Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani Piano Nazionale di Ripresa e Resilienza | Università degli Studi di Napoli Federico II | National Biodiversity Future Center

University of Naples

Federico II

DEPARTMENT OF BIOLOGY

Ph.D. in Biology

XXXVIII cycle

A.A. 2025-2026

Coordinator: Prof. Sergio Esposito

Adaptive genomic and transcriptomic variation in *Raphanus raphanistrum* in urban environments

Supervisor

Prof. Serena Aceto

Ph.D. candidate

Dr. Angela Carfora

INDEX

ABSTRACT

Urbanization represents one of the most transformative contemporary environmental changes, significantly altering biodiversity through habitat loss and fragmentation. Beyond their destructive impact, urban environments also serve as unique ecosystems that promote rapid evolutionary adaptation through novel selective pressures. Identifying which species prosper or decline as urbanization intensifies is essential for effective biodiversity management and conservation strategies.

Plants are continually exposed to adverse conditions and, although sessile, have evolved complex cellular, molecular, and biochemical mechanisms to withstand environmental challenges. However, many resistance strategies deployed under chronic urban stress remain insufficiently characterized. Addressing these knowledge gaps is critical for understanding the ecology and evolution of species in anthropogenic landscapes, making targeted urban biodiversity research necessary to clarify how species adapt and establish in these complex environments.

Raphanus raphanistrum (Brassicaceae) can be used as a model due to its notable plasticity and capacity to thrive in diverse and disturbed habitats. To investigate the molecular pathways underlying urban adaptation, samples were collected from sites selected based on impervious surface percentages and soil types, providing a robust framework to examine how urban-induced stressors affect the genetic and functional architecture of this species.

Although *R. raphanistrum* demonstrates ecological success in urban environments, the genetic and transcriptomic mechanisms underlying its resilience remain poorly understood. This study addressed these gaps using an integrated approach combining RNA sequencing, whole-genome sequencing, and genome-wide association studies.

Transcriptomic analysis revealed substantial gene expression reprogramming in urban populations, particularly within the flavonoid biosynthesis pathway. This coordinated upregulation indicates that enhanced antioxidant and UV-protective mechanisms represent a primary adaptive response to oxidative stress and altered light conditions typical of urban environments.

Building on these findings, genome-wide scans detected signatures of selection. Integration of multiple independent statistical methods enabled the identification of high-confidence candidate loci, where selection appears to promote the incorporation of specific variants into haplotype blocks that could enhance the regulation of genes involved in stress signaling and cellular homeostasis. To establish the relationship between molecular signals and organismal fitness, two morphological traits, leaf area and fruit weight, were evaluated through genome-wide association studies. The

identification of pleiotropic loci associated with reduced vegetative and reproductive organ size suggests an adaptive life-history shift toward size reduction, potentially improving survival under heat and water limitations characteristic of urban heat islands.

Population structure analysis of RNA-seq data confirmed the intra-population genetic variability observed in both genomic and transcriptomic datasets.

Finally, to capture the full spectrum of genetic diversity, a genus-level pangenome was constructed, that could be used in future studies to explore structural variations as drivers of genomic plasticity beyond single-nucleotide polymorphisms.

Collectively, these findings demonstrate that urban adaptation in *R. raphanistrum* is a multifaceted process involving coordinated modifications of both transcriptome and genome architecture. This research offers new insights into how resilient species modify their genetic identity in response to evolutionary forces. These results have direct implications for predicting species responses to ongoing urbanization and informing evidence-based conservation strategies in rapidly changing urban landscapes.

1. Introduction

1.1 Urbanization as an Evolutionary Force

Urbanization represents one of the most significant anthropogenic transformations in contemporary history (1), reshaping landscapes and ecosystems worldwide. As global urbanization accelerates, humans have become important agents of niche construction, profoundly influencing the ecological and evolutionary processes that drive biodiversity (2,3). Human activities have introduced novel selection pressures with significant evolutionary consequences, not only for humans but also for the broader natural world, resulting in complex feedbacks in both ecological and evolutionary processes (4–6). Although human activities have influenced ecological dynamics for millennia, the phenomenon of urbanization is marked by a substantial intensification in terms of scale, speed, and impact (7). Today, 55% of the world's population resides in urban areas, and urbanization is expected to continue growing (8).

The eco-evolutionary consequences of human-driven niche construction, however, have only recently begun to receive focused attention (9). Human activities fundamentally reshape ecosystems, acting as continuous agents of environmental transformation. The expansion of human populations into increasingly diverse environments has been made possible by sustained anthropogenic modifications of ecological systems. Cities, as concentrated hubs of human-induced change, have thus become important arenas for rapid evolutionary processes. Urban populations frequently exhibit shifts in genetic diversity and allele frequencies, which in turn lead to changes in phenotypic traits such as physiology, morphology, and behavior. These evolutionary responses, in turn, affect reproduction, survival, and population dynamics, with broader implications for nutrient cycling, productivity, and overall ecosystem functioning (10).

As species adapt to urban environments, they may develop behavioral, morphological, and genetic traits that differ substantially from their rural counterparts. This evolutionary divergence has practical implications for wildlife management, as strategies developed for rural populations may be ineffective in urban contexts, underscoring the need for location-specific biological research to inform effective conservation and management practices (11). The environmental changes impose novel selection pressures that can drive shifts in ecologically relevant traits, with cascading effects on species interactions and ecosystem stability (12,13) and can directly influence the physiology, growth, and behavior of both plants and animals (14–18), in particular, may alter natural selection on plant growth and development due to marked differences in temperature regimes, precipitation patterns,

and air quality compared to rural areas (19). While still limited, a growing body of research has documented evidence of adaptation and directional selection in response to urban conditions (20,21).

While these impacts can be profound and often disruptive, they are neither entirely detrimental nor entirely avoidable. Despite the ecological degradation these processes have frequently caused, they have also contributed to the creation of more sustainable systems in certain contexts (3,22).

Urban adaptation can occur through two main pathways: novel mutations after urban colonization or selection on existing genetic variations (20). A prominent example is the industrial melanism in peppered moths (*Biston betularia*), where a mutation arose in the early 1800s in response to pollution during the Industrial Revolution (23). In contrast, urban adaptations like reduced hydrogen cyanide production in white clover (*Trifolium repens*) suggest selection on standing genetic variation, particularly in warmer urban areas (17,18). Similar adaptations from pre-existing variation have also been observed in Atlantic killifish (19), tomcod (24), white-footed mice (25), and European blackbirds (26).

Urban evolution encompasses both adaptive and non-adaptive processes. While some species successfully adapt to urban conditions, habitat fragmentation and degradation fundamentally alter evolutionary dynamics by modifying patterns of genetic drift and gene flow among populations (27,28). Urban development converts natural habitats into human infrastructure (e.g., roads, canals, and dams), creating fragmented habitat patches that isolate wildlife populations and impede dispersal (29,30). This fragmentation increases genetic drift within populations while reducing gene flow between them, leading to decreased genetic diversity within populations and increased genetic differentiation among populations under neutral evolutionary processes (25). Isolated urban populations face elevated risks of inbreeding depression and genetic bottlenecks, which can potentially reduce their long-term evolutionary potential (31–33).

Dispersal patterns are particularly critical in determining species' ability to track changing climates and survive habitat loss (34,35). For species capable of dispersing across urban environments, or those transported by humans, urbanization can facilitate gene flow among populations through direct human-mediated movement (36) through two primary mechanisms: human-vectored dispersal (HVD), when humans directly transport organisms, or human-altered dispersal (HAD) indirect effects of humans including modifications to landscape structure, dispersal vectors, and animal behavior (37). Together, these processes can fundamentally reshape gene flow patterns in ways that either enhance or constrain evolutionary responses to urban environments.

Human-facilitated gene flow plays a crucial role in reducing the genetic isolation between distinct populations. This process enables the introduction of novel alleles into urban populations, thereby significantly enhancing their genetic diversity. Additionally, increased gene flow mitigates the effects of genetic drift, thereby fostering a more robust and adaptable urban ecosystem. While the introduction of new alleles can reduce the frequency of locally adapted genotypes (38), consistent selective pressures across urban environments may lead to the evolution of an "urban genotype" that spreads throughout urban areas (39–41).

1.2 Local Adaptation

Urban populations provide valuable insights into species' responses to environmental change, including the genes and traits involved, and the rate of local adaptation (42). Understanding which species can thrive under urban conditions and which are at risk of decline is crucial for informing conservation strategies (43).

Urban environments impose unique selective pressures driving local adaptation, resulting from the operation of selection on standing genetic variation, as opposed to novel mutations, over relatively short timescales (44,45). Local adaptation occurs when different populations of the same species evolve to develop traits that enhance their fitness in specific environments, regardless of how effective these traits are in other settings (46) (Fig. 1). The local environment significantly influences phenotypic differences across a species' range, affecting traits through phenotypic plasticity and ultimately impacting the evolution of local adaptation (47). Many of the traits under selection in urban contexts are quantitative, shaped by the combined effects of multiple genes of small effect (48,49), and may result in subtle but ecologically relevant phenotypic changes that are not immediately apparent (50). While urban adaptation frequently involves complex polygenic traits, simpler traits with more straightforward genetic architectures may also be subject to selection (51).

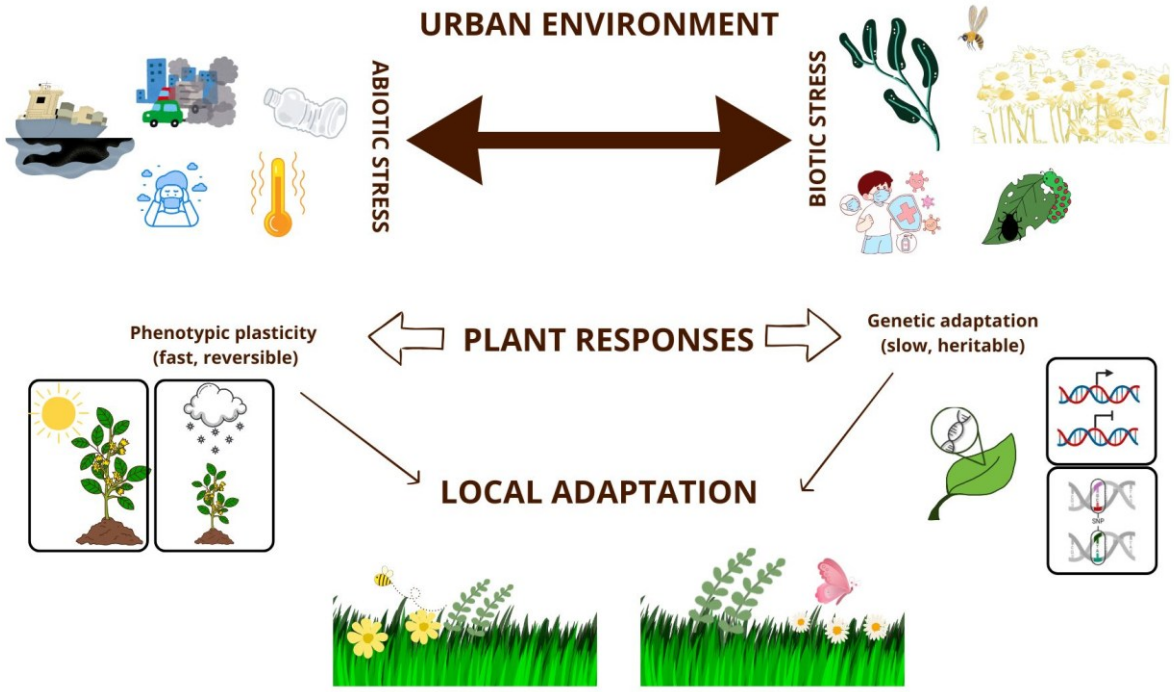

Figure 1. Overview of plant adaptation to urban environments. Urban environments expose plant populations to a range of abiotic (e.g., pollution, heat, drought) and biotic (e.g., altered competition, pathogens, herbivory) stressors. These pressures elicit two main types of responses: phenotypic plasticity (rapid and reversible adjustments mediated by gene expression) and genetic adaptation (long-term, heritable changes in allele frequencies). Together, these processes contribute to the emergence of local adaptation within urban ecosystems.

Urban stress adaptation proves especially crucial for non-mobile organisms such as plants, which cannot relocate when facing environmental pressures (52). Therefore, plant resilience in metropolitan environments constitutes a compelling research domain, particularly considering the varied circumstances under which vegetative areas are developed. Although the acquisition of novel phenotypic characteristics and the selection of stress-tolerant genetic variants require multiple generations, plants depend on immediate transcriptional regulation as a primary mechanism to withstand urban environmental challenges (20). Studies examining trait variation along urban–rural gradients reveal that plant responses involve both phenotypic plasticity and genetic divergence operating simultaneously. Research on Asian dayflower (*Commelina communis*) (53) populations across urbanization gradients demonstrates this dual mechanism: while increased plant height and leaf area in urban field populations might represent plastic responses to environmental conditions, reduced leaf production in urban plants suggests genetic divergence from rural ancestors.

Furthermore, urban populations evolved novel developmental coordination patterns, as evidenced by altered correlations between morphological traits that persisted even when plants were grown

under controlled greenhouse conditions, indicating that urbanization can reshape not only individual traits but also the integrated architecture of plant phenotypes. The review paper of Carfora et al. (54) outlines the molecular mechanisms underlying plant responses to urban stressors, highlighting both expression-based adjustments and genotypic changes associated with challenges such as heavy metal contamination, heightened CO_2 levels, air pollution, heat, drought, salinity, and herbivore pressure.

Investigating the genetic basis of local adaptation in cities provides valuable insights into key evolutionary processes, including speciation, the maintenance of genetic variation, and the responses of species to anthropogenic environmental change (55,56).

Given the complexity of urban adaptation and its relevance to conservation under global change, identifying adaptive markers and fitness-related genes in urban populations can inform effective conservation strategies. Such knowledge would enable the promotion of beneficial traits through assisted reproduction and planned hybridization, thereby maximizing a species' capacity to adapt to environmental changes (57–60). However, the polygenic nature of most adaptive traits makes detecting specific adaptive loci technically challenging (61,62).

1.3 Hypotheses and Study Objectives

My Ph.D. project has been funded by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU; Award Number: Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP H43C22000530001, Project title "National Biodiversity Future Center – NBFC".

An urban biodiversity research program that integrates approaches from evolutionary biology, genomics, and environmental sciences is essential to understand how species evolve and adapt in response to urbanization, especially for sessile organisms such as plants, which must cope with rapid and heterogeneous changes in their environment in order to persist and reproduce (63).

The general objectives of the Urban Biodiversity – Plants PNRR research line aim to investigate the reproductive success of plants in urban environments and to identify the ecological and biological causes underlying these patterns. At the same time, the project aims to clarify the factors that drive plant colonization in cities, with particular attention to the mechanisms that allow certain species to establish and reproduce successfully in highly anthropized contexts. A central question is whether these successful urban populations owe their persistence to genetic adaptation or whether their

performance is instead the result of phenotypic plasticity, allowing them to flexibly respond to urban pressures. Finally, the project seeks to understand whether adaptation to the urban environment exhibits a convergent pattern across different cities, suggesting that similar selective pressures may lead to repeatable evolutionary outcomes.

Within this broader framework, my Ph.D. research focuses specifically on the genetic of plant adaptation to urbanization. *Raphanus raphanistrum* is chosen as a biological model due to its agronomic value and its capability to grow in a wide range of environments, having colonized sterile, saline, and arid habitats, and developing a rich pool of resistance genes. Studying changes in gene expression profiles in this species, influenced by different urban and non-urban contexts, can provide insights into how plants adapt to diverse environments. At the same time, the project involves characterizing potential single-nucleotide polymorphisms (SNPs) associated with specific environments, aiming to determine whether it is possible to identify an "urban genotype" or if adaptation is primarily driven by gene expression regulation and phenotypic plasticity.

2. Study System and Background

2.1 Biology of *Raphanus raphanistrum* and suitability as a model for urban adaptation

Raphanus raphanistrum L. is a common broadleaf diploid weed (2n=18) that belongs to the Brassicaceae. This family has a worldwide distribution and approximately 3,500 species, most of which are herbaceous and primarily annual (64). The genus name *Raphanus* is possibly derived from the Greek word *ῥᾴ* (*ra*), meaning quickly, and *φαίνομαι* (*phainomai*) meaning to appear, referring to its rapid germination and seedling growth (65).

The genus *Raphanus* contains only two species and several sub-species native to western and central Europe, extending through the Mediterranean to central Asia. The wild radish, *R. raphanistrum*, has two recognized subspecies: *R. raphanistrum landra* and *R. raphanistrum raphanistrum*. The *landra* subspecies is native to the Mediterranean region, while *raphanistrum* is found worldwide (66).

R. raphanistrum is an annual herb, sometimes behaving as a biennial in cooler regions, that can reach heights of up to 1 m, reproducing solely by seed (65). The leaves initially form a low-growing cluster known as a basal rosette (67). The lower leaves are large (5 to 20 cm long) and elongated with deeply divided segments. The plant has a short hypocotyl and may or may not have an epicotyl. As it matures, the plant forms erect, branched stems covered in prickly hairs, with smaller, narrower, and

often undivided leaves higher on the stem. The flowers grow in clusters (racemes) at the ends of branches, featuring four regular petals (12–20 mm) alternating with four shorter sepals. Their colors range from white to purple, pink, or pale yellow, often with distinct veins. The fruit is a long, bead-like pod that constricts between seeds and ends in a slender, seedless point (Fig. 2).

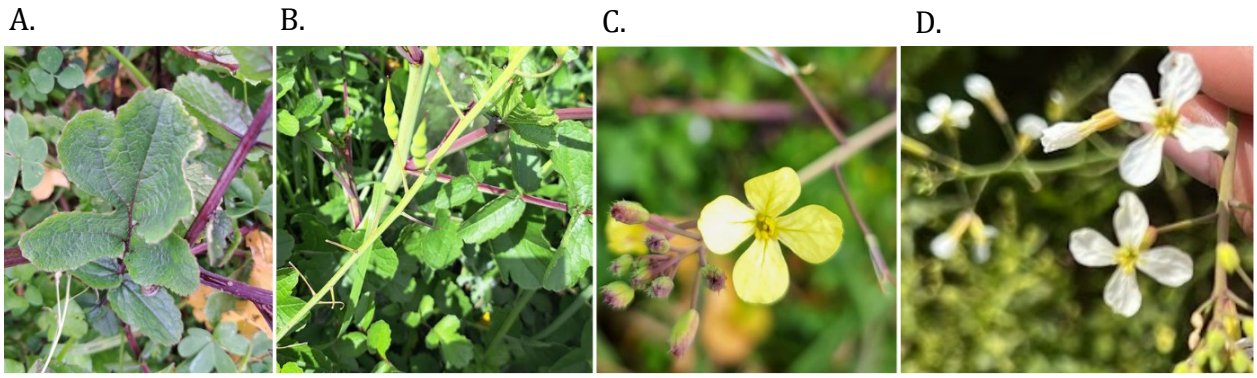

Figure 2. *Raphanus raphanistrum* morphological traits. (A) Basal leaf rosette; (B) silique; (C-D) floral variation showing yellow and white flower morphs.

When ripe, the yellowish-brown pod splits into segments, each containing a small (2–3 mm) round to oval seed, which has a yellowish to reddish-brown color and a net-like pattern (65). It is insect-pollinated and self-incompatible (67). Consequently, dispersal distances depend on local pollinator activity and are usually limited to a few hundred meters (68). Seed dispersal mainly occurs naturally when the pods dehisce, releasing segments up to 4.5 m from the parent plant (69).

R. raphanistrum is a pioneer species that competes strongly with cultivated plants, particularly winter cereals like wheat. It is one of the most problematic broad-leaved weeds due to its allelopathic properties, which can suppress the growth and germination of other plants (70,71). This effect is linked to its production of glucosinolates (GSLs), which, when hydrolyzed to isothiocyanates (ITCs), inhibit seed germination by disrupting key enzymatic processes (72–74). The seeds of this species are considered the most toxic part (75). The toxic compound is allyl isothiocyanate, derived from the glucosinolate sinigrin (76). Additionally, wild radish can act as an alternate host for several plant pests and diseases, including thrips (*Thrips* spp.), flea beetles (*Phyllotreta* spp.), and club root of brassicas (*Plasmodiophora brassicae*). Aqueous extracts of *R. raphanistrum* have demonstrated phytotoxic effects on a wide range of plant species (77,78). Malik (74) reported that extracts of *R. raphanistrum* inhibited the growth of various crops, including corn, wheat, and cotton, as well as specific weeds such as *Cyperus esculentus*, *Ipomoea lacunosa* L., *Sida spinosa* L., and *Senna obtusifolia* L. (79). Overall, the allelopathic potential of *R. raphanistrum*, especially through its aqueous extracts,

highlights its possible applications in integrated weed management programs for wheat cultivation (78).

The species is adapted to a wide range of habitats because of its robust competitive ability, fast growth rate, and rapid establishment (80). It has a high potential to be accidentally introduced to new areas, often acting as a contaminant of grain seed and other agricultural produce from various sources (81). Once established, it readily invades disturbed habitats such as agricultural land, abandoned fields, and roadsides (82).

The genus *Raphanus* is composed of just two primary species, alongside various subspecies, all sharing a karyotype characterized by nine chromosomes. This genetic uniformity renders the genus particularly valuable for conducting genomic comparisons. It provides an exceptional model for exploring genome evolution, especially regarding post-polyploid diploidization, a fascinating mechanism frequently observed in the plant kingdom (83).

As environmental changes accelerate, wild relatives of cultivated species are recognized as valuable genetic resources for crop improvement, providing traits to cope with new stresses. The cultivated radish *R. sativus*, or common radish, includes various forms like cherry belle, black radish, rat's tail radish, oilseed radish, and daikon. European varieties such as cherry belle and black radish originate from wild populations in Europe, while daikon has diversified across Asia, exhibiting significant morphological and phenotypic variations due to regional agricultural practices (84).

Recent studies have revealed that cultivated radish varieties have experienced several distinct domestication events (85,86), highlighting the dynamic interactions between human agriculture and natural ecosystems. Notably, species and varieties within the *Raphanus* genus are capable of inter-crossing, making this group a valuable system for investigating patterns of gene flow and co-evolution among crops, their wild progenitors, and associated weedy relatives (87). Such genetic exchanges may contribute to improving crop adaptability and promoting more sustainable agricultural systems in the context of global environmental change. For example, the California wild radish, a hybrid between wild and cultivated radish (*R. raphanistrum* × *R. sativus*), has largely replaced both parental species in the wild (88). This hybrid's unique combination of traits likely contributes to its ecological success in western North America, making it an important system for studying the evolutionary and functional genomics of invasive and hybrid species.

One notable trait is the variation in petal color, determined by the presence or absence of anthocyanin pigments derived from the phenylpropanoid pathway. Interestingly, pollinators prefer yellow flowers, while herbivores favor traits linked to chemical defenses like glucosinolate

production (89–91). These opposing selective forces, along with the genetic correlation between flower color and defense chemistry, help maintain stable polymorphisms in natural populations over time (92–94).

Collectively, these findings highlight the dynamic and reticulate nature of the *Raphanus* genome and its value as a model system for exploring the evolutionary impact of hybridization and human-mediated selection in changing environments.

2.2 Urban–Rural Gradient as an Environmental Variable

The distinction between urban and natural landscapes is often less clear than it appears. Urban growth tends to expand beyond city boundaries, blending built environments with remnants of natural and semi-natural habitats (95,96). As this process unfolds, the landscape becomes a mosaic of biological and physical elements shaped by human activity (97). During this process, the extensive agricultural and semi-natural areas have largely given way to smaller, fragmented green spaces such as parks, gardens, and residual natural patches. These fragments often remain ecologically isolated and subject to distinct urban pressures that influence species composition and ecological functions (98). To counterbalance this fragmentation, ecological corridors play a vital role in reconnecting habitats and promoting the movement of flora and fauna across the urban-rural continuum (99). Despite the increasing focus on urbanization as an ecological driver, there is no single, universally accepted definition of "urban". Various definitions emphasize different aspects of urban systems, each with its own advantages and disadvantages depending on the context in which it is applied.

Urban gradients are commonly described by the proportion of vegetation, buildings, and roads (100), which in turn correlate with changes in species composition, abundance, and richness along urbanization gradients (101–103).

Among these indicators, impervious surface (roads, buildings and parking areas) has emerged as one of the most widely used quantitative proxies for urban development intensity (104). Advances in remote sensing and image analysis now enable the efficient and consistent measurement of impervious cover across various scales (105,106), making it a robust tool for ecological gradient studies. Consistent with this approach, recent work by Liu et al. (107) showed that vegetation phenology shifts measurably along dynamic urban–rural gradients defined by impervious surface coverage, illustrating how ecological responses can be tightly linked to spatially quantifiable urbanization metrics.

Finally, given the lack of a single, universal definition of what constitutes “urban” (108), using measurable and reproducible indicators, such as impervious surface, enables researchers to define the

urbanization gradient in objective, comparable terms and provides a consistent framework for cross-site comparisons essential for global urban ecology.

3. Materials and Methods

3.1 Materials

During 2023 and 2024, leaf samples of *R. raphanistrum* were collected from 30 populations in the Campania and Abruzzo regions of Italy (Table 1). A subset of these samples was preserved in RNAlater (Ambion) and stored at -80 °C for subsequent RNA and/or DNA extraction. These samples were used to generate both RNA-seq data for gene-expression and population transcriptomics analyses and DNA-seq data for population-genomic analyses.

In parallel, individuals from several populations were grown in a common garden experiment, and some of these plants were also included in the DNA-seq dataset. Samples from a Swiss population were incorporated into selected analyses to broaden the geographic scope.

Quantitative phenotypic traits were measured in both common garden and wild populations. These phenotypic data, combined with genomic and transcriptomic datasets, were then used to explore genotype–phenotype associations.

Table 1. Complete dataset of populations used in this study. W, wild population; CG, common garden population.

POP_ID	LOCATION	ANALYSIS	TYPE
AGTE	Agnano Terme - Naples (NA)	Phenotype	W
ANNU	Annunziata - Massa Lubrense (NA)	RNA-seq; DNA-seq; Phenotype	W; CG
AROL	Arola - Vico Equense (NA)	Phenotype	W
ASTR	Astroni - Naples	RNA-seq; DNA-seq; Phenotype	W
CAPA	Via Vicinale Selva - Scisciano (NA)	Phenotype	W
CARR	Via Caravaggio - Naples	DNA-seq; Phenotype	W; CG
CATT	Via Cattolica - Naples	RNA-seq; DNA-seq; Phenotype	W; CG
CH	Zürich (Swiss)	DNA-seq	W
FAIT	Monte Faito - Massa Lubrense (NA)	DNA-seq; Phenotype	W; CG
FCAM	Fuorigrotta - Naples	Phenotype	W
GOCA	Gocce di Capri - Massa Lubrense (NA)	RNA-seq; DNA-seq; Phenotype	W; CG

MECA	Stazione metro Monte Sant'Angelo- Naples	Phenotype	W
MSA	Monte Sant'Angelo - Naples	RNA-seq; DNA-seq; Phenotype	WILD
MSCO	Monte San Costanzo - Massa Lubrense (NA)	RNA-seq; DNA-seq; Phenotype	W; CG
PALE	Palena (CH)	Phenotype	W; CG
POLV	Polvica (NA)	RNA-seq; DNA-seq; Phenotype	W
QUAR	Quarto (NA)	RNA-seq; Phenotype	W
SANG	Castel di Sangro (AQ)	Phenotype	W
SCHI	Schiazzano - Massa Lubrense (NA)	RNA-seq; DNA-seq; Phenotype	W
SCI	Scisciano (NA)	Phenotype	W
SGIO	San Giorgio a Cremano (NA)	RNA-seq; DNA-seq; Phenotype	W; CG
SGRM	San Gregorio Magno (SA)	Phenotype	W
SOCC	Soccavo - Naples	RNA-seq; DNA-seq; Phenotype	W; CG
TECH	Piazzale Tecchio - Naples	DNA-seq; Phenotype	W; CG
TERR	Via Terracina - Naples	RNA-seq; Phenotype	W
TREC	Trecase (NA)	RNA-seq; DNA-seq; Phenotype	W
TRIV	Trivio - Nola (NA)	RNA-seq; DNA-seq; Phenotype	W
VARG	Via Argine - Naples	RNA-seq; DNA-seq; Phenotype	W; CG
VESU	Vesuvio (NA)	Phenotype	W; CG
VMAR	Via Marina - Naples	Phenotype	W

3.2 Sampling

The classification of sampling sites was based on the percentage of impervious surface (I.S.), referred to the Copernicus Imperviousness Density 2021 dataset (10 m resolution) (109). The spatial visualization of land cover and the representation of sampling points were carried out using QGIS (v. 3.44.5) (110), by importing the coordinates of each population and overlaying them onto the imperviousness raster (Fig. 3).

