**Article**

# Signatures of selective sweeps in urban and rural white clover populations

James S. Santangelo[1], Marc T. J. Johnson[2,3,4], Rob W. Ness[2,3,4]

[1] Department of Integrative Biology, University of California Berkeley, Berkeley, California, USA, 94720

[2] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S 3G5, Canada

[3] Department of Biology, University of Toronto Mississauga, ON M5L 1C6, Canada

[4] Center for Urban Environments, University of Toronto Mississauga, ON M5L 1C6, Canada

*Correspondence: james.santangelo37@gmail.com*

## Data Accessibility

All code used to produce the manuscript's results can be found on the GitHub page for JSS (https://github.com/James-S-Santangelo/toronto_gwsd) and additionally archived on Zenodo (will provide final DOI upon acceptance). Raw reads for all samples have been deposited on GenBank (BioProject PRJNA1179961).

## Author Contributions

**JSS**: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - Original draft, Writing - Review & Editing, Visualization, Project administration **MTJJ**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Visualization, Supervision, Funding Acquisition **RWN**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Visualization, Supervision, Funding Acquisition

# Funding

This work was supported by a Natural Sciences and Engineering Research Council (NSERC) Discovery grant (RGPIN/06331-2016) and Canadian Foundation for Innovation John R. Evans Leaders fund (35591) to RWN. It was additionally supported by an NSERC Discovery Grant, Canada Research Chair, and Steacie Fellowship awarded to MTJJ.

# Acknowledgments

We would like to thank Ken A. Thompson for the initial plant collections from which the genomic data in this manuscript were produced. We would also like to thank the computational resources provided by the Digital Research Alliance of Canada.

# Abstract

Urbanization is increasingly recognized as a powerful force of evolutionary change. However, anthropogenic sources of selection can often be similarly strong and multifarious in rural habitats, but these are often ignored in studies of urban evolutionary ecology. Despite numerous examples of phenotypic differentiation between urban and rural populations, we still lack an understanding of the genes enabling adaptation to these contrasting habitats and the genetic architecture underlying urban and rural adaptation. In this study, we conducted whole genome sequencing of 120 urban, suburban, and rural white clover plants from Toronto, Canada. We used these data to identify signatures of selective sweeps across the genome using both SFS and haplotype-based approaches, and characterize the architecture of selective sweeps. We found evidence for selection in genomic regions involved in abiotic stress tolerance and growth/development in both urban and rural populations. Urban and rural populations did not differ in the proportion of hard vs. soft sweeps, though urban populations were characterized by wider sweeps, which may indicate differences in the strength or timescale of selection. In addition, patterns of allele frequency and haplotype differentiation suggest that most sweeps are incomplete. These results highlight how both urban and rural habitats are driving ongoing selection in white clover populations, and motivate future work disentangling the genetic architecture of ecologically important phenotypes, and estimating the strength and timescale of selection underlying adaptation to contemporary anthropogenic habitats.



**Keywords**: Anthropocene, Selection, *Trifolium repens*, WGS

# Introduction

Uncovering the genetic basis and architecture of adaptation to rapid environmental change remains a central goal in evolutionary biology. Urban environments are emerging as "living laboratories" for investigations of contemporary evolution, owing in part to their global replication and their pervasive environmental impacts that have the potential to drive rapid ecological and evolutionary change in natural populations (Diamond and Martin 2021; Johnson and Munshi-South 2017; Santangelo et al. 2018; Szulkin et al. 2020; Verrelli et al. 2022). While altered patterns of gene flow and genetic drift are some of the most frequently cited examples of urban evolutionary change (Carlen and Munshi-South 2021; Fusco et al. 2024; Harris et al. 2016; Lourenço et al. 2017; Miles et al. 2019; Mueller et al. 2018; Munshi-South 2012; Schmidt et al. 2020), there is growing evidence of adaptive phenotypic responses to urbanization. For example, urban heat islands have led to increased thermal tolerance and changes in life-history strategies (Brans et al. 2018; Brans and De Meester 2018; Brans et al. 2017; Diamond et al. 2017, 2018; Martin et al. 2021), while urban structures have driven changes in animal morphology, locomotion, and behaviour (Evans et al. 2012; Winchell et al. 2018; Winchell et al. 2017, 2016). In some cases, the genetic basis underlying phenotypic responses to urbanization are known—for example, the two-locus epistasis underlying urban-rural cyanogenesis clines in white clover (Santangelo et al. 2022), or regulatory changes to the Aryl hydrocarbon receptor (AHR) signaling pathway facilitating resistance of Gulf and Atlantic Killifish populations to toxic urban pollutants (Oziolor et al. 2019; Reid et al. 2016). Despite these examples, most studies have focused on candidate genes or reduced-representation sequencing (e.g., GBS) and thus we still lack a general understanding of the genetic architecture and genome-wide responses underlying adaptation to urban environments, and the strength and timescale of selection over which adaptation has occurred.

During the initial phases of adaptation to a new phenotypic optimum, rapid

adaptation is predicted to occur primarily via changes in loci of large effect (Hayward and Sella 2022). This is likely to be the case is urban environments given the substantial ecological changes they often impose. In addition, the relatively young age of cities suggests that most adaptation is likely to occur via selection on standing genetic variation, as new mutations are unlikely to have had time to accumulate. Genome-wide scans of natural selection—for example, using distortions in the site- or haplotype-frequency spectrum, or identifying regions of extended haplotype homozygosity generated by positive selection—and gene-environment association analyses (GEAs), can help generate hypotheses about putatively adaptive phenotypes, the selective agents underlying adaptive differentiation, and the number and effect sizes of alleles underlying local adaptation (Hoban et al. 2016; Lasky et al. 2023; Weigand and Leese 2018). In an urban context, recent work has identified signatures of selection among urban populations in genes related to metabolism and diet (Harpak et al. 2021; Harris and Munshi-South 2017; Ravinet et al. 2018), neural function and cognition (Mueller et al. 2020; Salmón et al. 2021), thermal physiology and environmental stress response (Campbell-Staton et al. 2020; Theodorou et al. 2018), and tolerance to pollution (Kamdem et al. 2017; Oziolor et al. 2019). While the number of studies investigating molecular adaptation to urban environments is rapidly increasing, very few of these have explicitly characterized the types of selective sweeps (i.e., hard, soft, incomplete, but see Harpak et al. 2021), and none have compared sweep signatures to nearby rural populations.

Studies of urban adaptation frequently compare the phenotypic or genomic responses of urban populations to those of nonurban, often rural, populations. Such studies are therefore primed for investigating urban-rural differences in the genetic basis and architecture of adaptation, though rural-specific signatures of selection are often ignored. Croplands and pastures currently account for approximately 40% of earth's ice-free land (Ellis and Ramankutty 2008; Foley et al. 2005), and anthropogenic sources of selection in these habitats can be strong and multifarious, often

rivaling those observed in urban environments. Indeed, agricultural landscapes have long been recognized as potential hotbeds of ecological changes that can drive the evolution of contemporary populations (Baker 1974; Lau and Funk 2023; Neve et al. 2009; Snaydon and Davies 1972; Turcotte et al. 2017). For example, widespread herbicide use has driven strong and rapid increases in resistance alleles among agriculturally problematic weeds over just the past few decades (Kreiner et al. 2022a,b). While the agents and targets of selection might differ between urban and rural habitats, ignoring signatures of adaptation in rural habitats risks missing an important source of anthropogenic selection and giving the misleading impression that the bulk of contemporary adaptation to anthropogenic disturbances is happening in cities, rather than the agricultural landscapes that support them.

In this paper, we perform genome-wide scans for signatures of selective sweeps to identify loci putatively involved in adaptation to urban and rural environments in and around Toronto, Canada in the plant white clover (*Trifolium repens*). White clover exhibits adaptive urban-rural clines hydrogen cyanide (HCN, an ecologically important plant defense) frequencies in multiple cities around the world (Santangelo et al. 2022). White clover also displays a multivariate cline across the city of Toronto, whereby numerous quantitative traits (e.g., germination time, flower size) are differentiated along an urbanization gradient (Santangelo et al. 2018). We address the following specific questions: (1) Is there evidence of population structure and differentiation among Toronto white clover populations? Addressing this question is a prerequisite to disentangling the signatures of selection from the genome-wide effects of demographic changes and population structure (Hoban et al. 2016). We then ask: (2) Is there evidence of selective sweeps in urban and rural populations, and if so, which genes show evidence of selection? Finally, we ask: (3) What types of sweeps (i.e., hard, soft, incomplete) underlie urban and rural signatures of selection?

# Methods

## Sampling, library preparation, and sequencing

We sequenced a total of 120 plants from 38 populations along three urbanization transects (Figure 1), all of which have been previously associated with phenotypic clines in cyanogenesis (Santangelo et al. 2020b; Thompson et al. 2016), and one of which shows a multivariate cline in multiple quantitative traits (Santangelo et al. 2020a). We sampled 50 plants from 5 urban populations and 45 plants across 21 rural populations, with plants and populations split equally among the three transects (Figure 1). We additionally sequenced 30 plants spread across 12 suburban populations, which ranked just below our urban-most populations in Human Influence Index (Figure 1 inset). For each plant, we extracted total DNA using a modified CTAB extraction protocol (Santangelo et al. 2022) and generated dual-indexed Illumina libraries using an in-house library preparation protocol (Santangelo et al. 2022). The 120 plants were randomly split into two batches and each batch was sequenced on a single lane of Novaseq S4 using $2 \times 150$ bp paired-end reads for an expected sequencing depth of $\approx$13X per sample.

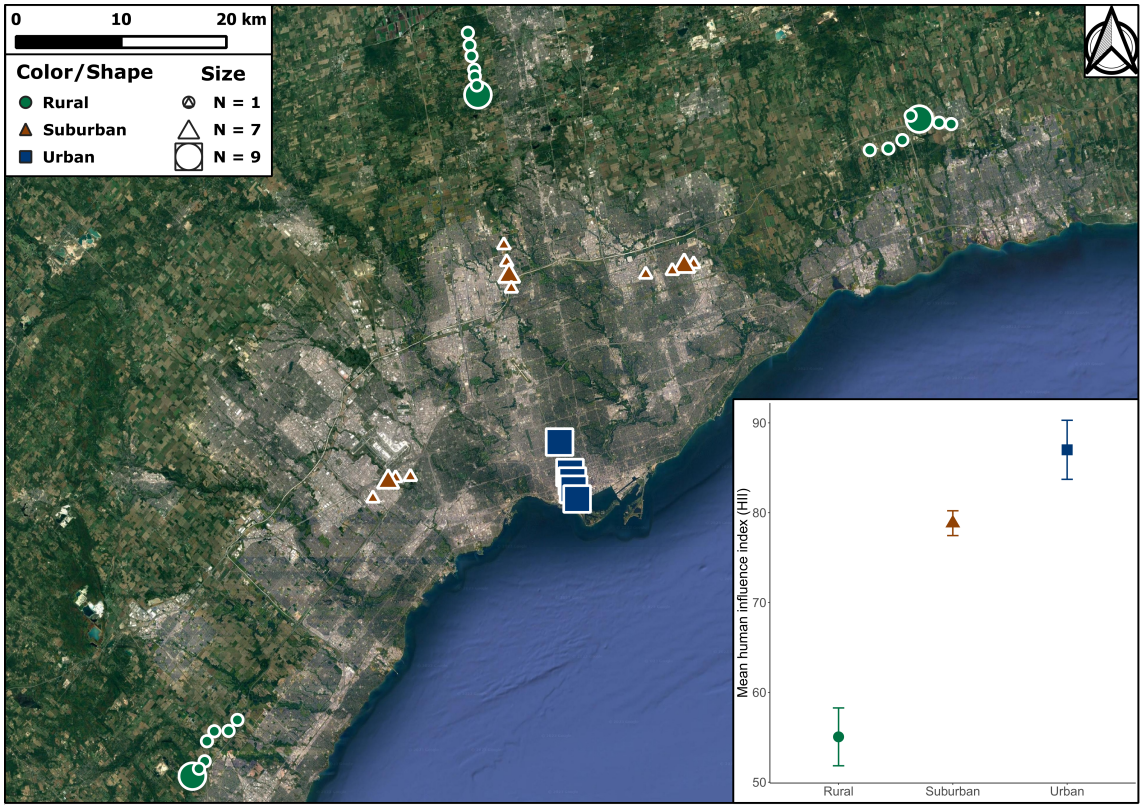


Figure 1: Map of 38 white clover populations sampled for sequencing in Toronto, ON. Sampled populations contained either 1, 7, or 9 samples from rural (green circles), suburban (brown triangles), or urban (blue squares) populations. Inset: Mean Human Influence Index (± 1SE) of rural, suburban, and urban populations.

## Sequence alignment and QC

All analyses that follow are available as a publicly-accessible Snakemake pipeline (v.7.21.0, Mölder et al. 2021) on the GitHub page for JSS (https://github.com/James-S-Santangelo/toronto_gwsd).

We trimmed raw reads using *fastp* v0.23.2 (Chen et al. 2018) with the *-trim_poly_g* argument to trim polyG tails that are commonly added to the ends of reads by Novaseq platforms. We performed per-sample QC of both raw and trimmed reads using *FastQC* v0.11.9 (Andrews et al. 2010) and mapped the trimmed reads to the the recently generated *Trifolium repens* haploid mapping assembly (GenBank accession GCA_030408175.1, Santangelo et al. 2023) using *BWA-MEM2* v2.2.1 (Li and Durbin 2009; Vasimuddin et al. 2019). Following read mapping, we marked duplicate reads and indexed and coordinate-sorted resulting BAM files using *SAMtools* v1.16.1 (Li et al. 2009) prior to performing QC of mapped reads using *Qualimap* v2.2.2 (Okonechnikov et al. 2016), *Bamtools* v2.5.2 (Barnett et al. 2011), and *multiQC* v1.14 (Ewels et al. 2016). QC of mapped reads revealed three samples with uncharacteristically high alignment error rates and analysis of *rbcL* and *matK* chloroplast genes confirmed that these were *Medicago lupulina*. These samples were removed from all downstream analyses. We removed an additional two samples that had very low sequencing depth (≈1-2X) and unusually high GC% for this species, resulting in 115 high quality samples remaining. Finally, analysis of pairwise relatedness among samples (see below) revealed six sample-pairs with high identity-by-descent ($r_{xy} > 0.5$); we removed one sample from each of these pairs from downstream analyses resulting in a total loss of 11 samples and final sample size of 109 (41 urban, 41 rural, and 27 suburban) with realized sequencing depth of 11.2X.

## Relatedness, population structure, & differentiation

We used genotype likelihoods, which incorporate uncertainty in genotype assignment, for inference of population structure and relatedness to avoid biases associated

185 with calling genotypes in low to medium coverage data (Han et al. 2013; Nielsen et al.
186 2012, 2011). In addition, we used LD-pruned, 4-fold degenerate sites as putatively
187 neutral sites for these analyses. We began by extracting genome-wide 4-fold sites
188 using *degenotate* v1.2.1 (https://github.com/harvardinformatics/degenotate). We
189 then used *ANGSD* v.0.938 (Korneliussen et al. 2014) to generate Beagle-formatted
190 genotype likelihoods (-doGlf 2) for all 4-fold SNPs (-sites 4-fold.sites, -SNP_pval
191 1e-6) using the *SAMtools* genotype likelihood model (-GL 1, Li 2011a) with base
192 alignment quality scores recalculated according to the "extended *SAMtools*" model
193 (-baq 2) to reduce false-positive SNP calls around INDELS (Li 2011b). We only
194 considered reads with a minimum phred-scaled mapping quality of 30 (-minMapQ
195 30) and sites with a mean base quality of 20 (-minQ 20), minor allele frequency
196 over 0.05 (-minMaf 0.05), and where at least 80% of sequenced individuals (-minInd
197 0.8*N) had at least three reads (-setMinDepthInd 3). We additionally ignored sites
198 with more than 2600X (-setMaxDepth 2600. $\approx 2 \times$ mean coverage $\times$ # samples)
199 to avoid calling SNPs in highly repetitive genomic regions. This resulted in a total
200 of 407,960 4-fold SNPs. We then estimated pairwise LD among these 4-fold SNPs
201 within a 20Kb distance—approximately twice the known breakdown of LD in this
202 system, (Inostroza et al. 2018; Kuo et al. 2024b)—using *ngsLD* v.1.1.1 (Fox et al.
203 2019) and pruned SNPs using the *prune_graph.pl* script packaged with *ngsLD* al-
204 lowing a max $r^2$ of 0.2 (–min_weight 0.2). This reduced the total number of 4-fold
205 SNPs to 199,297 ($\approx$49%), which were used for inference of population structure and
206 relatedness.

207 We estimated pairwise relatedness among all 115 high quality samples using
208 *ANGSD* and *NGSrelate* v.2.0 (Hanghøj et al. 2019). We first used *ANGSD* to re-
209 generate genotype likelihoods at all LD-pruned 4-fold SNPs in binary format (-doGlf
210 3) using the same filtering criteria as above. We then used *NGSrelate* to estimate
211 pairwise relatedness among all samples, with separate chromosomes processed in
212 parallel. We quantified relatedness using $r_{xy}$, which represents the proportion of ho-

mologous alleles shared between two individuals that are identical-by-descent, and is robust to the presence of inbreeding (if present) (Hedrick and Lacy 2015).

We examined patterns of population structure using principle components analysis (PCA), uniform manifold approximation projection (UMAP), and admixture analyses. First, we used *PCAngsd* v.0.99 (Meisner and Albrechtsen 2018) to estimate the variance-covariance matrix in allele frequencies using the LD-pruned 4-fold SNPs (MAF $> 0.05$) as input. We used this matrix as input to the *princomp* function in R v.4.1.3 (R Core Team 2020) to perform a PCA. Because the first two PCs explained $\approx$5% of the variation in allele frequencies and showed little evidence of clustering, we additionally performed a uniform manifold approximation and projection (UMAP, McInnes et al. 2018). This non-linear dimensionality reduction technique that is gaining popularity in population genetics for its ability to reveal fine-scale population structure and identify clusters in otherwise weakly structured datasets (Diaz-Papkovich et al. 2021). To perform the UMAP, we first converted the variance-covariance allele frequency matrix to a distance matrix using the *dist* function. We then used this as input to the *umap* function from the *umap* v.2.8 R-package with the recommended settings of 15 nearest neighbors, a minimum distance of 0.1, and 40 components, the latter of which corresponds to the number of significant PCs from the PCA according to a broken-stick model (Jackson 1993).

We performed admixture analyses analyses using *NGSAdmix* v0.933 (Skotte et al. 2013). Using the LD-pruned 4-fold SNPs as input, we ran *NGSAdmix* with ten different random seeds for each assumed number of ancestral clusters (K) between 2 and 10. We then used Evanno's $\Delta$K method (Evanno et al. 2005) implemented in CLUMPAK v.1.1 (Kopelman et al. 2015) to identify the uppermost level of population structure.

