwall, CT	MH	Rural	8	41.787584	-73.385292	5.7, 48.5, 112.2, 183.5	2.2, 1.7, 1.4, 0.9
Minnewaska Reserve	MR	Rural	8	41.727843	-74.260149	0, 0, 1.8, 9	0, 0, 0.1, 0.1
Wildwood State Park	WW	Rural	9	40.939296	-72.828573	13.1, 191.3, 1107, 2080	0.9, 3, 3.6, 3.5

917

918

919

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Summary genetic diversity statistics calculated by STACKS for nucleotide positions that were polymorphic in at least one population. *N* = average number of individuals genotyped at each locus; *Sites* = number of polymorphic nucleotide sites across the dataset; *%Poly* = percentage of polymorphic loci; *Private* = number of variable sites unique to each population; *P* = average frequency of the major allele; H_{obs} = average observed heterozygosity per locus; π = average nucleotide diversity.

Site	Type	N	Sites	%Poly	Private	P	H_{obs}	π
AP	City	10.4	14,859	0.388	154	0.944	0.080	0.089
CP	City	6.6	14,031	0.321	152	0.936	0.095	0.102
FM	City	7.7	14,812	0.229	90	0.953	0.061	0.071
FP	City	5.4	14,904	0.224	92	0.951	0.066	0.076
FT	City	3.7	14,880	0.235	78	0.948	0.082	0.086
IP	City	5.2	14,879	0.304	146	0.944	0.083	0.091
JB	City	4.3	14,750	0.230	54	0.952	0.075	0.079
KP	City	7.8	14,872	0.278	139	0.951	0.069	0.077
NYBG	City	7.9	12,657	0.321	138	0.925	0.117	0.106
PB	City	9.1	14,788	0.414	269	0.899	0.164	0.143
RR	City	7.7	14,648	0.209	88	0.957	0.054	0.066
VC	City	5.5	14,909	0.350	159	0.928	0.115	0.116
CPV	Suburb	6.0	14,712	0.383	273	0.903	0.164	0.142
LCC	Suburb	10.3	14,670	0.450	431	0.904	0.158	0.138
MRG	Suburb	5.1	14,830	0.371	198	0.903	0.165	0.146
SW	Suburb	5.7	14,698	0.303	126	0.945	0.081	0.090
CFP	Rural	7.0	14,814	0.380	269	0.914	0.141	0.122
CIE	Rural	5.6	14,393	0.300	134	0.946	0.078	0.088
HIP	Rural	7.0	14,886	0.416	515	0.910	0.148	0.129
HP	Rural	7.1	14,797	0.452	516	0.915	0.137	0.135
MH	Rural	6.9	14,906	0.432	362	0.894	0.181	0.150
MR	Rural	6.3	14,856	0.295	381	0.947	0.074	0.085
WW	Rural	6.6	14,630	0.327	131	0.920	0.132	0.113
Mean		6.7	14,660	0.331	212.8	0.930	0.110	0.106

928 **Table 3**. Results of selection among general linear models describing the influence of percent impervious surface cover (imprv) and
929 human population size (pop) on a) observed heterozygosity, b) nucleotide diversity, c) the number of private alleles within each
930 population, and d) the percentage of polymorphic loci within each population. The best models were chosen based on the second-order
931 Akaike Information Criterion (AIC_c), the rate of change in AIC_c (Δ_i), and the Akaike weights (w_i).
932