The classification of sampling sites was conducted in collaboration with Dr. Lucrezia Laccetti, from the research group of Prof. Giovanni Scopece. I gratefully acknowledge them for their support in GIS analyses. The areas were classified as follows:

1. Urban areas: I.S. $\geq$ 75%
2. Peri-urban areas: 15% < I.S. < 75%
3. Natural areas: I.S. $\leq$ 15%

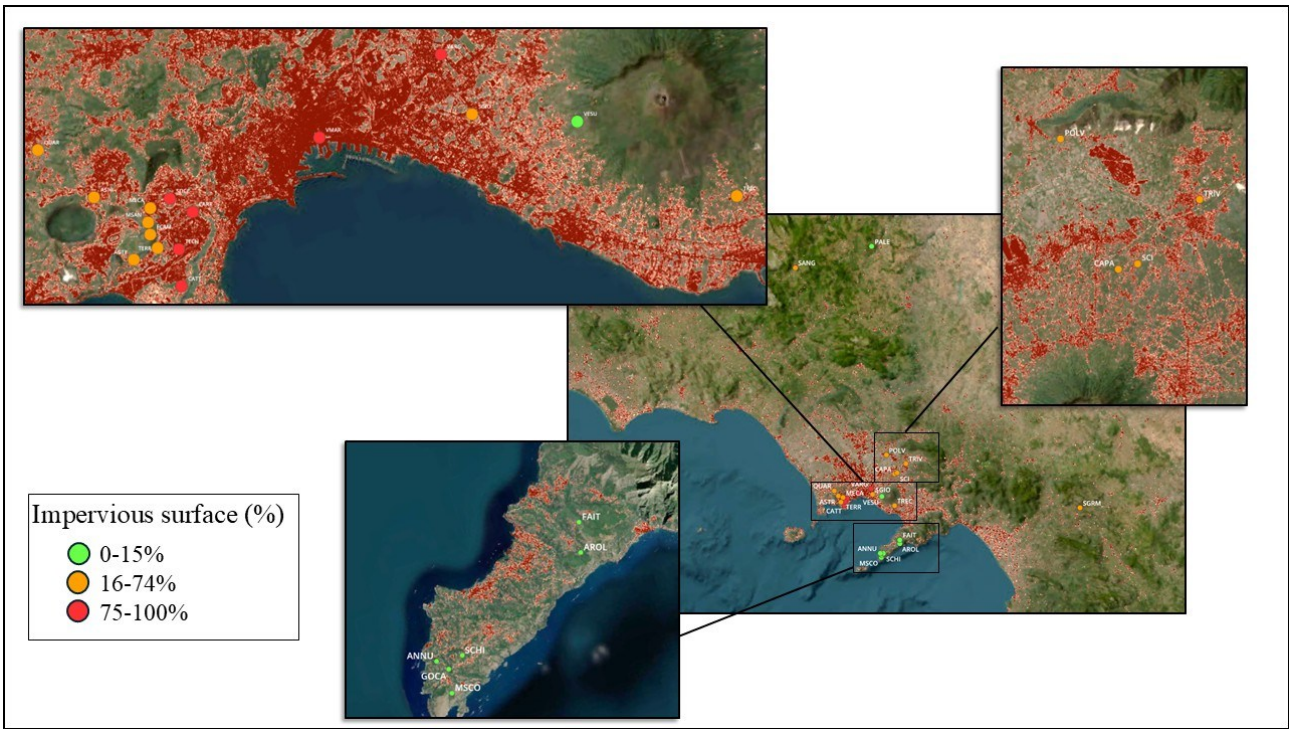

Figure 3. The map displays the study area along with the locations of the sampling sites. The sites are color-coded according to the percentage of impervious surfaces (I.S.): natural areas are represented in green, peri-urban areas in orange, and urban areas in red. This classification is based on the Copernicus Imperviousness Density 2021 dataset.

In addition, soil type was recorded during field sampling and classified into four geological categories: limestone, tuff, lava, and pyroclast. Although soil type was not treated as a primary ecological driver, it was included to assess potential environmental covariation with urbanization and to serve as a control factor in subsequent analyses.

3.3 Common garden experiment

The common garden experiment took place in April 2024, involving the selection of 59 maternal lines from both natural and urban populations. The experimental setup was supervised and coordinated by Dr. Yuan-Fu Chan, who ensured consistency across all procedures, and I thank him for kindly providing the plant material derived from the common garden experiment.

A total of 890 seeds were sown, with 20 seeds per maternal line when available. The seeds were stratified in water at 4°C for three days and then germinated in 4 cm wells filled with high-quality, non-composted soil (TerCompost; pH 6; 90% porosity) under a natural day/night cycle.

The seedlings were watered twice a week and monitored daily. Herbivores were manually removed as soon as they were spotted. From each maternal line, eight seedlings were randomly selected for transplantation; if germination was low, all available seedlings were used. By May 2024, 388 seedlings had been transplanted into 14 cm pots (1.1 L), with a nutrient-rich mixture of composted soil (TerCompost; pH 7.5; 85% porosity) and river sand. The plants were arranged in a randomized design across four benches in the sunlit experimental area of the Department of Biology at the University of Naples Federico II.

After transplantation, all seedlings received immediate watering, and uniform hydration was maintained by flooding each bench with the same amount of water. Pest management involved manual removal of caterpillars and, when necessary, the application of a commercial pesticide (Zepi Garden) for aphids. Natural pollinators had free access throughout the study.

3.4 Workflow Description

The workflow of my PhD project (Fig. 4) can be summarized into four main steps:

(1) RNA sequencing and gene expression analysis;

(2) DNA sequencing and population genomics analysis;

(3) Genome-wide association study;

(4) Population transcriptomics analysis.

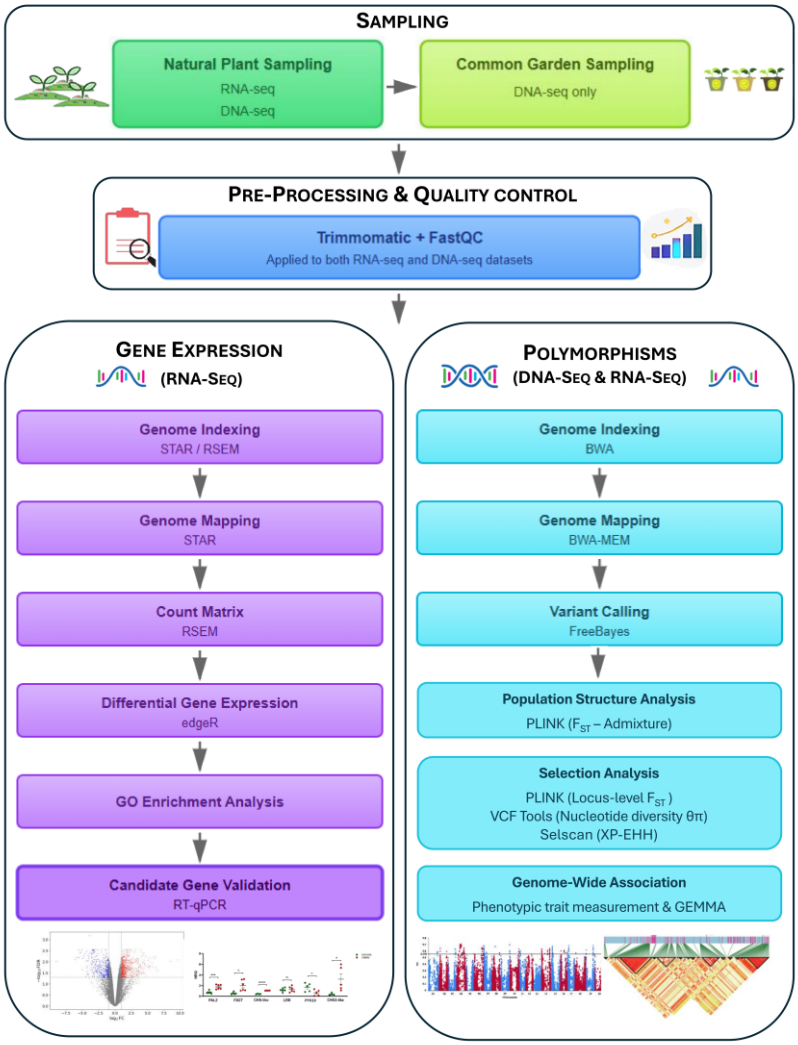

Figure 4. Schematic representation of the workflow and bioinformatic pipelines used for integrated RNA-seq and DNA-seq analyses, including sampling, preprocessing and quality control, gene expression profiling, variant calling, population structure, selection analyses, and genome-wide association studies.

3.5 RNA-seq and Gene Expression Analysis

The RNA-seq analysis was performed on 70 individuals from 15 wild populations. This included 38 individuals from peri-urban areas, 12 from natural areas, and 20 from urban areas defined by impervious surfaces. Additionally, the individuals were categorized based on soil type: 29 were from tuff soil, 22 from limestone, 16 from lava, and 3 from pyroclast (Fig. 5).

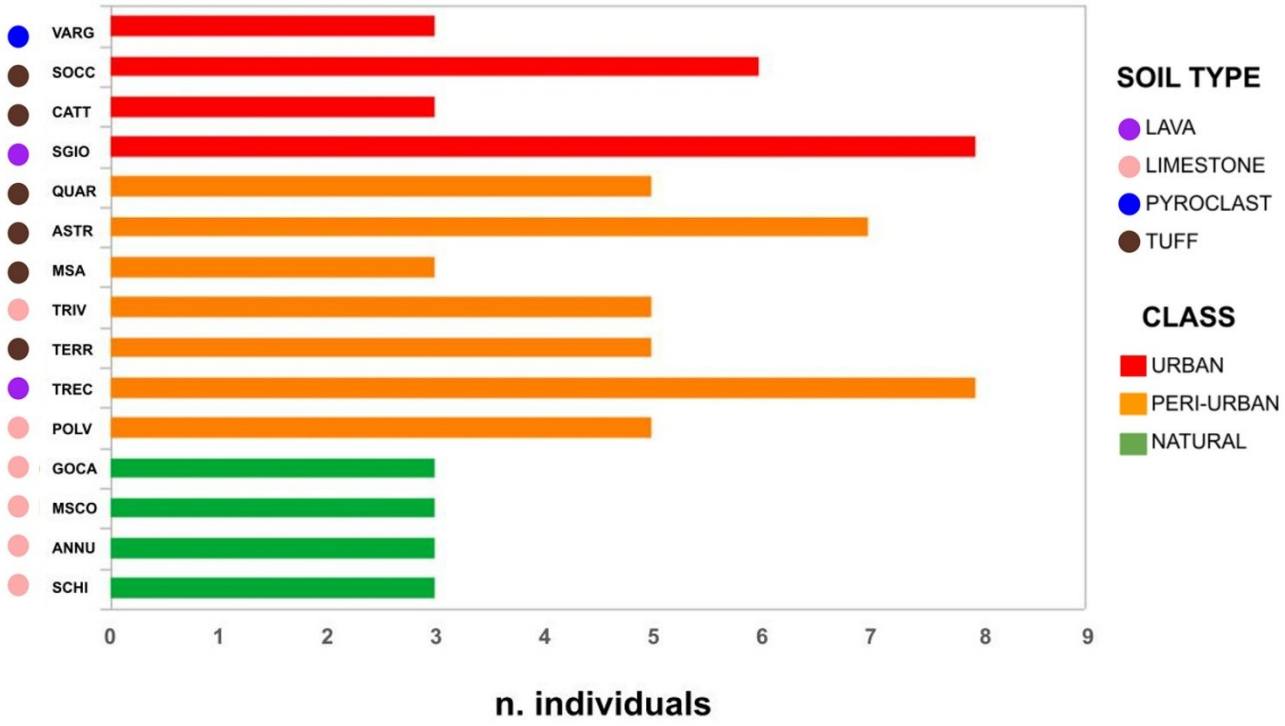

Figure 5. Summary of population classification and corresponding number of individuals included in the RNA-seq analysis.

From the same RNA samples, both transcriptomic analysis and variant calling were performed, allowing the investigation of gene expression profiles as well as genetic variation at the RNA level.

3.5.1. RNA Extraction and Library Preparation

Total RNA was extracted using the PureLink™ RNA Mini Kit (Ambion), following the manufacturer's protocol. The quality and concentration of the extracted RNA were assessed using the 2100 Bioanalyzer System (Agilent Technologies) and the Nanodrop 2000 (Thermo Scientific). Library preparation and sequencing were performed by Novogene (UK) using the Illumina NovaSeq X Plus (PE150, paired-end 150 bases), obtaining 9 Gb raw data per sample.

3.5.2. Indexing of the Reference Genome

The reference genome of *R. raphanistrum* (accession number GCA_019706035.1) was downloaded from the NCBI database, including both genome sequence (FASTA) and annotation (GFF) file. The annotation was converted to GTF format using *gffread* (111), as required by downstream tools.

Two reference datasets were then prepared: first, an RSEM-compatible transcriptome reference was generated using the *rsem-prepare-reference* command, integrating genome sequence and annotation into a structured dataset optimized for transcript quantification. Second, the reference genome was indexed with *STAR* (112) to enable spliced alignment of RNA-seq reads. The *STAR* indexing integrated both genome sequence and GTF annotation, with the *sjdbOverhang* parameter set to 100 to effectively model exon-exon junctions and improve spliced alignment accuracy.

3.5.3. RNA-Seq Data Processing

- **Pre-processing**

Raw paired-end reads were subjected to a quality control and trimming procedure before alignment. The quality assessment using *FastQC* (113) , provided a preliminary evaluation of sample quality, GC content, duplication levels, and potential adapter contamination. Reads were then processed with *Trimmomatic* (114) in paired-end mode to remove adapters and low-quality bases. Only paired reads passing quality filters were retained (*_P.fq.gz). Unpaired reads were generated but excluded from downstream analyses. After trimming, a second *FastQC* run was performed to confirm the improvement in read quality.

This pre-processing ensured high-quality input for alignment and quantification in subsequent steps.

- **Alignment and Gene Expression Quantification using RSEM**

Read alignment and expression quantification were then performed using *RSEM* with the *--star* option (115), which utilizes the *STAR* genome index prepared in the previous step for spliced alignment, while *RSEM* computes gene- and transcript-level abundance. *RSEM* provides a reliable and accurate statistical framework to estimate transcript abundance at both gene and isoform levels, even in the presence of overlapping transcripts or alternative splicing variants, making it well-suited for complex transcriptomes. The output files generated by *RSEM* include both raw expected counts and normalized expression values, such as TPM (Transcripts Per Million) and FPKM (Fragments Per Kilobase of transcript per Million mapped reads). TPM values are particularly suitable for cross-sample comparisons, as they are normalized for both gene length and sequencing depth.

- **Generation of the Gene and Isoform Expression Matrix**

Following transcript quantification with *RSEM*, each sample-specific *RSEM* output file (containing expected counts and normalized values) was used as input to generate a count matrix

containing gene-level expression values across all individuals. This was achieved using the *abundance_estimates_to_matrix.pl* script from the *Trinity* toolkit (116,117), which consolidates all expression estimates into a structured table suitable for downstream statistical analysis. The resulting matrix was subsequently used for differential expression analysis (DGE) and for all downstream transcriptomic analyses.

3.5.4. PCA Across Environmental Groups

To examine global transcriptional patterns and sample clustering by habitat, a Principal Component Analysis (PCA) was conducted on variance-stabilized expression data. Gene-level raw counts from *RSEM* were processed in R using *DESeq2* (118), with samples categorized by habitat and soil type. After normalization and variance-stabilizing transformation (VST), the 500 most variable genes were selected.

PCA was performed with the *prcomp()* function on the transposed expression matrix, visualizing the first two principal components (PC1 and PC2) using *ggplot2*, with samples colored by class and confidence ellipses added. To identify key genes influencing variance in the components, PCA loadings were examined, prioritizing the top 20 genes for each component based on their absolute loading values. Genes with high loadings on PC1 and PC2 revealed the main drivers of variation and clustering patterns in the PCA plots.

3.5.5. Differential Gene Expression (DGE) Analysis

To detect genes showing significant expression differences between environmental conditions, DGE analysis was performed using *edgeR* (119), implemented via the *Trinity* toolkit (120). Samples were classified according to impervious surface categories and soil type. *edgeR* models count-based data using a negative binomial distribution and estimates differential expression through several key statistics, including:

- log_2 *Fold Change* (logFC): measures the magnitude of expression change between conditions. Positive values indicate up-regulation in the first condition of the contrast, while negative values indicate down-regulation.
- *log Counts per Million* (logCPM): represents the normalized expression level of each gene, allowing comparison across samples with different sequencing depth.
- *P-value*: indicates the statistical significance of the observed expression change before multiple testing correction.

- *False Discovery Rate* (FDR): adjusted P-value used to control for false positives in multiple testing.

All downstream analyses on differentially expressed genes (DEGs) were performed for two environmental factors:

- urbanization gradients (urban, peri-urban, natural),
- soil categories (tuff, limestone, lava, pyroclast).

- **DEG Comparison Across Environmental Factors**

To compare sets of differentially expressed genes (DEGs) across environmental contrasts, Venn diagrams and Volcano plots were generated using filtered DEG lists (FDR < 0.05 and $|\log_2 FC| > 1$).

To further investigate the magnitude and statistical significance of expression changes for each gene, Volcano plots were produced using *ggplot2 (121)*. In these plots, each gene is positioned according to its $\log_2$ Fold Change (x-axis) and the $-\log_{10}$(FDR) (y-axis). This representation highlights the most strongly regulated genes and clearly distinguishes those with significant expression changes (FDR < 0.05 and $|\log_2 FC| > 1$) from genes not affected by the environmental contrast.

By integrating both types of visualization, it was possible to obtain an immediate, comprehensive view of how gene expression varies across environmental contexts. Venn diagrams highlight the relationships and overlaps between DEG sets, while volcano plots reveal the strength and direction of expression changes for each gene.

3.5.6. Functional Annotation and Enrichment Analysis

To investigate the functional implications of changes in gene expression, a Gene Ontology (GO) enrichment analysis was conducted using differentially expressed genes (DEGs) identified by *edgeR*. GO annotations were retrieved from the NGDC-CNCB database (122), specifically using the gene function file for *R. raphanistrum* sub. *raphanistrum*. These annotations were integrated with the DEG tables through a custom Python pipeline, which allowed us to associate each gene with its corresponding GO terms.

For each contrast, DEGs were categorized into up-regulated and down-regulated sets and tested for GO over-representation using Fisher's Exact Test. The results were summarized using dot plots.

This approach provided a functional interpretation of the transcriptomic responses to urbanization and soil type.

3.5.7. qPCR validation

- **Selection of Candidate Genes and qPCR Validation**

To validate the transcriptomic patterns observed in the RNA-seq dataset, a subset of genes was selected for *in vivo* validation through quantitative PCR (qPCR).

Candidate genes were identified using a combined approach based on:

- Differential expression analysis, retaining genes with FDR < 0.05 and $|\log_2 FC| > 1$;
- Gene Ontology enrichment, selecting genes associated with significantly enriched biological processes;
- Multivariate analysis, by identifying genes with high absolute loadings on PC1 and/or PC2 in the PCA, i.e. genes contributing most to the separation among samples.

Genes identified by both GO-based and PCA-based criteria were considered strong candidates for urbanization adaptation and selected for downstream qPCR validation. For each selected gene, *de novo* primers were designed using *Primer3* (123) (Table 2), ensuring optimal melting temperatures, amplicon sizes, and the absence of secondary structures. Primer specificity was verified by conventional PCR with *DreamTaq* (Thermo Scientific™) before qPCR.

Seven wild populations were selected to represent the extremes of the urban gradient. For each population, total RNA was extracted and cDNA was synthesised from pooled samples, each pool consisting of five individuals, in order to reduce inter-individual variability and capture population-level expression profiles. Reverse transcription was performed using 1000 ng of RNA with the LunaScript® RT SuperMix kit (New England Biolabs).

qPCR assays were performed using PowerUp™ SYBR™ Green Master Mix (Applied Biosystems™), following manufacturer's instructions. Reactions were run in technical duplicates, and 18s rRNA was used as a reference gene. Normalised Relative Quantity (NRQ ± SEM) was calculated for each gene relative to the internal control gene (124).

To assess differences in expression levels between urban and natural populations, multiple t-test was performed on NRQ values.

Table 2. List of the primer sequences used for quantitative real time PCR.

Gene_ID	Forward (5'-3')	Reverse(5'-3')
PAL2	GAAGTGATCCGTTACGCCAC	CCGAGAACTGAGCGAACATG

F3GT	AGTCGGAGTTGAAACGTTGC	TCGATTCCAACCCTTCTGCT
CHS-like	AGTGAACACATGACCGACCT	ACCACCACGATGTCTTGTCT
CHS3-like	CGCATGTGTGACAAGTCTAT	ATCTTTGACTTGGGCTGACC
LRR	GACTCGGATGCGCGATATTC	CCGGTGAGTTTGTTGTTGCT
PYK10	ACAAGGAGGAAATGGAGGCA	CAAACGGATCAGCTGGGATG

3.6 DNA-seq and Population genomic analyses

The DNA sequencing analysis was conducted on 88 individuals, which included 29 wild samples and 59 individuals from a common garden experiment, all belonging to 17 different populations. Among the wild samples, three samples from Zurich, Switzerland, were included as geographic controls to assess the extent of genetic differentiation between geographically distant populations compared to regional variation within the Italian sampling area. In total, the dataset comprised 43 individuals from natural areas, 34 from urban areas, and 8 from peri-urban areas, as classified by impervious surface percentage. The samples were collected in different soil types: 47 from limestone, 27 from tuff, 5 from lava, and 6 from pyroclast (Fig. 6).

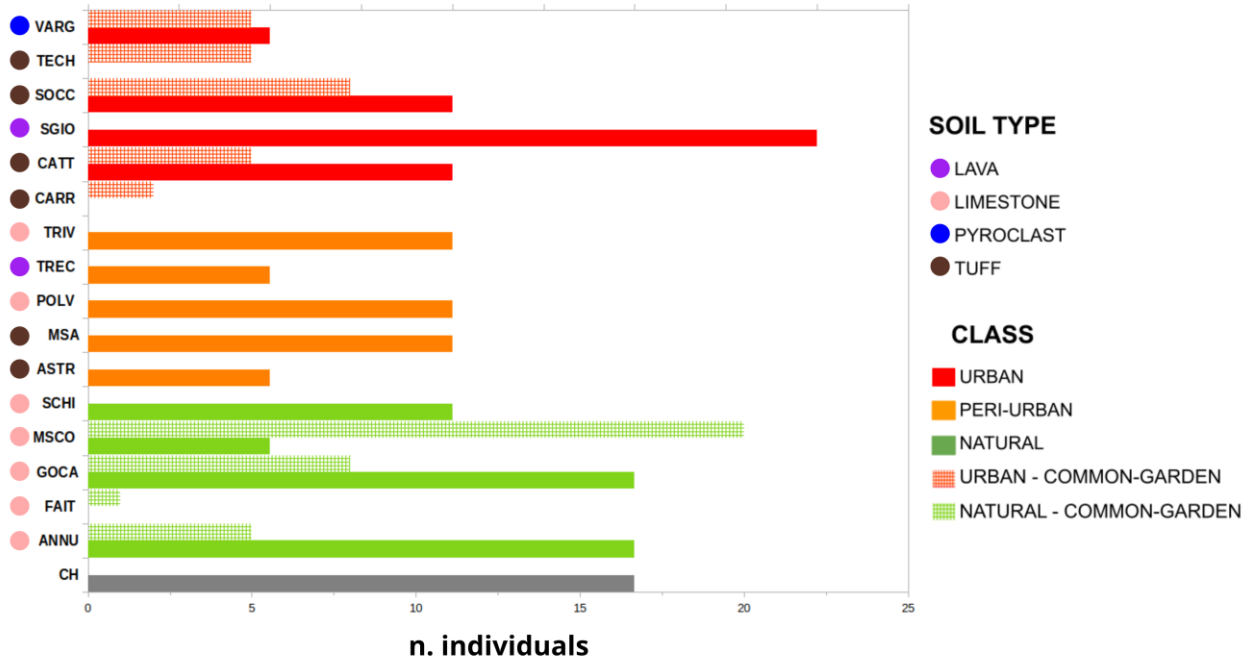

Figure 6. Summary of population classification and corresponding number of individuals included in the DNA-seq analysis.

3.6.1. DNA Extraction and Sequencing

Genomic DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN). The quality and concentration of the extracted DNA were assessed using the Qubit and the Nanodrop 2000 (Thermo Scientific). Whole-genome sequencing (PCR-free) was performed by Macrogen (South Korea) using TruSeq and Watchmaker library preparation methods on an Illumina sequencing platform, generating paired-end short reads (150 bp). In addition to the newly generated data, genomic sequences previously produced by the research group of Prof. Salvatore Cozzolino were included in the analyses. The author gratefully acknowledges Prof. Cozzolino for making these data available and for his valuable scientific contribution.

All steps involving quality assessment, preprocessing, and downstream analyses of genomic reads were carried out during the research stay abroad at the laboratory of Prof. Vincenza Colonna at the University of Tennessee Health Science Center (UTHSC) in Memphis, USA. The author gratefully acknowledges Prof. Colonna for her supervision, support, and valuable scientific guidance during this period.

3.6.2. Preparation and Indexing of the Reference Genome

For genomic analysis aimed at variant calling, the reference genome in FASTA format was downloaded from NCBI as described in the first section of paragraph 3.5.2.

Before alignment, the genome was indexed using *BWA* (Burrows-Wheeler Aligner) (125), which builds an FM-index structure from the FASTA sequence. This step enables efficient alignment using the BWA MEM algorithm.

3.6.3. DNA-Seq Data Processing

- **Pre-processing**

Raw genomic paired-end reads were subjected to the same quality control and trimming procedure described in paragraph 3.5.3. As in the RNA-seq workflow, only high-quality paired reads were retained for downstream analysis, while unpaired reads were discarded.

The pre-processing phase was explicitly designed to ensure high read quality, which was essential for minimizing false-positive SNP and indel detection during subsequent genomic analyses.

- **Reads mapping**

Cleaned paired-end reads were aligned to the indexed reference genome using the BWA-MEM algorithm, a widely adopted tool for short-read DNA sequencing data, particularly from Illumina

platforms (125). BWA-MEM is based on the algorithm Burrows–Wheeler transform (BWT) and efficiently performs similarity-based alignment while maintaining high accuracy, even in regions of moderate genomic complexity.

A read group was added to the alignment header to assign sample identity and sequencing metadata. This information is required by downstream tools for duplicate marking, quality recalibration, and accurate variant calling, particularly in multi-sample workflows.

The alignment generated a raw BAM file containing genomic mapping information for each sample, serving as the basis for post-alignment processing and variant detection.

- **Post-alignment Processing**

Following alignment, the mapping results were statistically evaluated using *SAMtools stats* (126), which provides a comprehensive summary of alignment quality, including read depth distribution, error rates, coverage metrics, and mapping efficiency. This quality assessment enables the identification of potential issues in the alignment and ensures the reliability of the data prior to variant calling.

The aligned BAM files were subsequently sorted and indexed using *Sambamba* (127), a high-performance tool specifically optimized for parallel BAM processing. Sorting by genomic coordinates is essential for downstream analyses, while indexing enables rapid access to specific genomic regions. Notably, S*ambamba sort* automatically generates the index file during sorting, streamlining the workflow and reducing computational time compared to traditional tools.

The result of this step was a set of sorted and indexed BAM files, ready for duplicate marking and subsequent variant calling.

3.6.4. Variant calling analysis

- **Variant Detection**

Variant detection was performed using *FreeBayes* v1.3.2 (128), a Bayesian haplotype-based variant caller specifically designed for population-level analyses. Unlike traditional alignment-dependent approaches, *FreeBayes* infers variants by directly analyzing the literal sequences of reads aligned to a genomic region, rather than relying exclusively on their precise alignment coordinates. This approach mitigates one of the significant limitations of conventional variant detection: identical sequences may align equally well to multiple genomic positions.

FreeBayes operates on short-read alignments (BAM files with Phred+33 encoded quality scores) across any number of individuals within a population, together with a reference genome in FASTA

format. It determines the most likely combination of genotypes at each genomic position by evaluating joint evidence across samples, making it particularly suitable for population-scale analyses. The tool reports putatively polymorphic sites in VCF format, including SNPs, indels, multi-nucleotide polymorphisms (MNPs), and complex composite events that are shorter than the read length.

For each sample, the sorted and indexed BAM file was used as input together with the reference genome. Variant calling was performed on selected genomic regions to optimize computational efficiency. The resulting VCF files contain all detected variants, along with their genomic positions, allele frequencies, genotype likelihoods, and associated quality metrics.

VCF is a text file format that contains genotype information for each sample at each position. It includes all detected variants, including their genomic positions, reference and alternative alleles, genotype calls for each sample, allele frequencies, genotype likelihoods, and various quality metrics.

The VCF structure includes both a header and a variant table: the header stores metadata (e.g., reference genome, software used, filter descriptions), while the tabular section lists one variant per row. The header line names the eight fixed, mandatory columns. These columns are as follows:

- #CHROM - chromosome identifier
- POS - genomic position of the variant
- ID - known identifiers from databases (e.g., dbSNP)
- REF / ALT - reference and alternative alleles
- QUAL - Phred-scaled quality score
- FILTER - filter status (e.g., PASS or filtered)
- INFO - additional annotations (e.g., AF, DP, annotations)

This standardized format facilitates interoperability across bioinformatic tools and enables downstream filtering, annotation, and population-level analyses.

- **VCF Compression and Indexing**

After variant calling, each VCF file VCF files were compressed using *bgzip*, a block-based compression method optimized for large genomic datasets and indexed with *tabix*, both implemented in the *HTSlib* toolkit (129). The indexing process generates a .tbi file, enabling fast retrieval of variants from specific genomic regions without decompressing the entire file. This step is essential for compatibility with many downstream tools, including variant-filtering software, annotation pipelines, and population-genetics analyses.