We estimated pairwise nucleotide diversity ($\theta_\pi$) within each habitat (urban, rural, suburban) and pairwise differentiation ($F_{ST}$) between habitat pairs to complement analyses of population structure. One- and two-dimensional folded SFS were es-

timated in *ANGSD*. We first generated separate urban, rural, and suburban Site Allele Frequency (SAF) likelihood files (-doSaf 1) using the same filtering criteria as for the estimation of genotype likelihoods, but with the following modifications: (1) We only required 60% of samples at a site to be covered by one read (-minInd 0.6 $\times$ N, -setMinDepthInd 1); (2) We did not impose a MAF cutoff to avoid biasing the SFS; and (3) We polarized alleles by forcing the reference base to be the major allele in each habitat (-doMajorMinor 4 with -ref) to ensure the same major and minor allele calls in each habitat for $F_{ST}$ estimates. We then used the *realSFS* script that is packaged with *ANGSD* to estimate folded (-fold 1) one-dimensional urban, suburban, and rural SFS,which were used to estimate per-habitat pairwise nucleotide diversity ($\theta_\pi$), and two-dimensional urban-rural, urban-suburban, or rural-suburban joint SFS, which were used for estimating Hudson's $F_{ST}$ (realSFS fst index -whichFst 1, Bhatia et al. 2013; Hudson et al. 1992).

## Signatures of selective sweeps

Given the weak signals of population structure (see Results and Discussion), we pooled samples by habitat when searching for signatures of selective sweeps. We focused on comparing selection between urban and rural populations because they represented the extremes of our environmental comparisons (Lotterhos and Whitlock 2015), though we note that urban sweep signatures were largely shared with suburban populations (Text S1, Figure S1). It is worth noting that the approaches described below are designed for detecting signatures of positive selection in either urban or rural habitats and are thus not meant to detect regions of divergent selection where alternative alleles in the same region are under selection in urban and rural populations.

To identify urban and rural signatures of selective sweeps, we used approaches based on extended haplotype homozygosity (XP-nSL, nSL, iHS, and iHH12), as well as those based on distortions in the site (i.e., $F_{ST}$) or haplotype (*saltiLassi*'s $\lambda$)

frequency spectrum. We began by calling variants with *freebayes* v.1.3.6 (Garrison and Marth 2012), ignoring regions with greater than 15,000X total coverage (–skip-coverage 15000) to avoid calling variants in highly repetitive regions. We then filtered VCFs using *bcftools* v.1.16 (Danecek et al. 2021) to include only biallelic SNPs with no missing data, phred-scaled quality scores over 30 ($QUAL > 30$), and additionally removed sites where the number (allele balance, $AB >= 0.25$ & $AB <= 0.75$ | $AB <= 0.01$), mapping quality ($MQM >= 30$ & $MQMR >= 30$), and paired status [$(PAIRED > 0.05$ & $PAIREDR > 0.05$ & $\frac{PAIREDR}{PAIRED} < 1.75$ & $\frac{PAIREDR}{PAIRED} > 0.25)$ | $(PAIRED < 0.05$ & $PAIREDR < 0.05)$] of reads supporting the reference and alternate alleles were substantially different. Post filtering, 4,507,505 biallelic SNPs remained for downstream analyses.

Following filtering, we phased SNPs for downstream haplotype-based scans of selective sweeps. We first performed read-backed phasing using *WhatsHap* v1.7 (Martin et al. 2016) with default parameters. We then used these phase sets with SHAPEIT4 v4.2.2 (Delaneau et al. 2019) to perform population level phasing (–use-PS 0.0001), which additionally used an interpolated recombination map as input. The recombination map was generated by mapping markers from two recently generated linkage maps (Olsen et al. 2021) to the white clover reference genome, removing any markers that mapped to the wrong chromosomes, and fitting shape-constrained cubic p-splines with 20 knots to the chromosomal Marey maps (Figure S2). Models were fit using the *scam* function in the *SCAM* v1.2 (Pya and Wood 2015) R package. These models were then used to interpolate the genetic distance for each biallelic SNP retained post-filtering.

We used *selscan* v1.3.0 (Szpiech and Hernandez 2014) to estimate urban-rural cross-population extended haplotype by length (XP-nSL), a haplotype-based statistic with improved power to detect partial and incomplete hard and soft selective sweeps, and which is more robust to genome-wide recombination rate variation than previous approaches (e.g., XP-EHH) (Szpiech et al. 2021). We used the rural habitat

296 as the "reference" population, and the urban habitat used as the "object" popula-
297 tion. As such, positive XP-nSL scores reflect positive selection occurring in the
298 urban habitat, whereas negative scores reflect selection in the rural habitat. Af-
299 ter estimating per-site XP-nSL scores, we used *norm* v1.3.0 that is packaged with
300 *selscan* to normalize genome-wide XP-nSL scores. In 50Kb non-overlapping win-
301 dows, we then estimated the mean normalized XP-nSL score and the proportion of
302 normalized scores $> 2$ (or $< -2$). Outlier windows for positive selection in urban
303 habitats were defined as those in the top 1% of the genome-wide empirical mean
304 XP-nSL distribution, and the top 1% of the distribution summarizing the proportion
305 of scores $> 2$ (Szpiech et al. 2021). Similarly, we defined outlier windows for positive
306 selection in rural habitats as those in the bottom 1% of the genome-wide empirical
307 mean XP-nSL distribution (i.e. the most negative scores), and the top 1% of the
308 distribution summarizing the proportion of scores $< -2$. To avoid spurious signals
309 driven by noisy XP-nSL estimates in regions of low diversity, we excluded windows
310 with fewer than 50 SNPs (14.1% of windows, Figure S3). We complemented the
311 XP-nSL analysis by additionally estimating single-population urban and rural seg-
312 regating sites by length (nSL, Ferrer-Admetlla et al. 2014), integrated haplotype
313 score (iHS, Voight et al. 2006), and iHH12 (Torres et al. 2018), an integrated ver-
314 sion of the H12 statistic that combines the two most frequent haplotypes and has
315 improved power for detecting soft sweeps (Garud et al. 2015).

316 Using the tails of empirical distributions as a way of identifying the genomic
317 regions that have experienced selection is common in the literature but has received
318 some criticism, in part, because there is no guarantee that selected loci will occur in
319 the upper tail of empirical distributions, or that loci in the upper tail have experi-
320 enced selection (Johri et al. 2022). A complimentary approach would be to simulate
321 a null distribution of XP-nSL under the most likely demographic model for our
322 samples and compare our observed values to this null. However, the near absence
323 of population structure between urban and rural white clover populations makes

identifying such a model challenging. Instead, we leverage this absence of structure and randomly permuted urban and rural samples 1,000 times, each time estimating XP-nSL in 50 kb windows across the genome, and compared our observed windowed XP-nSL values to this permuted null. We confirmed that the correlation between observed and permuted windowed mean XP-nSL scores were centered around zero across all 1000 permutations (Figure S4). For both the mean windowed XP-nSL score and the proportion of outlier scores within a window, the observed values were in the upper tail of the permuted null, providing further evidence that these are likely candidates of selection (Figure S5).

In addition to approaches based on extended haplotype homozygosity, we leveraged the fact that positive selection creates distortions in both the site and haplotype frequency spectra (i.e., SFS and HFS). We detected distortions in the SFS by estimating $F_{ST}$ in the same 50 Kb windows as our XP-nSL analyses, and considered windows in the top 1% of the genome-wide empirical $F_{ST}$ distribution as putative candidates of selection. $F_{ST}$ was estimated using the filtered VCF above in *Pixy* v.1.2.8.beta1 (Korunes and Samuk 2021), which correctly accounts for the presence of invariant sites and missing data in the calculation of $F_{ST}$.

Finally, we searched for evidence of positive selection and further characterized selective sweeps in urban and rural habitats using the single-population *saltiLassi* $\lambda$ statistic, a statistic based on distortions in the HFS (DeGiorgio and Szpiech 2022). Using the phased VCF above as input, we estimated $\lambda$ using *LASSI-Plus* v.1.1.1 (DeGiorgio and Szpiech 2022) with a window size of 201 SNPs since larger windows have more power to detect recent selection (DeGiorgio and Szpiech 2022), a step size of 100 SNPs, and using the top K = 20 haplotypes. We considered windows as putative outliers if they ranked in the top 1% of the genome-wide empirical $\lambda$ distribution. In addition to $\lambda$, *LASSI-Plus* reports the number of sweeping haplotypes in each window ($m$) and the inferred width of the sweep ($A$), which we leverage to characterize selective sweeps in urban and rural habitats. Specifically,

we intersected windows with evidence of positive selection in urban habitats from the above XP-nSL analysis with *saltiLassi* $\lambda$ windows showing evidence of selection in urban, but not rural, populations (i.e., windows in top 1% of urban distribution but not the top 1% of the rural distribution). Similarly, we intersected windows with evidence of positive selection in rural habitats with *saltiLassi* $\lambda$ windows showing evidence of selection in rural, but not urban, populations. Among these intersected urban and rural *saltiLassi* $\lambda$ windows, we classified sweep signatures as hard if the estimated number of sweeping haplotypes ($m$) was 1, or soft if it was greater than 1, as suggested by (DeGiorgio and Szpiech 2022) since the raw estimates of $m$ may be misleading. We then performed a Chi-squared test of independence to test if type of sweep (i.e., soft vs. hard) differed between urban and rural habitats. In addition, we compared the observed urban-rural difference in the width of selection sweeps (i.e., $log_{10}(A)$) to a null distribution generated by randomly permuting the urban and rural window labels 1000 times, each time estimating the urban-rural difference in $log_{10}(A)$.

## Functional characterization of selected regions

To examine candidate genes that may be targets of positive selection in urban and rural populations, we intersected our urban and rural XP-nSL outlier windows with the annotation file (GFF3) accompanying the recently published white clover reference genome (Santangelo et al. 2023). We focus only on regions that additionally fall in the top 5% of the urban-rural permuted XP-nSL distribution. In addition, we performed a Gene Ontology analysis for Biological Processes and Molecular Functions separately for urban and rural selected regions using the *topGO* v.2.50.0 (Alexa and Rahnenfuhrer 2016) R package with Fisher's exact test, the "weight01" algorithm and a *P-value* cutoff of 0.05.

# Results and Discussion

## Genetic diversity, differentiation, and population structure

Urban, rural, and suburban habitats showed identical levels of pairwise nucleotide diversity at four-fold degenerate sites ($\pi_{Rural} = \pi_{Suburban} = \pi_{Urban} = 0.0187$). There was little evidence of population structure from the PCA, whereby the first two PCs together accounted for ≈5.4% of genetic variation across samples (PC1 = 3.4%, PC2 = 2%) and, with the exception of one rural and one urban population, showed no evidence of clustering by habitat (Figure 2a). Similarly, the only evidence of substructure visible in the UMAP was one urban and one rural population that formed their own clusters, though most samples formed a single large, overlapping cluster (Figure 2b). Admixture analyses and Evanno's ΔK method identified K = 5 as the most likely number of ancestral genetic clusters, and with the exception of the same urban and rural population as in the above PCA, all clusters were well represented in each habitat regardless of the value of K (Figure 2c). The low levels of population structure among Toronto white clover samples was further supported by low levels of genetic differentiation at four-fold sites between habitat types (Rural-Suburban $F_{ST} = 0.0060$, Rural-Urban $F_{ST} = 0.0063$, Urban-Suburban $F_{ST} = 0.0062$).

Spatially-restricted gene flow combined with local genetic drift can lead to rapid population differentiation and strong patterns of population structure. Indeed, habitat fragmentation and reductions in population size associated with urbanization are frequently cited drivers of reduced genetic diversity in urban populations and/or increased urban-rural genetic differentiation in rodents (Combs et al. 2018; Harris et al. 2016; Munshi-South 2012; Munshi-South and Kharchenko 2010), lizards (Beninde et al. 2018, 2016; Littleford-Colquhoun et al. 2017; Lourenço et al. 2017; Noël et al. 2007), birds (Björklund et al. 2010; Mueller et al. 2018; Unfried et al. 2013), mammals writ-large (Schmidt et al. 2020), and plants (Bartlewicz et al. 2015;

Rivkin and Johnson 2022; Yakub and Tiffin 2017). However, a recent meta-analysis found no consistent effects of urbanization on within-population genetic diversity and between-population differentiation, and instead suggested that patterns are largely system-dependent since not all species are equally affected by urban fragmentation (Miles et al. 2019). Toronto white clover appears to be a particularly extreme case, with high genetic diversity (genome-wide four-fold $\pi \approx 0.0187$) in all habitats, overall weak urban-rural population structure, and high levels of admixture driving low differentiation (all pairwise $F_{ST} \approx 0.006$), consistent with previous work showing generally low levels of differentiation and urban-rural structure in this system (Caizergues et al. 2024; Johnson et al. 2018; Santangelo et al. 2022). White clover is a perennial, obligate outcrosser that maintains large populations, and is likely repeatedly introduced into cities from multiple sources due its use in agriculture and presence in turf mixtures (Kjærgaard 2003; Sincik and Acikgoz 2007). In addition, while gene flow via pollinator-mediated pollen dispersal likely occurs within a limited geographical range in this system (De Lucas et al. 2012; Løjtnant et al. 2012), white clover seeds are among the most commonly detected species found on human clothing, vehicles, and animals (Pickering and Mount 2010). Therefore, longer-distance human- or animal-mediated seed dispersal may be particularly important in this system, and this is likely more pronounced in and around cities where the dispersal agents (e.g., humans) are more common. In short, white clover is unlikely to have its population sizes reduced by urban fragmentation, and instead likely maintains high population connectivity and diversity in Toronto and other cities around the world.

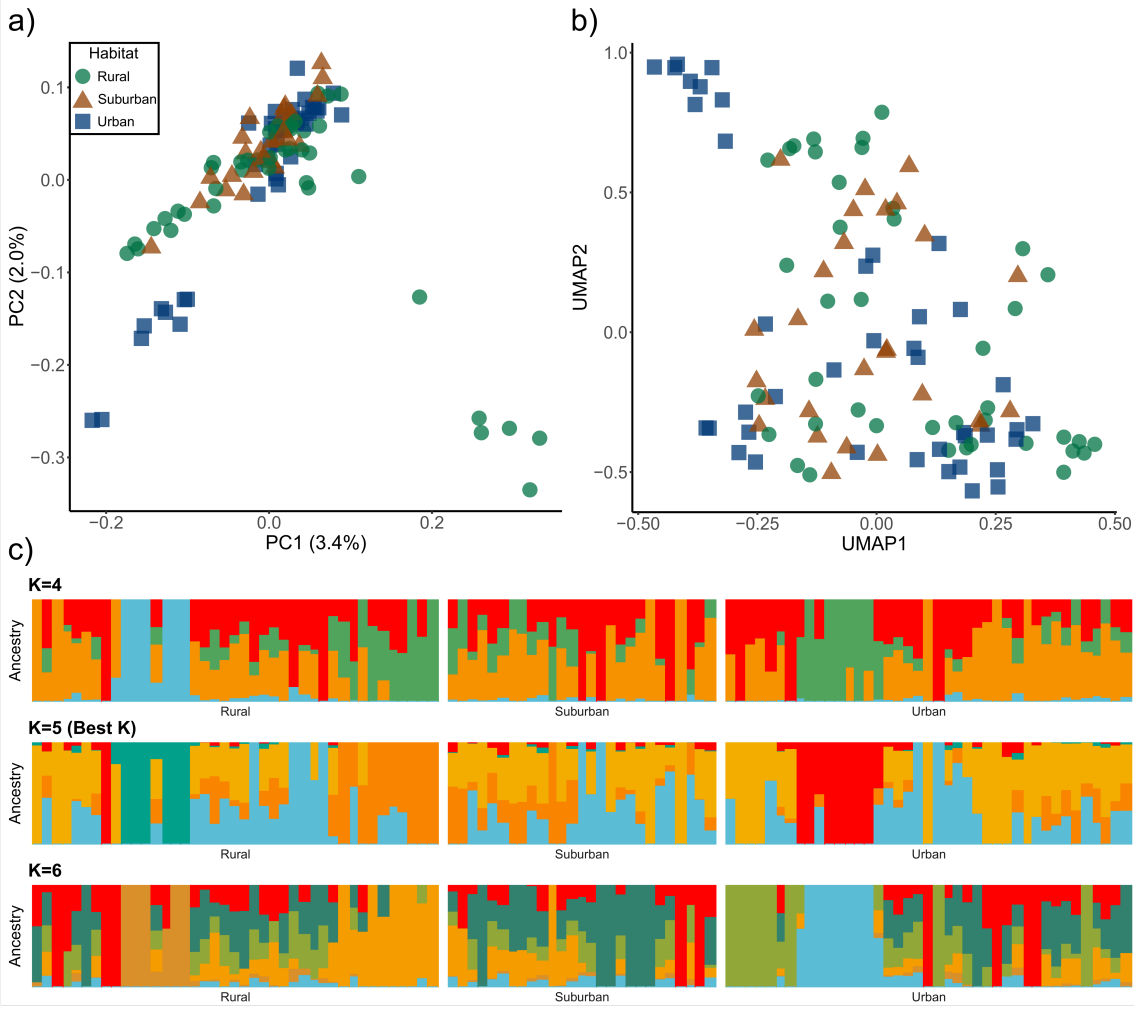


Figure 2: Population structure and admixture among Toronto white clover samples. (a) Principal components analysis (PCA), and (b) Unifold manifold approximation projection (UMAP) of 109 white clover samples from rural (green circles), suburban (brown triangles), and urban (blue squares) habitats. Each point in (a) and (b) represents a single individual. (c) Admixture proportions for Toronto while clover samples for K = 4, K = 5 (best K), and K = 6 assumed ancestral clusters.

## Signatures of selection in urban and rural populations

We identified 135 50 Kb XP-nSL outlier windows with evidence of positive selection in urban habitats, and 124 windows with evidence of positive selection in rural habitats (Figure 3), which reduce to 108 and 104 outlier regions in rural and urban populations after merging consecutive outlier windows, respectively. However, because outlier approaches can be enriched for false positive signals of selection, we focus our attention on outlier windows that are additionally in the upper 5% of the urban-rural permuted XP-nSL distribution (diamonds in Figure 3b, Figure S5), representing 20 rural and urban outlier windows combined into 9 and 8 outlier regions, respectively (Table 1).