Models	AIC_c	Δ_i	w_i	**Models**	AIC_c	Δ_i	w_i
a) observed heterozygosity				*b) nucleotide diversity (pi)*			
imprv2km	-85.98	0	0.25	imprv1km	-107.19	0	0.69
imprv1km	-85.33	0.65	0.18	pop1km + imprv1km	-104.25	2.93	0.16
imprv1.5km	-85.18	0.8	0.17	imprv2km	-101.54	5.65	0.04
pop1km + imprv1km	-84.63	1.36	0.13	imprv1.5km	-100.64	6.55	0.03
pop2km + imprv2km	-84.46	1.52	0.12	pop2km	-99.98	7.21	0.02
pop1.5km+ imprv1.5km	-84.23	1.75	0.1	imprv500m	-99.3	7.89	0.01
imprv500m	-79.89	6.09	0.01	pop1km	-98.99	8.19	0.01
pop2km	-79.47	6.52	0.01	pop1.5km	-98.89	8.29	0.01
pop500m + imprv500m	-78.95	7.04	0.01	pop2km + imprv2km	-98.84	8.34	0.01
intercept	-78.57	7.41	0.01	pop500m	-98.06	9.13	0.01
pop1.5km	-78.44	7.54	0.01	pop1.5km+ imprv1.5km	-97.93	9.25	0.01
pop1km	-78.25	7.73	0.01	pop500m + imprv500m	-97.72	9.47	0.01
pop500m	-76.65	9.34	0	intercept	-96.94	10.25	0
c) number of private alleles				*d) percentage of polymorphic loci*			
pop2km	283.12	0	0.37	imprv1km	-56.11	0	0.43
pop1.5km	283.88	0.76	0.25	pop1km + imprv1km	-53.18	2.93	0.1
imprv2km	285.66	2.54	0.1	pop2km	-53.13	2.98	0.1
pop2km + imprv2km	286.01	2.89	0.09	imprv2km	-52.89	3.22	0.09
imprv1.5km	286.08	2.97	0.08	imprv1.5km	-52.29	3.81	0.06
pop1.5km+ imprv1.5km	286.83	3.71	0.06	pop1.5km	-52.24	3.87	0.06

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pop1km	288.31	5.19	0.03	pop1km	-51.46	4.64	0.04
pop1km + imprv1km	290.47	7.35	0.01	imprv500m	-50.81	5.3	0.03
imprv1km	291.35	8.23	0.01	pop2km + imprv2km	-50.23	5.88	0.02
pop500m	293.88	10.77	0	intercept	-49.89	6.22	0.02
intercept	295.9	12.78	0	pop500m	-49.75	6.36	0.02
imprv500m	296.05	12.93	0	pop1.5km+ imprv1.5km	-49.5	6.61	0.02
pop500m + imprv500m	296.22	13.1	0	pop500m + imprv500m	-48.66	7.44	0.01

933

934

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

935 **FIGURE LEGENDS**

936 **Figure 1**

937 **a.** Geographic locations of the 23 sites at which we sampled white-footed mice for this study.
938 The map contains land cover categories and impervious surface levels from the 2011 National
939 Land Cover Database at 30m resolution. Red to purple colors represent increasing percentages
940 of impervious surface cover. Site abbreviations correspond to Table 1. **b.** Scatterplot of percent
941 impervious surface cover vs. human population size for the 23 sampling sites. Both variables
942 were log-transformed to improve linearity, and were measured using 2 km buffers around study
943 sites as described in the text. Red dots represent rural sites, green dots represent suburban sites,
944 and black dots represent urban sites.

945

946 **Figure 2.** Scatterplots and trend lines for best GLMs identified using AIC_c modeling that
947 describe the relationship between urbanization and **a)** heterozygosity, **b)** nucleotide diversity, **c)**
948 percent polymorphic loci, and **d)** number of private alleles for the 23 populations. Red dots
949 represent rural sites, green dots represent suburban sites, and black dots represent urban sites.
950 Site abbreviations correspond to Table 1.

951

952 **Figure 3.** Scatterplots resulting from Discriminant Analysis of Principal Components (DAPC)
953 for **a)** all 23 sampling sites, and **b)** a subset of sites occurring on Manhattan and mainland North
954 America. Insets represent the eigenvalues of retained principal components (top left), and the
955 eigenvalues of discriminant functions portrayed in the scatterplots (bottom left). Site
956 abbreviations correspond to Table 1.

957

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4. Bar plots resulting from evolutionary clustering analyses using a) fastSTRUCTURE assuming K = 2, and b-e) using ADMIXTURE assuming K= 2 – 5. Site abbreviations correspond to Table 1.

Figure 5. Posterior predictive sampling with 1,000 simulated datasets in BEDASSLE, using pairwise F_{ST} as a summary statistic for a) the simple model with presence within or outside NYC as the environmental variable; b) the beta binomial model for presence within or outside NYC; and c) the beta binomial model with resistance distances based on percent impervious surface as the environmental variable. d) Map of study sites with points scaled using the values of F_K estimated by the beta binomial model in BEDASSLE. F_K estimates the deviation of each population from global mean allele frequencies, and was highest for populations located within NYC.

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

971 **LIST OF SUPPLEMENTARY INFORMATION**

972 **Table S1.** Summary population genomic statistics calculated for all nucleotide positions (variant
973 and fixed).

974

975 **Figure S1**. Cross-validation (i.e. a-score optimization) to identify the optimal number of
976 principal components to retain for DAPC without overfitting4

977

978 **Figure S2.** Compoplot / bar plot result from DAPC analysis on all 23 populations.