- **VCF Normalization, Filtering and Merging**

All steps of normalization, filtering, merging, and concatenation were performed using the *BCFtools* package (126,130). The VCFs were normalized to ensure consistent representation of indels and to eliminate ambiguous variant annotations. Normalization was performed relative to the reference genome, and multi-allelic sites were decomposed into bi-allelic records to guarantee compatibility with downstream tools. Each normalized file was indexed before the quality-based filtering step, which retained only high-confidence variants. Specifically, a minimum QUAL threshold of 20 was used to remove low-quality variant calls, thereby reducing the risk of false positives and improving the reliability of subsequent analyses. After filtering, the VCF files were indexed again.

After normalization and filtering, VCF files from individual chromosomes were merged into a single multi-sample dataset.

For each of the nine chromosomes, all sample-specific VCF files were gathered and merged into chromosome-wise multi-sample VCF files. This allowed the inclusion of all individuals while preserving variant consistency and genotype information across samples. Then, the chromosome-wise VCF files were concatenated into a single genome-wide multi-sample VCF file. This step unified all genomic regions into a single dataset, representing the final variant matrix across all individuals and chromosomes.

3.6.5. Population Structure Analysis

Genotype data were processed using *PLINK 2.0* (131), a standard tool for large-scale genomic analyses. The merged VCF file was converted to the *PLINK* binary format (.bed/.bim/.fam), which is required as input for population structure analysis and GWAS (Genome Wide Association Study).

Only biallelic SNPs were retained, while variants with excessive missing data or extremely low allele frequency were removed. Specifically, SNPs with more than 2% missing genotypes and individuals with more than 5% missing data were discarded. Variants with minor allele frequency (MAF) lower than 0.10 were excluded to reduce the influence of rare alleles. Furthermore, SNPs departing from Hardy–Weinberg equilibrium (p-value $< 1\times10^{-6}$) were filtered out to remove genotyping errors and potential artefacts. Ambiguous base calls were also excluded, leaving only standard A/C/G/T SNPs.

To minimize redundancy in the dataset, linkage disequilibrium (LD) pruning was performed using a sliding-window approach (100-SNP window, step size of 10 SNPs, $r^2 < 0.3$).

Following quality control and LD-pruning, the final dataset comprised 1,093,320 high-quality SNPs across 88 individuals, which were used for all subsequent population structure analyses.

- **Principal Component Analysis (PCA)**

Principal component analysis (PCA) was performed in *PLINK 2.0*, using the LD-pruned SNP set. Only unlinked variants were retained to avoid bias from linkage disequilibrium, and allele frequencies were used to compute genetic distances among individuals. The PCA allows for visualizing the genetic structure of the samples and detecting possible outliers or stratification patterns related to sampling location or environmental conditions.

For this purpose, the pruned genotype dataset was used as input together with the allele frequency file, and the first principal components were computed to summarize the main axes of genetic variation. The resulting PCA plot was then annotated by coloring individuals according to both degree of urbanization and soil type, to assess whether genetic structure aligned with these environmental classifications. The analysis produced an eigenvector and an eigenvalue file, which were subsequently used for visualization.

- **Admixture analysis**

Population structure was investigated using *ADMIXTURE* (132), a software tool for maximum likelihood estimation of individual ancestries from multilocus SNP genotype datasets.

For the analysis, the LD-pruned genotype dataset in numeric BED format was used. Ten-fold cross-validation (CV) was performed for K values ranging from 1 to 3 to determine the optimal number of genetic clusters. The resulting admixture coefficients were used to visualize individual ancestry proportions and infer population structure among the sampled groups.

- **Fixation Index (F_{ST})**

Genetic differentiation among environmental groups was quantified using F_{ST}, computed with *PLINK 2.0* following the Hudson estimator.

In a first step, F_{ST} was calculated at the group level, considering all individuals partitioned by the three environmental categories and by soil type. This analysis provided a general assessment of genetic structuring across the complete set of ecological conditions represented in the dataset.

Subsequently, a second, more targeted analysis was carried out to obtain locus-specific F_{ST} values, using per-variant F_{ST} estimates with the 'report-variants' modifier; this generates a separate file for each population pair, focusing specifically on the environmental contrasts of primary interest. In this

case, for the environmental gradient, the comparison was restricted to individuals in the two extreme categories, urban and natural, which also provided the most balanced sample sizes, thereby maximizing the detection of divergence. For the soil variable, locus-level F_{ST} was estimated only between the two most-represented soil types, limestone and tuff, which also showed partial association with the urbanization gradient (limestone predominantly in natural areas, tuff in urban areas).

The locus-level F_{ST} results were then used to identify and characterize highly differentiated loci between environmental groups. Significant SNPs were first mapped to the genome using *BEDtools* (133) by intersecting their coordinates with the reference annotation. Their genomic context was subsequently visually inspected in the *Integrative Genomics Viewer* (*IGV*) (134) to verify whether they fell within or were in proximity to genes or regulatory regions. Finally, variants located in coding regions or near annotated genes were functionally characterized through similarity searches (BLAST) against protein databases, allowing the identification of nearby genes and the inference of their potential biological roles.

3.6.6. Detection of Selection Signals

To identify candidate genomic regions under selection associated with urbanization, multiple complementary statistics were computed and integrated to detect convergent signatures of adaptation between urban and natural populations. The genome-wide VCF file was first partitioned by environmental group, retaining only individuals classified as urban or natural based on the degree of urbanization of their sampling sites. Separate VCF files were generated for each population to enable independent computation of population-specific statistics and cross-population comparisons.

- **XP-EHH Analysis**

Cross-Population Extended Haplotype Homozygosity (XP-EHH) was computed using *Selscan* v2.0 (135), a high-performance tool for detecting recent positive selection by analyzing extended haplotype homozygosity. Unlike within-population methods (e.g., iHS), XP-EHH compares haplotype structure between two populations, making it particularly suited for identifying selective sweeps specific to one group.

XP-EHH scores were calculated for all SNPs in the dataset by comparing urban and natural populations using their respective VCF files. Since phased haplotype data were not available, the analysis was performed in unphased mode, which estimates haplotype structure directly from genotype data. A genetic map file was provided to account for recombination rate variation across the

genome. To reduce noise from low-frequency variants, only SNPs with minor allele frequency (MAF) ≥ 0.05 were included in the analysis.

Raw XP-EHH scores were then normalized within allele frequency bins to account for the dependency of haplotype homozygosity on allele frequency. Specifically, SNPs were grouped into 1% allele-frequency bins (0-1), and for each bin, XP-EHH values were standardized using a z-score transformation based on the bin's mean and standard deviation. This normalization approach reduces systematic biases associated with allele frequency and enables more accurate identification of selection signals across the allele frequency spectrum.

To facilitate the identification of genomic regions under selection, normalized XP-EHH values were aggregated into 50-kb sliding windows with a 20-kb step size across all chromosomes. For each window, the mean normalized XP-EHH score was computed, providing a genome-wide view of selection patterns. Positive XP-EHH values indicate selection in the urban population, while negative values suggest selection in the natural population.

- **Nucleotide Diversity (θπ) and π Ratio**

Nucleotide diversity (θπ) was estimated separately for urban and natural populations using *VCFtools* (136) with the *--window-pi* option in non-overlapping 10 kb windows across the genome. θπ quantifies the average number of pairwise differences per site within a population and serves as a measure of genetic variation.

To identify regions with contrasting diversity levels between populations, the ratio of nucleotide diversity (π_natural/π_urban) was calculated for each window and $\log_2$-transformed. A small constant ($\varepsilon = 10^{-6}$) was added to both numerator and denominator to avoid division by zero in windows with extremely low diversity. Positive $\log_2$ π ratio values indicate higher diversity in natural populations, whereas negative values suggest elevated diversity in urban populations. Regions with extreme π ratios may reflect differential selective pressures, demographic processes, or localized adaptive events.

- **Integration of Selection Statistics**

To identify high-confidence candidate regions under selection, the top 1% of genomic windows or SNPs were extracted independently for three statistics: normalized XP-EHH (aggregated in 50 kb windows), $\log_2$π Ratio (10 kb windows), and locus-specific F_{ST}.

Spatial overlap among the three selection statistics was assessed using a window-based approach. For F_{ST}, which is computed at individual SNP positions, a ±5 kb flanking region was defined around

each top 1% SNP to create candidate intervals. These intervals were then intersected with θπ ratio windows (10 kb) and XP-EHH windows (50 kb) using the *foverlaps* function from the *data.table* package in R, which performs efficient interval-based overlaps. Genomic regions showing convergent signals across all three statistics were prioritized as high-confidence candidate loci under selection, as the integration of complementary approaches reduces false positives and increases the reliability of selection inference.

Results were visualized using Manhattan plots to display the genome-wide distribution of each statistic across chromosomes. Candidate SNPs and windows identified through the integrated approach were subsequently annotated by mapping their genomic coordinates to the reference genome as described before.

- **Linkage Disequilibrium (LD)**

Genome-wide linkage disequilibrium (LD) was estimated separately for the Urban and Natural populations using *PLINK 2.0*, with the r^2 statistic (the squared correlation coefficient) calculated for all SNP pairs.

The genomic architecture and extent of selective sweeps on Chromosomes 2, 4, 5, and 9 were examined by estimating LD through calculation of r^2 between unphased genotypes. Analyses were conducted using *PLINK 2.0*, with Urban and Natural populations evaluated separately. A sliding window approach was applied, using a physical distance limit of 2,000 kb (*--ld-window-kb* 2000) and a maximum of 100 SNPs per window (*--ld-window* 100). To comprehensively represent local linkage structure, all pairwise r^2 values were retained (*--ld-window-r2* 0). For each chromosome, the extent of LD was assessed within specific SNP clusters by interpreting r^2 values according to established thresholds: $r^2 > 0.8$ indicated strong association, $0.5 < r^2 \leq 0.8$ indicated moderate association, and $r^2 < 0.2$ indicated weak or negligible association. On Chromosome 9, due to the presence of a cluster of 27 SNPs, data were aggregated into 100-SNP bins, and the mean r^2 for each bin pair was calculated to generate a whole-chromosome triangular heatmap. This approach facilitated identification of extended LD blocks, which are indicative of localized selective sweeps and reduced recombination rates associated with urban selection. Additionally, LD decay curves were visualized by plotting mean r^2 values against physical distance (bp), enabling direct comparison of background LD structure between populations and providing a baseline for interpreting the significance of the 27-SNP cluster.

3.7 Genome-wide association analysis

- **Phenotypic Trait Measurements**

To investigate plant variability across different environments, Dr. Yuan-Fu Chan and Dr. Laura Pellegrini measured a range of phenotypic traits of *R. raphanistrum* in both natural field conditions and a controlled common garden experiment. I deeply thank both Dr. Chan and Dr. Pellegrini for sharing data on phenotypic trait measurements.

A total of 68 traits were initially assessed, including morphometric measurements (fruit and flower dimensions, vegetative traits), reproductive output (flower and fruit counts), and fitness-related traits. From this initial set, traits were prioritized for genome-wide association analysis through a stepwise selection process: (i) 24 continuously distributed quantitative traits suitable for linear modeling were identified; (ii) traits showing significant variation among urbanization categories were retained (Wilcoxon rank-sum test, adjusted $p < 0.05$); (iii) traits with sufficient overlap between phenotypic measurements and available genomic data were selected, accounting for the fact that phenotyping and sequencing were conducted on partially overlapping sample sets. The phenotypic traits were specifically compared among urban, peri-urban, and natural populations to evaluate potential divergence along the urbanization gradient and to identify traits that may contribute to adaptive responses. Statistical comparisons were conducted using the Wilcoxon rank-sum test (137), and significant differences between groups were illustrated on the plots using adjusted p-values.

- **Genome-Wide Association Analysis**

GWAS was performed using a linear mixed model implemented in *GEMMA* (138), accounting for relatedness and population structure as random effects via a genetic similarity matrix, and including environmental category as a fixed-effect covariate. The analysis included 4,571,587 SNPs. p-values from the Wald test were calculated for each SNP. To account for multiple testing, a Bonferroni correction was applied ($p < 1.09 \times 10^{-8}$), and SNPs surpassing this threshold were considered genome-wide significant.

Manhattan and QQ plots were generated to visualize the distribution of p-values across the genome, with the Bonferroni threshold indicated on the Manhattan plot.

Significant SNPs were mapped to the reference genome using *BEDtools* and visually inspected in *IGV* to determine proximity to genes or regulatory regions. Coding and nearby genes were further characterized using BLAST against protein databases to infer potential biological roles associated with the identified loci.

3.8 Population transcriptomics

- **Pre-processing and alignment of RNA-seq Reads**

After the quality control and trimming procedure described in paragraph 3.5.3, only high-quality paired reads were retained. These filtered reads, obtained after the *FastQC* assessment and *Trimmomatic* trimming, were then processed through a Nextflow-based pipeline (*nf-core/rnavar*), which performed the following steps:

- A new quality control check of the processed reads (*FastQC*)
- Alignment to the reference genome (*STAR*)
- Sorting and indexing of alignments (*SAMtools*)
- Duplicate marking (*GATK4 MarkDuplicates*), applied to identify and flag PCR-induced duplicate reads. Marking these duplicates helps prevent artificial inflation of read depth. It reduces the risk of false-positive variant calls, ensuring a more accurate representation of the underlying genomic and transcriptomic variation.

- **Post-alignment Processing**

After alignment, reads containing Ns in their CIGAR strings were processed using *GATK4 SplitNCigarReads*. This tool identifies all N CIGAR operations, typically corresponding to splice junctions in RNA-seq data, and splits each affected read into multiple new reads (k+1, where k is the number of N elements). For each resulting read, the portion to the left of the first N is retained, while the segment spanning the spliced region (including the Ns) is hard-clipped, and the process is repeated for subsequent N blocks. This step is required for the proper post-processing of RNA-seq alignments against a genomic reference and enables downstream tools to handle spliced reads correctly.

Subsequently, base quality score recalibration was performed using *GATK4 BaseRecalibrator*, followed by *ApplyBQSR*, to correct systematic errors in base quality scores and improve variant calling accuracy.

- **Variant Detection and VCF Normalization**

Variant calling was performed with *GATK4 HaplotypeCaller*. The resulting VCF files for individual samples were then merged and normalized using *BCFtools* package (126,130), considering only the nine main chromosomes and excluding scaffolds, as previously done for the DNA variant calling.

3.8.1. Population Structure Analysis

Subsequently, population genetics analyses were carried out using the same approach as for DNA variant calling, including PCA, F_{ST} calculation, and *ADMIXTURE* clustering with *PLINK 2.0* as previous described in paragraph 3.6.5. These analyses were performed with the specific aim of comparing the population structure inferred from the RNA-derived variants with that obtained from the DNA dataset, leveraging the 22 samples that were shared between the two datasets. This allowed a direct assessment of the consistency between genomic and transcriptomic signals of genetic differentiation.

4. Results and Discussion

4.1 Gene expression analysis

The gene expression analysis across three urbanization levels (natural, peri-urban, and urban) revealed that urbanization leads to changes in gene expression (Fig. 7). Specifically, a total of 895 differentially expressed genes (DEGs) were identified between natural and urban sites (385 down-regulated, 509 up-regulated; Fig. 7B), 1,228 DEGs between peri-urban and natural sites (772 down-regulated, 456 up-regulated; Fig. 7C), and 601 DEGs between peri-urban and urban sites (274 down-regulated, 327 up-regulated; Fig. 7D). These findings suggest that urbanization affects the expression of hundreds of genes across various biological pathways.

The Venn diagram reveals that only 14 genes are shared across all three comparisons, indicating that each habitat transition triggers distinct gene expression changes (Fig. 7A). The peri-urban versus natural comparison shows the highest number of DEGs and unique genes (754), suggesting that peri-urban populations exhibit a distinct gene expression profile rather than simply representing an intermediate between natural and urban habitats. Peri-urban areas may impose unique selective pressures or induce transcriptomic instability not present in natural or urban environments, as they combine both types of stressors in close proximity (139). These areas may also support populations undergoing active adaptation, as evidenced by the high number of DEGs. In contrast, the peri-urban versus urban comparison yields fewer DEGs and unique genes, suggesting greater similarity in gene expression profiles between these two environments.

For subsequent analyses, we focused on the natural versus urban comparison. This contrast encompasses the full spectrum of environmental differences associated with urbanization, involves two clearly defined habitat types, and yielded a substantial yet manageable set of 895 DEGs for functional analysis.

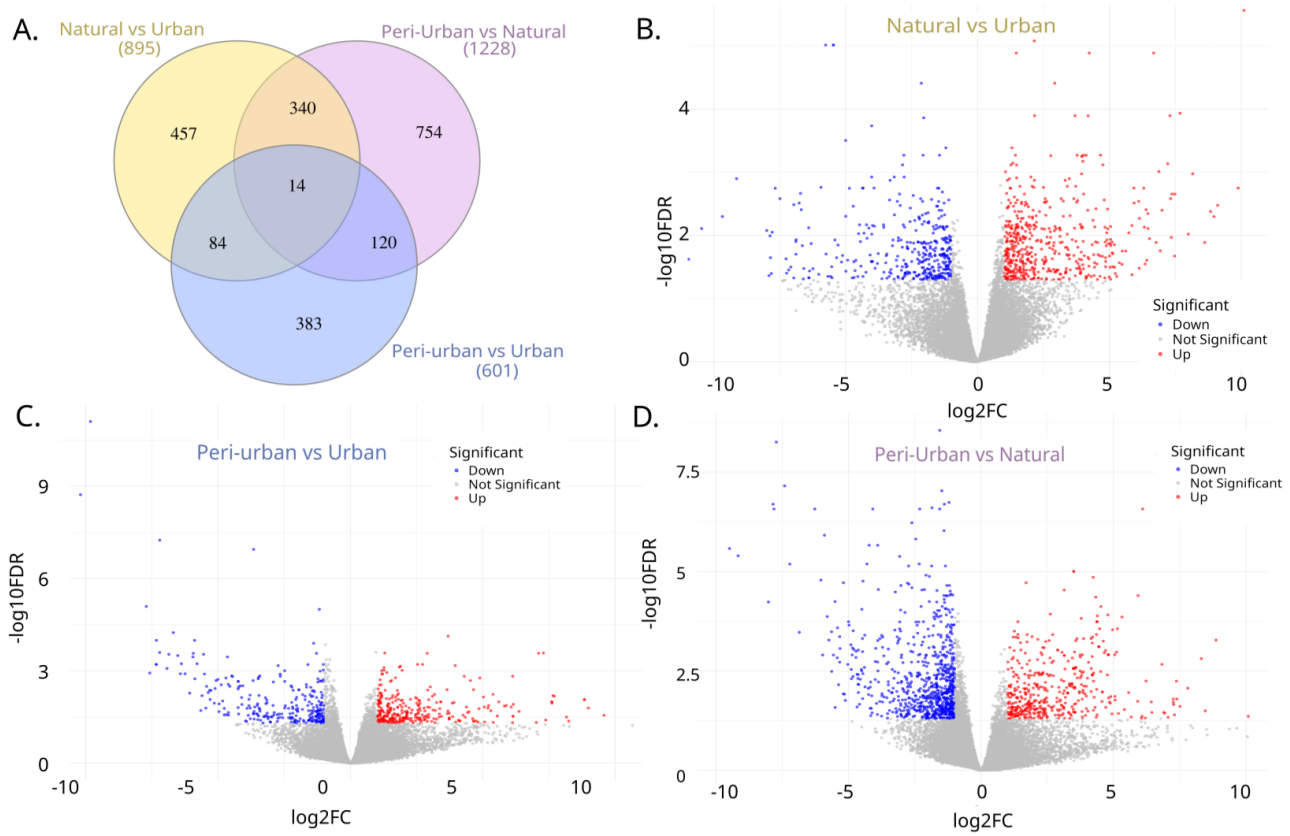

Figure 7. *In silico* differential gene expression analyses across an environmental gradient. (A) Venn diagram showing the overlap of differentially expressed genes (DEGs) among all pairwise comparisons. (B) Volcano plot of DEGs in the Natural versus Urban comparison. (C) Volcano plot of DEGs in the Peri-urban versus Natural comparison. (D) Volcano plot of DEGs in the Peri-urban versus Urban comparison. In the Volcano plots, red dots represent significantly upregulated genes, blue dots indicate significantly downregulated genes, and grey dots correspond to non-significant differentially expressed genes.

Comparisons among populations on different soil types revealed DEGs (Fig. 8A-B): 780 in limestone versus tuff (408 down-regulated, 372 up-regulated), 485 in lava versus tuff (251 down-regulated, 234 up-regulated), 376 in limestone versus lava (143 down-regulated, 233 up-regulated), 97 in lava versus pyroclast (90 down-regulated, 7 up-regulated), 67 in pyroclast versus tuff (3 down-regulated, 64 up-regulated), and 22 in limestone versus pyroclast (20 down-regulated, 2 up-regulated).

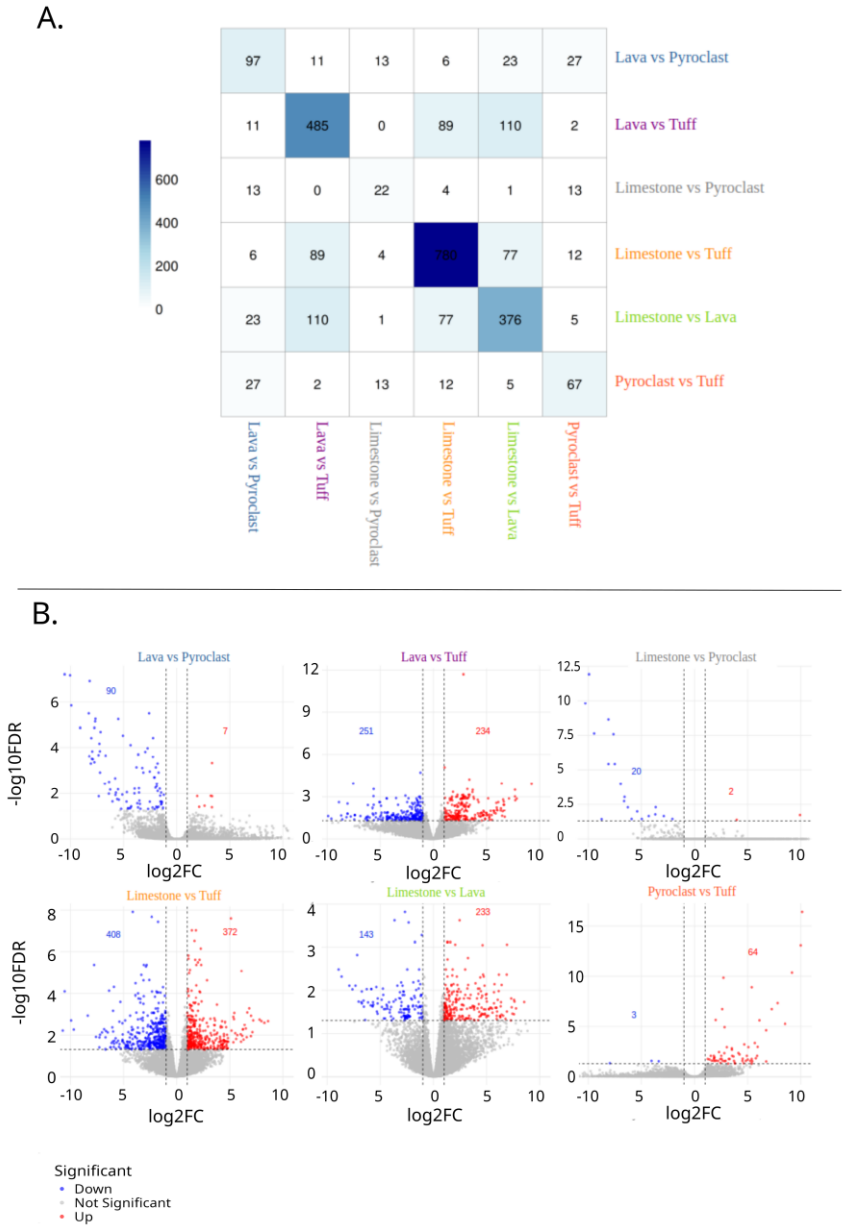

Figure 8. *In silico* differential gene expression analyses across different soil types. (A) Heatmap showing the number and overlap of DEGs among all pairwise soil-type comparisons; (B) Volcano plots of DEGs for each pairwise comparison between soil types (lava, tuff, pyroclast, and limestone). In the Volcano plots, red dots represent significantly upregulated genes, blue dots indicate significantly downregulated genes, and grey dots correspond to non-significant genes.

The sampling design matched the natural distribution of habitats in the study area, where urban development is limited by geology. As a result, all four natural populations (12 individuals) are found only on limestone, while urban populations are spread across lava (1 population), tuff (2 populations),

and pyroclastic substrates (1 population; 20 individuals total). This means that the natural versus urban comparison, which found 895 DEGs, mixes the effects of urbanization with changes in substrate, making it hard to separate their individual impacts. The limestone versus tuff comparison (780 DEGs) is similar in scale to the natural versus urban comparison, suggesting that substrate chemistry can have as much influence as urbanization. This overlap highlights the difficulty of separating these factors in studies that link urbanization and geology.

Ideally, the effects of urbanization and substrate should be separated through within-substrate or within-urbanization comparisons. However, limited replication at the population level prevents such analyses. For example, lava substrates are represented by only two populations: one peri-urban (TREC, n = 8) and one urban (SGIO, n = 8). Although the sample sizes are comparable, comparing only these two groups would constitute pseudo-replication, as observed differences could result from unique population histories or local factors rather than urbanization. Single-population comparisons lack sufficient replication to support broader inferences. Furthermore, because all natural populations occur on limestone, it is not possible to test substrate effects within natural habitats or to examine urbanization effects on non-limestone substrates. Due to these limitations, it is not possible to clearly distinguish the effects of urbanization and substrate on gene expression using standard pairwise comparisons.

4.1.1. Functional Categories Enriched in DE Genes

To characterize the functional nature of the differentially expressed genes between natural and urban populations, Gene Ontology (GO) enrichment analysis was conducted on the 895 DEGs identified in the Natural versus Urban comparison. This analysis revealed coordinated changes in functionally related gene sets, offering insight into the biological processes underlying adaptation to urban environments, whether driven by urbanization itself, correlated to substrate differences, or due to their interaction (Fig. 9).

GO enrichment analysis indicated that urban groups were enriched in biological processes associated with stress response and cellular restructuring. Specifically, urban populations showed enrichment in response to mechanical stimulus (gene ratio = 0.18), endocytic vesicle (gene ratio = 0.18), and regulation of programmed cell death (gene ratio = 0.07). In contrast, several terms related to secondary metabolism exhibited habitat-specific differences. Natural populations were enriched in the lignin biosynthetic process (gene ratio = 0.05), mucilage pectin biosynthetic process (gene ratio = 0.10), response to red light (gene ratio = 0.06), and leaf proximal/distal pattern formation (gene ratio = 0.20).

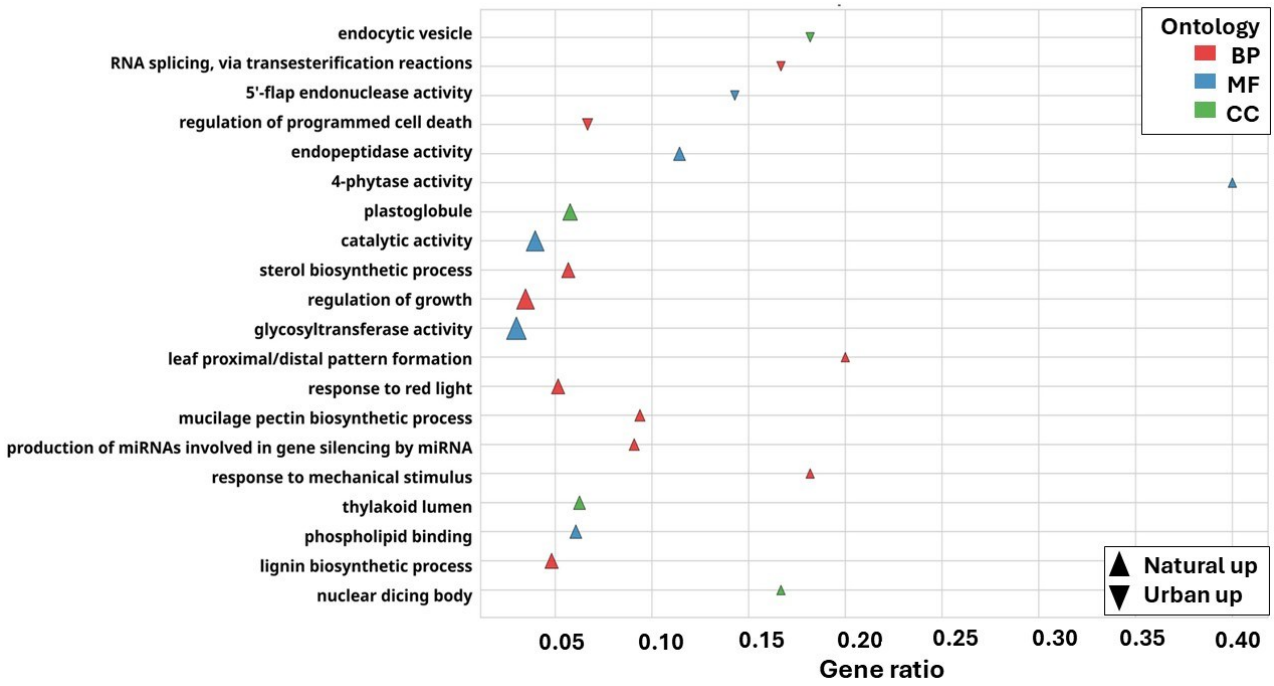

Figure 9. Gene Ontology (GO) enrichment analysis of differentially expressed genes in the Natural versus Urban comparison. The dot plot presents the top 20 significantly enriched GO terms (FDR < 0.05, |logFC| > 1) associated with differentially expressed genes. The x-axis displays the gene ratio, and the y-axis lists the GO terms. Colors represent GO ontology categories: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). Dot size indicates the number of genes associated with each GO term. Upward-pointing triangles indicate GO terms enriched among genes up-regulated in Natural populations (i.e., down-regulated in Urban), and downward-pointing triangles indicate GO terms enriched among genes up-regulated in Urban populations.