Among the strongest candidates of positive selection in urban populations is a 650 Kb region on chromosome 7 (Figure 3e). This region contained four genome-wide XP-nSL outlier regions, three of which were additionally outliers in the urban-rural permuted XP-nSL distribution (Table 1). In addition, these regions are overlapped by five $F_{ST}$ outlier windows, and three nSL, two iHS, and six iHH12 urban outlier windows (Figure 3e), and contain the maximum genome-wide XP-nSL score (yellow diamond in Figure 3e). Similarly, an approximately 3.3 Mb region on chromosome 7 (Figure 3d) contained eight outlier regions, three of which were in the top 5% of the urban-rural permuted XP-nSL distribution (Table 1). This region was additionally overlapped by three $F_{ST}$ outliers, and contain one nSL and one iHH12 urban outlier window. These regions on chromosomes 7 and 9 contain 47 and 42 genes, respectively, which are collectively involved in biotic and abiotic stress, such as plant pathogen immunity (*PBL19* and *TOM2A*, Li et al. 2022; Sun and Zhang 2020; Tsujimoto et al. 2003), salt stress tolerance (*APRR2*, Jiang et al. 2022), and drought/cold stress (*AAE15*, Shockey and Browse 2011), as well as in growth and metabolism such as starch degradation (*BAM3*, Monroe and Storm 2018), seedling germination and cotyledon growth (*RABH1B* and *LON2*, He et al. 2018; Tsitsekian et al. 2019), leaf formation and development (*PALM1*, Chen et al. 2010), root and

455 floral development (*TCP29*, Fang et al. 2021). The highest genome-wide XP-nSL
456 score occurred within the coding sequence of a gene encoding an Acetylornithine
457 deacetylase (NAOD), which plays a critical role in arginine and polyamine biosyn-
458 thesis and is known to impact flowering time and fruit ripening in *Arabidopsis*.

459 Rural populations showed similarly strong signatures of selection across the
460 genome. For example, a 650 Kb region on chromosome 7 (Figure 3c) contained
461 three outlier regions, two of which were additionally outliers in the urban-rural per-
462 muted XP-nSL distribution (Table 1). This region was additionally overlapped by
463 five $F_{ST}$ outlier windows, and two nSL and iHS rural outlier windows. Similarly,
464 a 400 Kb region on chromosome 16 (Figure 3f) contains three outlier regions, one
465 of which was additionally outliers in the urban-rural permuted XP-nSL distribution
466 (Table 1). This region is overlapped by two $F_{ST}$ outlier windows and three nSL and
467 iHS rural outlier windows, and contains the genome-wide minimum XP-nSL score
468 (yellow diamond in Figure 3f). These regions on chromosomes 7 and 16 contain
469 50 and 11 genes, respectively, and are involved in the development of plant struc-
470 tures such as leaves (*RER4*, Pérez-Pérez et al. 2013), pollen (*SEC31B*, Liu et al.
471 2021), and roots (*GPDHC1*, Hu et al. 2014), as well as resistance to abiotic stresses
472 such as arsenic (*HAC1*, Chao et al. 2014), shade (*CKX6*, Wang et al. 2020), heat
473 (*HSP22*, Yamaguchi et al. 2021), frost, (*MEX1*, Cvetkovic et al. 2021), and salt and
474 drought tolerance (Wu et al. 2019). Our results suggest that both urban and ru-
475 ral populations have experienced sweeps in genes underlying growth, development,
476 and response to biotic and abiotic stressors, consistent with expectation based on
477 Gene Ontology (GO) analysis of genes in the most strongly selected regions in each
478 habitat (Table S1, Table S2).

479 Both urban and rural land-use change have threatened biodiversity worldwide
480 and altered the ecological conditions experienced by natural populations (Johnson
481 and Munshi-South 2017; Newbold et al. 2015; Turcotte et al. 2017). Consistent with
482 this view, we identified signatures of positive selection in both habitats in genes

483 underlying tolerance to stressful biotic and abiotic conditions. Similarly, sweep sig-
484 natures in genes related to plant vegetative (e.g., leaves and roots) and floral growth
485 and development are consistent with previous common-garden work documenting
486 urban-rural differentiation in vegetative biomass, flowering time, germination time,
487 and flower size among plants from many of the same populations as those investi-
488 gated here (Santangelo et al. 2020a). White clover has long served as a model for
489 understanding the genetic basis of adaptation due to the presence of temperature-
490 mediated cyanogenesis clines across multiple continents (Daday 1954, 1965; Innes
491 et al. 2022; Kooyers and Olsen 2013), and recent detection of similar clines across
492 urban-rural gradients putatively driven by the same selective agents (Santangelo
493 et al. 2022). The recent release of multiple high quality reference genomes for the
494 species (Kuo et al. 2024a; Santangelo et al. 2023) has spurred investigations into
495 other genomic regions underlying local adaptation in white clover. For example,
496 recent work has shown that, while cyanogenesis frequencies are correlated with tem-
497 perature across North America, local adaptation is instead driven primarily by vari-
498 ation in genes underlying growth and flowering in response to seasonal temperature
499 cues (Kuo et al. 2024b). Similarly, white clover's invasion of the Americas from the
500 Mediterranean seems to be associated with variation around multiple, large, fitness-
501 associated structural variants that affect differential expression of genes underlying
502 growth, defense, and stress tolerance (Battlay et al. 2024). Three of these—two
503 on chromosome 7 and one on chromosome 9—form allele frequency clines in North
504 America (haploblocks HB7a2, HB7b, and HB9 Battlay et al. 2024)) and perfectly
505 overlap the strongest signals of positive selection reported here in rural (HB7a2:
506 Figure S3c) and urban (HB7b: Figure 3d, HB9: Figure 3e) populations. This
507 suggests that, in addition to facilitating white clover's rapid adaptation following
508 invasion, these regions may be similarly important in driving local adaptation along
509 the smaller spatial scale of urban-rural gradients. Ongoing and future work will
510 investigate the extent to which these sweep signatures are shared across cities, and

experimentally testing the agents and targets of selection underlying white clover's adaptation to urban and rural habitats.

We have focused on characterizing selective sweeps in urban and rural populations because these represent the extremes of our environmental gradient, and urban sweep signatures are largely shared with suburban populations. Nonetheless, it is worth noting that urban and suburban habitats show some evidence of differentially selected regions across the genome (section . Figure S1). While early work in urban evolutionary ecology tended to pool urban and suburban populations together, there has been increasing emphasis on the importance of considering heterogeneity in ecological and evolutionary responses among populations within urban spaces (Gorton et al. 2018; Prileson and Martin 2024; Santangelo et al. 2020a). Our results show that, although the strongest differences in genome-wide signatures of selective sweeps occur between urban and rural habitats, there is likely also fine-scale adaptation within occurring among populations within cities that warrants further consideration.

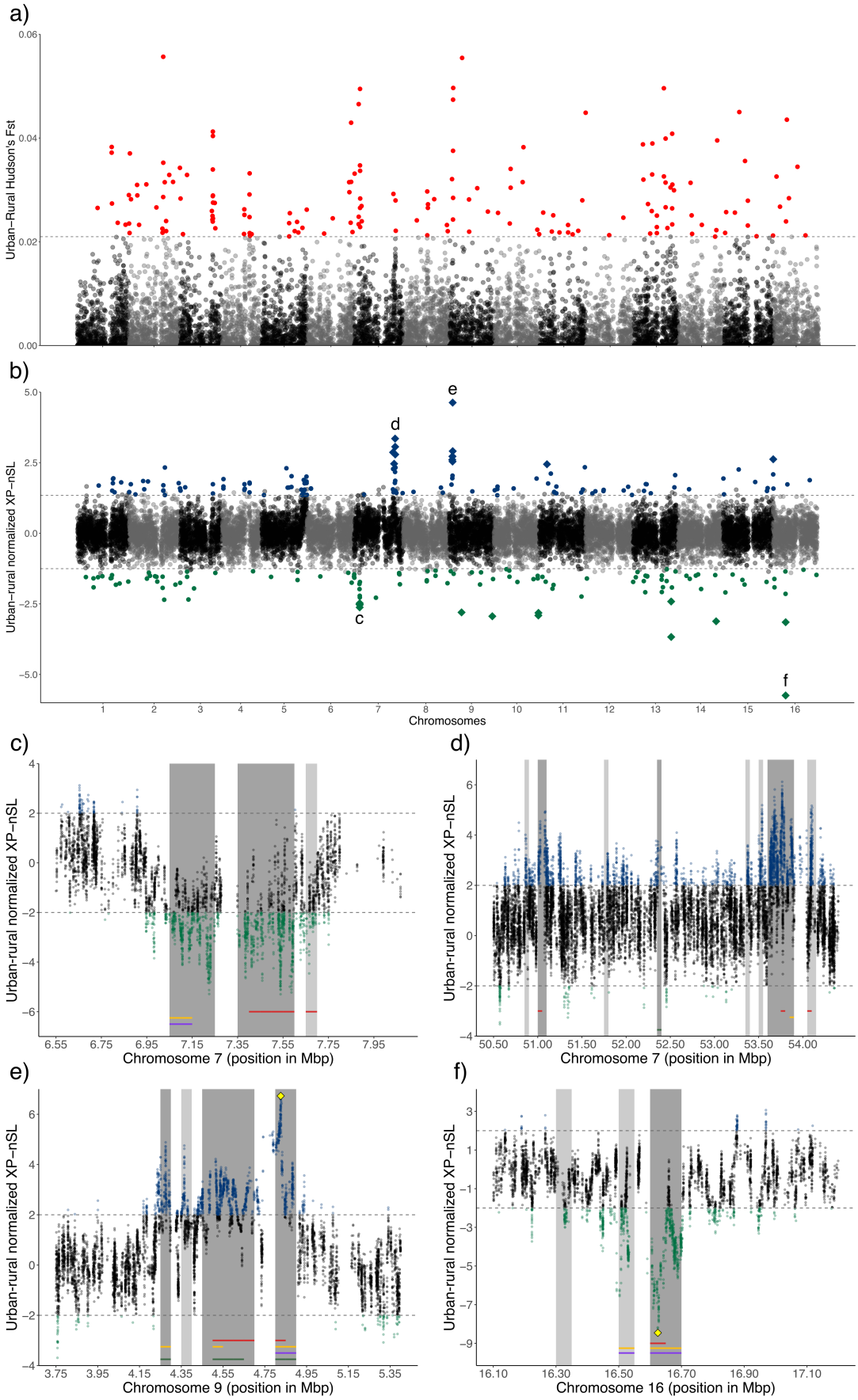


Figure 3: Figure caption on next page.

Figure 3: Selective sweeps in urban and rural Toronto white clover populations. (a) Urban-rural Hudson's $F_{ST}$ estimated in 50 kb windows across the genome. Red points indicate windows in the top 1% of the genome-wide empirical distribution. (b) Urban-rural normalized XP-nSL estimated in 50 Kb windows across the genome. Blue points represent outliers showing evidence of positive selection in urban habitats, while green points represent windows showing evidence of positive selection in rural habitats. Diamonds represent windows that are additionally in the top 5% of the urban-rural permuted XP-nSL distribution (Figure S5). Letters in (b) correspond to panels (c)-(f), where magnified Manhattan plots are shown for those regions. (c)-(f) Magnified Manhattan plots showing per-site normalized urban-rural XP-nSL scores for regions specified in (b). Coloured horizontal lines at the bottom of each panel show regions that are also in the top 1% of the genome-wide distribution for other SFS- and Haplotype-based statistics: $F_{ST}$ (red), nSL (orange), iHS (purple), and iHH12 (green). Light grey bands display XP-nSL outlier windows, while dark grey bands are XP-nSL windows that are additionally in the top 5% of the urban-rural permuted XP-nSL distribution. Yellow diamonds in (e) and (f) show the maximum and minimum genome-wide XP-nSL scores, respectively.

**526**

Table 1: Regions showing the strongest evidence of positive selection in urban and rural habitats (i.e., XP-nSL outliers additionally in the top 5% of the urban-rural permuted XP-nSL distributions). For each region, we show it's position (Chromosome, Start, End); size (as a multiple of 50 Kb where consecutive outlier windows were merged); the habitat under selection; the mean XP-nSL score; the minimum (for rural) or maximum (for urban) XP-nSL score; the proportion of raw scores greater than 2 (for urban) or less than -2 (for rural) for each window; the percentile of the urban-rural permuted null for mean XP-nSL scores in which the windows lie; the percentile of the urban-rural permuted null for the proportion of scores exceeding the threshold in which the windows lie; the rank of each outlier window in the concatenated region; any named annotated genes in the region; and the figures corresponding to that region. For regions containing multiple outlier windows, proportions, percentiles, and ranks report values for each window. Ranks are based on the "Percentile of proportion permuted XP-nSL distribution" column.

| Chromosome | Start | End | Region size | Selection | Mean XP-nSL | Min/Max XP-nSL | Proportion XP-nSL outliers | Percentile of mean permuted XP-nSL distribution | Percentile of proportion permuted XP-nSL distribution | XP-nSL ranks | Annotated genes | Figure |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 16600000 | 16700000 | 100000 | Rural | -4.445941317 | -8.45624 | 1,0.82 | 1,0.992 | 1,0.985 | 1,5 | ITPK1, SUVH2, CML18, SYP22, ABCB2 | Figure 3f, Figure S11 |
| 13 | 49750000 | 49900000 | 150000 | Rural | -2.823437767 | -5.60901 | 0.783,0.963,0.628 | 0.954,0.998,0.95 | 0.979,0.997,0.93 | 7,2,20 | MCM8, RPS23B, RH57, NFYB8, RPL32A | |
| 9 | 56600000 | 56650000 | 50000 | Rural | -2.938899839 | -4.54735 | 0.856 | 0.987 | 0.989 | 3 | ORC4 | |
| 7 | 7350000 | 7600000 | 250000 | Rural | -2.423756753 | -5.29576 | 0.848,0.795,0.581, 0.742,0.663 | 0.972,0.961,0.922, 0.965,0.933 | 0.988,0.981,0.902, 0.971,0.946 | 4,6,25,10,17 | ACA11, GPDHC1, CKX6, TIP51, MAN2, HSP22, MEX1, IQD11, UBP24, COG7, TOP1A | Figure 3c, Figure S8 |
| 10 | 58150000 | 58200000 | 50000 | Rural | -2.911004007 | -4.97952 | 0.76 | 0.986 | 0.975 | 8 | ABCG26, CBL3, HSC80 | |
| 10 | 58250000 | 58350000 | 100000 | Rural | -2.746389294 | -5.64982 | 0.615,0.754 | 0.976,0.983 | 0.923,0.973 | 22,9 | UKL4, SMC3, RFS1 | |
| 14 | 49450000 | 49500000 | 50000 | Rural | -3.116849852 | -5.83163 | 0.727 | 0.992 | 0.967 | 11 | RH57 | |
| 9 | 16400000 | 16450000 | 50000 | Rural | -2.801678579 | -5.61369 | 0.711 | 0.982 | 0.963 | 12 | IWS1 | |
| 7 | 7050000 | 7250000 | 200000 | Rural | -2.213846986 | -5.28792 | 0.524,0.699, 0.416,0.569 | 0.744,0.964, 0.85,0.968 | 0.848,0.959, 0.647,0.892 | 34,16,52,26 | RPL1, RER4, SCPL40, SEC31B, HAC1, ATG7 | Figure 3c, Figure S8 |
| 9 | 4800000 | 4900000 | 100000 | Urban | 3.765810123 | 6.72868 | 0.971,0.894 | 1,0.986 | 0.997,0.992 | 1,4 | PALM1 | Figure 3e, Figure S10 |
| 7 | 51000000 | 51100000 | 100000 | Urban | 2.570842211 | 4.93733 | 0.955,0.59 | 0.985,0.934 | 0.996,0.909 | 2,29 | FKBP13 | Figure 3d, Figure S9 |
| 7 | 53600000 | 53900000 | 300000 | Urban | 2.591351114 | 6.11845 | 0.655,0.92,0.637, 0.73,0.424,0.789 | 0.913,0.995,0.941, 0.991,0.788,0.982 | 0.944,0.994,0.936, 0.969,0.667,0.98 | 20,3,22,14,65,10 | AKHSDH1, AAE15, SHM4 | Figure 3d, Figure S9 |
| 9 | 4450000 | 4700000 | 250000 | Urban | 2.45259609 | 4.21664 | 0.684,0.859,0.81, 0.527,0.852 | 0.928,0.979,0.979, 0.874,0.967 | 0.955,0.989,0.984, 0.852,0.988 | 17,5,7,36,6 | TOM2A, APRR2, RABH1B, LON2, UPF1 | Figure 3e, Figure S10 |
| 16 | 350000 | 450000 | 100000 | Urban | 2.352235783 | 4.12488 | 0.657,0.798 | 0.889,0.973 | 0.945,0.982 | 19,9 | DJA6 | |
| 9 | 4250000 | 4300000 | 50000 | Urban | 2.601913643 | 4.88331 | 0.749 | 0.972 | 0.973 | 12 | DXO1, RPL37A, BAM3, CKB1, PBL19, BAM3 | Figure 3e, Figure S10 |
| 7 | 52350000 | 52400000 | 50000 | Urban | 2.460550909 | 3.73806 | 0.745 | 0.96 | 0.972 | 13 | TCP24, INT2 | Figure 3d, Figure S9 |
| 11 | 10100000 | 10150000 | 50000 | Urban | 2.443494079 | 3.84661 | 0.719 | 0.958 | 0.966 | 16 | | |

## Characterization of urban and rural sweeps signatures

Local adaptation requires that spatially divergent selection overcome the homogenizing effects of gene flow. For species with high levels of population connectivity and migration, local adaptation is predicted to be driven primarily by loci of large-effect (Hodgins and Yeaman 2019; Lenormand 2002; Yeaman 2022). This may be especially true in the early phases of adaptation following a rapid shift to a new environment (Hayward and Sella 2022), as may be expected when natural habitat is converted to cities or agricultural land. Our results support this view, as many of our strongest signatures of positive selection in both habitats overlapped known structural variants with large effects on fitness (Battlay et al. 2024). These structural variants are segregating at appreciable frequencies worldwide, suggesting adaptation is occurring via selection on standing genetic variation in these regions. However, we note that $F_{ST}$ values in these regions are generally low and that these regions likely represent incomplete sweeps with ongoing selection, which are better detected by haplotype-based statistics (e.g., XP-nSL) than $F_{ST}$ (Szpiech et al. 2021). Patterns of allele frequency and haplotype differentiation in these regions seem to support this, whereby haplotypes in the most strongly selected regions only weakly cluster by habitat (Figure S8-Figure S11), though clustering is still substantially greater than unselected regions of the genome (Figure S12).