979

980 **Figure S3**. Scatterplot of first two discriminant functions from DAPC for populations on Long
981 Island.

982

983 **Figure S4.** Cross-validation of results from ADMIXTURE for K = 1 – 12.

984

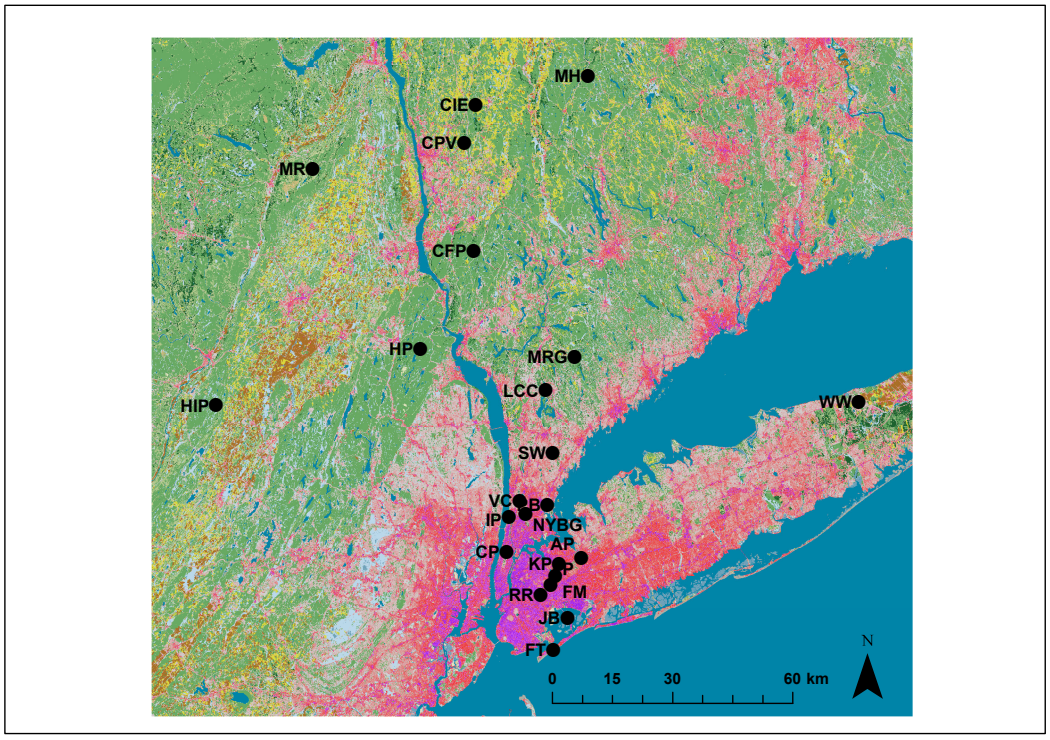
985 **Table S2.** Matrix of pairwise F_{ST} (above diagonal) and great-circle geographic distance (km;
986 below diagonal) calculated between all pairs of 23 populations.