Among the enriched pathways, flavonoid biosynthesis emerged as particularly significant in urban populations. Flavonoids are important compounds in plants, providing protection against both biotic and abiotic stresses (140). Elevated flavonoid levels mitigate oxidative stress by scavenging reactive oxygen species (ROS) and confer resistance to pathogens, herbivores, and various environmental challenges, including salinity, dehydration, extreme temperatures, intense light, and ultraviolet (UV) radiation (141–146). Consequently, the presence of flavonoid-related pathways may serve as indicators of multiple stress types, particularly in urban environments.

Urban populations exhibited significant upregulation of 4-phytase activity (gene ratio = 0.40) and glycosyltransferase activity (gene ratio = 0.05), both of which are essential for flavonoid modification and bioavailability. This pattern suggests that urban *R. raphanistrum* populations may enhance antioxidant and UV-protective mechanisms by increasing flavonoid production. Such an adaptation is likely a response to urban stressors, including elevated UV radiation due to reduced canopy cover, oxidative stress from pollution, and altered light quality (147).

4.1.2. Candidate Genes and qPCR validation

Among the DEGs identified in the GO enrichment analysis, four genes associated with flavonoid biosynthesis were selected for subsequent quantitative RT-PCR (qPCR) validation: *PAL2* (phenylalanine ammonia-lyase 2), *CHS3* (chalcone synthase 3), *CHS-like* (chalcone synthase-like) and *F3GT* (flavonol 3-O-glucosyltransferase). These enzymes play essential roles within the phenylpropanoid and flavonoid biosynthetic pathways and the coordinated *in silico* upregulation observed of *PAL2*, *CHS3*, *CHS-like,* and *F3GT* in urban populations suggests a change of the entire phenylpropanoid-flavonoid pathway, from initial substrate generation through commitment to flavonoid biosynthesis to final product stabilization. This pattern is consistent with the GO enrichment results and supports the interpretation that flavonoid biosynthesis is a central molecular mechanism underlying urban adaptation. Additionally, PCA analysis showed that *CHS-like*, *CHS3*, and *F3GT* were the top genes separating natural and urban populations along PC1 (15.4% variance explained), and all are part of the flavonoid biosynthesis pathway.

Beyond the flavonoid biosynthesis pathway, two genes were selected to represent distinct aspects of plant stress response: *LRR*, which encodes a leucine-rich repeat extensin-like protein, and *PYK10*, which encodes the PYK10-binding protein 1.

Differential expression analysis revealed upregulation of the *LRR* gene in urban populations. LRR proteins are integral components of the plant cell wall and contribute to pathogen recognition, cell wall reinforcement, and responses to physical stress (148). The observed increase in *LRR* expression in urban environments may indicate adaptation to multiple stressors, such as increased wind exposure due to reduced tree cover, enhanced cell wall strength to withstand heat and drought, or altered pathogen defense mechanisms. Additionally, LRR proteins are involved in plant development and hormone signaling (149), which may have several implications for adaptation to urban environments.

Conversely, the *PYK10* gene is downregulated in urban populations. *PYK10*, also referred to as *BGLU23* or β-glucosidase 23, along with its binding proteins, facilitates the formation of endoplasmic reticulum (ER) bodies. These specialized compartments in Brassicaceae plants store defense-related β-glucosidases and myrosinases (150). Upon tissue damage, these enzymes are released and catalyze the breakdown of glucosinolates into toxic isothiocyanates, providing protection against herbivores and pathogens (151). The reduced expression of *PYK10* in urban plants may reflect a decreased requirement for anti-herbivore defenses in urban environments, where herbivore pressure is lower, or a reallocation of resources from constitutive chemical defenses to alternative stress responses such as flavonoid production.

Quantitative real-time RT-PCR experiments were conducted on seven wild populations, comprising three natural populations (ANNU, GOCA, SCHI) and four urban populations (CATT, SGIO, SOCC, VARG) (Fig. 10). The qPCR expression patterns of five genes among six candidates confirm the *in silico* differential expression findings. *CHS-like*, *CHS3-like*, *F3GT,* and *PAL2* exhibited higher expression in urban populations compared to natural populations, whereas *PYK10* showed higher expression in natural populations. In contrast, *PRR* gene expression was similar across all populations.

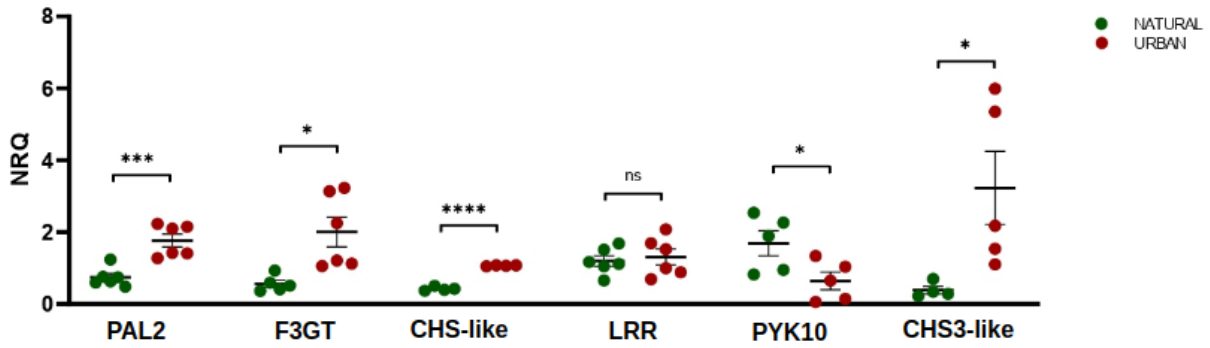

Figure 10. Relative expression of the different candidate genes in urban (red) and natural (green) populations. The expression is reported as Normalized Relative Quantity (NRQ ± SEM), normalized to 18S rRNA. The bars represent the SEM, and the asterisks indicate statistically significant differences in expression between urban and natural populations. * p-values < 0.05, ** p-values < 0.01, *** p-values < 0.001, **** p-values < 0.0001; ns = not significant.

Considerable intra-population variability in gene expression was observed across all genes in urban and natural sites. This variability indicates that individual plants within populations may employ diverse molecular strategies to adapt to urban environments. Such variation within populations may serve as the foundation for ongoing adaptive evolution as urbanization pressures increase.

4.2 Genomic analysis

4.2.1. Population Structure

A total of 1,093,320 biallelic single-nucleotide polymorphisms (SNPs) were identified across all 88 individuals. PCA revealed population stratification associated with both the urbanization gradient and soil type (Fig. 11).

The first two principal components accounted for 32.52% of the total variance (PC1: 32.43%; PC2: 0.09%). Populations from highly urbanized areas (I.S ≥ 75%; red) were primarily located in the right portion of the ordination space. Several individuals from these populations were isolated in the upper section, which indicates increased genetic differentiation. Moderately urbanized populations (15% < I.S. < 75%; yellow-orange) occupied an intermediate position. In contrast, populations from low urbanization areas (I.S. ≤ 15%; green) formed a compact cluster in the lower-right portion of the plane. The three wild individuals from Switzerland (grey circles) were mainly distributed among natural populations (Fig. 11A).

When populations were colored by soil type (Fig. 11B), the PCA revealed a partially distinct pattern. As highlighted before, natural populations occurred exclusively on limestone substrates. In contrast, urban populations were predominantly found on tuff, with one population on lava (SGIO) and one on pyroclast (VARG). Peri-urban populations occurred across limestone, lava, and tuff soil types. The partial overlap between urbanization and soil type patterns suggests a correlation between these variables in the study area, making it difficult to disentangle their independent effects.

There was considerable overlap between populations sampled from wild conditions (circles) and those from common garden conditions (triangles). This finding suggests that environmental origin primarily shaped the observed genetic structure.

The PCA results indicate that both the urbanization gradient and soil type contribute to the observed genetic structure, although with different intensities. The separation along PC1 of highly urbanized populations suggests that urbanization may exert selective pressure or serve as a source of genetic isolation.

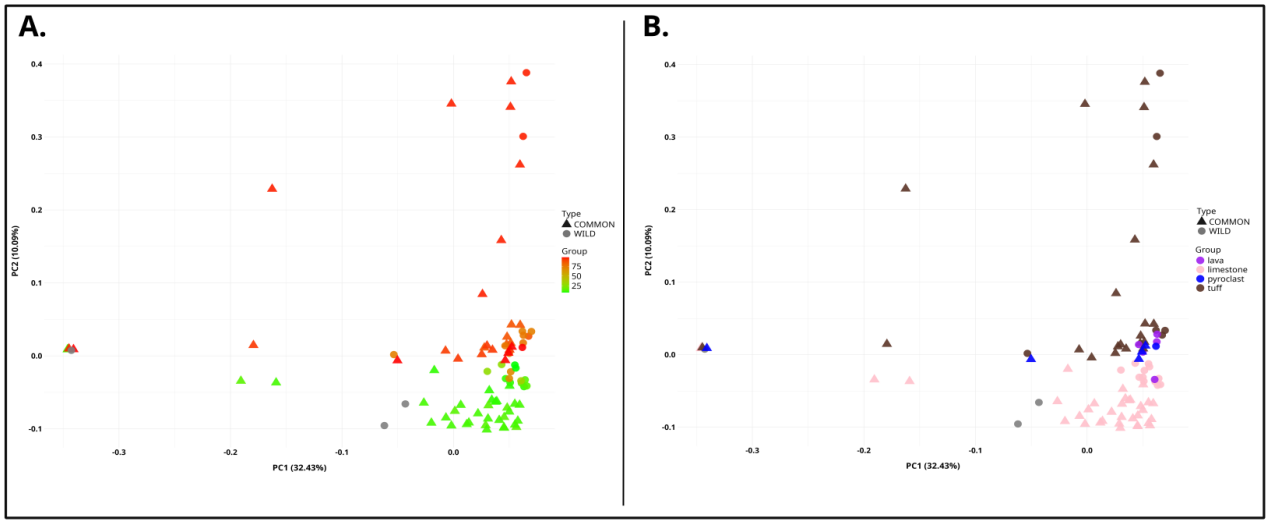

Figure 11. Principal Component Analysis of the DNA-seq data from common garden and wild individuals (A) colored according to the percentage of impervious surface at each sampling site (proxy for urbanization intensity); (B) grouped

according to substrate type (lava, limestone, pyroclast, and tuff). PC1 and PC2 explain 32.43% and 10.09% of the total variance, respectively.

Admixture analysis (Fig. 12) revealed a generally homogeneous genetic structure across most populations, as expected given their geographical proximity. At K=2 and K=3, a dominant ancestral component (blue) was prevalent throughout the study area, particularly in populations such as Astroni (ASTR, 99.84%) and Trecase (TREC, 96.91%).

This genetic similarity is likely attributable to the geographical proximity of the sampling sites, which facilitates continuous seed and pollen dispersal across the landscape. Although minor shifts in ancestry components were observed, such as a slight enrichment of the K3_1 component in urban tuff-associated populations (31.64%) compared to natural limestone populations (4.40%), the overall patterns indicate that neither anthropogenic nor geological factors have yet resulted in significant reproductive isolation. Marked differentiation was observed only in the Switzerland outgroup and the Soccavo population (SOCC, 69.68% K3_1), likely due to geographic distance in the former and localized founder effects or strong selection in the latter.

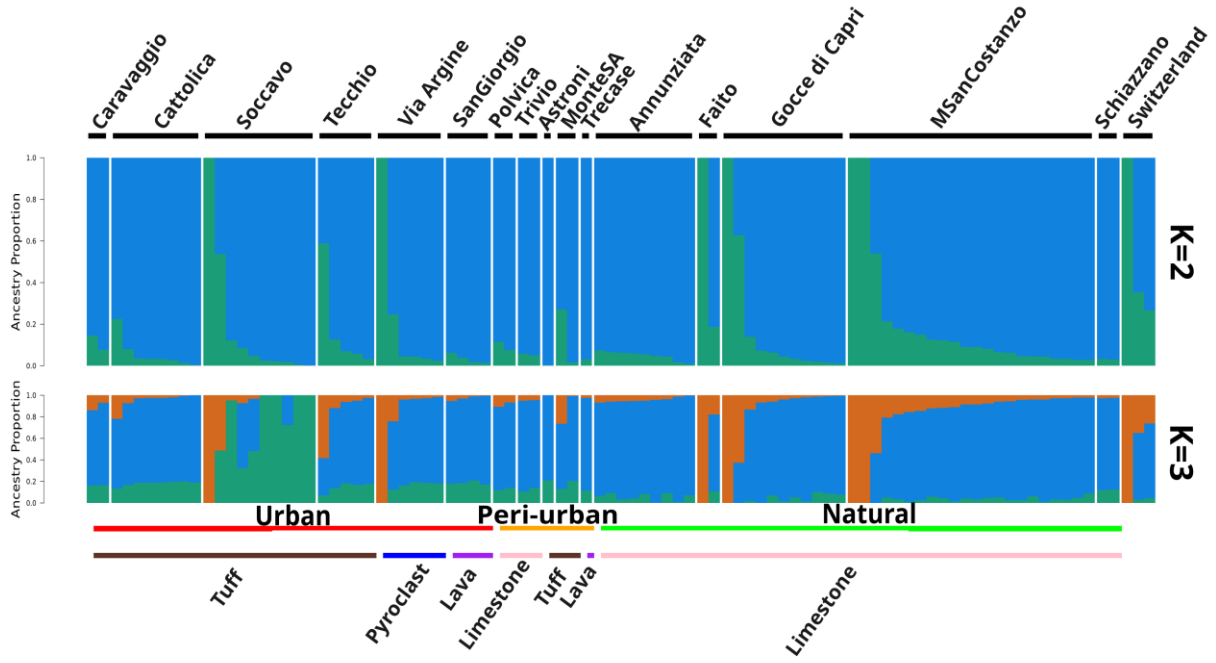

Figure 12. Admixture analysis was conducted on all DNA samples. Each vertical bar in the figure represents an individual, with colors indicating the proportions of ancestry for the inferred genetic clusters at K=2 (upper panel) and K=3 (lower panel). Individuals are grouped by population along the x-axis, and colored bars below denote population identity (black), urbanization class (red: Urban; orange: Peri-urban; green: Natural), and soil type (dark-brown: Tuff; violet: Lava; pink: Limestone; blue: Pyroclast).

The absence of strong stratification is in agreement with the low global F_{ST} values observed. Genome-wide F_{ST} analysis identified low genetic differentiation among populations across varying degrees of urbanization (Table 3) and geological substrates (Table 4). Pairwise F_{ST} values between environmental categories ranged from 0.018 to 0.025. The greatest differentiation occurred between Natural and Peri-urban environments (F_{ST} = 0.0245), followed by Peri-urban versus Urban (F_{ST} = 0.0227) and Natural versus Urban (F_{ST} = 0.0184). Populations from different soil types also showed low genome-wide differentiation (F_{ST} = 0.019-0.029 in most pairwise comparisons). Comparisons between populations and the Swiss outgroup revealed substantially higher F_{ST} values (0.10 to 0.13). These findings confirm strong phylogeographic structure at larger spatial scales and support the suitability of this reference for comparative analyses.

Table 3. Hudson F_{ST} of pairwise comparison between population based on impervious surface classification.

POP 1	POP 2	HUDSON_F_{ST}
Natural	CH	0.105679
Natural	Peri-urban	0.0245171
Natural	Urban	0.0183611
CH	Peri-urban	0.12695
CH	Urban	0.120179
Peri-urban	Urban	0.0227401

Table 4. Hudson F_{ST} of pairwise comparison between population based on soil type.

POP 1	POP 2	HUDSON_F_{ST}
lava	limestone	0.0202043
lava	CH	0.150448
lava	pyroclast	0.0286215
lava	tuff	0.0195865
limestone	CH	0.103641
limestone	pyroclast	0.0193051
limestone	tuff	0.0212901
CH	pyroclast	0.0923467
CH	tuff	0.129003

pyroclast	tuff	0.0228885

4.2.2. Detection of Selection Signals

- **SNPs with high F_{ST}**

Genome-wide F_{ST} values were generally low, indicating recent divergence or ongoing gene flow. However, loci exhibiting unusually high F_{ST} values compared to the genomic background may represent candidate regions influenced by selective processes or linked to selected sites.

In the comparison between Urban and Natural populations, 17 SNPs exhibited F_{ST} values greater than 0.45 (Fig. 13A; Table 1, Appendix). In the Limestone versus Tuff comparison (Fig. 13B), 81 SNPs showed F_{ST} values exceeding 0.45, nine of which overlapped with those identified in the Urban versus Natural comparison.

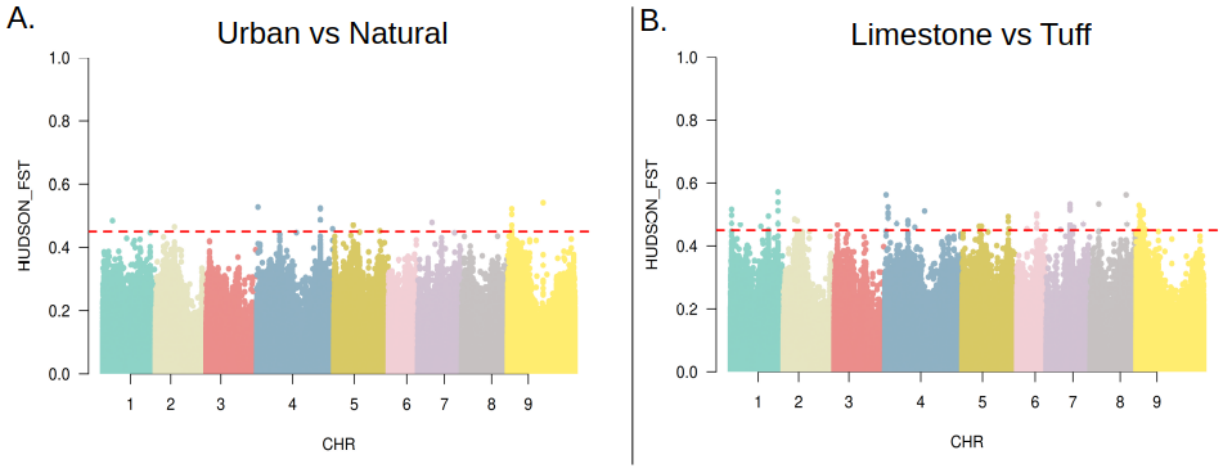

Figure 13. Manhattan plot illustrating genome-wide F_{ST} values: (A) comparison between natural and urban populations; (B) comparison between limestone and tuff soils. Single-nucleotide polymorphisms (SNPs) with F_{ST} values higher than 0.45 are positioned above the dashed red line.

Several loci of interest were identified among SNPs exceeding the F_{ST} threshold of 0.45, suggesting putative signals of positive selection in urban compared to natural populations. These candidates include genes involved in vascular development, transcriptional regulation, and stress response.

On chromosome 4, three SNPs were identified within the gene *bHLH155* encoding a transcription factor that contains a DNA-binding basic region and a helix-loop-helix (HLH) dimerization domain (152,153). The SNP 4:55057945:T:C is located within the intron 6, while the SNPs 4:55058281:G:T and 4:55058658:A:G, both in exon 8, result in the missense mutations M/I, and Q/R, respectively.

Both missense SNPs may affect protein function by altering protein stability, DNA binding, or dimerization capability (154). The low expression difference of *bHLH155* between urban and natural populations (logFC |0.15|) indicates that adaptation is likely mediated by changes in protein function rather than transcript abundance.

On chromosome 5, two candidate SNPs (5:16845155:A:T; 5:16850177:G:A) were identified: the first located 4,054 base pairs upstream and the second 9,076 base pairs upstream of *GSL7* gene (Callose Synthase 7). These variants may influence distal regulatory regions of *GSL7*. *GSL7* encodes a phloem-specific callose synthase that is involved in sieve element development, wound-induced callose deposition, and defense against pathogens (155,156). The SNP 5:16845155:A:T disrupts a TATA-binding protein site and introduces an HD-ZIP binding site, potentially influencing the expression of *GSL7*. HD-ZIP transcription factors regulate stress responses, root development, and vascular patterning in plants, which are processes relevant to both substrate adaptation and urban stress tolerance. *In silico* differential expression analysis indicated higher expression of *GSL7* in urban populations (logFC |1.09|), suggesting differences in defense mechanisms or resource allocation between urban and natural populations.

Another SNP (5:39885597:G:A) was found on chromosome 5 within the intron 1 of a *RICESLEEPER 1 (RIS1)* homolog, suggesting possible regulatory effects. The *SLEEPER* gene family, derived from domesticated hAT transposases, functions as a developmental regulator. SLEEPER proteins possess a BED zinc finger DNA-binding domain and an hATC dimerization domain. In *Arabidopsis*, *DAYSLEEPER* is essential for growth (157). The identification of a *RIS1* homolog under possible selection in urban populations of *R. raphanistrum* suggests potential adaptive developmental effects. The intronic SNP may influence *cis*-regulatory activity, although functional validation is necessary.

On chromosome 9, a cluster of SNPs within and adjacent to the gene Rr0R9c037999 exhibited high F_{ST} values in both the Urban-Natural (4 SNPs ; 9:3818087:G:C, 9:3818662:T:G, 9:3818679:A:C, 9:3818745:T:A) and Limestone-Tuff (11 SNPs) comparisons. The gene Rr0R9c037999 encodes a predicted protein 2,574 amino acids in length. Although this predicted protein lacks clear homologs in NCBI database, InterPro Scan analysis detected a reverse transcriptase zinc-binding domain, an endonuclease/exonuclease/phosphatase superfamily domain, and a DNA/RNA polymerase superfamily domain. These domains are characteristic of retrotransposons and suggest that this gene may represent a transposable element.

Additionally, on chromosome 9 the SNP 9:3773568:T:C falls 139 base pairs downstream of the *HDG5* gene (Homeobox-leucine zipper), within the 3′-UTR, and might influence mRNA stability and/or translation due to the creation of NF-Y and MYB binding sites.

- **Linkage Disequilibrium of the SNPs with high F_{ST}**

High F_{ST} values identify loci with substantial genetic differentiation, reflecting marked shifts in allele frequencies between environments, which are likely attributable to divergent selection. To determine whether these highly differentiated SNPs are components of shared genomic regions, F_{ST} analysis was combined with LD assessment. While F_{ST} highlights individual sites under selection, LD analysis shows the haplotypic structure of these regions. Increased LD in urban populations suggests the formation of consolidated blocks of adaptive alleles. In contrast, reduced LD relative to natural populations, despite high F_{ST}, may indicate that selection targets specific variants.

On chromosome 4, the clustering of three SNPs in the *bHLH155* gene (intron 6 and exon 8) suggests the presence of a critical functional region. LD analysis (Fig. 14) demonstrated higher pairwise r^2 values in urban populations than in natural populations across all three comparisons. Specifically, the linkage between the SNPs 4:55057945:T:C and 4:55058658:A:G (SNPs A and C in Fig. 14) was strong in the urban group ($r^2 = 0.50$) but lower in the natural population ($r^2 = 0.28$). The observation that all three SNP pairs exhibit significantly higher linkage in the urban population indicates that this genomic region is maintained as a consolidated haplotype block, representing a molecular signature of positive selection in the urban environment.

On chromosome 5, the two SNPs upstream the *GSL7* gene (5:16845155:A:T and 5:16850177:G:A, SNPs D and E in Fig. 14) displayed moderate linkage in natural populations ($r^2 = 0.15$), appearing nearly independent, but strong linkage in urban populations ($r^2 = 0.47$). This pattern suggests selection for a specific regulatory haplotype, potentially facilitating coordinated control of *GSL7* expression to enhance adaptation to urban environmental pressures.

In contrast, on chromosome 9, LD analysis of the cluster of SNPs 9:3818087:G:C, 9:3818662:T:G, 9:3818679:A:C, and 9:3818745:T:A (SNPs F, G, H, and I in Fig. 14) revealed higher r^2 values in natural samples (range: 0.61 to 0.90) compared to urban samples, which exhibited generally moderate to strong r^2 values (range: 0.43–0.86). This finding suggests that urban selective pressure is acting on an ancestral haplotype block, potentially leading to its progressive refinement.

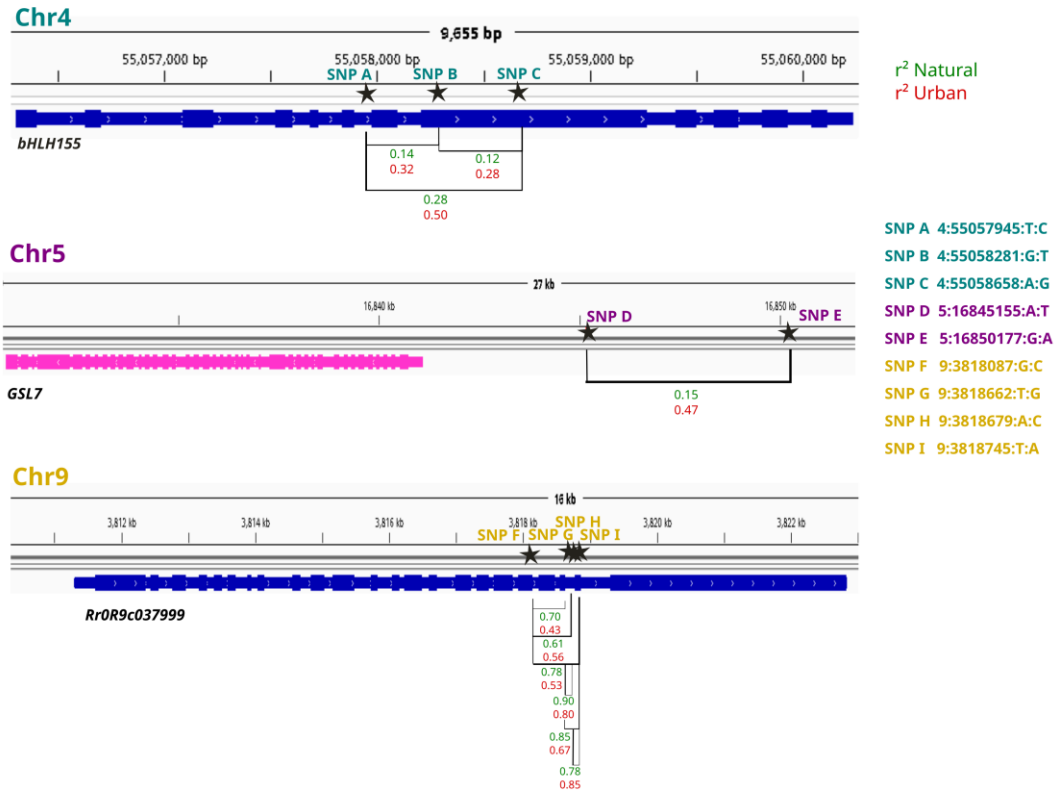

Figure 14. The genomic architecture and population-specific linkage disequilibrium (LD) patterns for three candidate loci: chromosome 4 *bHLH155* (positive strand), chromosome 5 downstream of *GSL7* (negative strand), and chromosome 9 unannotated gene RrOR9c037999 (positive strand). Blue and pink indicate positive and negative strand genes, respectively. Linkage disequilibrium (r^2) values between SNP pairs are presented for natural (green) and urban (red) populations.

- **Integrated Selection Scan**

Beyond F_{ST} analysis, selection signal analysis was performed using an integrated genomic dataset combining three complementary statistics: XP-EHH, the θπ ratio, and F_{ST} (Fig. 15). This approach aimed to identify high-confidence candidate regions under selection by requiring spatial convergence across multiple independent lines of evidence. Candidate loci with triple overlap were considered the most robust signals of urbanization-associated selection, as these regions independently satisfied stringent thresholds for extended haplotype homozygosity, population differentiation, and differential nucleotide diversity.

The dataset included 50,365,156 SNPs for XP-EHH analysis and 4,571,587 SNPs for F_{ST} calculations. Nucleotide diversity (θπ) was estimated in 40,698 and 40,689 non-overlapping 10 kb

windows for natural and urban populations, respectively. XP-EHH analysis identified 205 windows (top 1%) with extreme haplotype homozygosity, suggesting potential selective sweeps.

Population differentiation analysis identified 24,199 SNPs (top 1%) with elevated F_{ST} values, indicating pronounced allele-frequency divergence between urban and natural populations. Analysis of nucleotide diversity ratios revealed 814 windows (top 1%) with extreme $\log_2(\pi natural/\pi urban)$ values. The distribution of these windows encompassed both positive and negative extremes, indicating that certain genomic regions exhibit higher diversity in natural populations, while others display increased diversity in urban populations. These findings highlight the complex and heterogeneous selective pressures associated with urbanization.