The relatively young age of urban and rural habitats, particularly in parts of North America like Toronto, suggests that adaptation from standing genetic variation is likely common. Indeed, evidence from our most strongly selected regions suggests this is true (see above). Given evidence of selective sweeps in both urban and rural populations, we were interested in a simple characterization of the selective sweep signatures in both habitats. Leveraging all urban and rural XP-nSL outlier windows overlapping *saltiLassi* windows, we found no difference in hard vs. soft sweeps between urban and rural populations ($\chi^2 = 0.09$, P = 0.79, Figure 4a), suggesting that adaptation to both environments is occurring via similar underly-

ing processes. However, urban sweeps were wider than those in rural populations ($P < 0.001$, Figure 4b, Figure S7), which may suggest differences in the strength ($s$) or timescale ($t$) of selection between habitats. However, note that $s$ and $t$ are confounded in this analysis since multiple combinations $s$ and $t$ can lead to similar sweep widths (i.e. strong, recent selection vs. weaker, more historical selection) (DeGiorgio and Szpiech 2022). If clover populations in and around Toronto have been experiencing selection for similar amounts of time, this would suggest that—while the types of selective sweeps do not differ strongly between these habitats—the strength of selection may differ between urban and rural populations. However, we advise caution in interpreting these results, which are meant to motivate investigations into comparing the strength, timescale, and genetic architecture of urban and rural adaptation. Additional work developing well-defined demographic models in this system is needed, which would allow more thorough characterization of sweeps and estimation of the strength and timescale of selection using more recently developed machine-learning based approaches (e.g., Arnab et al. 2023; Kern and Schrider 2018). Similarly, the absence of phenotype data from sequenced plants has precluded a thorough investigation of the genetic architecture underlying local adaptation in this system, a shortcoming that ongoing work is seeking to address.

# Conclusion

Urbanization results in rapid and drastic environmental changes with profound consequences for the ecology and evolution of natural populations. Urban fragmentation frequently drives reduced population size and connectivity, leading to reduced diversity and increased differentiation and population structure for many species. However, white clover appears largely unaffected by urban fragmentation, instead maintaining high diversity across all habitats with little detectable structure between urban and rural populations, likely resulting from high dispersal across the

city in this obligately outcrossing perennial. While much focus has been placed on phenotypic and genomic responses to urbanization, rural habitats can similarly impose strong anthropogenic selection, and indeed we find signatures of selective sweeps in regions underlying growth, reproduction, and biotic and abiotic stress tolerance in both urban and rural Toronto white clover populations. Preliminary characterization of selective sweeps suggests no difference in the proportion of hard vs. soft sweeps in either habitat, and that selection is likely ongoing (i.e., incomplete sweeps). However, differences in the genomic width of sweeps between urban and rural habitats suggests potential differences in the strength and/or timescale of selection that warrants further investigation. Such investigations will be greatly facilitated as increased functional genomic information becomes available in this system to enable a more thorough characterization of the genetic architecture of adaptation to contemporary anthropogenic environments.

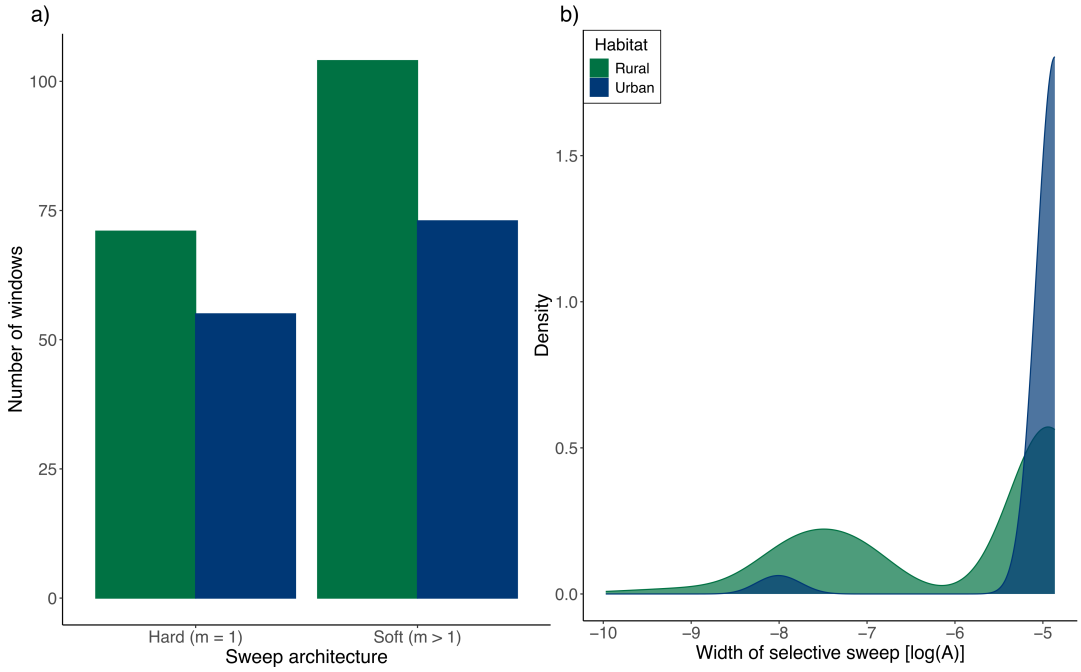


Figure 4: (a) Estimated number of sweeping haplotypes among urban and rural XP-nSL outlier windows overlapping outlier *saltiLassi* $\lambda$ widnows, with binary classification of sweeps into hard (m = 1) and soft (m > 1), as suggested by (DeGiorgio and Szpiech 2022) (see Figure S6 for version with raw number of estimated haplotypes). (b) Inferred width of selective sweeps, in urban and rural habitats in same regions as (a), estimated as $Log_{10}$ of the $A$ statistic reported by *saltiLassi* (DeGiorgio and Szpiech 2022). Permutation test for the urban-rural difference in selective sweep width is shown in Figure S7.

# References


Alexa, A. and J. Rahnenfuhrer. 2016. *topGO: Enrichment Analysis for Gene Ontology*. Version 2.50.0.

Andrews, S. et al. 2010. *FastQC: a quality control tool for high throughput sequence data*.

Arnab, S. P., M. R. Amin, and M. DeGiorgio. 2023. Uncovering footprints of natural selection through spectral analysis of genomic summary statistics. *Mol. Biol. Evol.* 40: msad157.

Baker, H. G. 1974. The evolution of weeds. *Annual Review of Ecology, Evolution, and Systematics* 5: 1–24.

Barnett, D. W., E. K. Garrison, A. R. Quinlan, M. P. Strömberg, and G. T. Marth. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27: 1691–1692.

Bartlewicz, J., K. Vandepitte, H. Jacquemyn, and O. Honnay. 2015. Population genetic diversity of the clonal self-incompatible herbaceous plant *Linaria vulgaris* along an urbanization gradient. *Biol. J. Linn. Soc. Lond.* 116: 603–613.

Battlay, P., B. T. Hendrickson, J. I. Mendez-Reneau, J. S. Santangelo, L. J. Albano, J. Wilson, A. E. Caizergues, N. King, A. Puentes, A. Tudoran, C. Violle, F. Vasseur, C. M. Patterson, M. J. Foster, C. Stamps, S. G. Innes, R. Allio, F. Angeoletto, D. N. Anstett, J. Anstett, A. Bucharova, M. S. Comerford, S. David, M. Falahati Anbaran, W. Godsoe, C. Gonzalez-Lagos, P. E. Gundel, G. R. Hood, R. Karousou, C. Lampei, C. Lara, A. Lazaro-Lobo, D. Leandro, T. J. S. Merritt, N. Mitchell, M. Mohammadi Bazargani, A. Moles, M. Murua, J. Paule, V. Pfeiffer, J. A. M. Raeymaekers, D. J. Rennison, R. S. Rios, J. K. Rowntree, A. C. Schneider, K. Stack Whitney, I. Tamburrino, A. VanWallendael, P. Y. Kim, R. W. Ness, M. T. J. Johnson, K. A. Hodgins, and N. J. Kooyers. 2024. Structural variants underlie parallel adaptation following global invasion. *Evolutionary Biology*.

Beninde, J., S. Feldmeier, M. Veith, and A. Hochkirch. 2018. Admixture of hybrid swarms of native and introduced lizards in cities is determined by the cityscape structure and invasion history. *Proceedings of the Royal Society B: Biological Sciences* 285: 20180143.

Beninde, J., S. Feldmeier, M. Werner, D. Peroverde, U. Schulte, A. Hochkirch, and M. Veith. 2016. Cityscape genetics: structural vs. functional connectivity of an urban lizard population. *Mol. Ecol.* 25: 4984–5000.

Bhatia, G., N. Patterson, S. Sankararaman, and A. L. Price. 2013. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* 23: 1514–1521.

Björklund, M., I. Ruiz, and J. C. Senar. 2010. Genetic differentiation in the urban habitat: the great tits (*Parus major*) of the parks of Barcelona city. *Biol. J. Linn. Soc. Lond.* 99: 9–19.

Brans, K., R. Stoks, and L. De Meester. 2018. Urbanization drives genetic differentiation in physiology and structures the evolution of pace-of-life syndromes in the water flea *Daphnia magna*. *Proceedings of the Royal Society B: Biological Sciences* 285: 20180169.

Brans, K. I. and L. De Meester. 2018. City life on fast lanes: Urbanization induces an evolutionary shift towards a faster lifestyle in the water flea Daphnia. *Funct. Ecol.* 32: 2225–2240.

Brans, K. I., M. Jansen, J. Vanoverbeke, N. Tüzün, R. Stoks, and L. De Meester. 2017. The heat is on: Genetic adaptation to urbanization mediated by thermal tolerance and body size. *Glob. Chang. Biol.* 23: 5218–5227.

Caizergues, A. E., J. S. Santangelo, R. W. Ness, F. Angeoletto, D. N. Anstett, J. Anstett, F. Baena-Diaz, E. J. Carlen, J. A. Chaves, M. S. Comerford, K. Dyson, M. Falahati-Anbaran, M. D. E. Fellowes, K. A. Hodgins, G. R. Hood, C. Iñiguez-Armijos, N. J. Kooyers, A. Lázaro-Lobo, A. T. Moles, J. Munshi-South, J. Paule, I. M. Porth, L. Y. Santiago-Rosario, K. S. Whitney, A. J. M. Tack, and M. T. J.

Johnson. 2024. Does urbanisation lead to parallel demographic shifts across the world in a cosmopolitan plant? *Mol. Ecol.* 33: e17311.

Campbell-Staton, S. C., K. M. Winchell, N. C. Rochette, J. Fredette, I. Maayan, R. M. Schweizer, and J. Catchen. 2020. Parallel selection on thermal physiology facilitates repeated adaptation of city lizards to urban heat islands. *Nat Ecol Evol* 4: 652–658.

Carlen, E. and J. Munshi-South. 2021. Widespread genetic connectivity of feral pigeons across the Northeastern megacity. *Evol. Appl.* 14: 150–162.

Chao, D.-Y., Y. Chen, J. Chen, S. Shi, Z. Chen, C. Wang, J. M. Danku, F.-J. Zhao, and D. E. Salt. 2014. Genome-wide association mapping identifies a new arsenate reductase enzyme critical for limiting arsenic accumulation in plants. *PLoS Biol.* 12: e1002009.

Chen, J., J. Yu, L. Ge, H. Wang, A. Berbel, Y. Liu, Y. Chen, G. Li, M. Tadege, J. Wen, V. Cosson, K. S. Mysore, P. Ratet, F. Madueño, G. Bai, and R. Chen. 2010. Control of dissected leaf morphology by a Cys(2)His(2) zinc finger transcription factor in the model legume *Medicago truncatula*. *Proc. Natl. Acad. Sci. U. S. A.* 107: 10754–10759.

Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34: i884–i890.

Combs, M., K. Byers, B. M. Ghersi, M. J. Blum, A. Caccone, F. Costa, C. G. Himsworth, J. L. Richardson, and J. Munshi-South. 2018. Urban rat races: spatial population genomics of brown rats (Rattus norvegicus) compared across multiple cities. *Proceedings of the Royal Society B: Biological Sciences* 285: 20180245.

Cvetkovic, J., I. Haferkamp, R. Rode, I. Keller, B. Pommerrenig, O. Trentmann, J. Altensell, M. Fischer-Stettler, S. Eicke, S. C. Zeeman, and H. E. Neuhaus. 2021. Ectopic maltase alleviates dwarf phenotype and improves plant frost tolerance of maltose transporter mutants. *Plant Physiol.* 186: 315–329.

Daday, H. 1954. Gene frequencies in wild populations of *Trifolium repens* I. Distribution by latitude. *Heredity* 8: 61–78.

— 1965. Gene frequencies in wild populations of *Trifolium repens* L IV. Mechanism of natural selection. *Heredity* 20: 355–365.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10: giab008.

De Lucas, J. A., J. W. Forster, K. F. Smith, and G. C. Spangenberg. 2012. Assessment of gene flow in white clover (*Trifolium repens* L.) under field conditions in Australia using phenotypic and genetic markers. *Crop Pasture Sci.* 63: 155–163.

DeGiorgio, M. and Z. A. Szpiech. 2022. A spatially aware likelihood test to detect sweeps from haplotype distributions. *PLoS Genet.* 18: e1010134.

Delaneau, O., J.-F. Zagury, M. R. Robinson, J. L. Marchini, and E. T. Dermitzakis. 2019. Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* 10: 5436.

Diamond, S. E., L. Chick, A. B. E. Perez, S. A. Strickler, and R. A. Martin. 2017. Rapid evolution of ant thermal tolerance across an urban-rural temperature cline. *Biol. J. Linn. Soc. Lond.* 121: 248–257.

Diamond, S. E., L. D. Chick, A. Perez, S. A. Strickler, and R. A. Martin. 2018. Evolution of thermal tolerance and its fitness consequences: parallel and non-parallel responses to urban heat islands across three cities. *Proceedings of the Royal Society B: Biological Sciences* 285: 20180036.

Diamond, S. E. and R. A. Martin. 2021. Evolution in Cities. *Annu. Rev. Ecol. Evol. Syst.*

Diaz-Papkovich, A., L. Anderson-Trocmé, and S. Gravel. 2021. A review of UMAP in population genetics. *J. Hum. Genet.* 66: 85–91.

Ellis, E. C. and N. Ramankutty. 2008. Putting people in the map: anthropogenic biomes of the world. *Front. Ecol. Environ.* 6: 439–447.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611–2620.

Evans, K. L., J. Newton, K. J. Gaston, S. P. Sharp, A. Mcgowan, and B. J. Hatchwell. 2012. Colonisation of urban environments is associated with reduced migratory behaviour, facilitating divergence from ancestral populations. *Oikos* 121: 634–640.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32: 3047–3048.

Fang, Y., Y. Zheng, W. Lu, J. Li, Y. Duan, S. Zhang, and Y. Wang. 2021. Roles of miR319-regulated TCPs in plant development and response to abiotic stress. *Crop J.* 9: 17–28.

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

Foley, J. A., R. Defries, G. P. Asner, C. Barford, G. Bonan, S. R. Carpenter, F. S. Chapin, M. T. Coe, G. C. Daily, H. K. Gibbs, J. H. Helkowski, T. Holloway, E. A. Howard, C. J. Kucharik, C. Monfreda, J. A. Patz, I. C. Prentice, N. Ramankutty, and P. K. Snyder. 2005. Global consequences of land use. *Science* 309: 570–574.

Fox, E. A., A. E. Wright, M. Fumagalli, and F. G. Vieira. 2019. ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* 35: 3855–3856.

Fusco, N. A., B. J. Cosentino, J. P. Gibbs, M. L. Allen, A. J. Blumenfeld, G. H. Boettner, E. J. Carlen, M. Collins, C. Dennison, D. DiGiacopo, A.-P. Drapeau Picard, J. Edmonson, M. C. Fisher-Reid, R. Fyffe, T. Gallo, A. Grant, W. Harbold, S. B. Heard, D. J. R. Lafferty, R. M. Lehtinen, S. Marino, J. E. McDonald, A. Mortelliti, M. Murray, A. Newman, K. N. Oswald, C. Ott-Conn, J. L. Richardson, R. Rimbach, P. Salaman, M. Steele, M. R. Stothart, M. C. Urban, K. Vandegrift,

J. P. Vanek, S. N. Vanderluit, L. Vezina, and A. Caccone. 2024. Population genomic structure of a widespread, urban-dwelling mammal: The eastern grey squirrel (*Sciurus carolinensis*). *Mol. Ecol.* 33: e17230.

Garrison, E. and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]*: 1–9.

Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genet.* 11: e1005004.

Gorton, A. J., D. A. Moeller, and P. Tiffin. 2018. Little plant, big city: a test of adaptation to urban environments in common ragweed (*Ambrosia artemisiifolia*). *Proceedings of the Royal Society B: Biological Sciences* 285: 20180968.

Han, E., J. S. Sinsheimer, and J. Novembre. 2013. Characterizing bias in population genetic inferences from low-coverage sequencing data. *Mol. Biol. Evol.* 31: 723–735.

Hanghøj, K., I. Moltke, P. A. Andersen, A. Manica, and T. S. Korneliussen. 2019. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* 8.

Harpak, A., N. Garud, N. A. Rosenberg, D. A. Petrov, M. Combs, P. S. Pennings, and J. Munshi-South. 2021. Genetic adaptation in New York City rats. *Genome Biol. Evol.* 13.

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

Harris, S. E., A. T. Xue, D. Alvarado-Serrano, J. T. Boehm, T. Joseph, M. J. Hickerson, and J. Munshi-South. 2016. Urbanization shapes the demographic history of a native rodent (the white-footed mouse, *Peromyscus leucopus*) in New York City. *Biol. Lett.* 12: 20150983–.

Hayward, L. K. and G. Sella. 2022. Polygenic adaptation after a sudden change in environment. *Elife* 11.

He, M., M. Lan, B. Zhang, Y. Zhou, Y. Wang, L. Zhu, M. Yuan, and Y. Fu. 2018. Rab-H1b is essential for trafficking of cellulose synthase and for hypocotyl growth in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* 60: 1051–1069.

Hedrick, P. W. and R. C. Lacy. 2015. Measuring relatedness between inbred individuals. *J. Hered.* 106: 20–25.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L. Poss, L. K. Reed, A. Storfer, and M. C. Whitlock. 2016. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am. Nat.* 188: 379–397.

Hodgins, K. A. and S. Yeaman. 2019. Mating system impacts the genetic architecture of adaptation to heterogeneous environments. *New Phytol.* 224: 1201–1214.

Hu, J., Y. Zhang, J. Wang, and Y. Zhou. 2014. Glycerol affects root development through regulation of multiple pathways in *Arabidopsis*. *PLoS One* 9: e86269.

Hudson, R. R., M. Slatkin, and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132: 583–589.

Innes, S. G., J. S. Santangelo, N. J. Kooyers, K. M. Olsen, and M. T. J. Johnson. 2022. Evolution in response to climate in the native and introduced ranges of a globally distributed plant. *Evolution* 76: 1495–1511.

Inostroza, L., M. Bhakta, H. Acuña, C. Vásquez, J. Ibáñez, G. Tapia, W. Mei, M. Kirst, M. Resende, and P. Munoz. 2018. Understanding the complexity of cold tolerance in white clover using temperature gradient locations and a GWAS approach. *Plant Genome* 11: 170096.

Jackson, D. A. 1993. Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology* 74: 2204–2214.

Jiang, L., Y. Fu, X. Tian, Y. Ma, F. Chen, and G. Wang. 2022. The *Anthurium* APRR2-like gene promotes photosynthetic pigment accumulation in response to salt stress. *Trop. Plant Biol.* 15: 12–21.