987

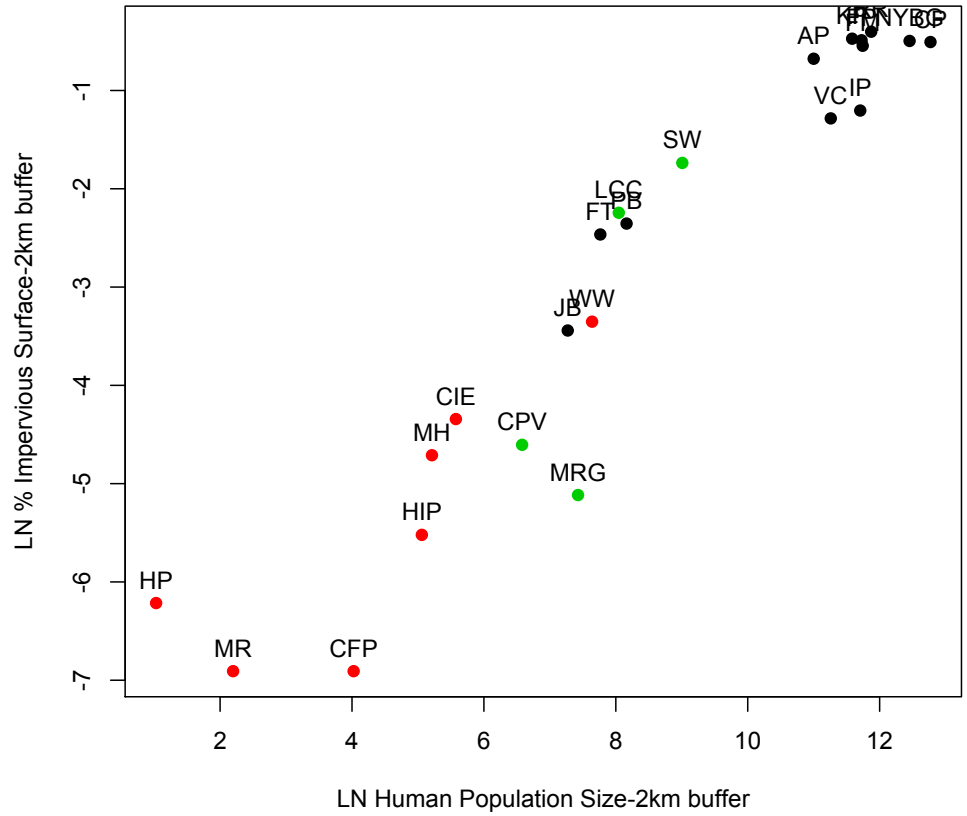
988 **Figure S5**. Cumulative current map from isolation by resistance modeling in Circuitscape.
989 Lighter areas represent landscape cells with higher cumulative predicted current (i.e. higher
990 movement). All landscape cells with impervious surface % > 70% were assigned resistance =
991 100, and all cells with impervious surface < %70 were assigned a resistance equal to their
992 percent impervious surface (e.g. 50% impervious surface results in 50X higher resistance than
993 1% impervious surface).

bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

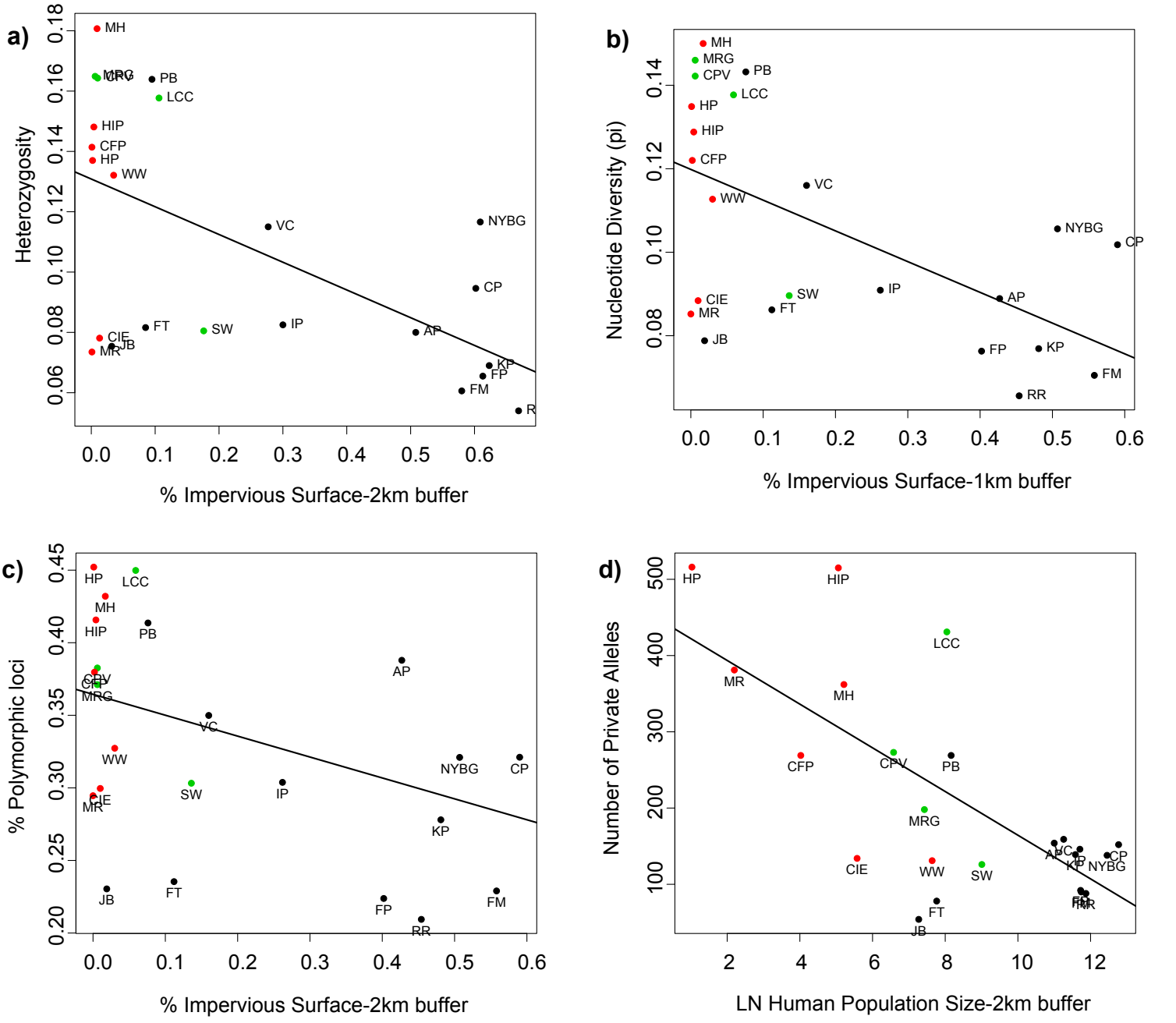
a)