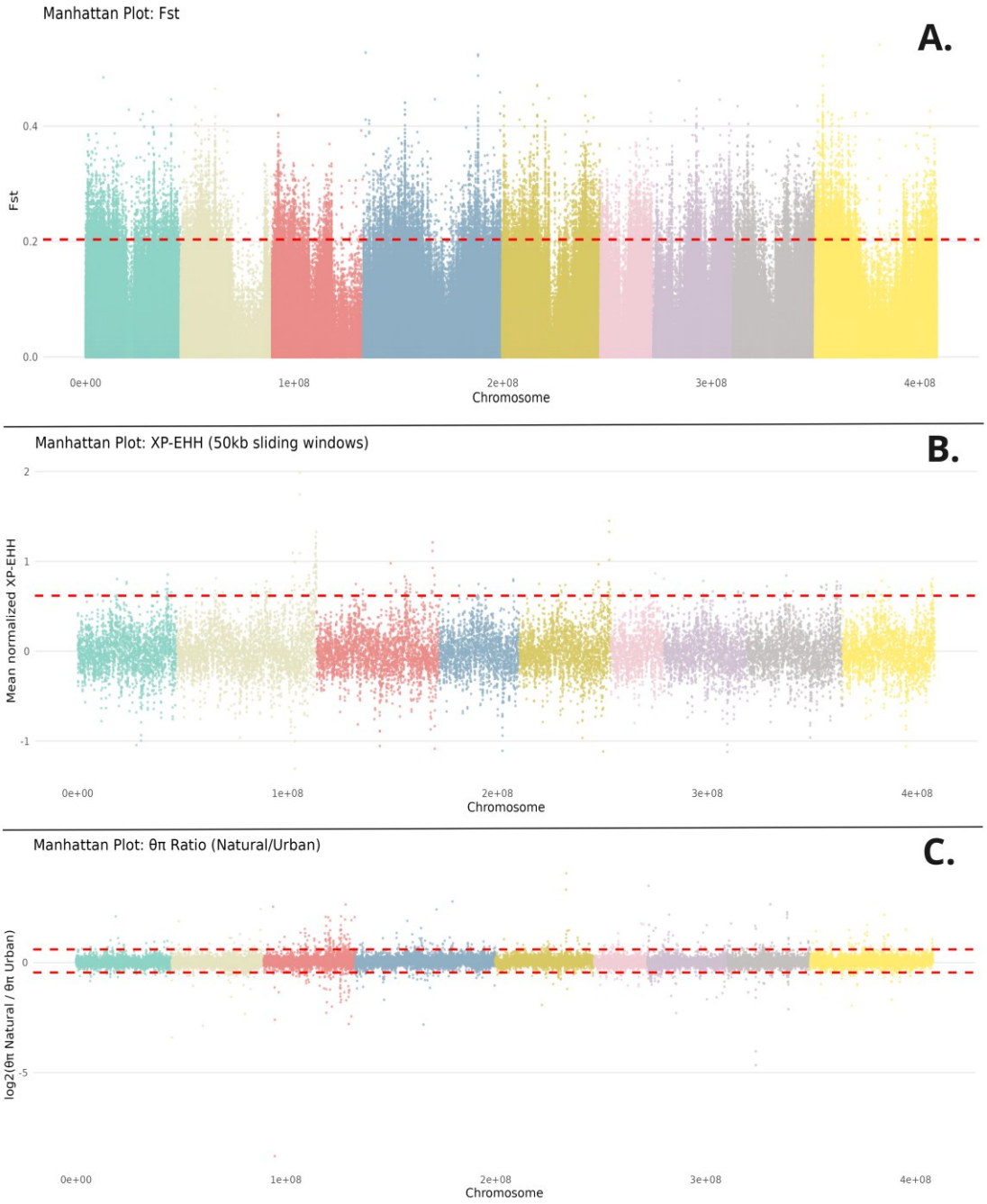

Figure 15. Genome-wide selection scan identifying candidate regions under selection. The genomic landscape of selection between urban and natural populations of *R. raphanistrum* is shown as follows: (A) Manhattan plot of F_{ST} values, which indicate the degree of genetic differentiation; the red dashed line marks the significance threshold (F_{ST} > 0.2). (B) Manhattan plot of normalized XP-EHH values (50 kb sliding windows), which detect selective sweep based on extended haplotype homozygosity; values above the red dashed line indicate potential directional selection. (C) Manhattan plot of nucleotide diversity θπ ratio, expressed as log2(π_natural/π_urban); positive values indicate higher diversity in natural populations, while negative values indicate reduced diversity in urban populations. Horizontal red dashed lines represent the top 1% of the distribution for each statistic.

The integrated selection scan (triple overlap) identified candidate SNPs on chromosomes 2, 4 and 9.

Among the SNP identified in the overlap analysis (Table 2, Appendix), a particularly compelling candidate (2:12139680:C:T) was located on chromosome 2 within the intron 1 of the *Flowering Locus C* gene (*FLC*), a key regulator of flowering time in response to environmental cues. *FLC* functions as a major flowering repressor in Brassicaceae, integrating environmental temperature signals, especially vernalization (prolonged cold exposure), to ensure optimal reproductive timing (158,159). Vernalization induces stable epigenetic silencing of *FLC* via Polycomb-mediated histone modifications, thereby establishing a molecular memory of winter (160). Furthermore, environmental temperature fluctuations influence flowering through both *FLC*-dependent and independent pathways, with effects that vary across genotypes and thermal contexts (161,162). Urban environments impose distinctive thermal regimes that differ substantially from those of natural habitats. In *R. raphanistrum*, these thermal modifications constitute persistent selective pressures that may favor altered *FLC* regulation to adjust flowering time. Urban warming can effectively shorten winter duration, modify vernalization requirements, or alter thermal cues that signal optimal reproductive timing. The intronic position of this selected SNP is functionally plausible because *FLC* regulation relies on extensive intronic elements. *FLC* introns contain regulatory sequences that control alternative splicing (163), chromatin remodeling sites required for vernalization response (164), and generate long non-coding RNAs (*COOLAIR*, *COLDAIR*) that epigenetically modulate *FLC* expression (165). Thus, intronic variants may alter *FLC* function through multiple post-transcriptional mechanisms without necessarily affecting basal transcription levels.

Three additional SNPs (2:12147145:A:C, 2:12148604:T:C, 2:12149585:C:A) were identified on chromosome 2, in exon 3, exon 5, and intron six of a gene encoding a TPR repeat-containing thioredoxin TTL3-like protein. Two of these polymorphisms are missense mutations, specifically E/A (2:12147145:A:C) and F/S substitution (2:12148604:T:C). The TTL gene family is essential for plant adaptation to abiotic stress. *TTL3*, in particular, serves as a critical regulator of lateral root development under high osmotic stress by modulating cell wall biophysical properties to facilitate root emergence (166). The occurrence of these missense mutations in urban populations indicates adaptive fine-tuning of root architecture, which may enable *R. raphanistrum* to sustain root growth and water acquisition efficiency in compacted or osmotically challenging urban soils.

On chromosome 4, a cluster of six candidate SNPs (4:44545214:T:C, 4:44545284:T:C, 4:44547067:A:G, 4:44551913:A:G, 4:44557895:C:T, 4:44558728:T:C) was identified within a non-

coding intergenic region spanning approximately 13.5 kilobases between the *PHD finger protein ING2* (inhibitor of growth protein) and Rr0R4c018433, an uncharacterized gene. *ING2* encodes an epigenetic regulator that participates in chromatin remodeling, growth control, and stress response by recognizing histone modifications (167,168). The intergenic position of these variants indicates localization within potential regulatory elements that may influence one or both adjacent genes.

- **Linkage Disequilibrium of the SNPs from the integrated selection scan**

LD analysis indicated that the cluster of three SNPs on chromosome 2 exhibited low to moderate r^2 values in both natural and urban populations. The highest LD was observed between the SNPs 2:12147145:A:C and 2:12148604:T:C (SNP J and K in Fig. 16) in natural populations ($r^2 = 0.41$), compared to urban populations ($r^2 = 0.31$). This pattern suggests a process of fine-mapping, in which selective pressure targets a specific site within the intergenic region, and the gradual erosion of the surrounding ancestral linkage.

On chromosome 4, the SNPs 4:44545214:T:C, 4:44545284:T:C, and 4:44547067:A:G (SNPs M, N, and O in Fig. 16) display high to moderate r^2 values in both natural and urban populations. In contrast, the SNPs 4:44551913:A:G, 4:44557895:C:T, and 4:44558728:T:C (SNPs P, Q, and R in Fig. 16) exhibit higher LD in urban than in natural populations. Notably, the central SNP 4:44551913:A:G (SNP P in Fig. 16) functions as a molecular anchor, connecting both upstream and downstream sites, with r^2 values increasing from approximately 0.50 to 0.79 (upstream) and 0.83 (downstream).

This pattern suggests that urban selection is driving the integration of previously independent variants into a single, extended functional unit.

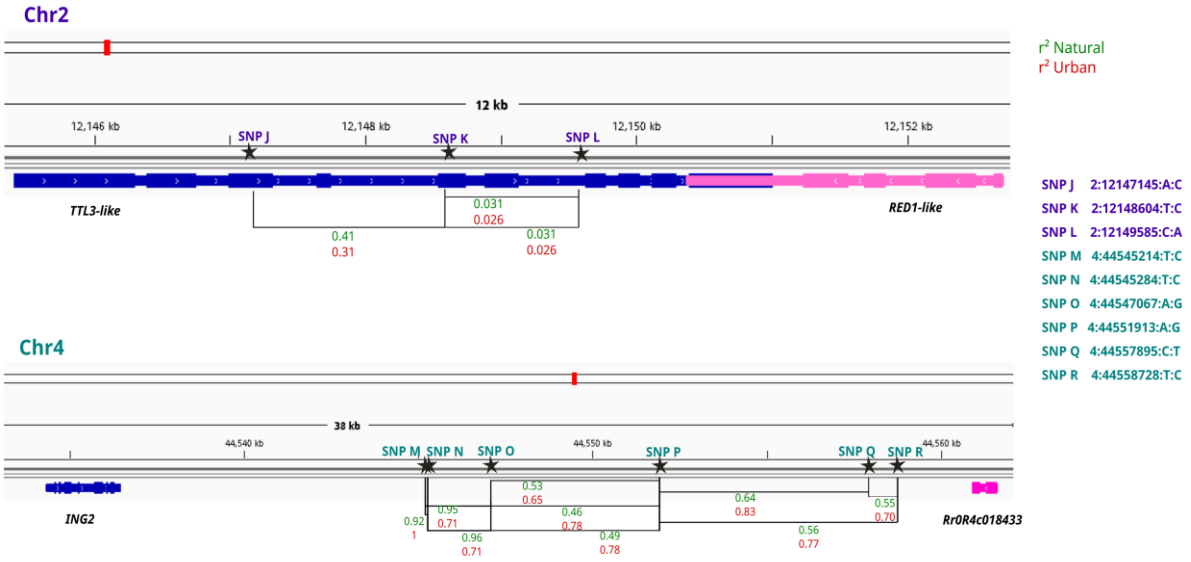

Figure 16. The genomic architecture and population-specific linkage disequilibrium (LD) patterns for two candidate loci: chromosome 2 *TTL-like* (positive strand) and chromosome 4 between ING2 (positive strand) and unannotated gene *RrOR4c018433* (negative strand). Blue and pink indicate positive and negative strand genes, respectively. Linkage disequilibrium (r^2) values between SNP pairs are presented for natural (green) and urban (red) populations.

The integrated selection scan identified a critical region on chromosome 9 containing 27 high-confidence SNPs under possible divergent selection between urban and natural populations. To characterize the evolutionary dynamics of this genomic hotspot, LD patterns were analyzed at both chromosome-wide and fine-scale resolutions, comparing haplotype structure between urban and natural populations.

At the whole-chromosome scale, LD patterns were broadly similar between urban and natural populations, reflecting their shared evolutionary history and ongoing gene flow. Genome-wide mean r^2 values were low in both groups (Natural: 0.1018; Urban: 0.1144). Strong LD ($r^2 > 0.8$) was observed in less than 1.11% of SNP pairs, and moderate LD ($r^2 > 0.5$) affected only 4.34–5.33% of comparisons. The rate of LD decay was extremely rapid, with r^2 falling below the 0.2 threshold within the first few kilobases (Fig. 1, Appendix). This rapid decay indicates that recombination has effectively reduced LD across most of the chromosome, providing a low-noise baseline that accentuates the significance of localized clusters. The genome-wide similarity further suggests that the populations have not undergone global differentiation.

In contrast, high-resolution analysis of the target region spanning from 42.66 to 42.69 Mb on chromosome 9 revealed distinct local haplotype architectures between natural and urban populations

(Fig. 17A-B). Notably, 41.38% of SNP pairs in the urban target region exhibited strong LD ($r^2 > 0.8$), compared to only 6.21% in natural populations. The observed high LD in urban populations likely represents the molecular signature of a recent and strong selective sweep. The 27 SNPs within this LD block may represent causal variants that directly contribute to urban adaptation or neutral variants that hitchhike with the selected haplotype. The clustering of selection signals (F_{ST}, XP-EHH, $\theta\pi$) across this region, together with the haplotype structure, designates this locus as a high-priority target for functional validation. Identifying the specific variants, coding, regulatory, or both, that confer a fitness advantage in urban environments will require complementary approaches, including association mapping in larger samples, experimental manipulation of candidate genes, and phenotypic characterization of haplotype classes.

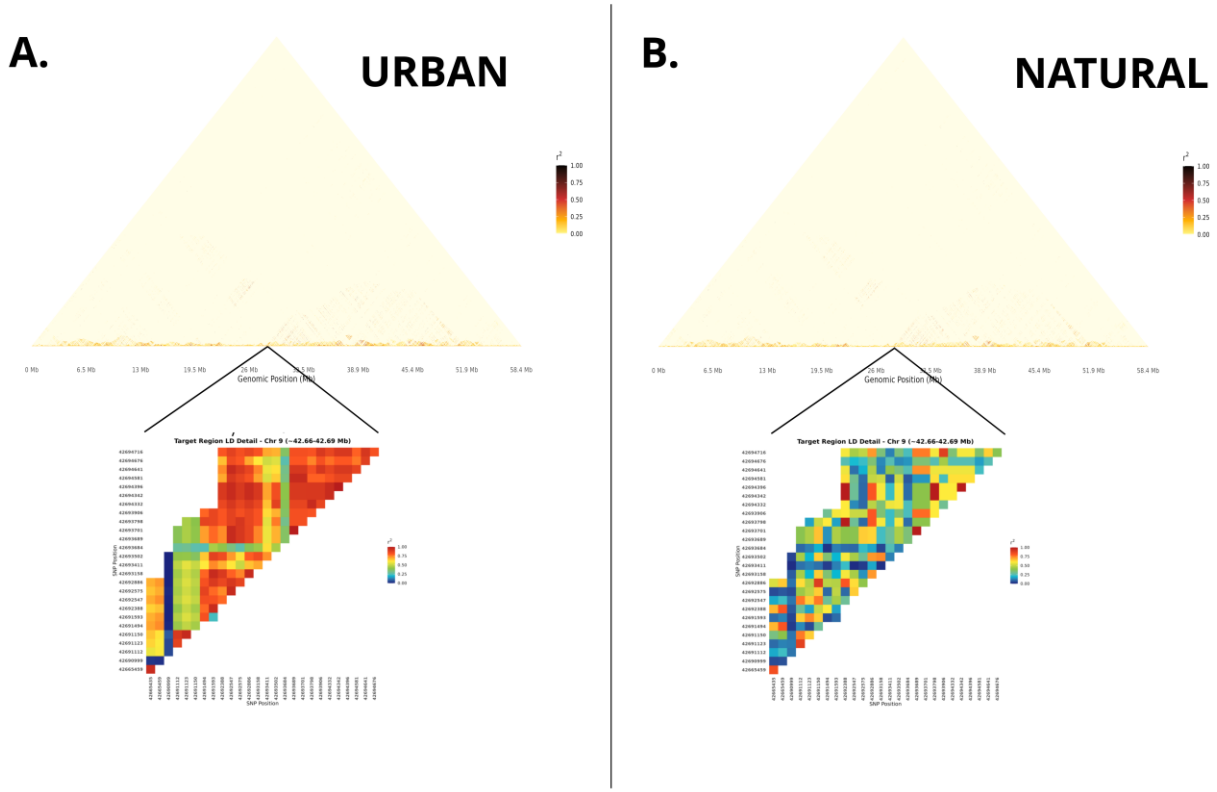

Figure 17. Comparison of LD patterns on Chromosome 9 in urban and natural populations. The figure illustrates the genomic architecture of LD in (A) urban and (B) natural populations of *R. raphanistrum*. In each panel, the top heatmap depicts the whole-chromosome LD structure, with SNPs aggregated into 100-SNP bins to represent mean r^2 values. The bottom heatmap offers a high-resolution view of LD blocks within the 42.66–42.69 Mb target region, as identified by the triple overlap selection scan.

4.3 Genome-wide association analysis

Among the phenotypic traits assessed in the common garden experiment, Leaf Area and Fruit Weight were selected for Genome-Wide Association Study (GWAS) analysis due to the robustness of the dataset, adequate statistical power for mixed-model implementation, and their biological relevance to urban adaptation hypotheses. Both traits demonstrated significant association peaks with well-distributed p-values, as confirmed by quantile-quantile plots, which indicate well-calibrated test statistics and effective control for population structure (Fig. 18, 19).

GWAS was performed on 55 common garden individuals for leaf area and 50 individuals for fruit weight, analyzing 4,571,587 SNPs using a univariate linear mixed model (LMM) implemented in *GEMMA*. The LMM accounts for population structure and cryptic relatedness by partitioning phenotypic variance into genetic (Vg, modeled through a kinship matrix) and residual environmental (Ve) components, while testing each SNP for marginal association with the phenotype. Both traits yielded nearly identical narrow-sense heritability estimates (Leaf Area: PVE = 56.07% ± 14.21%; Fruit Weight: PVE = 56.01% ± 16.73%), indicating that genetic factors explain more than half of the observed phenotypic variation and providing a robust foundation for identifying adaptive loci.

For the leaf area trait, the association analysis identified 55 genomic regions that exceeded the Bonferroni significance threshold (Fig. 18, Table 3, Appendix).

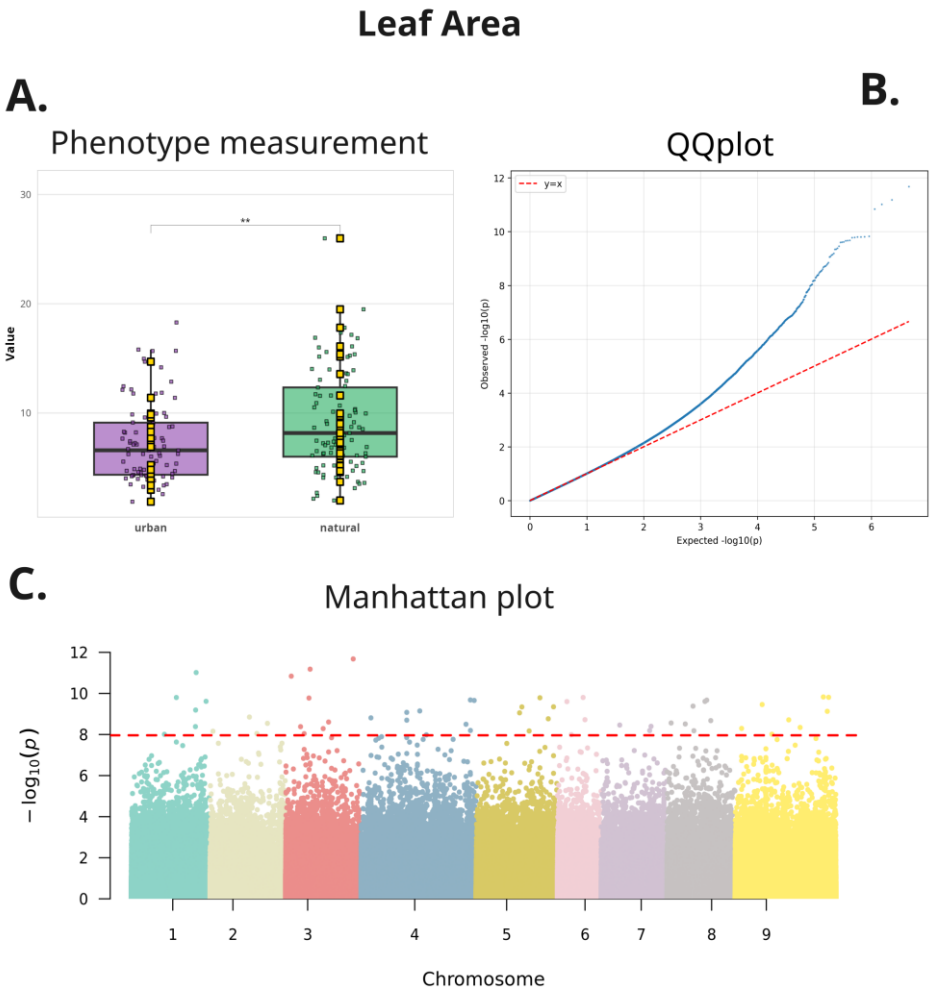

Figure 18. (A) Measurement of leaf area sampled across urban (purple) and natural (green) environments. Asterisks indicate statistical significance **p<0.01; (B) Quantile–quantile (Q-Q) plots generated from the *GEMMA* mixed model, illustrating the observed vs. expected p-value distributions and accounting for population structure; (C) Manhattan plot displaying the significance of SNP associations across the genome. The horizontal dashed line represents the Bonferroni-corrected threshold.

For the fruit weight trait, the association analysis identified 57 genomic regions that exceeded the Bonferroni significance threshold (Fig. 19, Table 4, Appendix).

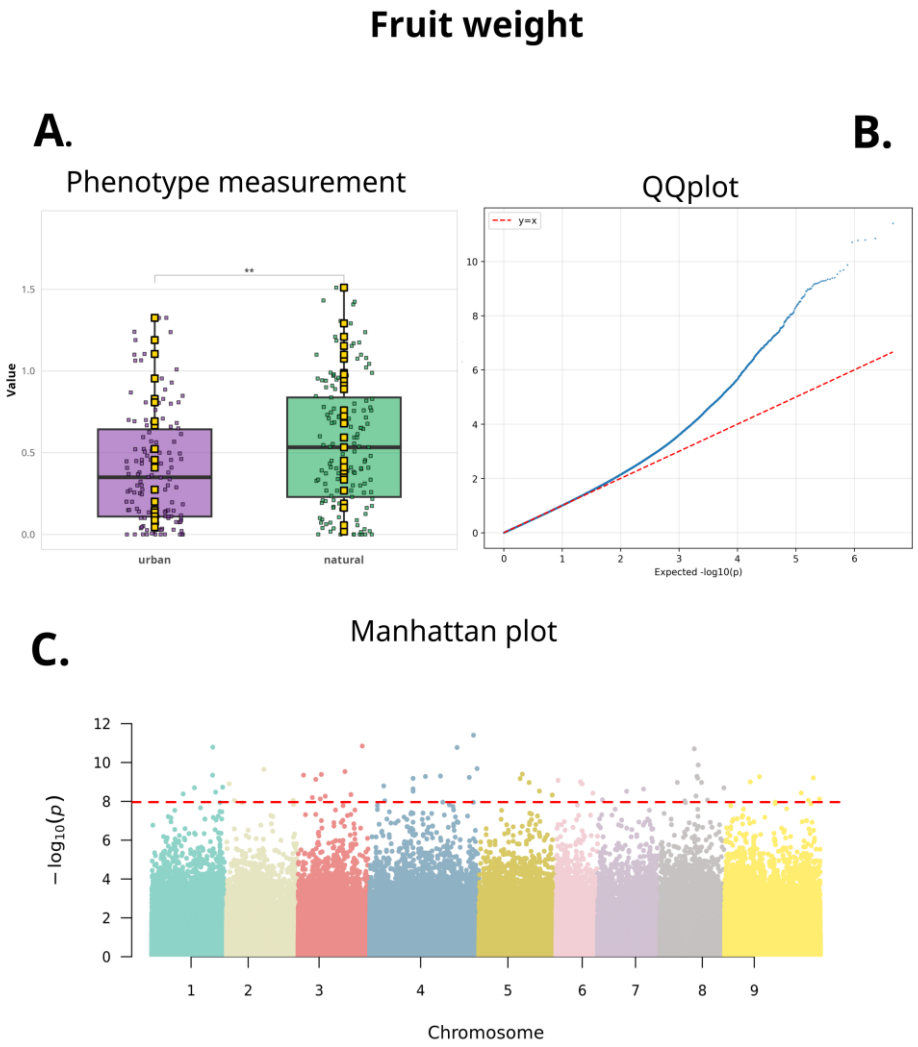

Figure 19. (A) Measurement of Fruit weight sampled across urban (purple) and natural (green) environments. Asterisks indicate statistical significance **p<0.01; (B) Quantile–quantile (Q-Q) plots generated from the *GEMMA* mixed model, illustrating the observed vs. expected p-value distributions and accounting for population structure; (C) Manhattan plot displaying the significance of SNP associations across the genome. The horizontal dashed line represents the Bonferroni-corrected threshold.

Among the candidate genes identified through GWAS for leaf area and fruit weight, four loci were particularly notable because they were associated with identical SNPs showing concordant effect directions for both traits, indicating pleiotropic effects. These loci (*PILS3*, *HSP70-BiP2*, *TORNADO2*, and the *cell wall protein GP1*) were detected with genome-wide significance in both GWAS analyses and are known for their established roles in growth regulation, stress responses, and direct mechanistic involvement in organ size determination. The observed pleiotropy suggests that urban

adaptation has proceeded via selection on core developmental and physiological pathways that simultaneously influence multiple morphological traits.

On chromosome 1, a genome-wide significant association was identified at SNP 1:26070237:G:A for both leaf area ($p = 1.58\times10^{-10}$, β = -13.23 cm²) and fruit weight ($p = 1.58\times10^{-9}$, β = -13.23 g), located 7.7 kb downstream of *PIN-LIKE 3* (*PILS3*). This SNP represents the strongest association in both GWAS analyses, with highly consistent effect sizes across traits. *PILS3* functions as an intracellular auxin transporter localized to the endoplasmic reticulum, where it regulates cytoplasmic auxin homeostasis and directly controls cell expansion and organ size (169). The PILS protein family modulates cytoplasmic auxin availability by sequestering or releasing auxin from endoplasmic reticulum compartments, thereby fine-tuning auxin-mediated growth responses. Auxin-driven cell expansion operates universally across all plant organs (leaves, stems, roots, and fruits). High cytoplasmic auxin level activates plasma membrane H^+-ATPases, leading to apoplastic acidification, expansin activation, and cell wall loosening, ultimately driving cell expansion. Because this pathway functions in all developing tissues, genetic variation in *PILS3* activity necessarily affects both vegetative (leaf) and reproductive (fruit/silique) organ sizes. Auxin sensitivity is temperature-dependent, and urban heat stress may select for altered *PILS3* activity to sustain optimal growth under elevated temperatures. The 7.7 kb downstream SNP may indicate the presence of a regulatory element or distal enhancer influencing *PILS3* expression. Urban environments impose distinct temperature regimes that affect auxin transport dynamics and cell expansion rates. The observed large negative effect for both traits suggest that urban-enriched alleles associate with coordinated size reduction across vegetative and reproductive structures, potentially representing an adaptive response to multiple urban stressors: reduced transpirational surface under heat and water stress (smaller leaves), and potentially earlier flowering or reduced reproductive investment per fruit (smaller fruits) as part of a fast life-history strategy.

On chromosome 5, highly significant associations were identified at SNP 5:30730277:G:A for both leaf area ($p = 6.67\times10^{-9}$, β = -9.06 cm²) and fruit weight ($p = 6.73\times10^{-9}$, β = -9.06 g), located approximately 2.0 kb upstream of a gene encoding heat shock protein 70 kDa BiP2. *HSP70* genes are essential for plant development and act as key mediators of responses to diverse abiotic stresses (170). Under environmental stress, specific transcription factors bind to *cis*-regulatory elements in the promoters of these genes to initiate targeted transcription (171). Previous research has demonstrated that HSP70s respond to various abiotic factors, including salt, drought, heat, and cold, as well as hormonal treatments such as ABA, IAA, GA3, and SA (172,173). In *R. sativus*, most

RsHSP70s exhibit tissue-specific expression profiles, particularly during early seedling development in 20-day-old leaves and roots (174). The *BiP2* isoform, a member of the HSP70 family, is localized to the endoplasmic reticulum and is significantly upregulated not only by abiotic stressors but also during biotic interactions with pathogens such as *Sclerotinia sclerotiorum* and *Plasmodiophora syringae*. Subcellular localization studies of the RsHSP70–23 protein in *Arabidopsis* protoplasts have confirmed its cytoplasmic presence. These findings indicate that while the core functions of the RsHSP70 family may be evolutionarily conserved between radish and *Arabidopsis*, these proteins may have also developed unique functional roles through changes in cellular localization. The pleiotropic effects of HSP70-BiP2 on both leaf area and fruit weight underscore its central role in safeguarding growth processes during thermal stress. Urban heat islands consistently elevate ambient temperatures by 2-3°C, thereby impacting the development of all plant organs. HSP70 preserves the functional integrity of growth-regulatory proteins, including cyclins, kinases, and transcription factors, across various tissues, thereby supporting continued development under heat stress. The presence of the associated SNP 2.0 kb upstream of the gene (within the putative promoter) suggests a *cis*-regulatory variant that may influence HSP70 expression levels, thermal responsiveness, or tissue specificity. The observed negative effect sizes indicate that alleles enriched in urban environments are associated with reduced organ sizes for both leaves and fruits. This pattern may represent an adaptive trade-off, in which smaller organs decrease heat load and resource requirements, while increased HSP70 expression maintains essential cellular functions during prolonged thermal stress.