Johnson, M. T. J. and J. Munshi-South. 2017. Evolution of life in urban environments. *Science* 358: eaam8327.

Johnson, M. T. J., C. Prashad, M. Lavoignat, and H. S. Saini. 2018. Contrasting the effects of natural selection, genetic drift and gene flow on urban evolution in white clover (*Trifolium repens*). *Proceedings of the Royal Society B: Biological Sciences* 285: 20181019.

Johri, P., C. F. Aquadro, M. Beaumont, B. Charlesworth, L. Excoffier, A. Eyre-Walker, P. D. Keightley, M. Lynch, G. McVean, B. A. Payseur, S. P. Pfeifer, W. Stephan, and J. D. Jensen. 2022. Recommendations for improving statistical inference in population genomics. *PLoS Biol.* 20: e3001669.

Kamdem, C., C. Fouet, S. Gamez, and B. J. White. 2017. Pollutants and insecticides drive local adaptation in african malaria mosquitoes. *Mol. Biol. Evol.* 34: 1261–1275.

Kern, A. D. and D. R. Schrider. 2018. DiploS/HIC: An updated approach to classifying selective sweeps. *G3 (Bethesda)* 8: 1959–1970.

Kjærgaard, T. 2003. A plant that changed the world: the rise and fall of clover 1000-2000. *Landscape Res.* 28: 41–49.

Kooyers, N. J. and K. M. Olsen. 2013. Searching for the bull's eye: agents and targets of selection vary among geographically disparate cyanogenesis clines in white clover (*Trifolium repens* L.) *Heredity* 111: 495–504.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15: 1179–1191.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15: 1–13.

Korunes, K. L. and K. Samuk. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* 21: 1359–1368.

Kreiner, J. M., S. M. Latorre, H. A. Burbano, J. R. Stinchcombe, S. P. Otto, D. Weigel, and S. I. Wright. 2022a. Rapid weed adaptation and range expansion in response to agriculture over the past two centuries. *Science* 378: 1079–1085.

Kreiner, J. M., G. Sandler, A. J. Stern, P. J. Tranel, D. Weigel, J. Stinchcombe, and S. I. Wright. 2022b. Repeated origins, widespread gene flow, and allelic interactions of target-site herbicide resistance mutations. *Elife* 11.

Kuo, W.-H., S. J. Wright, L. L. Small, and K. M. Olsen. 2024a. De novo genome assembly of white clover (*Trifolium repens* L.) reveals the role of Copy Number Variation in rapid environmental adaptation.

Kuo, W.-H., L. Zhong, S. J. Wright, D. M. Goad, and K. M. Olsen. 2024b. Beyond cyanogenesis: Temperature gradients drive environmental adaptation in North American white clover (Trifolium repens L.) *Mol. Ecol.*: e17484.

Lasky, J. R., E. B. Josephs, and G. P. Morris. 2023. Genotype-environment associations to reveal the molecular basis of environmental adaptation. *Plant Cell* 35: 125–138.

Lau, J. A. and J. L. Funk. 2023. How ecological and evolutionary theory expanded the 'ideal weed' concept. *Oecologia* 203: 251–266.

Lenormand, T. 2002. Gene flow and the limits to natural selection. *Trends Ecol. Evol.* 17: 183–189.

Li, H. 2011a. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993.

— 2011b. Improving SNP discovery by base alignment quality. *Bioinformatics* 27: 1157–1158.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Li, Y., J. Xue, F.-Z. Wang, X. Huang, B.-Q. Gong, Y. Tao, W. Shen, K. Tao, N. Yao, S. Xiao, J.-M. Zhou, and J.-F. Li. 2022. Plasma membrane-nucleo-cytoplasmic coordination of a receptor-like cytoplasmic kinase promotes EDS1-dependent plant immunity. *Nat. Plants* 8: 802–816.

Littleford-Colquhoun, B. L., C. Clemente, M. J. Whiting, D. Ortiz-Barrientos, and C. H. Frère. 2017. Archipelagos of the Anthropocene: rapid and extensive differentiation of native terrestrial vertebrates in a single metropolis. *Mol. Ecol.* 26: 2466–2481.

Liu, X., M. Tong, A. Zhang, M. Liu, B. Zhao, Z. Liu, Z. Li, X. Zhu, Y. Guo, and R. Li. 2021. COPII genes SEC31A/B are essential for gametogenesis and interchangeable in pollen development in *Arabidopsis*. *Plant J.* 105: 1600–1614.

Løjtnant, C. L., B. Boelt, S. K. Clausen, C. Damgaard, P. Kryger, A. Novy, M. Philipp, C. H. Ingvordsen, and R. B. Jørgensen. 2012. Modelling gene flow between fields of white clover with honeybees as pollen vectors. *Environ. Model. Assess.* 17: 421–430.

Lotterhos, K. E. and M. C. Whitlock. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* 24: 1031–1046.

Lourenço, A., D. Álvarez, I. J. Wang, and G. Velo-Antón. 2017. Trapped within the city: integrating demography, time since isolation and population-specific traits to assess the genetic effects of urbanization. *Mol. Ecol.* 26: 1498–1514.

Martin, M., M. Patterson, S. Garg, S. O. Fischer, N. Pisanti, G. W. Klau, A. Schöenhuth, and T. Marschall. 2016. WhatsHap: fast and accurate read-based phasing. *bioRxiv*: 085050.

Martin, R. A., L. D. Chick, M. L. Garvin, and S. E. Diamond. 2021. In a nutshell, a reciprocal transplant experiment reveals local adaptation and fitness trade-offs in response to urban evolution in an acorn-dwelling ant. *Evolution* 75: 876–887.

McInnes, L., J. Healy, and J. Melville. 2018. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv [stat.ML]*.

Meisner, J. and A. Albrechtsen. 2018. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* 210: 719–731.

Miles, L. S., L. R. Rivkin, M. T. J. Johnson, J. Munshi-South, and B. C. Verrelli. 2019. Gene flow and genetic drift in urban environments. *Mol. Ecol.* 28: 4138–4151.

Mölder, F., K. P. Jablonski, B. Letcher, M. B. Hall, C. H. Tomkins-Tinch, V. Sochat, J. Forster, S. Lee, S. O. Twardziok, A. Kanitz, A. Wilm, M. Holtgrewe, S. Rahmann, S. Nahnsen, and J. Köster. 2021. Sustainable data analysis with Snakemake. *F1000Res.* 10: 33.

Monroe, J. D. and A. R. Storm. 2018. Review: The *Arabidopsis* $\beta$-amylase (BAM) gene family: Diversity of form and function. *Plant Sci.* 276: 163–170.

Mueller, J., K. Heiner, S. Boerno, J. Tella, M. Carrete, and B. Kempenaers. 2018. Evolution of genomic variation in the burrowing owl in response to recent colonization of urban areas. *Proceedings of the Royal Society B: Biological Sciences* 285: 20180206.

Mueller, J. C., M. Carrete, S. Boerno, H. Kuhl, J. L. Tella, and B. Kempenaers. 2020. Genes acting in synapses and neuron projections are early targets of selection during urban colonization. *Mol. Ecol.* 29: 3403–3412.

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

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

Neve, P., M. Vila-Aiub, and F. Roux. 2009. Evolutionary-thinking in agricultural weed management. *New Phytol.* 184: 783–793.

Newbold, T., L. N. Hudson, S. L. L. Hill, S. Contu, I. Lysenko, R. A. Senior, L. Börger, D. J. Bennett, A. Choimes, B. Collen, J. Day, A. De Palma, S. Díaz, S. Echeverria-Londoño, M. J. Edgar, A. Feldman, M. Garon, M. L. K. Harrison, T. Alhusseini, D. J. Ingram, Y. Itescu, J. Kattge, V. Kemp, L. Kirkpatrick, M. Kleyer, D. L. P. Correia, C. D. Martin, S. Meiri, M. Novosolov, Y. Pan, H. R. P. Phillips, D. W. Purves, A. Robinson, J. Simpson, S. L. Tuck, E. Weiher, H. J. White, R. M. Ewers, G. M. Mace, J. P. W. Scharlemann, and A. Purvis. 2015. Global effects of land use on local terrestrial biodiversity. *Nature* 520: 45–50.

Nielsen, R., T. Korneliussen, A. Albrechtsen, Y. Li, and J. Wang. 2012. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS One* 7: e37558.

Nielsen, R., J. S. Paul, A. Albrechtsen, and Y. S. Song. 2011. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12: 443–451.

Noël, S., M. Ouellet, P. Galois, and F. J. Lapointe. 2007. Impact of urban fragmentation on the genetic structure of the eastern red-backed salamander. *Conserv. Genet.* 8: 599–606.

Okonechnikov, K., A. Conesa, and F. García-Alcalde. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32: 292–294.

Olsen, K. M., D. M. Goad, S. J. Wright, M. L. Dutta, S. R. Myers, L. L. Small, and L.-F. Li. 2021. Dual-species origin of an adaptive chemical defense polymorphism. *New Phytol.* 232: 1477–1487.

Oziolor, E. M., N. M. Reid, S. Yair, K. M. Lee, S. G. Verploeg, P. C. Bruns, J. R. Shaw, A. Whitehead, and C. W. Matson. 2019. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science* 126: 455–457.

Pérez-Pérez, J. M., D. Esteve-Bruna, R. González-Bayón, S. Kangasjärvi, C. Caldana, M. A. Hannah, L. Willmitzer, M. R. Ponce, and J. L. Micol. 2013. Functional redundancy and divergence within the *Arabidopsis* RETICULATA-RELATED gene family. *Plant Physiol.* 162: 589–603.

Pickering, C. and A. Mount. 2010. Do tourists disperse weed seed? A global review of unintentional human-mediated terrestrial seed dispersal on clothing, vehicles and horses. *J. Sustainable Tour.* 18: 239–256.

Prileson, E. G. and R. A. Martin. 2024. Evolution and plasticity of physiological traits in the collembolan *Orchesella villosa* at fine spatial scales within the city. *Biol. J. Linn. Soc. Lond.*: blae038.

Pya, N. and S. N. Wood. 2015. Shape constrained additive models. *Stat. Comput.* 25: 543–559.

R Core Team. 2020. *R: A language and environment for statistical computing.* Version 3.6.3. Vienna, Austria.

Ravinet, M., T. Elgvin, C. Trier, M. Aliabadian, A. Gavrilov, and G.-P. Saetre. 2018. Signatures of human-commensalism in the house sparrow genome. *Proceedings of the Royal Society B: Biological Sciences* 285: 20181246.

Reid, N. M., D. A. Proestou, B. W. Clark, W. C. Warren, J. K. Colbourne, J. R. Shaw, S. I. Karchner, M. E. Hahn, D. Nacci, M. F. Oleksiak, D. L. Crawford, and A. Whitehead. 2016. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* 354: 1305 LP–1308.

Rivkin, L. R. and M. T. J. Johnson. 2022. The impact of urbanization on outcrossing rate and population genetic variation in the native wildflower, *Impatiens capensis*. *J Urban Ecol* 8.

Salmón, P., A. Jacobs, D. Ahrén, C. Biard, N. J. Dingemanse, D. M. Dominoni, B. Helm, M. Lundberg, J. C. Senar, P. Sprau, M. E. Visser, and C. Isaksson. 2021. Continent-wide genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nat. Commun.* 12: 2983.

Santangelo, J. S., P. Battlay, B. T. Hendrickson, W.-H. Kuo, K. M. Olsen, N. J. Kooyers, M. T. J. Johnson, K. A. Hodgins, and R. W. Ness. 2023. Haplotype-Resolved, Chromosome-Level Assembly of White Clover (*Trifolium repens* L., Fabaceae). *Genome Biol. Evol.* 15.

Santangelo, J. S., L. R. Rivkin, C. Advenard, and K. A. Thompson. 2020a. Multivariate phenotypic divergence along an urbanization gradient. *Biol. Lett.* 16: 20200511.

Santangelo, J. S., L. R. Rivkin, and M. T. J. Johnson. 2018. The evolution of city life. *Proceedings of the Royal Society B: Biological Sciences* 285: 20181529.

Santangelo, J. S., K. A. Thompson, B. Cohan, J. Syed, R. W. Ness, and M. T. J. Johnson. 2020b. Predicting the strength of urban-rural clines in a Mendelian polymorphism. *Evolution Letters* 4: 212–225.

Santangelo, J. S. et al. 2022. Global urban environmental change drives adaptation in white clover. *Science* 375: 1275–1281.

Schmidt, C., M. Domaratzki, R. P. Kinnunen, J. Bowman, and C. J. Garroway. 2020. Continent-wide effects of urbanization on bird and mammal genetic diversity. *Proc. Biol. Sci.* 287: 20192497.

Shockey, J. and J. Browse. 2011. Genome-level and biochemical diversity of the acyl-activating enzyme superfamily in plants: Biochemistry and evolution of plant AAE proteins. *Plant J.* 66: 143–160.

Sincik, M. and E. Acikgoz. 2007. Effects of white clover inclusion on turf characteristics, nitrogen fixation, and nitrogen transfer from white clover to grass species in turf mixtures. *Commun. Soil Sci. Plant Anal.* 38: 1861–1877.

Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195: 693–702.

Snaydon, R. W. and M. S. Davies. 1972. Rapid population differentiation in a mosiac environment II: Morphological variation in *Anthoxanthum odoratum*. *Evolution* 26: 390–405.

Sun, L. and J. Zhang. 2020. Regulatory role of receptor-like cytoplasmic kinases in early immune signaling events in plants. *FEMS Microbiol. Rev.* 44: 845–856.

Szpiech, Z. A. and R. D. Hernandez. 2014. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* 31: 2824–2827.

Szpiech, Z. A., T. E. Novak, N. P. Bailey, and L. S. Stevison. 2021. Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. *Evol. Lett.* 5: 408–421.

Szulkin, M., J. Munshi-South, and A. Charmantier. 2020. *Urban Evolutionary Biology*. Oxford University Press, Oxford, United Kingdom.

Theodorou, P., R. Radzevičiūtė, B. Kahnt, A. Soro, I. Grosse, and R. J. Paxton. 2018. Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidarius* L.) *Proceedings of the Royal Society B: Biological Sciences* 285: 20172806.

Thompson, K. A., M. Renaudin, and M. T. J. Johnson. 2016. Urbanization drives the evolution of parallel clines in plant populations. *Proceedings of the Royal Society B: Biological Sciences* 283: 20162180.

Torres, R., Z. A. Szpiech, and R. D. Hernandez. 2018. Human demographic history has amplified the effects of background selection across the genome. *PLoS Genet.* 14: e1007387.

Tsitsekian, D., G. Daras, A. Alatzas, D. Templalexis, P. Hatzopoulos, and S. Rigas. 2019. Comprehensive analysis of Lon proteases in plants highlights independent gene duplication events. *J. Exp. Bot.* 70: 2185–2197.

Tsujimoto, Y., T. Numaga, K. Ohshima, M.-A. Yano, R. Ohsawa, D. B. Goto, S. Naito, and M. Ishikawa. 2003. *Arabidopsis* TOBAMOVIRUS MULTIPLICATION (TOM) 2 locus encodes a transmembrane protein that interacts with TOM1. *EMBO J.* 22: 335–343.

Turcotte, M. M., H. Araki, D. S. Karp, K. Poveda, and S. R. Whitehead. 2017. The eco-evolutionary impacts of domestication and agricultural practices on wild species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372.

Unfried, T. M., L. Hauser, and J. M. Marzluff. 2013. Effects of urbanization on Song Sparrow (*Melospiza melodia*) population connectivity. *Conserv. Genet.* 14: 41–53.

Vasimuddin, M., S. Misra, H. Li, and S. Aluru. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*: pp. 314–324.

Verrelli, B. C., M. Alberti, S. Des Roches, N. C. Harris, A. P. Hendry, M. T. J. Johnson, A. M. Savage, A. Charmantier, K. M. Gotanda, L. Govaert, L. S. Miles, L. R. Rivkin, K. M. Winchell, K. I. Brans, C. Correa, S. E. Diamond, B. Fitzhugh, N. B. Grimm, S. Hughes, J. M. Marzluff, J. Munshi-South, C. Rojas, J. S. Santangelo, C. J. Schell, J. A. Schweitzer, M. Szulkin, M. C. Urban, Y. Zhou, and C. Ziter. 2022. A global horizon scan for urban evolutionary ecology. *Trends Ecol. Evol.*

Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4: e72.

Wang, X., J. Ding, S. Lin, D. Liu, T. Gu, H. Wu, R. N. Trigiano, R. McAvoy, J. Huang, and Y. Li. 2020. Evolution and roles of cytokinin genes in angiosperms 2: Do ancient CKXs play housekeeping roles while non-ancient CKXs play regulatory roles? *Hortic. Res.* 7: 29.

Weigand, H. and F. Leese. 2018. Detecting signatures of positive selection in non-model species using genomic data. *Zool J Linn Soc* 184: 528–583.

Whiting, J. R. 2022. *JimWhiting91/genotype_plot: Genotype Plot*. Version v0.2.1.

Winchell, K. M., I. Maayan, J. Fredette, and L. J. Revell. 2018. Linking locomotor performance to morphological shifts in urban lizards. *Proceedings of the Royal Society Biological Sciences* 285: 20180229.

Winchell, K. M., E. J. Carlen, A. R. Puente-Rolón, and L. J. Revell. 2017. Divergent habitat use of two urban lizard species. *Ecol. Evol.* 8: 25–35.

Winchell, K. M., R. G. Reynolds, S. R. Prado-Irwin, A. R. Puente-Rolón, and L. J. Revell. 2016. Phenotypic shifts in urban areas in the tropical lizard *Anolis cristatellus*. *Evolution* 70: 1009–1022.

Wu, R., W. Zheng, J. Tan, R. Sammer, L. Du, and C. Lu. 2019. Protein partners of plant ubiquitin-specific proteases (UBPs). *Plant Physiol. Biochem.* 145: 227–236.

Yakub, M. and P. Tiffin. 2017. Living in the city: urban environments shape the evolution of a native annual plant. *Glob. Chang. Biol.* 23: 2082–2089.

Yamaguchi, N., S. Matsubara, K. Yoshimizu, M. Seki, K. Hamada, M. Kamitani, Y. Kurita, Y. Nomura, K. Nagashima, S. Inagaki, T. Suzuki, E.-S. Gan, T. To, T. Kakutani, A. J. Nagano, A. Satake, and T. Ito. 2021. H3K27me3 demethylases alter HSP22 and HSP17.6C expression in response to recurring heat in *Arabidopsis*. *Nat. Commun.* 12: 3480.

Yeaman, S. 2022. Evolution of polygenic traits under global vs local adaptation. *Genetics* 220.