b)

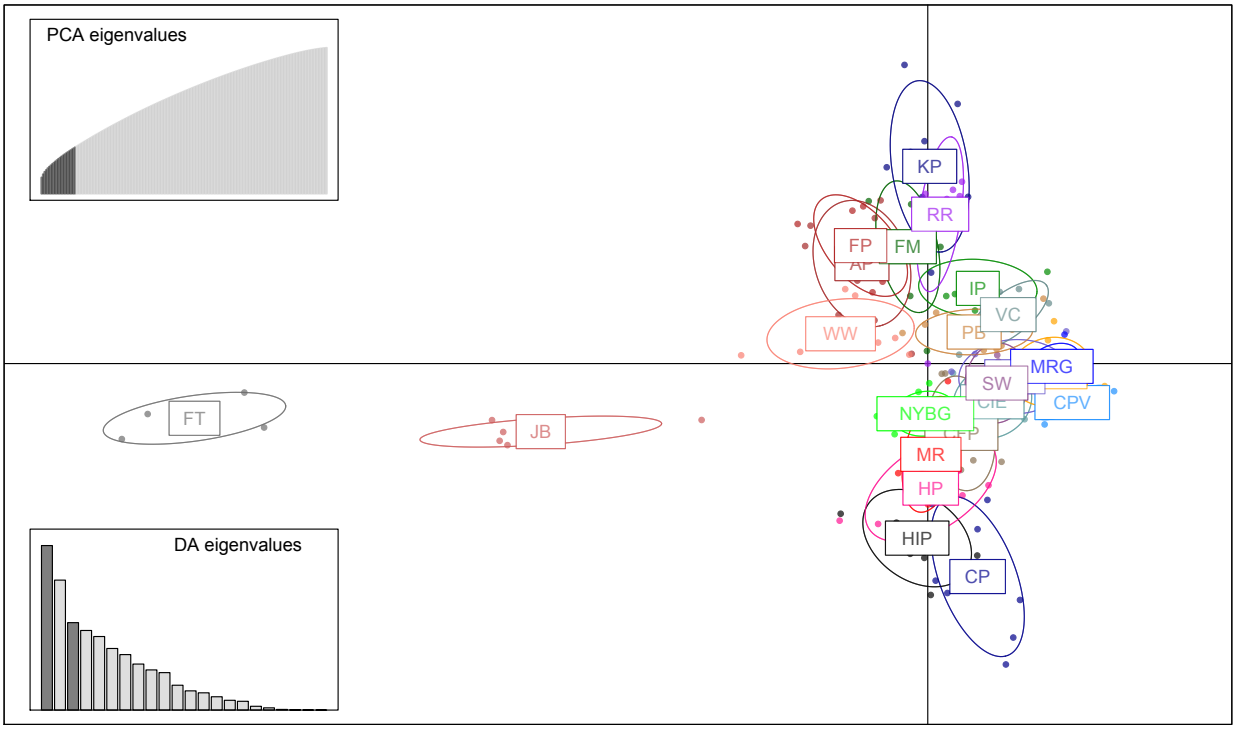


bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

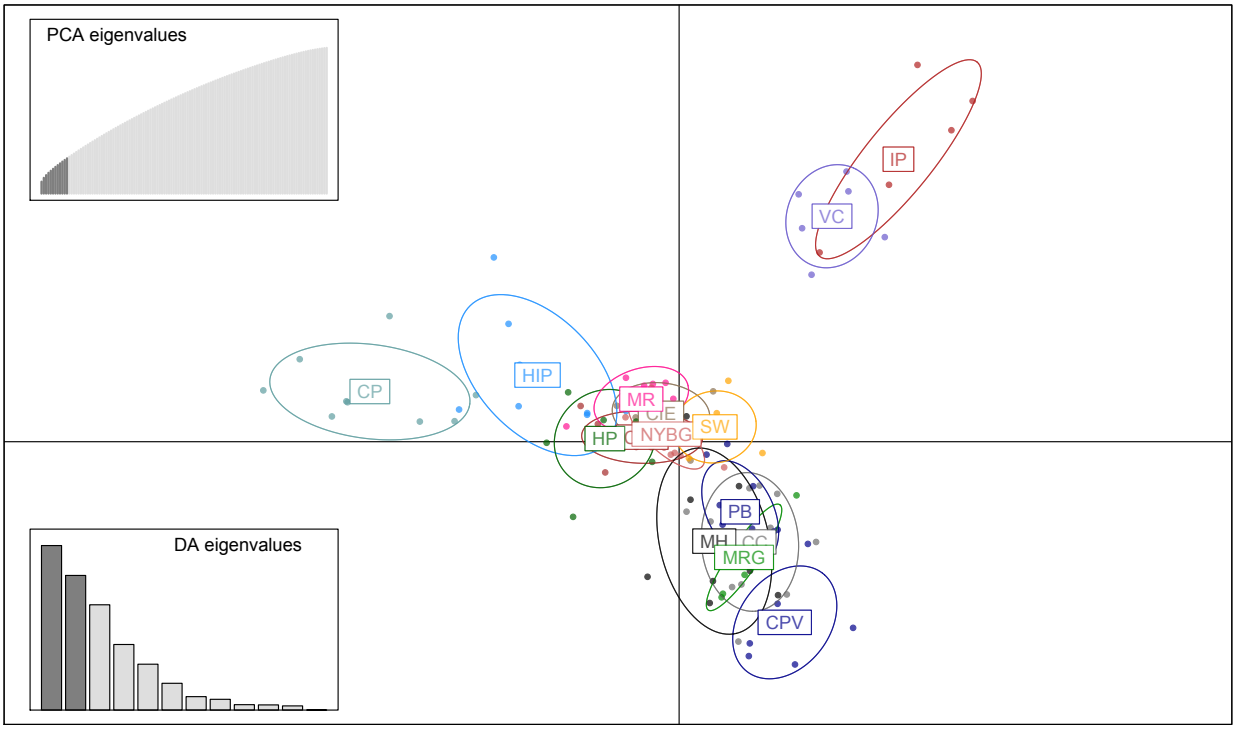


bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

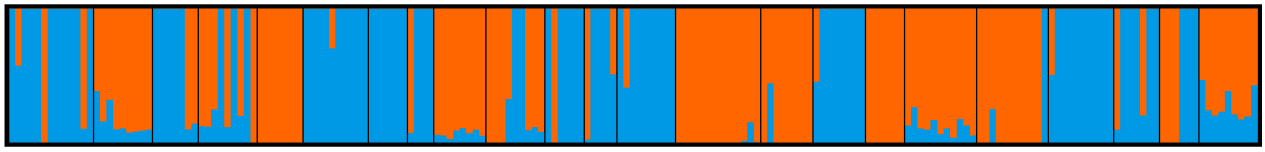
a)