Two genome-wide significant SNPs on chromosome 4, 4:26461451:T:A (leaf area: $p = 2.00\times10^{-8}$, β = -13.20 cm²; fruit weight: $p = 2.01\times10^{-9}$, β = -13.21 g) and 4:26460722:A:T (leaf area: $p = 8.37\times10^{-9}$, β = -13.14 cm²; fruit weight: $p = 2.93\times10^{-8}$, β = -6.66 g), are located approximately 2.1 kb and 1.4 kb from the *TORNADO 2-like* (*TRN2*) gene, respectively. In *A. thaliana*, the *TRN1* and *TRN2* genes play roles in leaf patterning, including lamina venation, symmetry, and lateral growth (175). These genes may act early in leaf primordium development, influencing how cells interpret positional information. In urban plants, loss of *TRN2* may result in smaller leaves due to defective early patterning and impaired tissue polarity, which adaptively reduces the cell proliferation window in developing leaves. This represents a proactive, adaptive strategy: producing smaller organs rather than building large ones, and subsequently responding to stress. The negative effect size suggests that urban alleles shorten the proliferation phase, resulting in smaller but potentially more stress-tolerant leaves. No pairwise LD estimate was obtained between the two SNPs, likely due to their large physical separation (~7.5 Mb), which exceeds the LD window applied in the analysis.

4.4 Variant calling on RNA-seq data

4.4.1. Population structure

The PCA based on RNA variant calling of 1,315,815 variants reveals distinct patterns of genetic variation among the samples. PC3 and PC4 explain 9.64% and 7.95% of the total variance, respectively (Fig. 20).

Natural populations are located on the left side of the plot, whereas Urban and Peri-urban populations are primarily distributed in the central and right regions, indicating genetic differentiation. In terms of substrate type, limestone (pink) samples exhibit high dispersion, which suggests greater intra-group genetic heterogeneity; these samples include both Natural and Peri-urban populations. In contrast, lava (purple), pyroclast (blue), and tuff (dark brown) samples occupy intermediate positions and form compact clusters, indicating more homogeneous expression profiles.

This pattern differs from the population structure observed in the genomic analysis; however, a separation between Natural and Urban or Peri-urban populations remains evident.

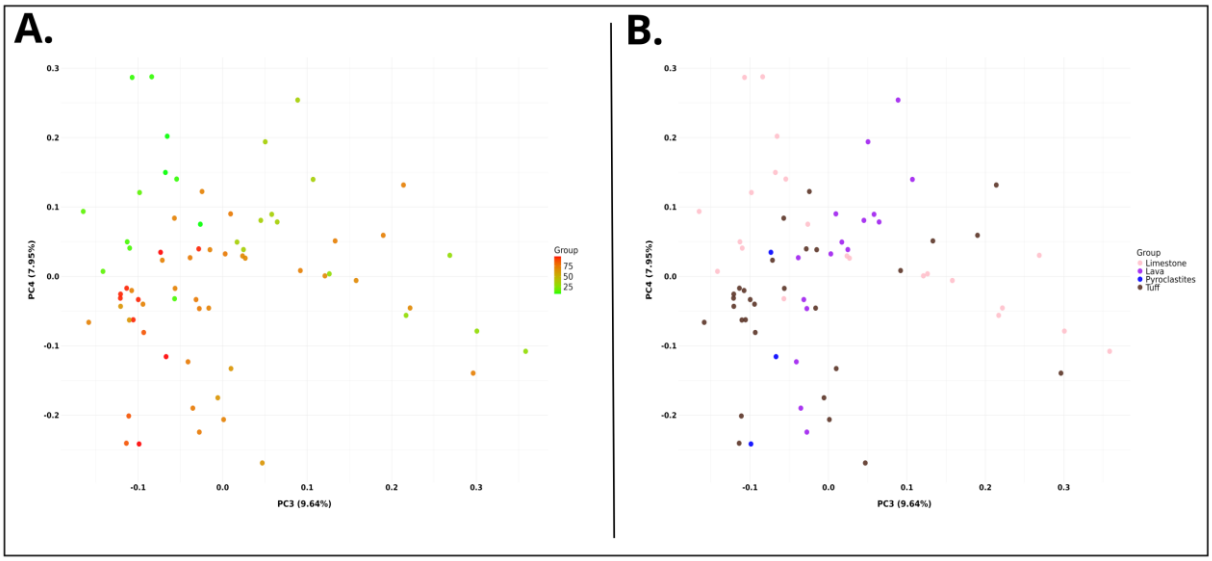

Figure 20. PCA on RNA-seq data. (A) Individuals are colored according to the percentage of impervious surface at each sampling site (proxy for urbanization intensity); (B) individuals are colored according to substrate type (lava, limestone, pyroclast, and tuff). PC3 and PC4 explain 9.64% and 7.95% of the total variance, respectively.

ADMIXTURE analysis of RNA-seq data revealed a largely homogeneous genetic structure across both urbanization gradients and substrate types, with only minor differences observed (Fig. 21).

At K=2, no single ancestral component dominated; however, the k2_2 component (green) was more prevalent in Urban populations (mean value 0.574), while the alternative component (blue) was

more common in Peri-urban populations (0.612) and Natural populations (0.771). Terracina was almost entirely assigned to the blue component.

At K=3, Urban populations exhibited elevated K3_3 ancestry (orange), particularly in Soccavo (0.709) and MonteSA (0.738). Peri-urban populations displayed intermediate ancestry proportions across all three components, with Terracina remaining an outlier, and Natural populations were predominantly K3_2 (blue). Although there was a small change in ancestral groups across soil types, the overall results suggest that differences in the land have not yet led to major separation between groups.

The population structure inferred from RNA variants does not differ from that observed in genomic DNA data.

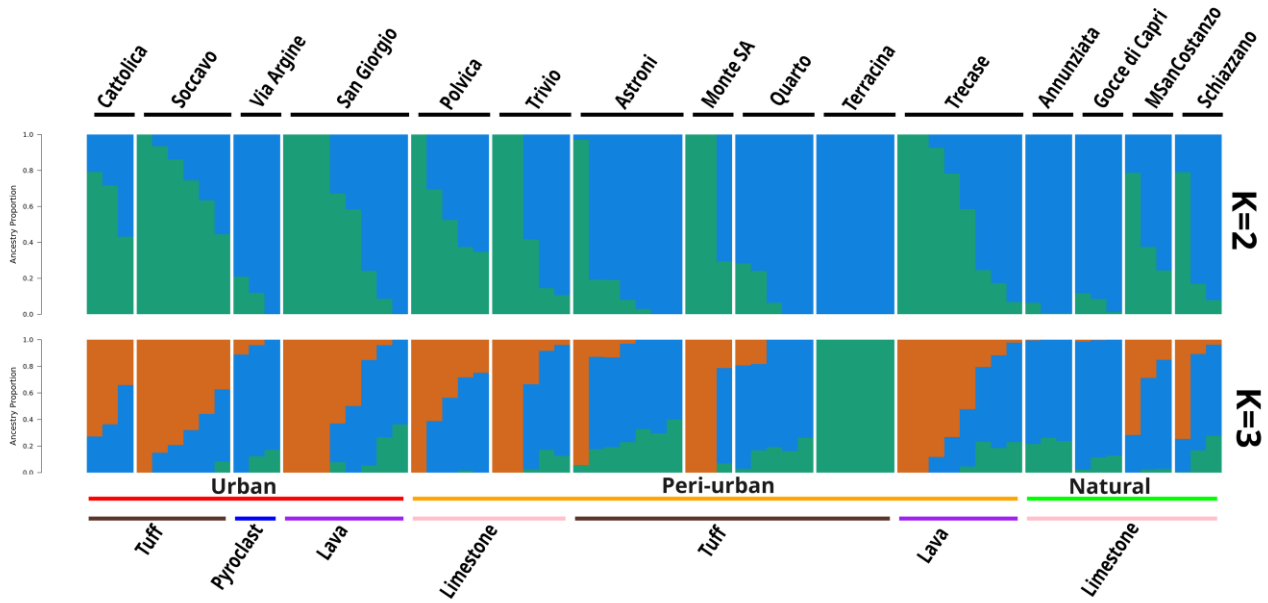

Figure 21. Admixture analysis on RNA-seq data. Each vertical bar in the figure represents an individual, with colors indicating the proportions of ancestry for the inferred genetic clusters at K=2 (upper panel) and K=3 (lower panel). Individuals are grouped by population along the x-axis, and colored bars below denote population identity (black), urbanization class (red: Urban; orange: Peri-urban; green: Natural), and soil type (dark-brown: Tuff; violet: Lava; pink: Limestone; blue: Pyroclast).

F_{ST} values obtained from RNA-seq were low, revealing absence of population structure both based on impervious surface and soil composition (Table 5-6), in agreement with the F_{ST} values obtained from DNA-seq data.

Table 5. Hudson F_{ST} on RNA-data of pairwise comparison between population based on urban classification.

POP 1	POP 2	HUDSON_F_{ST}
Natural	Peri-urban	0.0452657
Natural	Urban	0.0667226
Peri-urban	Urban	0.0363978

Table 6. Hudson F_{ST} on RNA-data of pairwise comparison between population based on soil type.

POP 1	POP 2	HUDSON_F_{ST}
lava	limestone	0.0366418
lava	pyroclast	0.164769
lava	tuff	0.0406721
limestone	pyroclast	0.116538
limestone	tuff	0.0261575
pyroclast	tuff	0.113062

4.4.2. Comparison of F_{ST} derived from DNA and RNA Data

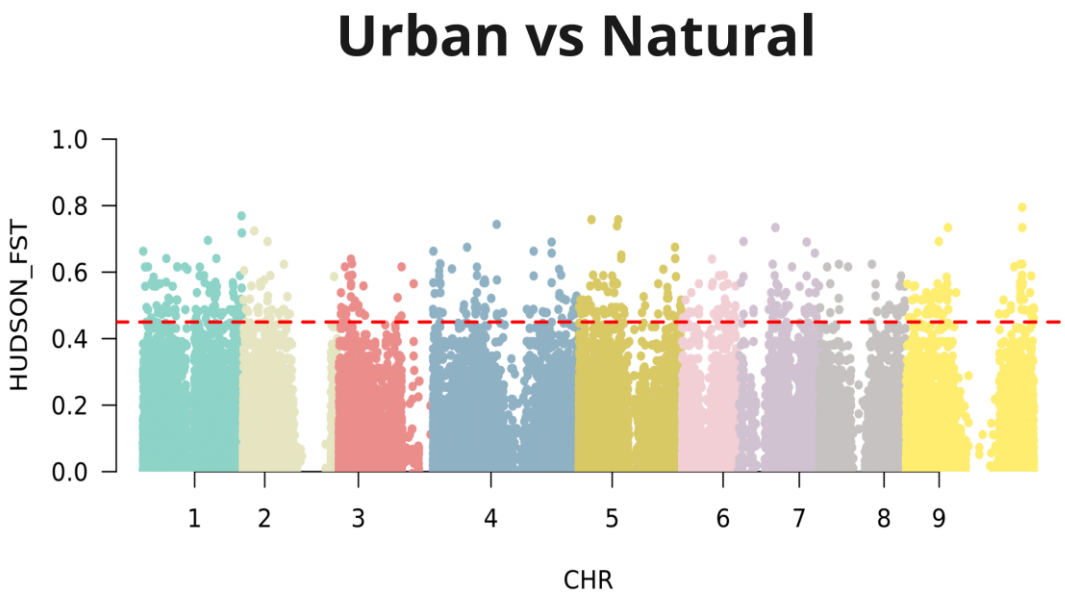

Figure 22. Manhattan plot illustrating genome-wide F_{ST} values in a comparison of natural and urban populations. SNPs with F_{ST} values exceeding 0.45 are shown above the dashed red line.

A pronounced disparity was observed in the number of highly differentiated variants: approximately 500 SNPs exhibited F_{ST} values greater than 0.45 in RNA variant data (Fig. 22), whereas only 17 such SNPs were identified in genomic DNA analysis.

No SNPs with F_{ST} values greater than 0.45 in RNA data overlapped with those meeting the same threshold in genomic DNA analysis. However, seven SNPs with F_{ST} values exceeding 0.1 in genomic DNA were also present in the RNA high-F_{ST} set (Table 7). These seven SNPs are distributed across five chromosomes (1, 4, 7, 8, and 9), with a notable cluster of three SNPs on chromosome 4. All seven shared SNPs display 1.5- to 3-fold higher F_{ST} values in RNA data (range: 0.471–0.565) compared to their genomic DNA counterparts (range: 0.183–0.332).

The seven shared SNPs exhibiting elevated differentiation in both datasets ($F_{ST} > 0.1$ in DNA, $F_{ST} > 0.45$ in RNA) likely represent functionally significant variants subjected to selective pressure associated with local adaptation. These loci may harbor coding or regulatory variants that directly affect gene expression levels or protein function, rendering them targets of environmental selection. The markedly greater differentiation observed in these RNA variants ($F_{ST} > 0.45$) compared to their genomic signatures ($F_{ST} > 0.1$) could be mainly related to the differences in the number of samples between the two data-sets (less samples in the RNA-seq data).

Table 7. Common significant SNPs (F_{ST} > 0.45 in RNA data; F_{ST} > 0.1 in DNA data) identified in the comparison between urban and natural environments.

SNP ID (chr:pos)	HUDSON_F_{ST} RNA-data	HUDSON_F_{ST} DNA-data	Gene_ID	SNP position & Type	Annotation
1:29458346:C:G	0.478261	0.258824	Rr0R1c003507	Exon 3 missense L → V	PREDICTED: *Raphanus sativus* protochlorophyllide-dependent translocon component 52, chloroplastic (LOC108839495), mRNA
4:206564:C:G	0.474189	0.313208	Rr0R4c014162	Exon 2 missense A → G	PREDICTED: *Raphanus sativus* zinc finger protein GIS2 (LOC108854723), transcript variant X1, mRNA
4:19973747:A:G	0.481164	0.235364	Rr0R4c016501	Exon 6 synonymous (L)	PREDICTED: *Raphanus sativus* G-type lectin S-receptor-like serine/threonine-protein kinase At5g24080 (LOC108854093), transcript variant X2, mRNA
7:20103004:G:A	0.474189	0.302701	Rr0R7c031253	Exon 6 synonymous (E)	PREDICTED: *Raphanus sativus* guanylate kinase 2 (LOC108814608), mRNA
8:35238655:A:G	0.488118	0.18381	Rr0R8c036598	Exon 9 synonymous (A)	Transmembrane protein 214
8:37251579:T:C	0.565217	0.310289	Rr0R8c036958	Exon 1 missense N → D	PREDICTED: *Raphanus sativus* uncharacterized LOC108822414 (LOC108822414), ncRNA
9:16316254:G:T	0.474189	0.267542	Rr0R9c039583	Exon 2 synonymous (V)	SH3 domain-containing protein YSC84-like

Functional annotation of the seven shared SNPs exhibiting elevated differentiation in both genomic DNA and RNA variants revealed diverse roles in cellular signaling, metabolism, and organellar function.

On chromosome 1, the SNP 1:29458346:C:G is located within exon 3 of the *protochlorophyllide-dependent translocon component 52* (*PTC52*) gene, which is involved in chloroplast protein import and is essential for the assembly of photosynthetic machinery and chloroplast biogenesis. This gene is part of a translocon complex most abundantly expressed in etiolated plants and is responsible for the protochlorophyllide-dependent import of the precursor NADPH:protochlorophyllide oxidoreductase A (pPORA). Additionally, *PTC52* has been characterized as a redox-sensitive component that modulates TOC channel activity in response to the plant's metabolic and environmental condition (176). The SNP is a missense mutation (L/V) and may represent a structural adaptation that optimizes chloroplast biogenesis in response to the specific stressors of urban heat islands and altered light regimes.

On chromosome 4, two highly differentiated SNPs were identified: 4:206564:C:G and 4:19973747:A:G, both located in genes involved in stress tolerance and developmental regulation.

The first SNP is located in exon 2 of two transcripts annotated as the Zinc finger protein GIS2 (*GIS2*) gene and in both cases is a missense mutation (A/G) which may alter the protein's functional properties. *GIS2* is a well-characterized transcriptional regulator that integrates hormonal signaling, such as gibberellins and cytokinins, with environmental cues. It plays a primary role in trichome development, an adaptive trait that helps manage transpiration and thermal stress. Furthermore, *GIS2* has been implicated in the regulation of specialized metabolic pathways: it may positively regulate Squalene Epoxidase 5 (*SQE5*), linking sterol biosynthesis to drought tolerance, while simultaneously negatively influencing Senescence-Associated Gene 1 (*SEN1*), thereby delaying stress-induced leaf senescence via ABA signaling pathways (177) The second SNP is situated within exon 6 of the G-type lectin S-receptor-like serine/threonine-protein kinase (B120) and it is a synonymous mutation. Research by Sun et al. (178) demonstrated that B120 is a critical component in the plant's response to abiotic stress, specifically conferring salt stress tolerance across various developmental stages. Receptor-like kinases (RLKs) such as B120 function as environmental sensors, initiating signaling cascades that enable the plant to adjust its physiology to osmotic and ionic imbalances, which are often intensified in urban soils due to high salinity and pollution. The identification of these variants suggests a coordinated adaptive strategy in urban environments. The variation in *B120* indicates optimization of early-warning stress signaling, while the polymorphism in *GIS2* suggests structural and physiological reorganization. By potentially modulating trichome density and delaying senescence, urban populations may sustain photosynthetic activity for longer periods under the drought and heat conditions characteristic of urban landscapes.

On chromosome 7, the SNP 7:20103004:G:A is located within *exon 6* of Guanylate Kinase 2 (*GK2*) an enzyme essential for nucleotide metabolism and energy homeostasis and the SNP is a synonimous mutation. *GK2* catalyzes the conversion of GMP to GDP (179). The catalytic activity of this enzyme is highly conserved across eukaryotes. While plants possess two distinct types of GKs essential for growth and plastid development (180) both yeast and humans rely on a single, essential GK for GDP biosynthesis (181,182). In plants, *GK2* is critical for maintaining nucleotide homeostasis, which supports energy-intensive processes such as DNA repair, protein synthesis, and chloroplast biogenesis. This activity is particularly important during environmental stress, when plants must rapidly reallocate metabolic resources to sustain energetic efficiency. The presence of a highly differentiated variant in urban populations indicates potential adaptive fine-tuning of GTP production. Optimizing these metabolic flux rates may enable urban plants to better support rapid

signaling and physiological adjustments necessary for survival under chronic abiotic stressors, including the urban heat island effect and pollution.

On chromosome 8, two highly differentiated SNPs were identified: 8:35238655:A:G, a synonymous mutation located within exon 9 of Transmembrane Protein 214 (*TMEM214*), and 8:37251579:T:C, a missense mutation (N/D), situated in exon 1 of an uncharacterized gene. Transmembrane proteins are integral components of the cell membrane, facilitating communication and molecular exchange between the cytoplasm and the external environment. In plants, these proteins serve as the primary interface for environmental sensing and are essential for detecting external stimuli, such as chemical pollutants or pathogen-associated molecular patterns (PAMPs). Specifically, TMEM214 has been implicated in various biological processes, including mRNA degradation and viral replication pathways (183). TMEM214 is localized on the endoplasmic reticulum (ER) membrane, where it mediates ER stress-induced apoptosis. Li et al. (184) report that TMEM214 acts as a molecular anchor, recruiting procaspase-4 (or its functional plant analogs) to the ER, thereby facilitating activation and the subsequent programmed cell death pathway. Urban environments expose plants to chronic abiotic stressors, such as heat and pollutants, which result in the accumulation of misfolded proteins and trigger ER stress. Optimization of TMEM214-mediated pathways may enable urban plants to manage cellular damage more efficiently, potentially by modulating the threshold for autophagy or apoptosis.

On chromosome 9, SNP 9:16316254:G:T is situated in exon 2 of a gene identified by InterPro Scan as encoding a SH3 domain-containing protein YSC84-like. The SH3 domain represents a conserved structural motif that mediates protein-protein interactions essential for cytoskeletal organization, fluid-phase endocytosis, and vesicle trafficking. As a plant homologue of the yeast protein Ysc84, this protein family plays a crucial role in the early stages of endocytosis (185). Ysc84/SH3YL1-like proteins interact with actin-regulatory components to facilitate actin cytoskeleton remodeling, a process necessary for membrane vesicle internalization. Recent RNA-sequencing studies in oilseed sunflower have identified *SH3YL1* as a key differentially expressed gene under salinity stress (186). In urban environments, where soil salinity and osmotic stress are common, the presence of a differentiated SNP at this locus indicates that urban populations may have adapted their endocytic pathways to more effectively modulate membrane composition and signal transduction, thereby enhancing survival under variable osmotic conditions. Furthermore, SH3-domain proteins are critical regulators of effector-triggered immunity (ETI) (187). For instance, the rice homologue *SH3P2* interacts with the *Pib* gene, which mediates resistance to the rice blast fungus,

acting as a molecular switch. This interaction can suppress or modulate immune responses to prevent an over-activation of the defense system, which would otherwise lead to significant growth penalties. The identification of this variant suggests a sophisticated adaptive strategy in urban plants, balancing the metabolic costs of immunity with the physiological demands of a high-stress environment.

5. Conclusions and Future Perspectives

This study offers insights into the genetic mechanisms underlying plant adaptation to urbanization in *R. raphanistrum*. By integrating RNA sequencing, whole-genome sequencing, and genome-wide association analyses, the research identifies multiple adaptive signatures linked to urban environments, thereby elucidating the molecular basis of contemporary evolution in anthropogenic landscapes.

In silico gene expression analysis demonstrated significant gene expression reprogramming in urban populations, with flavonoid biosynthesis identified as a central pathway in urban adaptation. The coordinated upregulation of key genes, including *PAL2*, *CHS3*, *CHS-like*, and *F3GT*, suggests that enhanced antioxidant and UV-protective mechanisms constitute primary adaptive responses to urban stressors such as increased UV radiation, pollution-induced oxidative stress, and altered light regimes beneath urban canopies.

Population genomic analyses revealed moderate genetic differentiation between urban and natural populations, with localized signals of strong selection at specific loci. Several genes were identified as key targets of urban adaptation, particularly those involved in flowering time regulation, development, and stress response. A distinct signature of recent positive selection was observed in a large cluster on chromosome 9, where a 30 kb region containing 27 SNPs exhibited convergent signals across all selection statistics and elevated linkage disequilibrium in urban populations relative to natural populations. In contrast, chromosomes 2 and 4 displayed heterogeneous linkage disequilibrium patterns, potentially reflecting balancing selection or localized recombination variation, which underscores the complexity of adaptive genomic architecture.

Genome-wide association studies of morphological traits measured under common-garden conditions identified candidate pleiotropic loci influencing both leaf area and fruit weight. Key candidate genes include *PILS3* (auxin transporter), *HSP70-BiP2* (heat shock protein), and *TORNADO2* (leaf patterning), each exhibiting coordinated effects on vegetative and reproductive organ size. The negative effect sizes of urban-enriched alleles indicate that adaptive size reduction may enhance survival under heat and water limitations typical of urban heat islands, potentially

reflecting a shift toward a rapid life-history strategy that favors early reproduction over vegetative growth.

Population structure analysis of RNA variants confirmed intra-population genetic variability, with seven SNPs showing elevated differentiation in both genomic and transcriptomic datasets. These variants affect genes involved in chloroplast biogenesis, stress signaling, and cellular homeostasis, representing high-confidence adaptive loci where genetic and transcriptomic signals converge.

Although this study establishes a robust foundation for understanding urban adaptation in *R. raphanistrum*, several avenues for future research could further strengthen and extend these findings. Functional validation experiments are necessary to confirm the causal roles of differentially expressed genes and genomic variants. Approaches such as gene silencing, overexpression studies, or CRISPR-mediated knockout of key candidates would directly test their contributions to urban stress tolerance and clarify whether observed transcriptional changes reflect plastic responses or genetically fixed differences. Fine-mapping and haplotype analysis would help identify causal variants within candidate regions. Sequencing larger population samples would enhance statistical power to distinguish between linked neutral variants and functional polymorphisms, especially in regions with complex linkage disequilibrium patterns. Temporal sampling across multiple generations could monitor allele-frequency dynamics at candidate loci, providing direct estimates of selection coefficients and validating the adaptive significance of identified variants.

The current GWAS framework would benefit from larger sample sizes and expanded phenotypic characterization. Preliminary data on herbivory resistance are under analysis and may yield valuable insights into loci underlying biotic interactions in urban environments. Expanding the common-garden experiment to include additional maternal lines and populations would improve detection of small-effect loci and enable more robust genotype-phenotype associations.

Reciprocal transplant experiments are being established to complement the common garden approach. These experiments will transplant individuals from natural populations into urban environments and, conversely, move urban-adapted individuals into natural settings, enabling direct measurement of fitness consequences under field conditions. This approach is essential for distinguishing adaptive genetic differentiation from phenotypic plasticity and for quantifying the selective advantage conferred by urban-associated genotypes within their respective environments, to the growing body of evidence demonstrating that urbanization functions as a significant evolutionary force capable of driving rapid adaptive changes over contemporary timescales. The identification of specific adaptive markers and pathways establishes a foundation for comparative

analyses across multiple urban centers, enabling tests of parallel evolution and the predictability of urban adaptation.

R. raphanistrum serves as a valuable model system for investigating rapid adaptation to human-modified environments, demonstrating the utility of integrative multi-omics approaches in dissecting the genetic architecture of contemporary evolution. The molecular signatures identified in this study offer testable hypotheses regarding the mechanisms by which organisms cope with anthropogenic environmental change.

From a conservation perspective, understanding the genetic basis of urban tolerance has practical applications. Identifying adaptive loci and pathways can inform strategies for predicting which species are likely to persist in increasingly urbanized landscapes and may guide assisted adaptation efforts for vulnerable taxa. Furthermore, the methodological framework established in this study, which integrates population genomics and transcriptomics analyses, provides a transferable approach for studying adaptive potential in other species facing rapid environmental change.

Finally, to address the limitations of single-reference genome approaches, a *Raphanus* genus-level pangenome has been constructed using genomic data from Zhang et al. (84) (Fig. 23). This pan-genome will serve as a reference for a novel alignment pipeline developed by the team of Prof. Vincenza Colonna and prof. Erik Garrison, enabling the mapping of short-read DNA sequences against the complete genomic diversity of the genus. Constructing a pan-genome graph comprising sequences from multiple individuals will facilitate high-resolution analysis of structural variations (SVs) and segmental duplications (SDs), thereby uncovering their roles in genomic plasticity and adaptation. SVs and SDs, which are large genomic alterations, play critical roles in shaping genetic diversity and influencing gene expression, potentially driving adaptation to diverse environments.

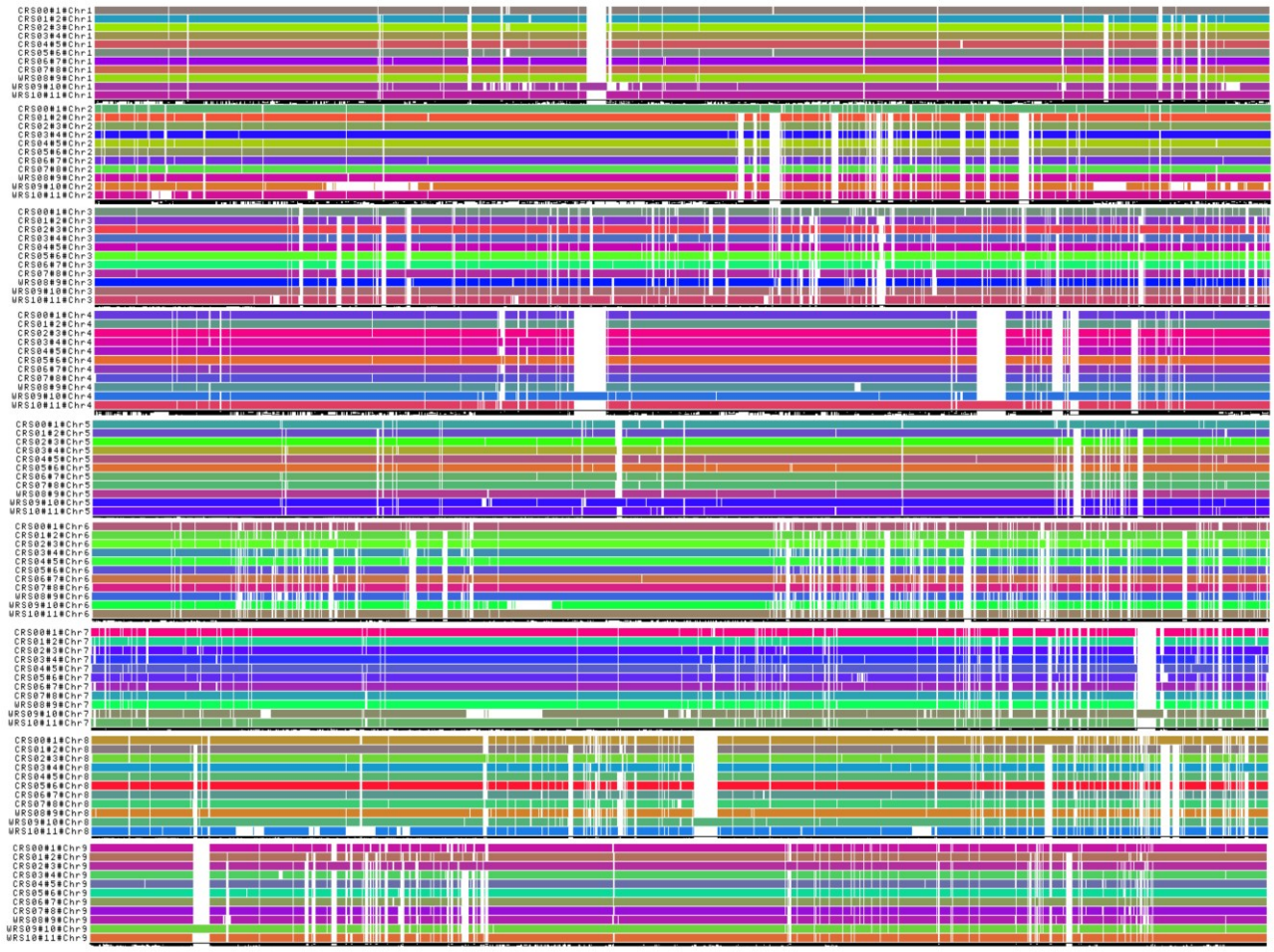

Figure 23. *Raphanus* genus pan-genome graph. Structural genomic comparison between wild (W) and domesticated (C) accessions at the scale of the nine chromosomes.