# Supplementary Text

## Text S1: Sweep signatures involving suburban populations

In addition to testing for signatures of positive selection between urban and rural populations, we used our phased haplotype data to compare compare selective sweep signatures between suburban and rural populations, and between urban and suburban populations. Specifically, we used *selscan* with suburban populations as the "object" population and rural populations as the "reference" (Figure S1a), and again with urban populations as the "object" and suburban populations as the "reference" (Figure S1b). Of the 135 genome-wide windows with evidence of positive selection in urban populations from the urban-rural comparison (Figure 3b), 15.5% were shared with suburban populations (Figure S1a). By contrast, of the 124 genome-wide windows with evidence of positive selection in rural populations from the urban-rural comparison (Figure 3b), only 0.8% were shared with suburban populations (Figure S1a). In addition, XP-nSL scores appeared less clustered and lower in magnitude among the urban-suburban comparison (Figure S1b) than the suburban-rural (Figure S1a) or urban-rural (Figure 3b) comparisons.

# Supplementary Figures

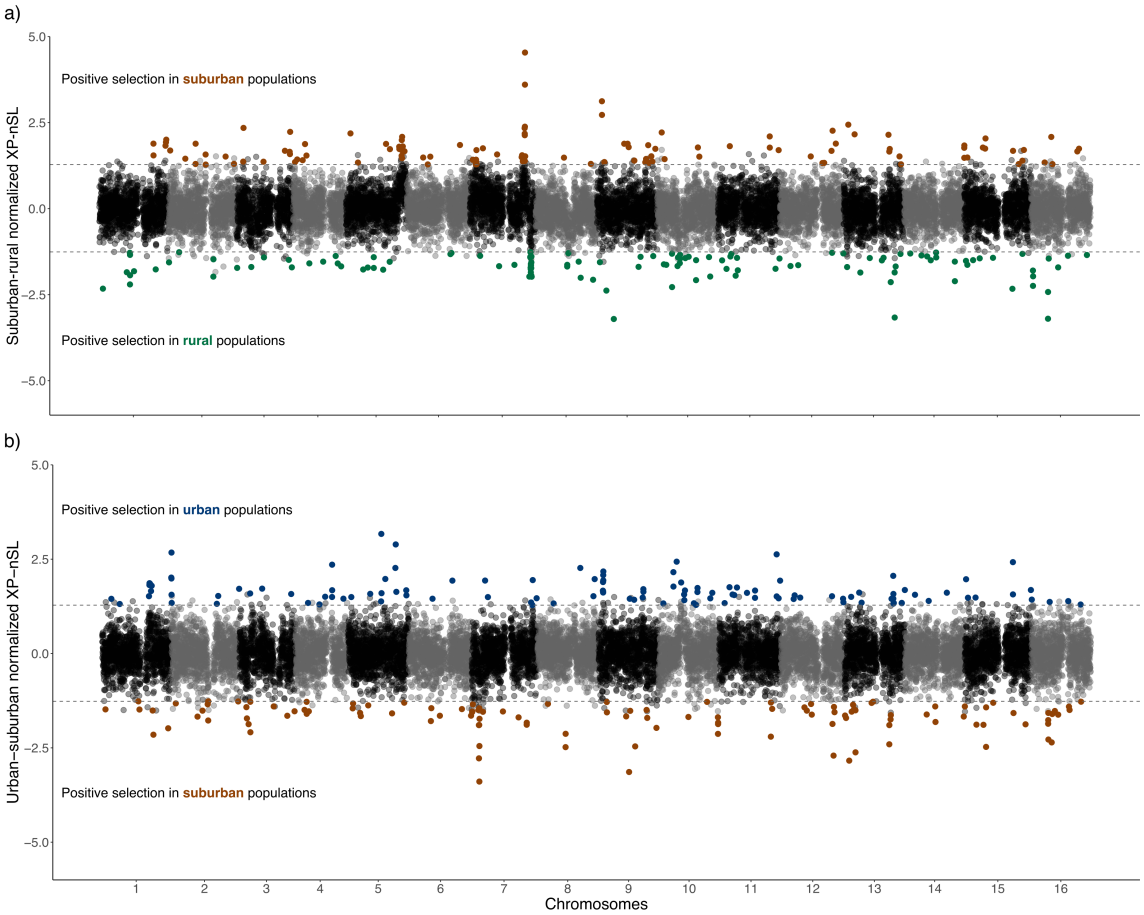


Figure S1: (a) Suburban-rural normalized XP-nSL estimated in 50 Kb windows across the genome. Brown points represent outliers showing evidence of positive selection in suburban habitats, while green points represent windows showing evidence of positive selection in rural habitats. (b) Urban-suburban normalized XP-nSL estimated in 50 Kb windows across the genome. Blue points represent outliers showing evidence of positive selection in urban habitats, while brown points represent windows showing evidence of positive selection in suburban habitats.

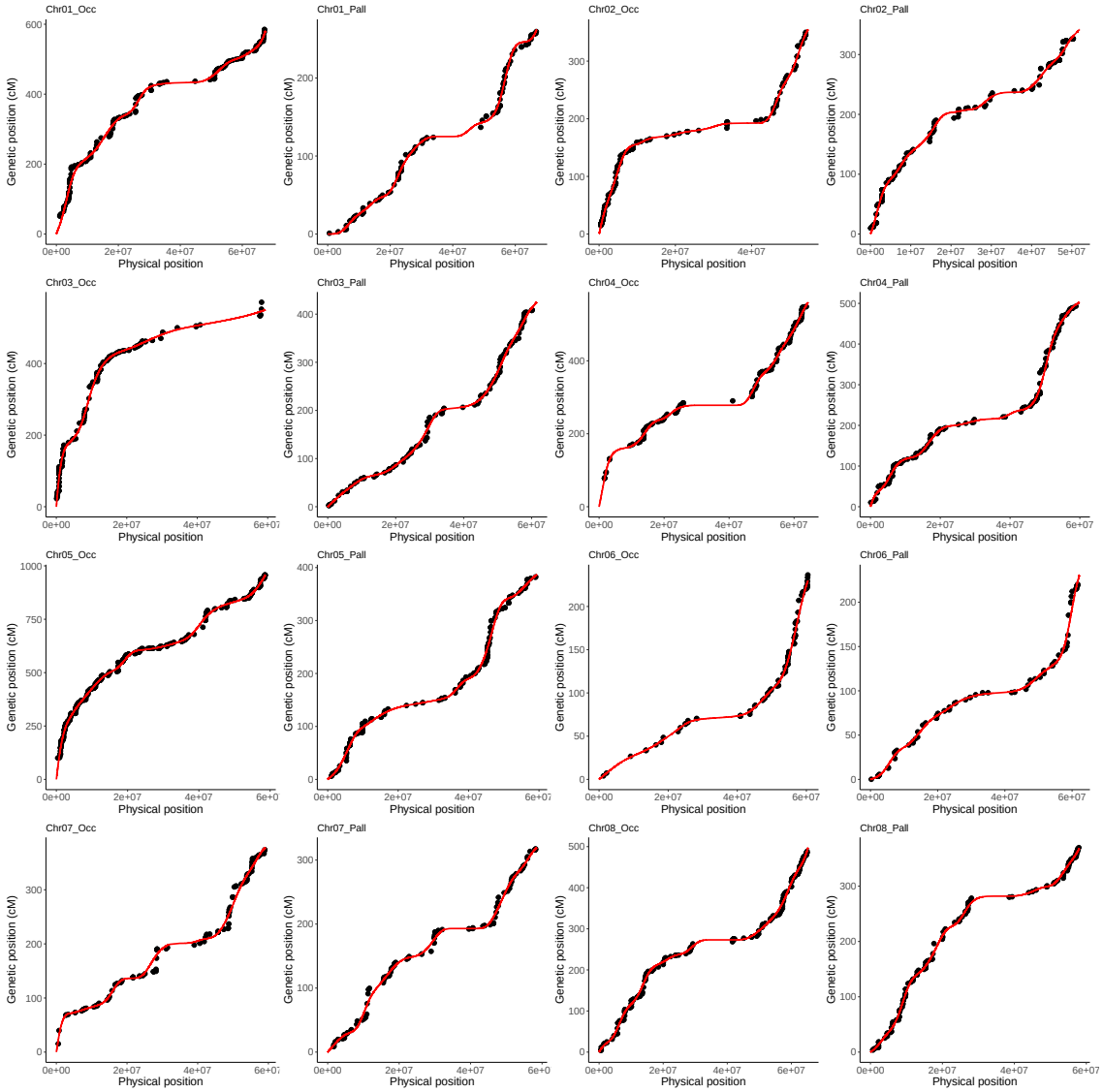


Figure S2: Chromosomal Marey Maps for all 16 white clover chromosomes, fit with interpolated and monotonically increasing cubic p-splines with 20 knots.

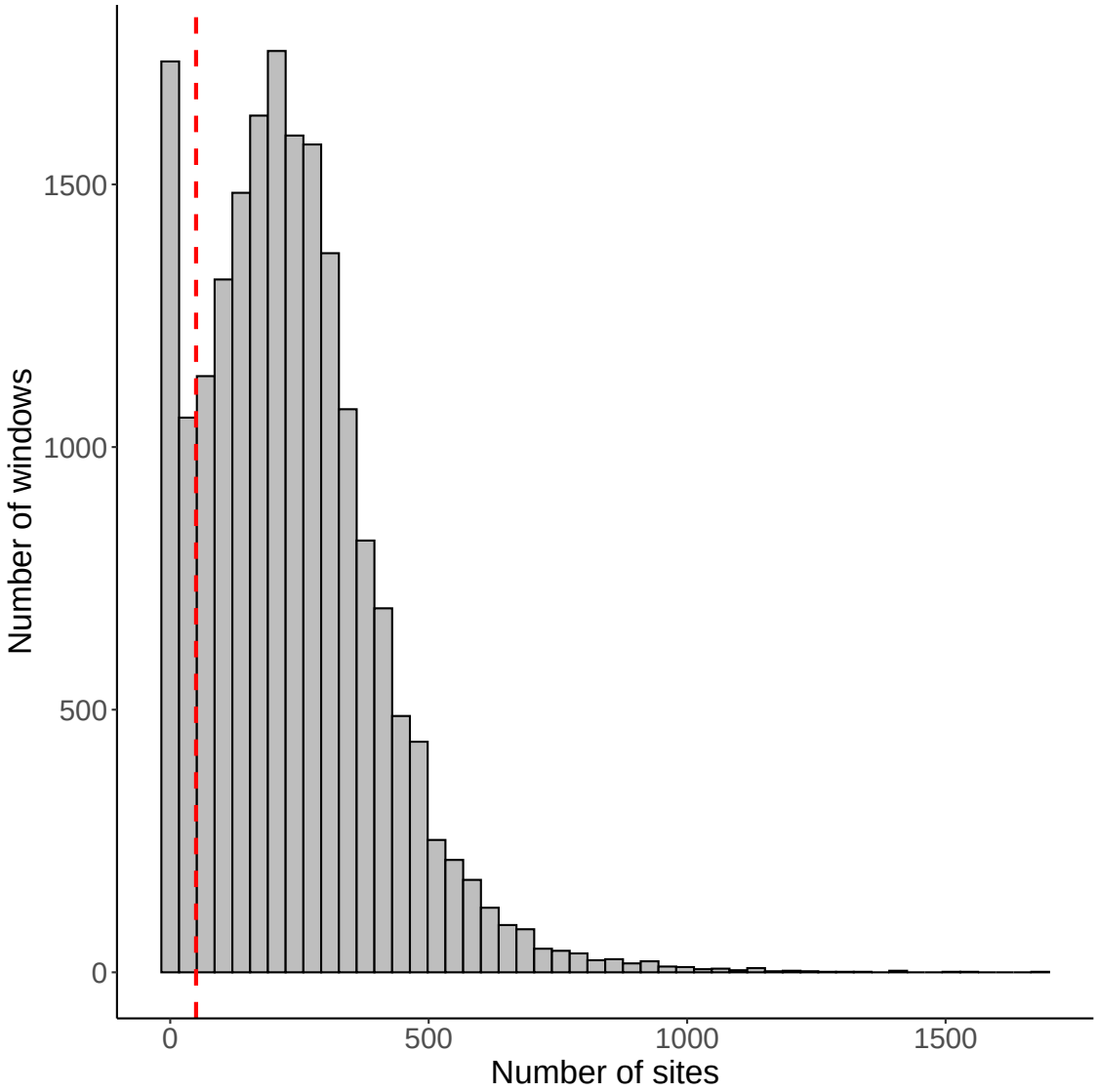


Figure S3: Histogram showing the number of SNPs in all 19,372, 50 Kb urban-rural XP-nSL windows. The vertical red dashed line is at 50 SNPs and represents our cutoff for inclusion of windows. Windows with fewer than 50 SNPs (i.e., left of red line, N = 2,733 or 14.1%) were removed to avoid noisy XP-nSL estimates in regions of low diversity.

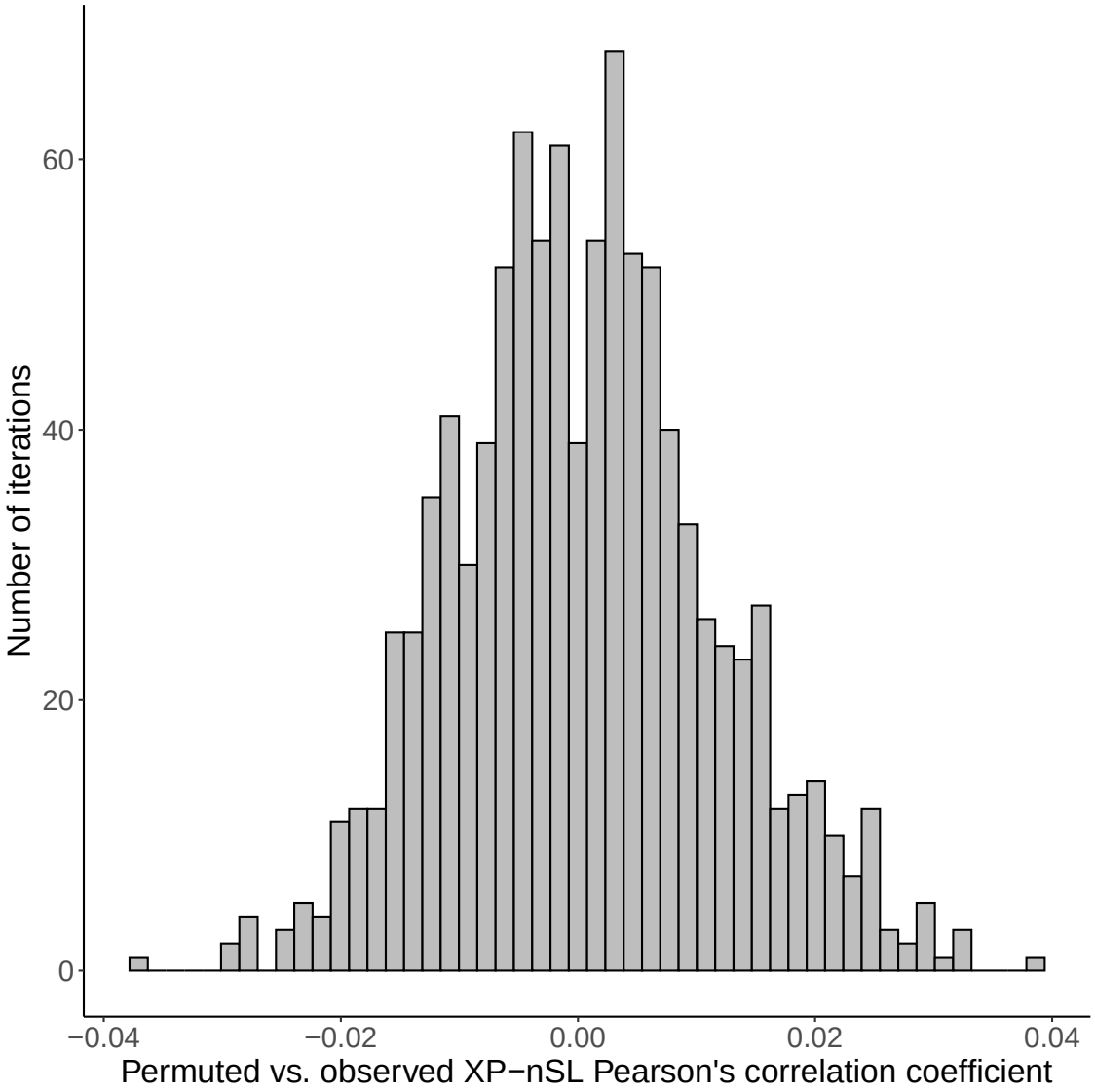


Figure S4: Histogram showing the Pearson's correlation coefficient comparing observed vs. permuted windowed mean XP-nSL scores across all 1000 permutations of urban and rural samples. As expected, the distribution is centered around zero.

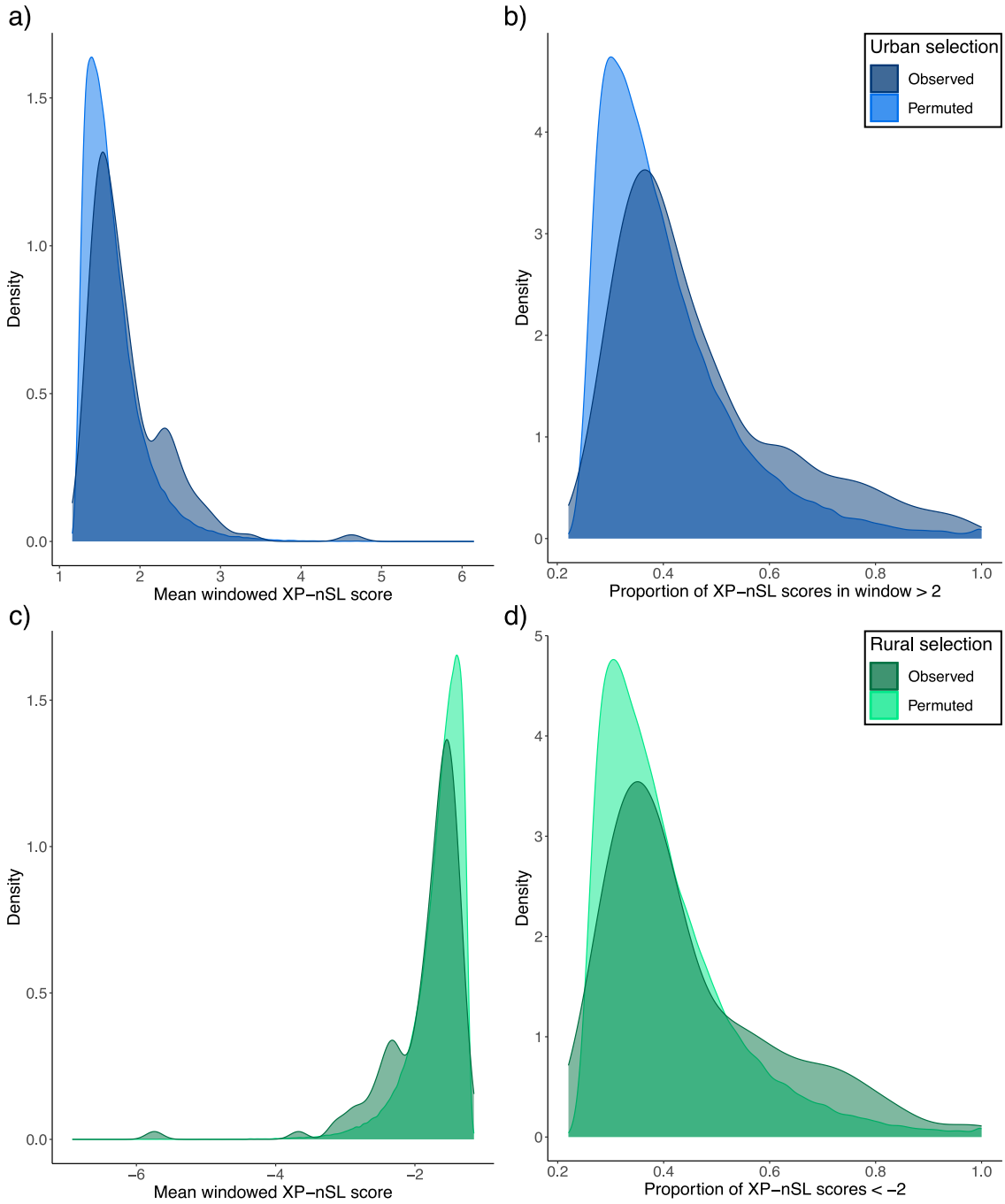


Figure S5: Comparison of observed urban-rural XP-nSL scores to those generated following 1000 permutations of urban and rural samples. (a) Comparison of mean windowed XP-nSL scores for windows with evidence of positive selection in urban habitats. (b) Comparison of the proportion of raw XP-nSL scores with values > 2 for windows with evidence of positive selection in urban habitats. Panels (c) and (d) are the same as (a) and (b), but for windows showing evidence of positive selection in rural habitats.