b)

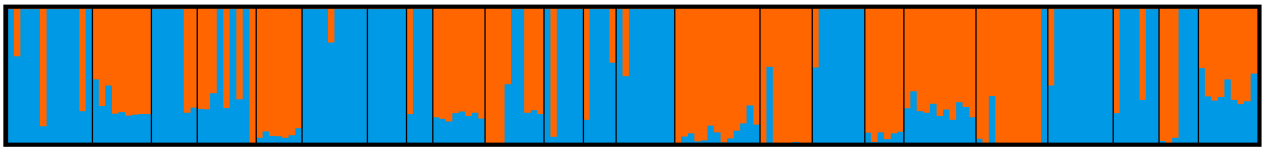


a) K=2 fastSTRUCTURE

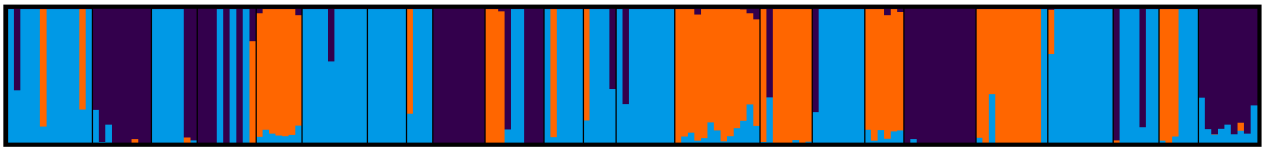


bioRxiv preprint doi: https://doi.org/10.1101/025007; this version posted August 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