REFERENCES

1. World population prospects 2019: highlights. New York: United Nations; 2019.

2. Odling-Smee J, Erwin DH, Palkovacs EP, Feldman MW, Laland KN. Niche construction theory: a practical guide for ecologists. Q Rev Biol. 2013;88(1):3–28.

3. Boivin NL, Zeder MA, Fuller DQ, Crowther A, Larson G, Erlandson JM, et al. Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. Proc Natl Acad Sci. 2016;113(23):6388–96.

4. Palumbi SR. Humans as the world's greatest evolutionary force. Science. 2001;293(5536):1786–90.

5. Palkovacs EP, Post DM. Experimental evidence that phenotypic divergence in predators drives community divergence in prey. Ecology. 2009;90(2):300–5.

6. Start D, Barbour MA, Bonner C. Urbanization reshapes a food web. Rodriguez-Cabal M, editor. J Anim Ecol. 2020 Mar;89(3):808–16.

7. Ellis EC. Ecology in an anthropogenic biosphere. Ecol Monogr. 2015;85(3):287–331.

8. Desa U. United Nations Department of economic and social affairs. Popul Div World Popul Prospects. 2019;

9. Palkovacs EP, Kinnison MT, Correa C, Dalton CM, Hendry AP. Fates beyond traits: Ecological consequences of human-induced trait change. Evol Appl. 2012;5(2):183–91.

10. Alberti M, Palkovacs EP, Roches SD, Meester LD, Brans KI, Govaert L, et al. The complexity of urban eco-evolutionary dynamics. BioScience. 2020;70(9):772–93.

11. Ditchkoff SS, Saalfeld ST, Gibson CJ. Animal behavior in urban ecosystems: modifications due to human-induced stress. Urban Ecosyst. 2006;9(1):5–12.

12. van de Leemput IA, Dakos V, Scheffer M, van Nes EH. Slow recovery from local disturbances as an indicator for loss of ecosystem resilience. Ecosystems. 2018;21(1):141–52.

13. Alberti M, Correa C, Marzluff JM, Hendry AP, Palkovacs EP, Gotanda KM, et al. Global urban signatures of phenotypic change in animal and plant populations. Proc Natl Acad Sci. 2017;114(34):8951–6.

14. Gregg JW, Jones CG, Dawson TE. Urbanization effects on tree growth in the vicinity of New York City. Nature. 2003;424(6945):183–7.

15. Neil K, Wu J. Effects of urbanization on plant flowering phenology: a review. Urban Ecosyst. 2006;9(3):243–57.

16. Grimm NB, Faeth SH, Golubiewski NE, Redman CL, Wu J, Bai X, et al. Global change and the ecology of cities. science. 2008;319(5864):756–60.

17. Alberti M. Eco-evolutionary dynamics in an urbanizing planet. Trends Ecol Evol. 2015 Feb;30(2):114–26.

18. Johnson MT, Thompson KA, Saini HS. Plant evolution in the urban jungle. Am J Bot. 2015;102(12):1951–3.

19. Reid NM, Proestou DA, Clark BW, Warren WC, Colbourne JK, Shaw JR, et al. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science. 2016;354(6317):1305–8.

20. Johnson MT, Munshi-South J. Evolution of life in urban environments. Science. 2017;358(6363):eaam8327.

21. Santangelo JS, Rivkin LR, Johnson MT. The evolution of city life. Vol. 285, Proceedings of the Royal Society B. The Royal Society; 2018. p. 20181529.

22. Kareiva P, Watts S, McDonald R, Boucher T. Domesticated nature: shaping landscapes and ecosystems for human welfare. Science. 2007;316(5833):1866–9.

23. Hof AE van't, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534(7605):102–5.

24. Wirgin I, Roy NK, Loftus M, Chambers RC, Franks DG, Hahn ME. Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson River. Science. 2011;331(6022):1322–5.

25. Munshi-South J, Zolnik CP, Harris SE. Population genomics of the Anthropocene: Urbanization is negatively associated with genome-wide variation in white-footed mouse populations. Evol Appl. 2016;9(4):546–64.

26. Müller JC, Partecke J, Hatchwell BJ, Gaston KJ, Evans KL. Candidate gene polymorphisms for behavioural adaptations during urbanization in blackbirds. Mol Ecol. 2013;22(13):3629–37.

27. Miles LS, Rivkin LR, Johnson MT, Munshi-South J, Verrelli BC. Gene flow and genetic drift in urban environments. Mol Ecol. 2019;28(18):4138–51.

28. Rivkin LR, Santangelo JS, Alberti M, Aronson MF, de Keyzer CW, Diamond SE, et al. A roadmap for urban evolutionary ecology. Evol Appl. 2019;12(3):384–98.

29. Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. Landscape genetics: where are we now? Mol Ecol. 2010;19(17):3496–514.

30. Holderegger R, Di Giulio M. The genetic effects of roads: a review of empirical evidence. Basic Appl Ecol. 2010;11(6):522–31.

31. Brady SP. Road to evolution? Local adaptation to road adjacency in an amphibian (Ambystoma maculatum). Sci Rep. 2012;2(1):235.

32. Cheptou PO, Carrue O, Rouifed S, Cantarel A. Rapid evolution of seed dispersal in an urban environment in the weed Crepis sancta. Proc Natl Acad Sci. 2008;105(10):3796–9.

33. Leimu R, Mutikainen P, Koricheva J, Fischer M. How general are positive relationships between plant population size, fitness and genetic variation? J Ecol. 2006;94(5):942–52.

34. Bullock JM, White SM, Prudhomme C, Tansey C, Perea R, Hooftman DA. Modelling spread of British wind-dispersed plants under future wind speeds in a changing climate. J Ecol. 2012;100(1):104–15.

35. Santini L, Cornulier T, Bullock JM, Palmer SC, White SM, Hodgson JA, et al. A trait-based approach for predicting species responses to environmental change from sparse data: how well might terrestrial mammals track climate change? Glob Change Biol. 2016;22(7):2415–24.

36. Crispo E, Moore JS, Lee-Yaw JA, Gray SM, Haller BC. Broken barriers: Human-induced changes to gene flow and introgression in animals: An examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. BioEssays. 2011;33(7):508–18.

37. Bullock JM, Bonte D, Pufal G, da Silva Carvalho C, Chapman DS, García C, et al. Human-mediated dispersal and the rewiring of spatial networks. Trends Ecol Evol. 2018;33(12):958–70.

38. Wright S. The shifting balance theory and macroevolution. Annu Rev Genet. 1982;16(1):1–20.

39. Krtinić B, Ludoški J, Milankov V. Study on siphonal measurements and usefulness in delimitation of" rural" and" urban" ecotypes of Culex pipiens. 2012;

40. Schapira A, Boutsika K. Malaria ecotypes and stratification. Adv Parasitol. 2012;78:97–167.

41. Yakub M, Tiffin P. Living in the city: urban environments shape the evolution of a native annual plant. Glob Change Biol. 2017 May;23(5):2082–9.

42. Harris SE, Munshi-South J. Signatures of positive selection and local adaptation to urbanization in white-footed mice (Peromyscus leucopus). Mol Ecol. 2017;26(22):6336–50.

43. Williams NS, Hahs AK, Vesk PA. Urbanisation, plant traits and the composition of urban floras. Perspect Plant Ecol Evol Syst. 2015;17(1):78–86.

44. Stapley J, Reger J, Feulner PG, Smadja C, Galindo J, Ekblom R, et al. Adaptation genomics: the next generation. Trends Ecol Evol. 2010;25(12):705–12.

45. Barrett RD, Schluter D. Adaptation from standing genetic variation. Trends Ecol Evol. 2008;23(1):38–44.

46. Kawecki TJ, Ebert D. Conceptual issues in local adaptation. Ecol Lett. 2004;7(12):1225–41.

47. Sork VL. Genomic studies of local adaptation in natural plant populations. J Hered. 2018;109(1):3–15.

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

49. Orr HA. The genetic theory of adaptation: a brief history. Nat Rev Genet. 2005;6(2):119–27.

50. Sih A, Ferrari MC, Harris DJ. Evolution and behavioural responses to human-induced rapid environmental change. Evol Appl. 2011;4(2):367–87.

51. Thompson KA, Renaudin M, Johnson MT. Urbanization drives the evolution of parallel clines in plant populations. Proc R Soc B Biol Sci. 2016;283(1845):20162180.

52. Ashapkin VV, Kutueva LI, Aleksandrushkina NI, Vanyushin BF. Epigenetic mechanisms of plant adaptation to biotic and abiotic stresses. Int J Mol Sci. 2020;21(20):7457.

53. Taichi N, Ushimaru A. Trait variation along an urban–rural gradient in Asian dayflower: The contribution of phenotypic plasticity and genetic divergence. Plant Biol. 2024;26(1):74–81.

54. Carfora A, Lucibelli F, Di Lillo P, Mazzucchiello SM, Saccone G, Salvemini M, et al. Genetic responses of plants to urban environmental challenges. Planta. 2025;261(5):1–21.

55. Tiffin P, Ross-Ibarra J. Advances and limits of using population genetics to understand local adaptation. Trends Ecol Evol. 2014;29(12):673–80.

56. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14(11):807–20.

57. Crandall KA, Bininda-Emonds OR, Mace GM, Wayne RK. Considering evolutionary processes in conservation biology. Trends Ecol Evol. 2000;15(7):290–5.

58. Hoffmann AA, Willi Y. Detecting genetic responses to environmental change. Nat Rev Genet. 2008;9(6):421–32.

59. Barton NH. The role of hybridization in evolution. Mol Ecol. 2008;10(3):551–68.

60. Nolte AW, Tautz D. Understanding the onset of hybrid speciation. Trends Genet. 2010;26(2):54–8.

61. Santure AW, De Cauwer I, Robinson MR, Poissant J, Sheldon BC, Slate J. Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus major) population. Mol Ecol. 2013;22(15):3949–62.

62. Robinson MR, Santure AW, DeCauwer I, Sheldon BC, Slate J. Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. Mol Ecol. 2013;22(15):3963–80.

63. Alberti M, Marzluff JM, Shulenberger E, Bradley G, Ryan C, Zumbrunnen C. Integrating humans into ecology: opportunities and challenges for studying urban ecosystems. BioScience. 2003;53(12):1169–79.

64. Rich TCG. Crucifers of Great Britain and Ireland: BSBI Handbook No. 6. Vol. 6. Botanical Society of Britain and Ireland; 2006.

65. Cheam A, Code G. The biology of Australian weeds 24. Raphanus raphanistrum L. Plant Prot Q. 1995;10:2–2.

66. Sahli HF, Conner JK, Shaw FH, Howe S, Lale A. Adaptive differentiation of quantitative traits in the globally distributed weed, wild radish (Raphanus raphanistrum). Genetics. 2008;180(2):945–55.

67. Warwick SI, Francis A. The biology of Canadian weeds. 132. Raphanus raphanistrum L. Can J Plant Sci. 2005;85(3):709–33.

68. Albrecht M, Duelli P, Obrist MK, Kleijn D, Schmid B. Effective long-distance pollen dispersal in Centaurea jacea. PloS One. 2009;4(8):e6751.

69. Cousens R, Wiegand T, Taghizadeh M. Small-scale spatial structure within patterns of seed dispersal. Oecologia. 2008;158(3):437–48.

70. Walsh MJ, Powles SB. Impact of crop-topping and swathing on the viable seed production of wild radish (Raphanus raphanistrum). Crop Pasture Sci. 2009;60(7):667–74.

71. Eslami SV, Gill GS, Bellotti B, McDonald G. Wild radish (Raphanus raphanistrum) interference in wheat. Weed Sci. 2006;54(4):749–56.

72. Petersen J, Belz R, Walker F, Hurle K. Weed suppression by release of isothiocyanates from turnip-rape mulch. Agron J. 2001;93(1):37–43.

73. Walsh MJ, Duane RD, Powles SB. High frequency of chlorsulfuron-resistant wild radish (Raphanus raphanistrum) populations across the Western Australian wheatbelt. Weed Technol. 2001;15(2):199–203.

74. Malik MS. Biology and ecology of wild radish (Raphanus raphanistrum) [PhD Thesis]. Clemson University; 2009.

75. Connor HE. The poisonous plants in New Zealand. 1977;

76. Cooper MR, Johnson AW. Poisonous plants in Britain and their effects on animals and man. 1984.

77. Norsworthy JK. Allelopathic Potential of Wild Radish (Raphanus raphanistrum) 1. Weed Technol. 2003;17(2):307–13.

78. Malik MS, Riley MB, Norsworthy JK, Bridges Jr W. Glucosinolate profile variation of growth stages of wild radish (Raphanus raphanistrum). J Agric Food Chem. 2010;58(6):3309–15.

79. Haramoto ER, Gallandt ER. Brassica cover cropping for weed management: A review. Renew Agric Food Syst. 2004;19(4):187–98.

80. Reeves T, Code G, Piggin C. Seed production and longevity, seasonal emergence and phenology of wild radish (Raphanus raphanistrum L.). Aust J Exp Agric. 1981;21(112):524–30.

81. Woolcock JL, Cousens R. A mathematical analysis of factors affecting the rate of spread of patches of annual weeds in an arable field. Weed Sci. 2000;48(1):27–34.

82. McGeocha M, Kalwija J, Rhodesa J. A spatial assessment of Brassica napus gene flow potential to wild and weedy relatives in the Fynbos Biome. South Afr J Sci. 2009;105(3):109–15.

83. Mandáková T, Lysak MA. Post-polyploid diploidization and diversification through dysploid changes. Curr Opin Plant Biol. 2018;42:55–65.

84. Zhang X, Liu T, Wang J, Wang P, Qiu Y, Zhao W, et al. Pan-genome of Raphanus highlights genetic variation and introgression among domesticated, wild, and weedy radishes. Mol Plant. 2021;14(12):2032–55.

85. Yamagishi H, Terachi T. Multiple origins of cultivated radishes as evidenced by a comparison of the structural variations in mitochondrial DNA of Raphanus. Genome. 2003;46(1):89–94.

86. Kobayashi H, Shirasawa K, Fukino N, Hirakawa H, Akanuma T, Kitashiba H. Identification of genome-wide single-nucleotide polymorphisms among geographically diverse radish accessions. DNA Res. 2020;27(1):dsaa001.

87. Campbell LG, Snow AA, Sweeney PM. When divergent life histories hybridize: insights into adaptive life-history traits in an annual weed. New Phytol. 2009;184(4):806–18.

88. Hegde SG, Nason JD, Clegg JM, Ellstrand NC. The evolution of California's wild radish has resulted in the extinction of its progenitors. Evolution. 2006;60(6):1187–97.

89. Irwin RE, Strauss SY, Storz S, Emerson A, Guibert G. The role of herbivores in the maintenance of a flower color polymorphism in wild radish. Ecology. 2003;84(7):1733–43.

90. Irwin RE, Strauss SY. Flower color microevolution in wild radish: evolutionary response to pollinator-mediated selection. Am Nat. 2005;165(2):225–37.

91. Stanton ML, Snow AA, Handel SN, Bereczky J. The impact of a flower-color polymorphism on mating patterns in experimental populations of wild radish (Raphanus raphanistrum L.). Evolution. 1989;43(2):335–46.

92. Fineblum WL, Rausher MD. Do floral pigmentation genes also influence resistance to enemies? The W locus in Ipomoea purpurea. Ecology. 1997;78(6):1646–54.

93. Armbruster W. Can indirect selection and genetic context contribute to trait diversification? A transition-probability study of blossom-colour evolution in two genera. J Evol Biol. 2002;15(3):468–86.

94. Brachi B, Meyer CG, Villoutreix R, Platt A, Morton TC, Roux F, et al. Coselected genes determine adaptive variation in herbivore resistance throughout the native range of Arabidopsis thaliana. Proc Natl Acad Sci. 2015;112(13):4032–7.

95. Clergeau P, Jokimäki J, Snep R. Using hierarchical levels for urban ecology. Trends Ecol Evol. 2006;21(12):660–1.

96. McDonnell MJ, Pickett ST. Ecosystem structure and function along urban-rural gradients: an unexploited opportunity for ecology. Ecology. 1990;71(4):1232–7.

97. Zipperer WC, Wu J, Pouyat RV, Pickett ST. The application of ecological principles to urban and urbanizing landscapes. Ecol Appl. 2000;685–8.

98. Williams NS, Schwartz MW, Vesk PA, McCarthy MA, Hahs AK, Clemants SE, et al. A conceptual framework for predicting the effects of urban environments on floras. J Ecol. 2009;97(1):4–9.

99. Hess GR, Fischer RA. Communicating clearly about conservation corridors. Landsc Urban Plan. 2001;55(3):195–208.

100. McDonnell MJ, Hahs AK. The use of gradient analysis studies in advancing our understanding of the ecology of urbanizing landscapes: current status and future directions. Landsc Ecol. 2008;23(10):1143–55.

101. Clergeau P, Savard JPL, Mennechez G, Falardeau G. Bird abundance and diversity along an urban-rural gradient: a comparative study between two cities on different continents. The condor. 1998;100(3):413–25.

102. Germaine SS, Wakeling BF. Lizard species distributions and habitat occupation along an urban gradient in Tucson, Arizona, USA. Biol Conserv. 2001;97(2):229–37.

103. Pillsbury FC, Miller JR. Habitat and landscape characteristics underlying anuran community structure along an urban–rural gradient. Ecol Appl. 2008;18(5):1107–18.

104. Stewart ID, Oke TR. Local climate zones for urban temperature studies. Bull Am Meteorol Soc. 2012;93(12):1879–900.

105. Schneider A, Friedl MA, Potere D. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions.' Remote Sens Environ. 2010;114(8):1733–46.

106. Herold M, Scepan J, Clarke KC. The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. Environ Plan A. 2002;34(8):1443–58.

107. Liu Z, Zhou Y, Feng Z. Response of vegetation phenology to urbanization in urban agglomeration areas: A dynamic urban–rural gradient perspective. Sci Total Environ. 2023;864:161109.

108. Wu J. Urban ecology and sustainability: The state-of-the-science and future directions. Landsc Urban Plan. 2014;125:209–21.

109. Copernicus Land Monitoring Service, Copernicus Land Monitoring Service helpdesk. Imperviousness Density 2021 (raster 10 m), Europe, 3-yearly, Jun 2025 [Internet]. Copernicus Land Monitoring Service; 2025 [cited 2025 Nov 27]. Available from: https://sdi.eea.europa.eu/catalogue/srv/api/records/34ef6334-d432-4041-a3da-67e156d6501d?language=all

110. QGIS Development Team. QGIS Geographic Information System [Internet]. QGIS Association; 2025. Available from: https://www.qgis.org

111. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. F1000Research. 2020;9:ISCB-Comm.

112. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

113. FastQC AS. Babraham bioinformatics. Publ Online. 2010;

114. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

115. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):323.

116. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.

117. Trinity RNA-Seq Team. Trinity RNA-seq. [cited 2025 Nov 28]. Trinity Transcript Quantification – GitHub Wiki. Available from: https://github.com/trinityrnaseq/trinityrnaseq/wiki/Trinity-Transcript-Quantification

118. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

119. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. bioinformatics. 2010;26(1):139–40.

120. Trinity RNA-Seq Team. Trinity Differential Expression with edgeR [Internet]. [cited 2025 Nov 28]. Available from: https://github.com/trinityrnaseq/trinityrnaseq/wiki/Trinity-Differential-Expression

121. Wickham H. ggplot2. Wiley Interdiscip Rev Comput Stat. 2011;3(2):180–5.

122. NGDC–CNCB: National Genomics Data Center, China National Center for Bioinformation. [Internet]. NGDC–CNCB: National Genomics Data Center, China National Center for Bioinformation. Available from: https://ngdc.cncb.ac.cn/

123. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics. 2007;23(10):1289–91.

124. Rieu I, Powers SJ. Real-time quantitative RT-PCR: design, calculations, and statistics. Plant Cell. 2009;21(4):1031–3.

125. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics. 2009;25(14):1754–60.

126. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.

127. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 2015;31(12):2032–4.

128. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. ArXiv Prepr ArXiv12073907. 2012;

129. Bonfield JK, Marshall J, Danecek P, Li H, Ohan V, Whitwham A, et al. HTSlib: C library for reading/writing high-throughput sequencing data. Gigascience. 2021;10(2):giab007.

130. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–93.

131. PLINK 2.0 [Internet]. 2025. Available from: https://www.cog-genomics.org/plink/2.0/

132. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.

133. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.

134. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.

135. Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014;31(10):2824–7.

136. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.

137. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat. 1947;50–60.

138. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44(7):821–4.

139. Cusin F, Lefebvre H, Sigaud T. The Peri-urban question: A study of growth and diversity in peripheral areas in France. Rev Fr Sociol Engl Ed. 2016;57(4):437–71.

140. Hernández I, Alegre L, Van Breusegem F, Munné-Bosch S. How relevant are flavonoids as antioxidants in plants? Trends Plant Sci. 2009;14(3):125–32.

141. Nakabayashi R, Yonekura-Sakakibara K, Urano K, Suzuki M, Yamada Y, Nishizawa T, et al. Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. Plant J. 2014;77(3):367–79.

142. Nakabayashi R, Saito K. Integrated metabolomics for abiotic stress responses in plants. Curr Opin Plant Biol. 2015;24:10–6.

143. Ma S, Lv L, Meng C, Zhang C, Li Y. Integrative analysis of the metabolome and transcriptome of Sorghum bicolor reveals dynamic changes in flavonoids accumulation under saline–alkali stress. J Agric Food Chem. 2020;68(50):14781–9.

144. Xu Y, Li J, Lin Z, Liang W, Qin L, Ding J, et al. Isorhamnetin alleviates airway inflammation by regulating the Nrf2/Keap1 pathway in a mouse model of COPD. Front Pharmacol. 2022;13:860362.

145. Dong NQ, Lin HX. Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. J Integr Plant Biol. 2021;63(1):180–209.

146. Böttner L, Grabe V, Gablenz S, Böhme N, Appenroth KJ, Gershenzon J, et al. Differential localization of flavonoid glucosides in an aquatic plant implicates different functions under abiotic stress. Plant Cell Environ. 2021;44(3):900–14.

147. Jing Y, Watanabe M, Aarabi F, Fernie AR, Borghi M, Tohge T. Cross-species metabolomic analyses in the Brassicaceae reveals common responses to ultraviolet-B exposure. Plant Cell Physiol. 2023;64(12):1523–33.

148. Kobe B, Kajava AV. The leucine-rich repeat as a protein recognition motif. Curr Opin Struct Biol. 2001;11(6):725–32.

149. Sun X, Wang GL. Genome-wide identification, characterization and phylogenetic analysis of the rice LRR-kinases. PloS One. 2011;6(3):e16079.

150. Nakano RT, Piślewska-Bednarek M, Yamada K, Edger PP, Miyahara M, Kondo M, et al. PYK10 myrosinase reveals a functional coordination between endoplasmic reticulum bodies and glucosinolates in Arabidopsis thaliana. Plant J. 2017;89(2):204–20.

151. Wittstock U, Burow M. Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. Arab BookAmerican Soc Plant Biol. 2010;8:e0134.

152. Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, Robertson DL. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. Plant Physiol. 2010;153(3):1398–412.

153. Gao F, Dubos C. The arabidopsis bHLH transcription factor family. Trends Plant Sci. 2024;29(6):668–80.

154. Massari ME, Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol. 2000;20(2):429–40.

155. Li N, Lin Z, Yu P, Zeng Y, Du S, Huang LJ. The multifarious role of callose and callose synthase in plant development and environment interactions. Front Plant Sci. 2023;14:1183402.

156. Barratt DP, Kölling K, Graf A, Pike M, Calder G, Findlay K, et al. Callose synthase GSL7 is necessary for normal phloem transport and inflorescence growth in Arabidopsis. Plant Physiol. 2011;155(1):328–41.

157. Bundock P, Hooykaas P. An Arabidopsis hAT-like transposase is essential for plant development. Nature. 2005;436(7048):282–4.

158. Michaels SD, Amasino RM. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell. 1999;11(5):949–56.

159. Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, et al. The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell. 1999;11(3):445–58.

160. Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C. Vernalization requires epigenetic silencing of FLC by histone methylation. Nature. 2004;427(6970):164–7.

161. Capovilla G, Schmid M, Posé D. Control of flowering by ambient temperature. J Exp Bot. 2015;66(1):59–69.

162. Verhage L, Angenent GC, Immink RG. Research on floral timing by ambient temperature comes into blossom. Trends Plant Sci. 2014;19(9):583–91.

163. Marquardt S, Raitskin O, Wu Z, Liu F, Sun Q, Dean C. Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. Mol Cell. 2014;54(1):156–65.

164. Csorba T, Questa JI, Sun Q, Dean C. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. Proc Natl Acad Sci. 2014;111(45):16160–5.

165. Swiezewski S, Liu F, Magusin A, Dean C. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature. 2009;462(7274):799–802.

166. Xin P, Schier J, Šefrnová Y, Kulich I, Dubrovsky JG, Vielle-Calzada JP, et al. The Arabidopsis TETRATRICOPEPTIDE-REPEAT THIOREDOXIN-LIKE (TTL) family members are involved in root system formation via their interaction with cytoskeleton and cell wall remodeling. Plant J. 2022;112(4):946–65.

167. Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, Lugovskoy AA, et al. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell. 2003;114(1):99–111.

168. Doyon Y, Cayrou C, Ullah M, Landry AJ, Côté V, Selleck W, et al. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell. 2006;21(1):51–64.

169. Barbez E, Kubeš M, Rolčík J, Béziat C, Pěnčík A, Wang B, et al. A novel putative auxin carrier family regulates intracellular auxin homeostasis in plants. Nature. 2012;485(7396):119–22.

170. Wang W, Vinocur B, Shoseyov O, Altman A. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci. 2004;9(5):244–52.

171. Tang T, Yu A, Li P, Yang H, Liu G, Liu L. Sequence analysis of the Hsp70 family in moss and evaluation of their functions in abiotic stress responses. Sci Rep. 2016;6(1):33650.

172. Guo M, Lu JP, Zhai YF, Chai WG, Gong ZH, Lu MH. Genome-wide analysis, expression profile of heat shock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (Capsicum annuum L.). BMC Plant Biol. 2015;15(1):151.

173. Tang T, Yu A, Li P, Yang H, Liu G, Liu L. Sequence analysis of the Hsp70 family in moss and evaluation of their functions in abiotic stress responses. Sci Rep. 2016;6(1):33650.

174. Pan X, Zheng Y, Lei K, Tao W, Zhou N. Systematic analysis of Heat Shock Protein 70 (HSP70) gene family in radish and potential roles in stress tolerance. BMC Plant Biol. 2024;24(1):2.

175. Cnops G, Neyt P, Raes J, Petrarulo M, Nelissen H, Malenica N, et al. The TORNADO1 and TORNADO2 genes function in several patterning processes during early leaf development in Arabidopsis thaliana. Plant Cell. 2006;18(4):852–66.

176. Reinbothe S, Bartsch S, Rossig C, Davis MY, Yuan S, Reinbothe C, et al. A protochlorophyllide (Pchlide) a oxygenase for plant viability. Front Plant Sci. 2019;10:593.

177. Yasin MU, Sun L, Yang C, Liu B, Gan Y. C2H2 Zinc Finger Proteins GIS2 and ZFP8 Regulate Trichome Development via Hormone Signaling in Arabidopsis. Int J Mol Sci. 2025;26(15):7265.

178. Sun XL, Yu QY, Tang LL, Ji W, Bai X, Cai H, et al. GsSRK, a G-type lectin S-receptor-like serine/threonine protein kinase, is a positive regulator of plant tolerance to salt stress. J Plant Physiol. 2013;170(5):505–15.

179. Tavares GA, Panepucci EH, Brunger AT. Structural characterization of the intramolecular interaction between the SH3 and guanylate kinase domains of PSD-95. Mol Cell. 2001;8(6):1313–25.

180. Nomura Y, Izumi A, Fukunaga Y, Kusumi K, Iba K, Watanabe S, et al. Diversity in guanosine 3′, 5′-bisdiphosphate (ppGpp) sensitivity among guanylate kinases of bacteria and plants. J Biol Chem. 2014;289(22):15631–41.