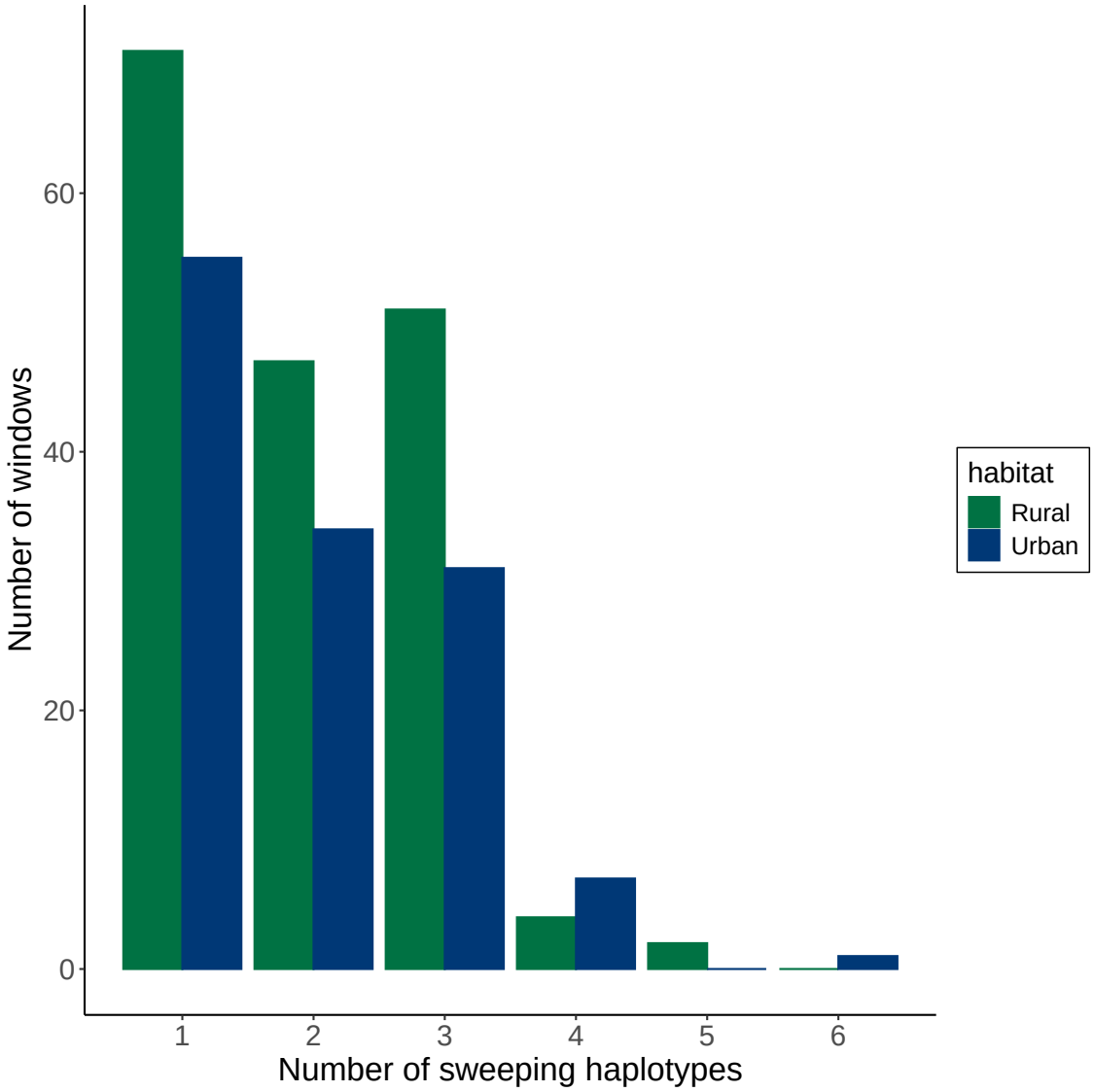


Figure S6: Estimated number of sweeping haplotypes among urban and rural XP-nSL outlier windows overlapping outlier *saltiLassi* $\lambda$ windows.

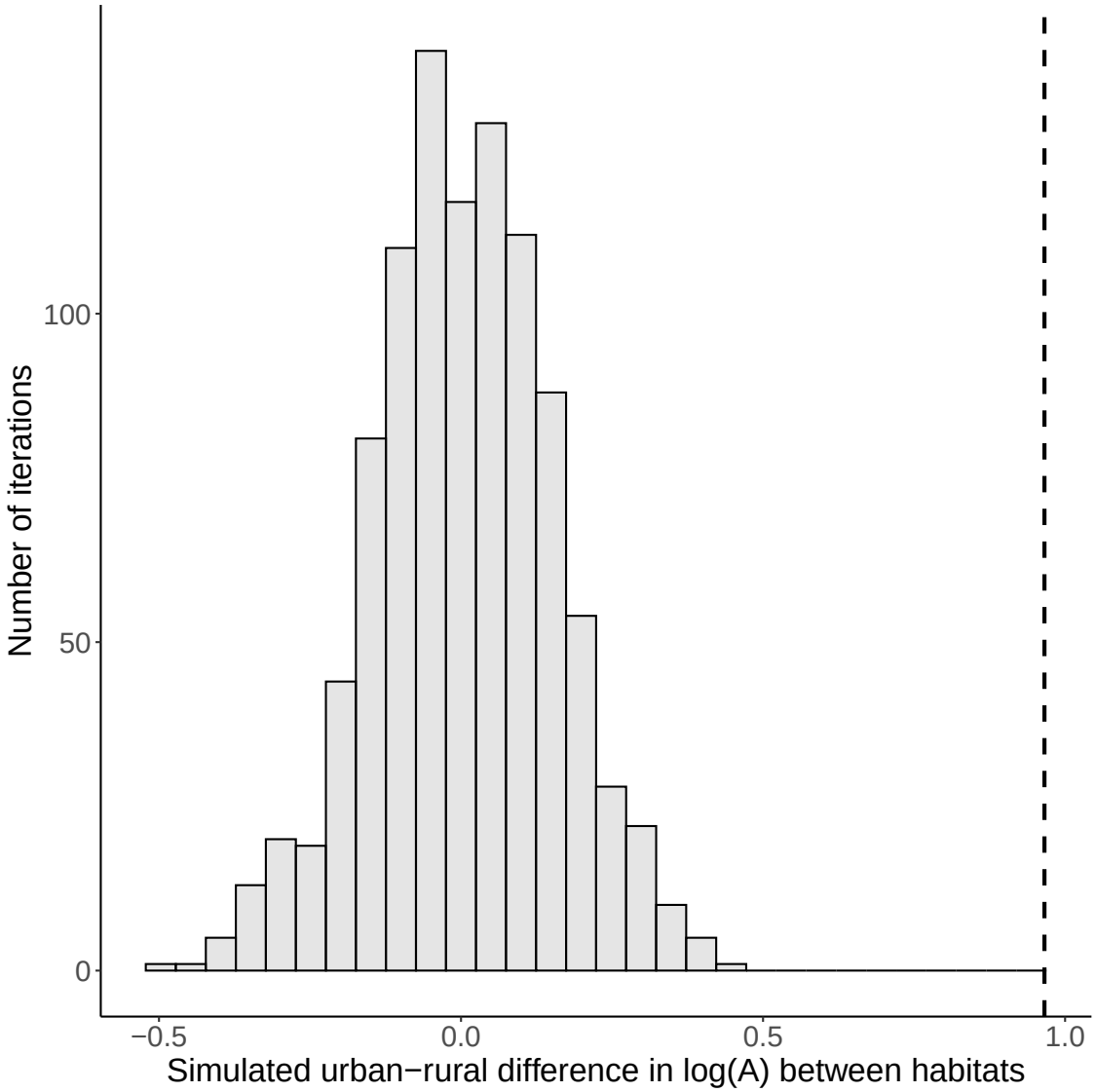


Figure S7: Permutation test of urban-rural difference in width of selective sweeps. Histogram shows simulated mean urban-rural difference in selective sweep widths following 1000 random permutations of urban and rural outlier window labels. Vertical dashed line shows the observed mean urban-rural difference in selective sweep widths.

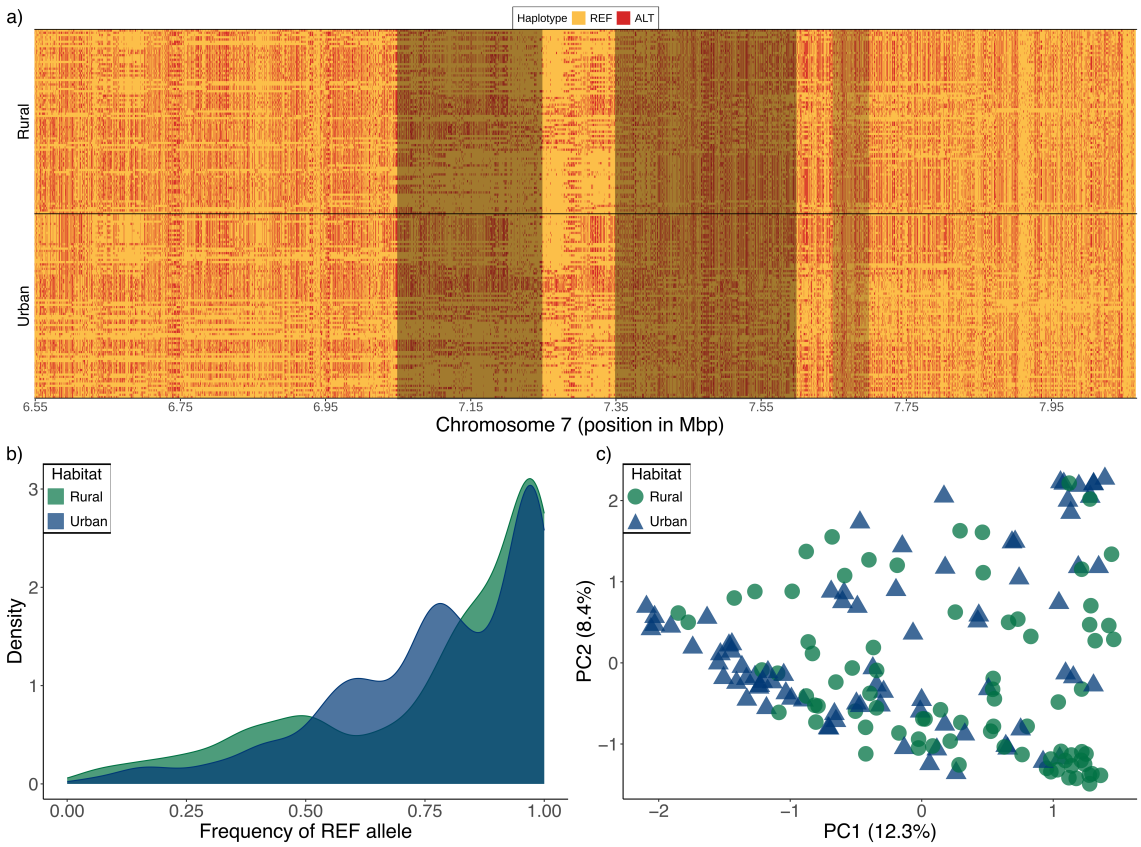


Figure S8: (a) Raw haplotypes in urban and rural habitats for region shown in Figure 3c with evidence of positive selection in rural habitats. Each individual is represented by two rows (one for each haplotype), with reference alleles coloured yellow and alternative alleles colored red. Light grey bands display XP-nSL outlier windows, while dark grey bands are XP-nSL windows that are additionally in the top 5% of the urban-rural permuted XP-nSL distribution. Figure generated using *genotype_plot.R* (Whiting 2022). (b) Frequency of the REF allele in urban and rural populations across all sites highlighted in grey from (a). (c) Principal components analysis performed on haplotype distances for regions highlighted in grey in (a). In the absence of selection, we expect urban and rural haplotype distances to be overlapping as in Figure S12b with little variation explained by the PCs. Selection drives haplotype differentiation leading to the formation of clusters in the PCA, though the number and shape of these clusters depends on the number of sweeping haplotypes, and the strength and time scale of selection. For example, a recent hard sweep in a single habitat would be expected to lead to a single cluster of samples in that habitat to most individuals sharing the selected haplotype, though cluster formation is expected to be incomplete in the case of ongoing selection.

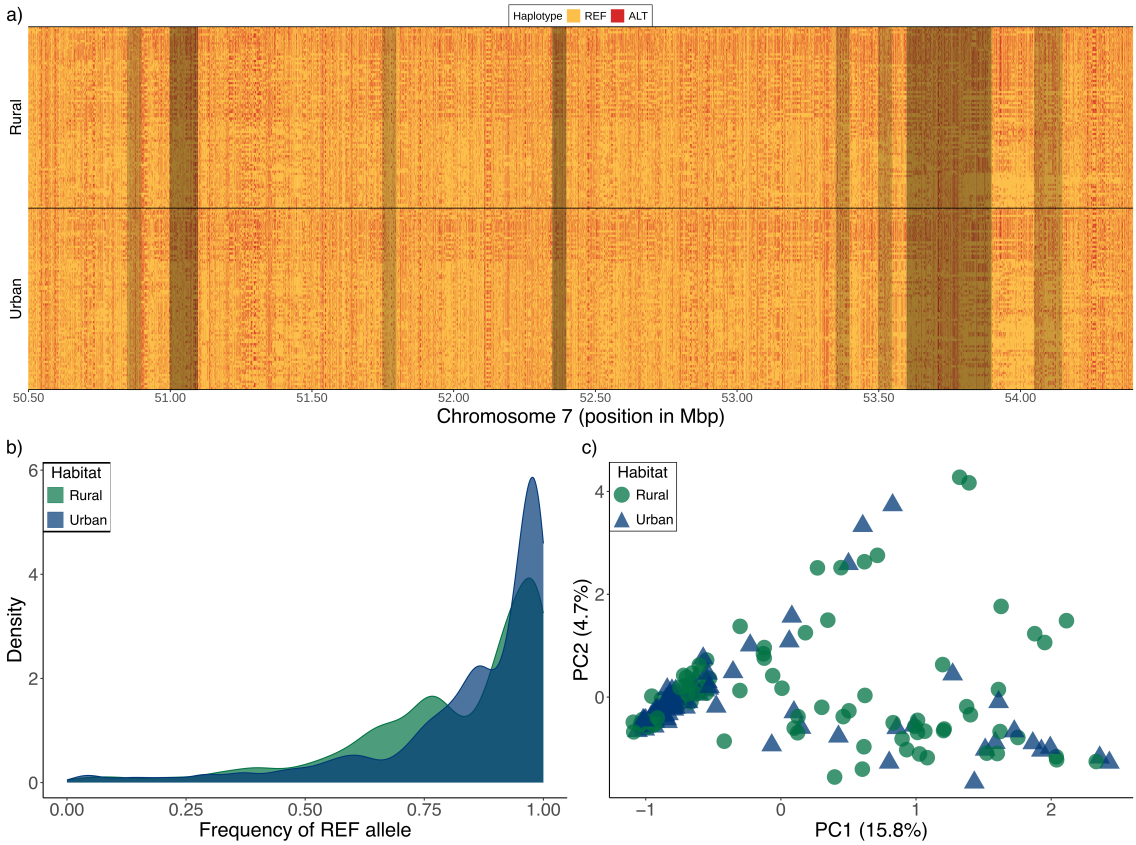


Figure S9: (a) Raw haplotypes in urban and rural habitats for region shown in Figure 3d with evidence of positive selection in urban habitats. Each individual is represented by two rows (one for each haplotype), with reference alleles coloured yellow and alternative alleles colored red. Light grey bands display XP-nSL outlier windows, while dark grey bands are XP-nSL windows that are additionally in the top 5% of the urban-rural permuted XP-nSL distribution. Figure generated using *genotype_plot.R* (Whiting 2022). (b) Frequency of the REF allele in urban and rural populations across all sites highlighted in grey from (a). (c) Principal components analysis performed on haplotype distances for regions highlighted in grey in (a).