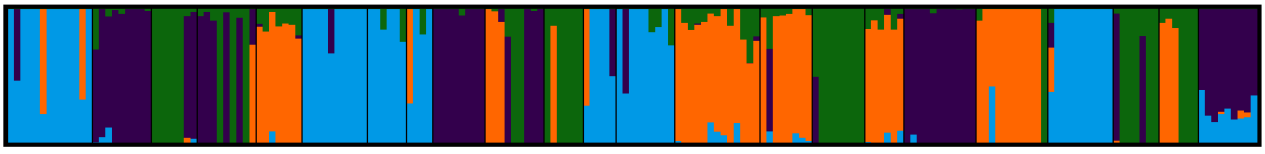
b) K=2 ADMIXTURE



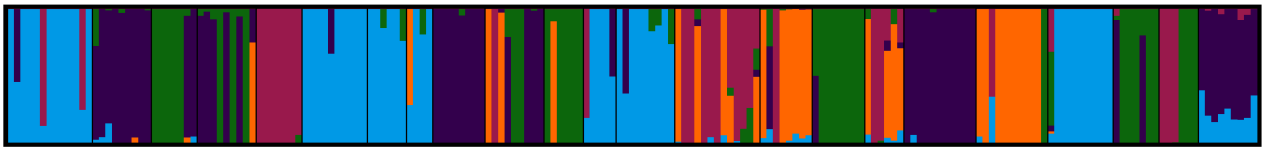
c) K=3

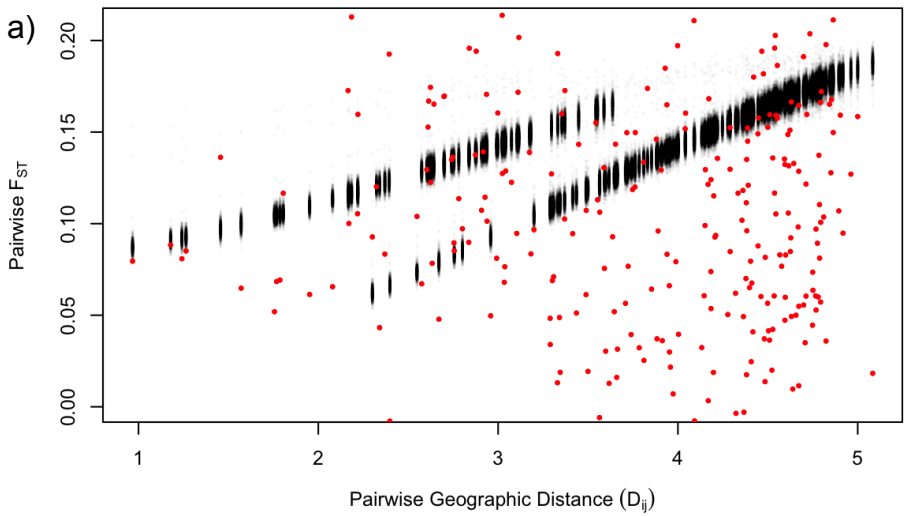


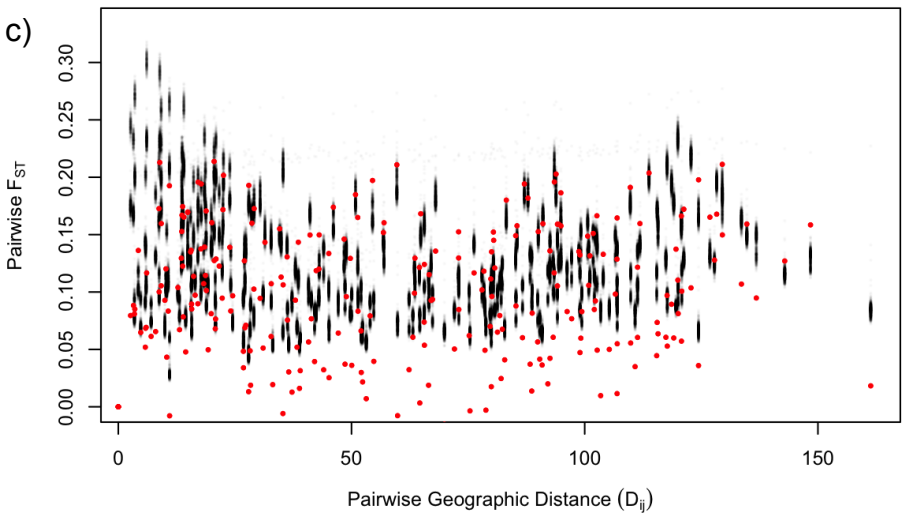
d) K=4

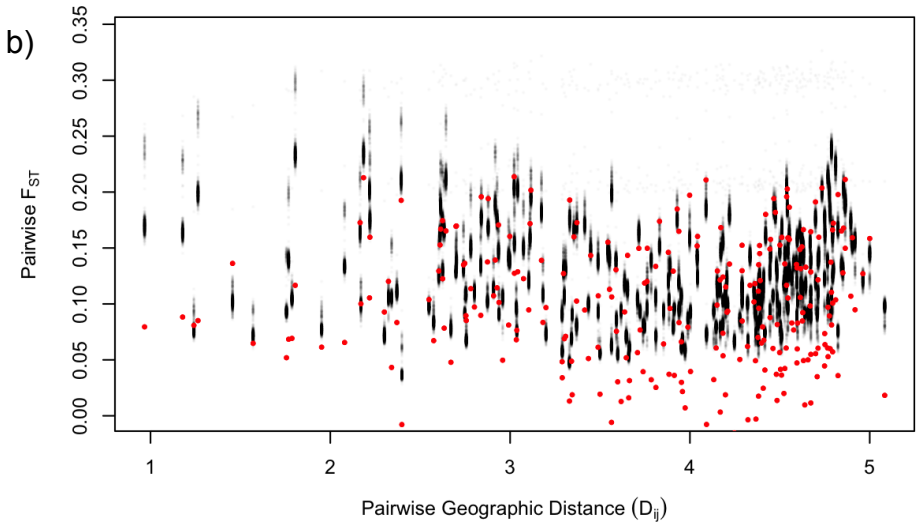


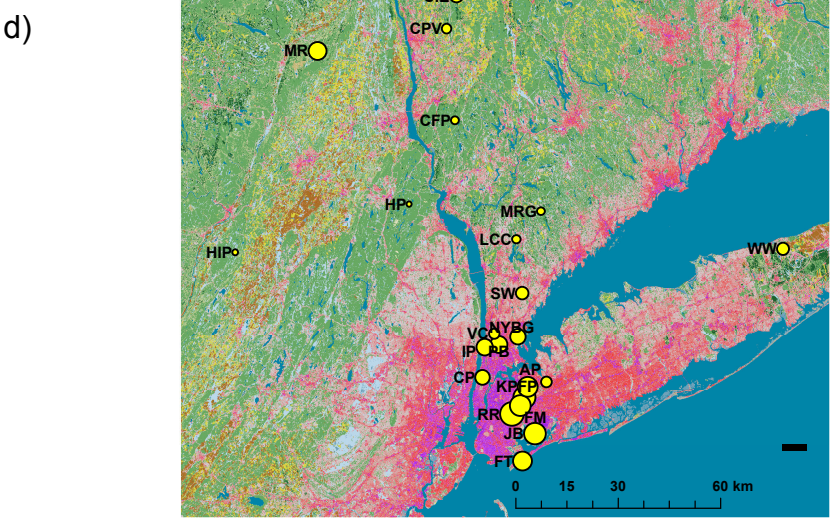
e) K=5




a)
Pairwise F_{ST}
0.00
0.05
0.10
0.15
0.20
1
2
3
4
5
Pairwise Geographic Distance (D_{ij})


c)
Pairwise F_{ST}
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
50
100
150
Pairwise Geographic Distance (D_{ij})


b)
Pairwise F_{ST}
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
1
2
3
4
5
Pairwise Geographic Distance (D_{ij})


d)
MH
CIE
CPV
MR
CFP
HP
MRG
LCC
HIP
WW
SW
VC
NYBG
IP
PB
AP
CP
KP
RR
FM
JB
FT
0
15
30
60 km