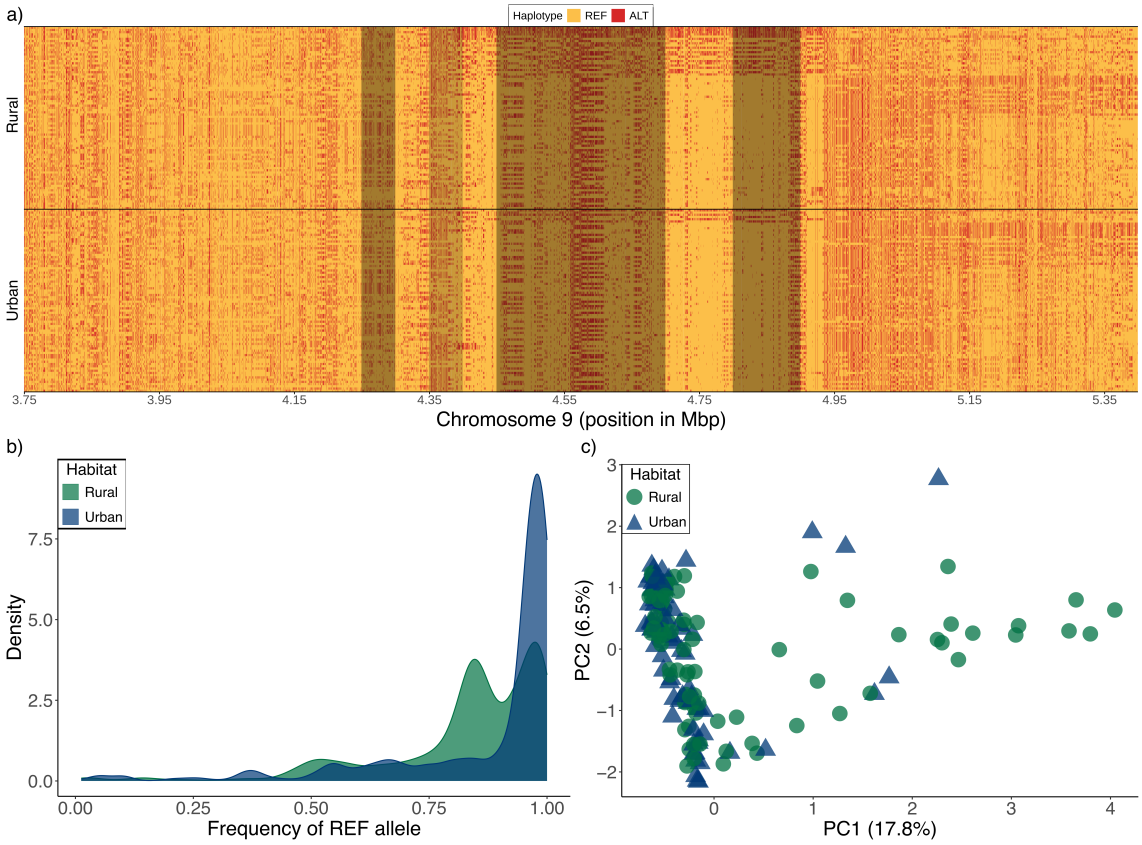


Figure S10: (a) Raw haplotypes in urban and rural habitats for region shown in Figure 3e with evidence of positive selection in urban habitats. Each individual is represented by two rows (one for each haplotype), with reference alleles coloured yellow and alternative alleles colored red. Light grey bands display XP-nSL outlier windows, while dark grey bands are XP-nSL windows that are additionally in the top 5% of the urban-rural permuted XP-nSL distribution. Figure generated using *genotype_plot.R* (Whiting 2022). (b) Frequency of the REF allele in urban and rural populations across all sites highlighted in grey from (a). (c) Principal components analysis performed on haplotype distances for regions highlighted in grey in (a).

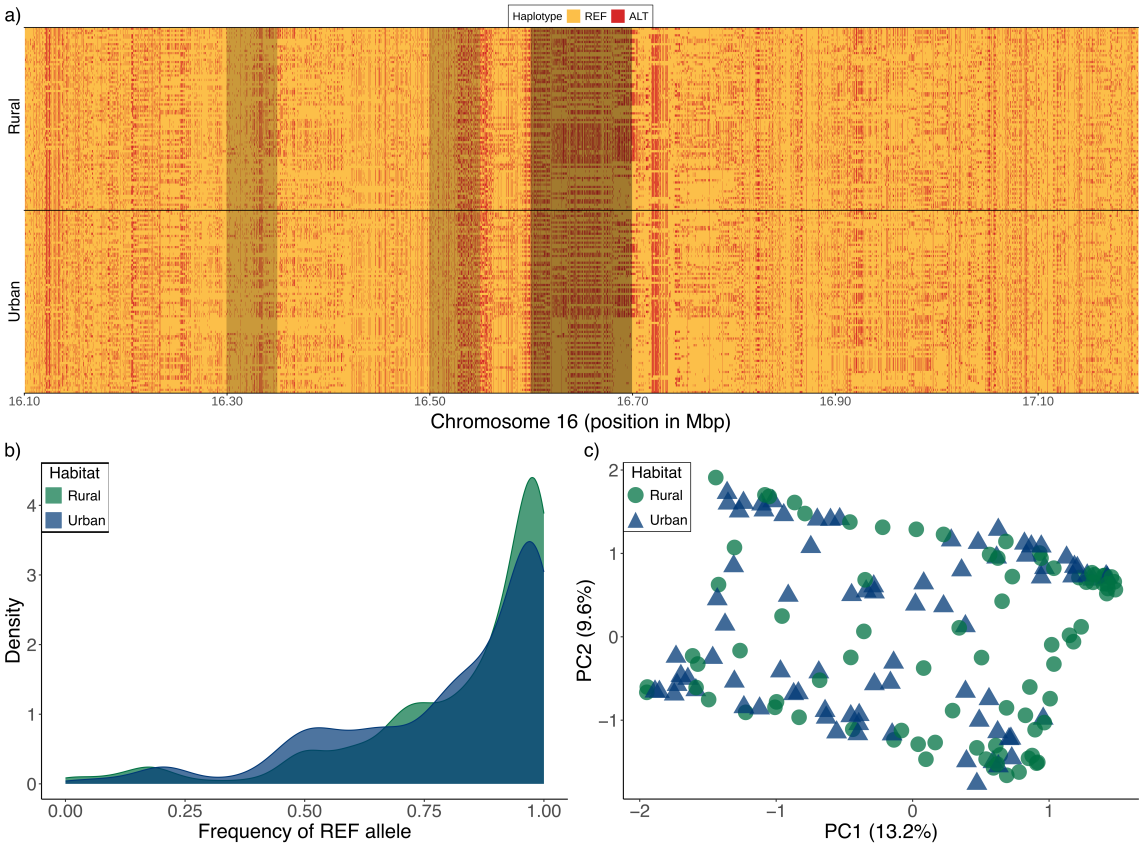


Figure S11: (a) Raw haplotypes in urban and rural habitats for region shown in Figure 3f with evidence of positive selection in urban habitats. Each individual is represented by two rows (one for each haplotype), with reference alleles coloured yellow and alternative alleles colored red. Light grey bands display XP-nSL outlier windows, while dark grey bands are XP-nSL windows that are additionally in the top 5% of the urban-rural permuted XP-nSL distribution. Figure generated using *genotype_plot.R* (Whiting 2022). (b) Frequency of the REF allele in urban and rural populations across all sites highlighted in grey from (a). (c) Principal components analysis performed on haplotype distances for regions highlighted in grey in (a).

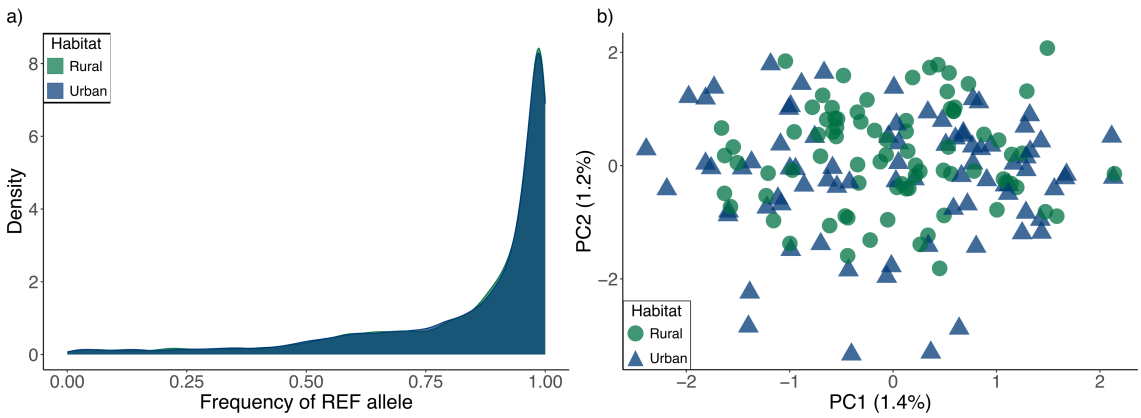


Figure S12: Urban and rural allele frequencies and haplotype differentiation across 100 randomly-selected, unselected regions (i.e., not XP-nSL outliers) along chromosome 1. (a) Frequency of the REF allele in urban and rural populations across all sites. Note that urban and rural allele frequency histograms are perfectly overlapping. (b) Principal components analysis performed on haplotype distances across regions.

# 1070 Supplementary Tables

Table S1: Significant Gene Ontology (GO) categories for Biological Processes and Molecular Function for genes within regions with evidence of selective sweeps in urban populations.

| Ontology | GO.ID | Term | Annotated | Significant | Expected | Pval |
|---|---|---|---|---|---|---|
| Biological Process | GO:0000972 | transcription-dependent tethering of RNA... | 29 | 2 | 0.04 | 0.00064 |
| Biological Process | GO:0000272 | polysaccharide catabolic process | 39 | 2 | 0.05 | 0.00116 |
| Biological Process | GO:0006405 | RNA export from nucleus | 108 | 2 | 0.14 | 0.00856 |
| Biological Process | GO:0006357 | regulation of transcription by RNA polym... | 956 | 5 | 1.23 | 0.00901 |
| Biological Process | GO:0009065 | glutamine family amino acid catabolic pr... | 9 | 1 | 0.01 | 0.01152 |
| Biological Process | GO:0006606 | protein import into nucleus | 127 | 2 | 0.16 | 0.01167 |
| Biological Process | GO:0034067 | protein localization to Golgi apparatus | 10 | 1 | 0.01 | 0.01279 |
| Biological Process | GO:0010089 | xylem development | 11 | 1 | 0.01 | 0.01406 |
| Biological Process | GO:0035999 | tetrahydrofolate interconversion | 16 | 1 | 0.02 | 0.02039 |
| Biological Process | GO:0019264 | glycine biosynthetic process from serine | 16 | 1 | 0.02 | 0.02039 |
| Biological Process | GO:0071586 | CAAX-box protein processing | 16 | 1 | 0.02 | 0.02039 |
| Biological Process | GO:0006047 | UDP-N-acetylglucosamine metabolic proces... | 19 | 1 | 0.02 | 0.02417 |
| Biological Process | GO:0006525 | arginine metabolic process | 21 | 1 | 0.03 | 0.02668 |
| Biological Process | GO:0001763 | morphogenesis of a branching structure | 21 | 1 | 0.03 | 0.02668 |
| Biological Process | GO:0015693 | magnesium ion transport | 22 | 1 | 0.03 | 0.02793 |
| Biological Process | GO:0048646 | anatomical structure formation involved ... | 23 | 1 | 0.03 | 0.02918 |
| Biological Process | GO:2000112 | regulation of cellular macromolecule bio... | 1861 | 6 | 2.39 | 0.0306 |
| Biological Process | GO:0110154 | RNA decapping | 28 | 1 | 0.04 | 0.03541 |
| Biological Process | GO:0006623 | protein targeting to vacuole | 29 | 1 | 0.04 | 0.03666 |
| Biological Process | GO:0009063 | cellular amino acid catabolic process | 30 | 1 | 0.04 | 0.0379 |
| Biological Process | GO:0017038 | protein import | 243 | 2 | 0.31 | 0.03907 |
| Biological Process | GO:0000184 | nuclear-transcribed mRNA catabolic proce... | 34 | 1 | 0.04 | 0.04284 |
| Biological Process | GO:0006536 | glutamate metabolic process | 35 | 1 | 0.05 | 0.04407 |
| Biological Process | GO:0006625 | protein targeting to peroxisome | 38 | 1 | 0.05 | 0.04776 |
| Molecular Function | GO:0016161 | beta-amylase activity | 20 | 2 | 0.03 | 0.0004 |
| Molecular Function | GO:0000978 | RNA polymerase II cis-regulatory region ... | 439 | 5 | 0.65 | 0.00049 |
| Molecular Function | GO:0017056 | structural constituent of nuclear pore | 47 | 2 | 0.07 | 0.00223 |
| Molecular Function | GO:0046983 | protein dimerization activity | 982 | 6 | 1.45 | 0.00328 |
| Molecular Function | GO:0033612 | receptor serine/threonine kinase binding | 70 | 2 | 0.1 | 0.00488 |
| Molecular Function | GO:0004412 | homoserine dehydrogenase activity | 9 | 1 | 0.01 | 0.01325 |
| Molecular Function | GO:0004072 | aspartate kinase activity | 13 | 1 | 0.02 | 0.01908 |
| Molecular Function | GO:0004372 | glycine hydroxymethyltransferase activit... | 16 | 1 | 0.02 | 0.02344 |
| Molecular Function | GO:0015095 | magnesium ion transmembrane transporter ... | 22 | 1 | 0.03 | 0.03209 |
| Molecular Function | GO:0030170 | pyridoxal phosphate binding | 197 | 2 | 0.29 | 0.03463 |
| Molecular Function | GO:0030544 | Hsp70 protein binding | 25 | 1 | 0.04 | 0.03638 |
| Molecular Function | GO:0008194 | UDP-glycosyltransferase activity | 525 | 3 | 0.78 | 0.043 |
| Molecular Function | GO:0004497 | monooxygenase activity | 902 | 4 | 1.34 | 0.04467 |
| Molecular Function | GO:0016868 | intramolecular transferase activity, pho... | 31 | 1 | 0.05 | 0.04492 |

Table S2: Significant Gene Ontology (GO) categories for Biological Processes and Molecular Function for genes within regions with evidence of selective sweeps in rural populations.

| Ontology | GO.ID | Term | Annotated | Significant | Expected | Pval |
|---|---|---|---|---|---|---|
| Biological Process | GO:0007021 | tubulin complex assembly | 7 | 2 | 0.01 | 7.8e-05 |
| Biological Process | GO:0007023 | post-chaperonin tubulin folding pathway | 10 | 2 | 0.02 | 0.00017 |
| Biological Process | GO:0045492 | xylan biosynthetic process | 42 | 2 | 0.08 | 0.00305 |
| Biological Process | GO:0009057 | macromolecule catabolic process | 1618 | 4 | 3.15 | 0.00998 |
| Biological Process | GO:0016559 | peroxisome fission | 6 | 1 | 0.01 | 0.01162 |
| Biological Process | GO:0018130 | heterocycle biosynthetic process | 4797 | 11 | 9.33 | 0.01304 |
| Biological Process | GO:0051083 | 'de novo' cotranslational protein foldin... | 7 | 1 | 0.01 | 0.01354 |
| Biological Process | GO:0032784 | regulation of DNA-templated transcriptio... | 91 | 2 | 0.18 | 0.01364 |
| Biological Process | GO:0006450 | regulation of translational fidelity | 9 | 1 | 0.02 | 0.01737 |
| Biological Process | GO:0044271 | cellular nitrogen compound biosynthetic ... | 6247 | 14 | 12.15 | 0.01911 |
| Biological Process | GO:0032957 | inositol trisphosphate metabolic process | 11 | 1 | 0.02 | 0.02119 |
| Biological Process | GO:0046168 | glycerol-3-phosphate catabolic process | 12 | 1 | 0.02 | 0.0231 |
| Biological Process | GO:0018410 | C-terminal protein amino acid modificati... | 12 | 1 | 0.02 | 0.0231 |
| Biological Process | GO:0034727 | piecemeal microautophagy of the nucleus | 14 | 1 | 0.03 | 0.0269 |
| Biological Process | GO:0006376 | mRNA splice site selection | 15 | 1 | 0.03 | 0.02879 |
| Biological Process | GO:0030490 | maturation of SSU-rRNA | 138 | 2 | 0.27 | 0.02971 |
| Biological Process | GO:0007018 | microtubule-based movement | 140 | 2 | 0.27 | 0.03051 |
| Biological Process | GO:0007059 | chromosome segregation | 249 | 2 | 0.48 | 0.04342 |
| Biological Process | GO:0045039 | protein insertion into mitochondrial inn... | 24 | 1 | 0.05 | 0.04567 |
| Molecular Function | GO:0048487 | beta-tubulin binding | 5 | 2 | 0.01 | 3.3e-05 |
| Molecular Function | GO:0015018 | galactosylgalactosylxylosylprotein 3-bet... | 10 | 2 | 0.02 | 0.00015 |
| Molecular Function | GO:0042285 | xylosyltransferase activity | 14 | 2 | 0.03 | 0.0003 |
| Molecular Function | GO:0005524 | ATP binding | 4687 | 17 | 8.56 | 0.00407 |
| Molecular Function | GO:0016887 | ATP hydrolysis activity | 1201 | 7 | 2.19 | 0.00634 |
| Molecular Function | GO:0010333 | terpene synthase activity | 68 | 2 | 0.12 | 0.00694 |
| Molecular Function | GO:0000287 | magnesium ion binding | 265 | 3 | 0.48 | 0.01275 |
| Molecular Function | GO:0047325 | inositol tetrakisphosphate 1-kinase acti... | 11 | 1 | 0.02 | 0.01991 |
| Molecular Function | GO:0052725 | inositol-1,3,4-trisphosphate 6-kinase ac... | 11 | 1 | 0.02 | 0.01991 |
| Molecular Function | GO:0052726 | inositol-1,3,4-trisphosphate 5-kinase ac... | 11 | 1 | 0.02 | 0.01991 |
| Molecular Function | GO:0047952 | glycerol-3-phosphate dehydrogenase [NAD(... | 12 | 1 | 0.02 | 0.0217 |
| Molecular Function | GO:0003724 | RNA helicase activity | 125 | 2 | 0.23 | 0.02207 |
| Molecular Function | GO:0016765 | transferase activity, transferring alkyl... | 332 | 3 | 0.61 | 0.02305 |
| Molecular Function | GO:0003777 | microtubule motor activity | 140 | 2 | 0.26 | 0.02723 |
| Molecular Function | GO:0003917 | DNA topoisomerase type I (single strand ... | 16 | 1 | 0.03 | 0.02883 |
| Molecular Function | GO:0008017 | microtubule binding | 365 | 3 | 0.67 | 0.02939 |
| Molecular Function | GO:0019139 | cytokinin dehydrogenase activity | 19 | 1 | 0.03 | 0.03415 |
| Molecular Function | GO:0008641 | ubiquitin-like modifier activating enzym... | 22 | 1 | 0.04 | 0.03943 |
| Molecular Function | GO:0005452 | solute:inorganic anion antiporter activi... | 22 | 1 | 0.04 | 0.03943 |
| Molecular Function | GO:0003688 | DNA replication origin binding | 22 | 1 | 0.04 | 0.03943 |
| Molecular Function | GO:0030544 | Hsp70 protein binding | 25 | 1 | 0.05 | 0.04469 |
| Molecular Function | GO:0000978 | RNA polymerase II cis-regulatory region ... | 439 | 3 | 0.8 | 0.04658 |
| Molecular Function | GO:0005388 | P-type calcium transporter activity | 28 | 1 | 0.05 | 0.04992 |