181. Konrad M. Cloning and expression of the essential gene for guanylate kinase from yeast. J Biol Chem. 1992;267(36):25652–5.

182. Jain R, Khan N, Menzel A, Rajkovic I, Konrad M, Techert S. Insights into open/closed conformations of the catalytically active human guanylate kinase as investigated by small-angle X-ray scattering. Eur Biophys J. 2016;45(1):81–9.

183. Lehner B, Sanderson CM. A protein interaction framework for human mRNA degradation. Genome Res. 2004;14(7):1315–23.

184. Li C, Wei J, Li Y, He X, Zhou Q, Yan J, et al. Transmembrane Protein 214 (TMEM214) mediates endoplasmic reticulum stress-induced caspase 4 enzyme activation and apoptosis. J Biol Chem. 2013;288(24):17908–17.

185. Robertson AS, Allwood EG, Smith AP, Gardiner FC, Costa R, Winder SJ, et al. The WASP homologue Las17 activates the novel actin-regulatory activity of Ysc84 to promote endocytosis in yeast. Mol Biol Cell. 2009;20(6):1618–28.

186. Sharifi Alishah M, Darvishzadeh R, Ahmadabadi M, Piri Kashtiban Y, Hasanpur K. Identification of differentially expressed genes in salt-tolerant oilseed sunflower (Helianthus annuus L.) genotype by RNA sequencing. Mol Biol Rep. 2022;49(5):3583–96.

187. Xie Y, Wang Y, Yu X, Lin Y, Zhu Y, Chen J, et al. SH3P2, an SH3 domain-containing protein that interacts with both Pib and AvrPib, suppresses effector-triggered, Pib-mediated immunity in rice. Mol Plant. 2022;15(12):1931–46.

APPENDIX

Table 1. List of significant SNPs (F_{ST}> 0.45) in the Urban versus Natural comparison.

SNP ID (chr:pos)	HUDSON_F_{ST}	Gene_ID	SNP position & Type	Annotation- BLAST
1:8511201:G:A	0.484529	Rr0R1c001354	Downstream 1,277 bp	PREDICTED: *Raphanus sativus* hexokinase-1 (LOC108823542)
2:16530342:A:C	0.464706	Rr0R2c008541	Upstream 4,304 bp	PREDICTED: *Raphanus sativus* uncharacterized LOC108837599 (LOC108837599)
4:1169691:T:A	0.527395	Rr0R4c014299	Downstream 6,637 bp	PREDICTED: *Raphanus sativus* homeobox-leucine zipper protein HAT22 (LOC108855736)
4:55057945:T:C	0.521719	Rr0R4c019526	Intron (exons 6 - 7)	PREDICTED: *Raphanus sativus* transcription factor bHLH155 (LOC108854757)
4:55058281:G:T	0.524192	Rr0R4c019526	Exon 8 - missense M→I	PREDICTED: *Raphanus sativus* transcription factor bHLH155 (LOC108854757)
4:55058658:A:G	0.48752	Rr0R4c019526	Exon 8 - missense Q→R	PREDICTED: *Raphanus sativus* transcription factor bHLH155 (LOC108854757)
4:65651587:T:C	0.458494	Rr0R4c020935	Upstream 10,286 bp	PREDICTED: *Raphanus sativus* protein RGF1 INDUCIBLE TRANSCRIPTION FACTOR 1 (LOC108853092)
5:16845155:A:T	0.471286	Rr0R5c023891	Upstream 4,054 bp	PREDICTED: *Raphanus sativus* callose synthase 7 (LOC108862791)
5:16850177:G:A	0.469397	Rr0R5c023891	Upstream 9,076 bp	PREDICTED: *Raphanus sativus* callose synthase 7 (LOC108862791)
5:39885597:G:A	0.452185	Rr0R5c025765	Intron (exons 1 - 2)	PREDICTED: Brassica napus zinc finger BED domain-containing protein RICESLEEPER 1 (LOC106454000)
7:12459785:T:C	0.478813	Rr0R7c030454	Downstream 93,028 bp	PREDICTED: *Raphanus sativus* BEL1-like homeodomain protein 4 (
9:30933808:T:C	0.541176	Rr0R9c040314	Downstream 44,447 bp	R.sativa gene for 25S (partial) and 18S (partial) ribosomal RNA
9:3773568:T:C	0.454036	Rr0R9c037997	Downstream 139 bp	PREDICTED: *Raphanus sativus* homeobox-leucine zipper protein HDG5 (LOC108827035)
9:3818087:G:C	0.452983	Rr0R9c037999	Exon 20	PREDICTED: *Raphanus sativus* uncharacterized LOC108827372 (LOC108827372)
9:3818662:T:G	0.521767	Rr0R9c037999	Intron (exons 21 - 22)	PREDICTED: *Raphanus sativus* uncharacterized LOC108827372 (LOC108827372)
9:3818679:A:C	0.504442	Rr0R9c037999	Intron (exons 21 - 22)	PREDICTED: *Raphanus sativus* uncharacterized LOC108827372 (LOC10882737
9:3818745:T:A	0.469894	Rr0R9c037999	Intron (exons 21 - 22)	PREDICTED: *Raphanus sativus* uncharacterized LOC108827372 (LOC108827372)

Table 2. List of significant SNPs in XP-EHH, θπ ratio, and F_{ST} in the Urban versus Natural comparison.

SNP ID (chr:pos)	HUDSON_FST	log2_theta_pi_ratio	mean_xpehh	Gene_ID	SNP position & Type	Annotation
2:12139680:C:T	3.14E-01	6.82E-01	6.62E-01	Rr0R2c007972	Intron (exons 1 – 2)	PREDICTED: *Raphanus sativus* protein DOWNSTREAM OF FLC (LOC108841258), mRNA
2:12147145:A:C	3.12E-01	6.82E-01	6.62E-01	Rr0R2c007973	Exon 3 missense E → A	*Raphanus sativus* inactive TPR repeat-containing thioredoxin TTL3-like (LOC108841519), mRNA
2:12148604:T:C	2.07E-01	6.82E-01	6.62E-01	Rr0R2c007973	Exon 5 missense F → S	*Raphanus sativus* inactive TPR repeat-containing thioredoxin TTL3-like (LOC108841519), mRNA
2:12149585:C:A	2.18E-01	6.82E-01	6.62E-01	Rr0R2c007973	Intron (exons 6 – 7)	*Raphanus sativus* inactive TPR repeat-containing thioredoxin TTL3-like (LOC108841519), mRNA
4:44545214:T:C	2.43E-01	6.02E-01	6.93E-01	Rr0R4c018432 - Rr0R4c018433	Downstream 8733 bp – Upstream 15,690 bp	PREDICTED: *Raphanus sativus* PHD finger protein ING2 (LOC108830853), mRNA
4:44545284:T:C	2.16E-01	6.02E-01	6.93E-01	Rr0R4c018432 - Rr0R4c018433	Downstream 8803 bp – Upstream 15,620 bp	PREDICTED: *Raphanus sativus* PHD finger protein ING2 (LOC108830853), mRNA
4:44547067:A:G	2.13E-01	6.02E-01	6.93E-01	Rr0R4c018432 - Rr0R4c018433	Downstream 10589 bp – Upstream 13,837 bp	PREDICTED: *Raphanus sativus* PHD finger protein ING2 (LOC108830853), mRNA
4:44551913:A:G	2.16E-01	6.02E-01	6.93E-01	Rr0R4c018432 - Rr0R4c018433	Downstram 15432 bp – Upstream 8,991 bp	PREDICTED: *Raphanus sativus* PHD finger protein ING2 (LOC108830853), mRNA
4:44557895:C:T	2.19E-01	6.02E-01	6.93E-01	Rr0R4c018432 - Rr0R4c018433	Downstram 21414 bp – Upstream 3009 bp	PREDICTED: *Raphanus sativus* PHD finger protein ING2 (LOC108830853), mRNA
4:44558728:T:C	2.49E-01	6.02E-01	6.93E-01	Rr0R4c018432 - Rr0R4c018433	Downstram 22247 bp – Upstream 2176 bp	PREDICTED: *Raphanus sativus* PHD finger protein ING2 (LOC108830853), mRNA
9:42665435:C:A	2.15E-01	-8.79E-01	7.09E-01	Rr0R9c040544	Downstream 11,149 bp	Glutamic endopeptidase
9:42665459:A:T	2.19E-01	-8.79E-01	7.09E-01	Rr0R9c040544	Downstream 11,173 bp	Glutamic endopeptidase
9:42690999:T:C	2.31E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42691112:C:T	2.15E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42691123:A:C	2.50E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42691150:G:A	2.15E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42691494:C:T	2.39E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42691593:C:G	2.19E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42692388:A:T	2.24E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42692547:G:A	2.30E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42692575:C:A	2.76E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42692886:T:C	2.19E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42693158:G:A	2.04E-01	-6.60E-01	7.09E-01	NA	non coding region	NA

9:42693411:G:A	2.53E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42693502:T:A	2.34E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42693684:T:C	2.06E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42693689:G:T	2.76E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42693701:C:T	2.76E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42693798:C:T	2.24E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42693906:G:T	2.24E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42694332:A:G	2.39E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42694342:G:A	2.55E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42694396:G:T	2.55E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42694581:A:C	2.30E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42694641:C:T	2.99E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42694676:C:T	2.42E-01	-6.60E-01	7.09E-01	NA	non coding region	NA
9:42694716:T:G	2.04E-01	-6.60E-01	7.09E-01	NA	non coding region	NA

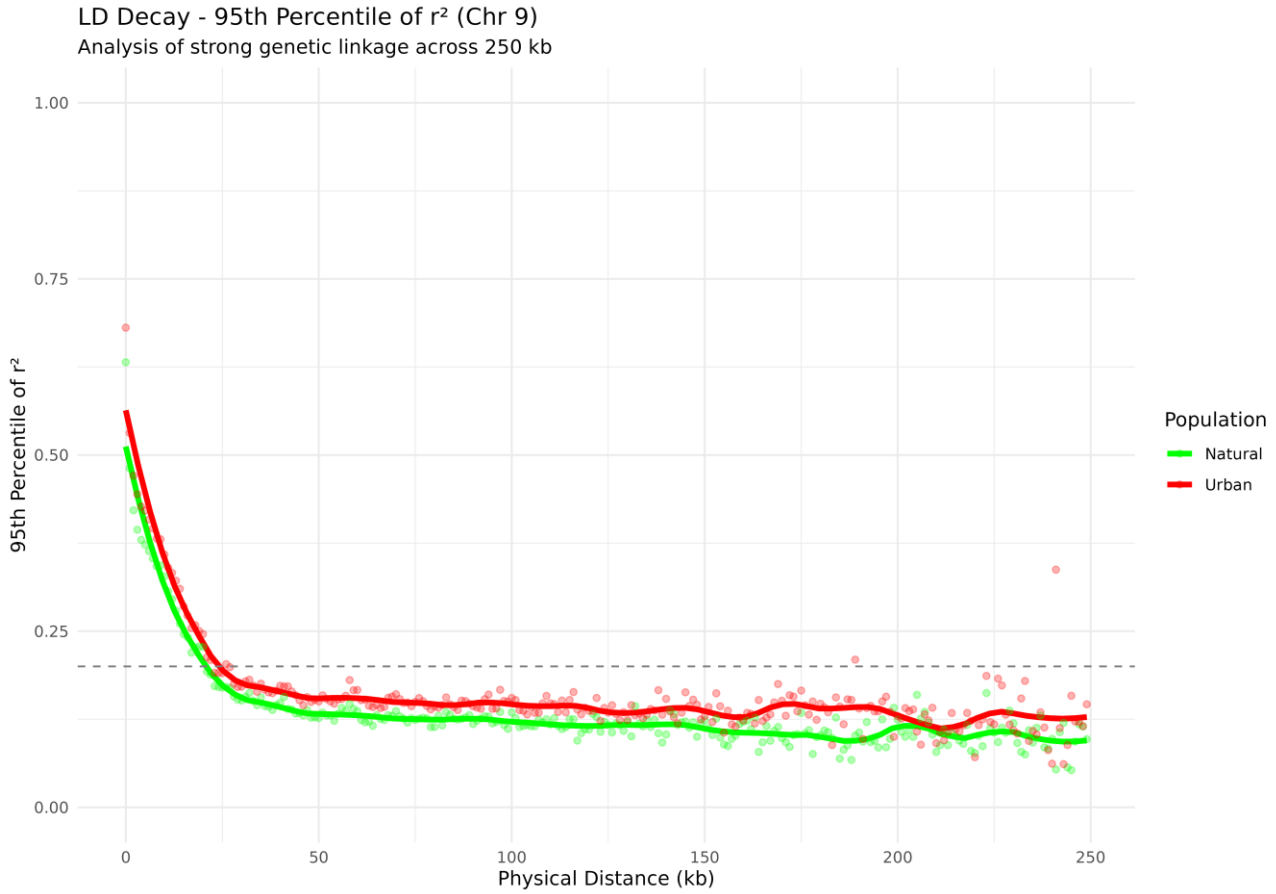

Figure 1. Linkage disequilibrium (LD) decay on Chromosome 9. The plot illustrates the decline of 95th percentile r^2 values as a function of physical distance in kilobases (kb). The dashed horizontal

line indicates the $r^2 = 0.2$ threshold, which is commonly used to define the extent of significant linkage. Data are presented for the Urban population (red) and the Natural population (green).

Table 3. List of genomic regions associated to leaf area phenotype, that exceeded the Bonferroni significance threshold

SNP ID (chr:pos)	p_wald	beta	af	se	Gene_ID	Nearest gene (bb)	Annotation- BLAST
1:19138482:G:A	9.80E-09	-8.33	0.16	1.31	Rr0R1c002680	-1145	PREDICTED: *Brassica napus* transcription factor MYB1R1-like
1:26070237:G:A	1.58E-10	-13.23	0.13	1.82	Rr0R1c003180	7737	PREDICTED: *Raphanus sativus* protein PIN-LIKES 3
1:37128222:C:T	4.13E-09	-7.66	0.21	1.17	Rr0R1c004556	-947	PREDICTED: *Raphanus sativus* amino acid permease 1
1:37237524:A:T	6.45E-10	-12.66	0.14	1.82	Rr0R1c004568	2904	PREDICTED: *Raphanus sativus* protein NRT1/ PTR FAMILY 4.3-like
1:37542255:T:C	9.64E-12	-8.6	0.23	1.09	Rr0R1c004603	5943	PREDICTED: *Raphanus sativus* putative potassium transporter 12
1:43277172:A:G	2.42E-10	-12.5	0.12	1.74	Rr0R1c005403	2297	PREDICTED: *Raphanus sativus* probable thionin-2.4
2:1778823:C:A	6.97E-09	-12.13	0.14	1.89	Rr0R2c006136	-1429	PREDICTED: *Raphanus sativus* putative F-box/kelch-repeat protein At4g35120
2:22901962:C:T	1.41E-09	-8.17	0.18	1.2	Rr0R2c009213	-8408	
2:27186499:C:G	8.93E-09	-12.74	0.11	2	Rr0R2c009361	-121126	*Raphanus sativus* isolate Tibet-B mitochondrion, complete genome
2:33333467:C:A	2.86E-09	-12.11	0.15	1.83	Rr0R2c009377	-443125	
3:10737665:A:G	9.13E-09	-8.62	0.17	1.35	Rr0R3c011571	-1358	PREDICTED: *Raphanus sativus* UDP-glycosyltransferase 73C12
3:13504273:T:C	1.68E-10	-9.06	0.2	1.25	Rr0R3c011893	-8343	PREDICTED: *Raphanus sativus* 14-3-3-like protein GF14 mu (LOC108847404), transcript variant X2
3:14174054:A:G	6.59E-12	-13.63	0.13	1.71	Rr0R3c011957	19156	PREDICTED: *Brassica rapa* inactive endochitinase At2g43600
3:21733349:G:A	5.16E-09	-9.61	0.21	1.48	Rr0R3c012543	8994	REDICTED: *Raphanus sativus* B3 domain-containing protein REM12 (LOC108847104), transcript variant X1,
3:24872599:T:G	2.47E-09	-11.48	0.17	1.72	Rr0R3c012787	-16444	PREDICTED: *Brassica napus* tropomyosin-1, isoforms 33/34 (LOC106346199), transcript variant X2, mRNA
3:3268097:A:C	1.45E-11	-8.25	0.21	1.06	Rr0R3c010436	850	PREDICTED: *Raphanus sativus* uncharacterized LOC108845599 (LOC108845599), mRNA

3:39135325:A:T	2.10E-12	-13.12	0.13	1.6	Rr0R3c014094	-98834	PREDICTED*: Raphanus sativus* uncharacterized protein At3g43530-like (LOC130506040), mRNA
3:8634742:G:T	4.15E-09	-11.68	0.13	1.78	Rr0R3c011300	-606	PREDICTED: *Raphanus sativus* E3 ubiquitin-protein ligase SINA-like 2
4:26460722:A:T	8.37E-10	-13.14	0.12	1.9	Rr0R4c017391	2149	PREDICTED*: Raphanus sativus* protein TORNADO 2-like
4:26461451:T:A	2.01E-09	-13.21	0.11	1.97	Rr0R4c017391	1420	PREDICTED: *Raphanus sativus* protein TORNADO 2-like
4:34007442:A:G	7.13E-10	-9.73	0.15	1.4	Rr0R4c018014	6528	PREDICTED: *Raphanus sativus* uncharacterized LOC130498716 (LOC130498716), mRNA
4:37828420:T:G	1.04E-08	10.33	0.28	1.63	Rr0R4c018170	3511	Unknown
4:5713267:C:T	1.55E-09	-9.62	0.18	1.42	Rr0R4c014787	752	PREDICTED*: Raphanus sativus* poly [ADP-ribose] polymerase 2
4:60774455:G:T	3.18E-09	-11.21	0.15	1.69	Rr0R4c020376	5288	PREDICTED*: Raphanus sativus* piezo-type mechanosensitive ion channel homolog
4:63311616:C:A	2.08E-10	-11.29	0.3	1.56	Rr0R4c020653	1028	PREDICTED*: Brassica napus* phytyl ester synthase 2, chloroplastic
4:63350484:G:A	6.36E-09	-9.55	0.16	1.48	Rr0R4c020661	1184	PREDICTED*: Raphanus sativus* protein IQ-DOMAIN 5-like
4:65525138:G:T	2.20E-10	-11.62	0.13	1.61	Rr0R4c020921	-1211	PREDICTED: *Brassica rapa* flavanone 3-dioxygenase 3
5:25179192:G:T	8.82E-10	-13.46	0.11	1.95	Rr0R5c024526	38543	PREDICTED: *Raphanus sativus* uncharacterized LOC108810197 (LOC108810197), transcript variant X1, mRNA
5:26525770:A:C	4.55E-10	-12.41	0.12	1.76	Rr0R5c024541	-58523	PREDICTED*: Raphanus sativus* uncharacterized LOC130503230 (LOC130503230), mRNA
5:30730277:G:A	6.73E-09	-9.06	0.17	1.41	Rr0R5c024843	-2098	PREDICTED*: Raphanus sativus* heat shock 70 kDa protein BIP2
5:36961574:T:C	1.63E-10	-8.08	0.22	1.11	Rr0R5c025354	-2764	PREDICTED: *Raphanus sativus* protein PAT1 homolog 1
5:41805095:G:T	1.72E-09	-9.3	0.17	1.38	Rr0R5c026034	3317	PREDICTED*: Raphanus sativus* 3-epi-6-deoxocathasterone 23-monooxygenase CYP90D1
5:44837510:C:T	4.49E-10	-9.2	0.19	1.31	Rr0R5c026526	11240	PREDICTED*: Raphanus sativus* rab escort protein 1
6:14855547:T:C	1.58E-10	-12.16	0.14	1.67	Rr0R6c028045	-2906	PREDICTED: *Raphanus sativus* receptor-like protein kinase HSL1
6:15931253:G:A	1.90E-09	-9.19	0.16	1.37	Rr0R6c028197	801	PREDICTED*: Raphanus sativus* proline-rich receptor-like protein kinase PERK10
6:5472434:A:G	2.50E-10	-11.67	0.19	1.63	Rr0R6c027456	-10545	Unknown
6:8028487:C:T	1.02E-08	-9.16	0.21	1.44	Rr0R6c027684	-19362	Unknown

7:10744038:T:C	3.48E-09	-12.28	0.13	1.86	Rr0R7c030448	-55347	PREDICTED: *Raphanus sativus* T-complex protein 1 subunit epsilon-like
7:28077965:T:A	6.45E-09	-11.17	0.14	1.73	Rr0R7c031976	796	PREDICTED: *Raphanus sativus* transmembrane E3 ubiquitin-protein ligase FLY2
7:28883324:G:A	4.00E-09	-10.69	0.15	1.63	Rr0R7c032080	-1717	PREDICTED: *Raphanus sativus* putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g61610
8:10664996:T:A	3.07E-09	-8.43	0.17	1.27	Rr0R8c034553	16945	PREDICTED: *Raphanus sativus* WAT1-related protein At1g44800
8:15155304:G:C	4.16E-10	-13.14	0.12	1.86	Rr0R8c034892	15927	Unknown
8:15647036:G:A	6.57E-09	-12.46	0.12	1.93	Rr0R8c034912	-23343	Unknown
8:22053223:A:G	2.45E-10	-10.82	0.14	1.51	Rr0R8c035010	-5266	Unknown
8:23103665:G:A	2.12E-10	-12.83	0.14	1.78	Rr0R8c035098	5577	Unknown
8:2469506:G:A	2.75E-09	-8.18	0.24	1.23	Rr0R8c033801	-1249	Unknown
8:25435199:C:T	2.11E-09	-7.97	0.18	1.19	Rr0R8c035324	-839	PREDICTED: *Raphanus sativus* vegetative cell wall protein gp1
9:15697853:A:G	3.47E-10	-10.2	0.15	1.44	Rr0R9c039508	-10663	PREDICTED: *Brassica napus* protein DETOXIFICATION 8
9:21292065:G:T	9.73E-09	-12.11	0.13	1.9	Rr0R9c040039	11132	PREDICTED: *Brassica napus* cytochrome P450 72A15-like
9:31087490:T:C	1.94E-09	-12.98	0.11	1.93	Rr0R9c040349	0	Unknown
9:37691107:A:T	4.54E-09	-10.73	0.14	1.64	Rr0R9c040392	-150748	Unknown
9:3785953:C:T	5.07E-09	-8.28	0.19	1.27	Rr0R9c037998	11325	PREDICTED: *Raphanus sativus* lysine-specific demethylase JMJ13-like
9:51123460:G:A	1.49E-10	-7.94	0.22	1.09	Rr0R9c041181	-951	Unknown
9:53422517:A:G	7.45E-10	-9.91	0.21	1.43	Rr0R9c041508	-904	PREDICTED: *Brassica oleracea* var. oleracea lipid transfer-like protein VAS (LOC106299404), transcript variant X1
9:54205187:A:G	1.55E-10	-13.81	0.11	1.9	Rr0R9c041599	-2171	PREDICTED: *Raphanus sativus* receptor homology region, transmembrane domain- and RING domain-containing protein 2

Table 4. List of genomic regions associated to fruit weigh phenotype, that exceeded the Bonferroni significance threshold

SNP ID (chr:pos)	p_wald	beta	af	se	Gene_ID	Nearest gene (bb)	Annotation
1:19138482:G:A	4.14E-09	-4.37	0.16	0.67	Rr0R1c002680	-1145	PREDICTED: *Brassica napus* transcription factor MYB1R1-like
1:26070237:G:A	2.00E-09	-6.7	0.13	1	Rr0R1c003180	7737	PREDICTED: *Raphanus sativus* protein PIN-LIKES 3

1:37128222:C:T	4.51E-10	-4.19	0.21	0.59	Rr0R1c004556	-947	PREDICTED: *Raphanus sativus* amino acid permease 1
1:37237524:A:T	1.62E-11	-7.21	0.14	0.93	Rr0R1c004568	2904	PREDICTED: *Raphanus sativus* protein NRT1/ PTR FAMILY 4.3-like
1:38981819:C:A	3.34E-09	-4.21	0.18	0.64	Rr0R1c004767	-8502	PREDICTED: *Raphanus sativus* ethylene-responsive transcription factor 1B
1:43277172:A:G	1.87E-09	-6.37	0.12	0.95	Rr0R1c005403	2297	PREDICTED: *Raphanus sativus* probable thionin-2.4
2:1778823:C:A	1.25E-09	-6.69	0.14	0.98	Rr0R2c006136	-1429	PREDICTED: *Raphanus sativus* putative F-box/kelch-repeat protein At4g35120
2:22901962:C:T	2.25E-10	-4.45	0.18	0.62	Rr0R2c009213	6146	Unknown
2:40785868:T:A	8.90E-09	-6.9	0.1	1.08	Rr0R2c009472	-8408	PREDICTED: *Raphanus sativus* ubiquitin C-terminal hydrolase 12
2:4918566:G:A	9.03E-09	-6.25	0.11	0.98	Rr0R2c006746	967	PREDICTED: *Raphanus sativus* kinesin-like protein KIN-14M (LOC108843169), transcript variant X3, mRNA
3:10737665:A:G	7.41E-10	-4.81	0.17	0.69	Rr0R3c011571	850	PREDICTED: *Raphanus sativus* UDP-glycosyltransferase 73C12
3:13504273:T:C	7.51E-09	-4.34	0.2	0.68	Rr0R3c011893	-606	PREDICTED: *Raphanus sativus* 14-3-3-like protein GF14 mu (LOC108847404), transcript variant X2
3:14174054:A:G	4.11E-10	-6.89	0.13	0.97	Rr0R3c011957	-1358	PREDICTED: *Brassica rapa* inactive endochitinase At2g43600
3:16552729:C:T	5.43E-09	-6.33	0.14	0.98	Rr0R3c012175	-8343	Unknown
3:28559137:T:G	2.92E-10	-5.04	0.17	0.71	Rr0R3c013168	19156	PREDICTED: *Brassica oleracea* var. oleracea mediator of RNA polymerase II transcription subunit 12
3:32208790:G:T	4.48E-09	-6.39	0.14	0.98	Rr0R3c013753	-579	PREDICTED: *Raphanus sativus* leucine-rich repeat receptor-like serine/threonine-protein kinase SKM1
3:3268097:A:C	4.49E-10	-4.03	0.21	0.57	Rr0R3c010436	-2808	Unknown
3:39135325:A:T	1.42E-11	-6.62	0.13	0.85	Rr0R3c014094	27904	Unknown
3:8634742:G:T	6.37E-09	-5.98	0.13	0.93	Rr0R3c011300	-98834	PREDICTED: *Raphanus sativus* E3 ubiquitin-protein ligase SINA-like 2
4:26438604:C:T	2.28E-09	-4.8	0.15	0.72	Rr0R4c017388	-782	PREDICTED: *Raphanus sativus* protein RGF1 INDUCIBLE TRANSCRIPTION FACTOR 1-like
4:26460722:A:T	2.99E-09	-6.67	0.12	1.01	Rr0R4c017391	1174	PREDICTED: *Raphanus sativus* protein TORNADO 2-like

4:26461451:T:A	6.59E-10	-6.97	0.11	1	Rr0R4c017391	8212	PREDICTED: *Raphanus sativus* protein TORNADO 2-like
4:34007442:A:G	5.22E-10	-5.15	0.15	0.73	Rr0R4c018014	2149	Unknown
4:43199943:A:C	4.96E-10	-4.11	0.21	0.59	Rr0R4c018363	1420	PREDICTED: *Raphanus sativus* F-box/LRR-repeat protein At4g14103-like
4:53319373:G:T	1.68E-11	-6.94	0.15	0.89	Rr0R4c019294	6528	PREDICTED: *Raphanus sativus* NDR1/HIN1-like protein 10
4:60774455:G:T	5.85E-10	-5.99	0.15	0.86	Rr0R4c020376	1104	PREDICTED: *Raphanus sativus* piezo-type mechanosensitive ion channel homolog
4:63350484:G:A	3.93E-12	-5.65	0.16	0.7	Rr0R4c020661	1330	PREDICTED: *Raphanus sativus* protein IQ-DOMAIN 5-like
4:65525138:G:T	2.06E-10	-6.01	0.13	0.83	Rr0R4c020921	5288	PREDICTED: *Brassica rapa* flavanone 3-dioxygenase 3
4:8544976:A:G	1.61E-09	-6.49	0.14	0.96	Rr0R4c014997	1184	PREDICTED: *Raphanus sativus* uncharacterized LOC108813793
4:9209800:T:C	9.36E-09	-5.45	0.22	0.86	Rr0R4c015062	-1211	PREDICTED: *Raphanus sativus* probable aminotransferase ACS12 (
5:25179192:G:T	6.76E-10	-7.23	0.11	1.04	Rr0R5c024526	38543	Unknown
5:26525770:A:C	3.96E-10	-6.42	0.12	0.91	Rr0R5c024541	58523	Unknown
5:30671230:A:C	1.07E-09	-3.77	0.26	0.55	Rr0R5c024839	14104	PREDICTED: *Raphanus sativus* heat stress transcription factor C-1
5:36961574:T:C	2.97E-09	-3.94	0.22	0.59	Rr0R5c025354	2764	PREDICTED: *Raphanus sativus* protein PAT1 homolog 1
5:44837510:C:T	4.71E-09	-4.61	0.19	0.71	Rr0R5c026526	11240	PREDICTED: *Raphanus sativus* rab escort protein 1
6:1261238:A:C	8.30E-10	-4.44	0.22	0.64	Rr0R6c027059	12665	PREDICTED: *Raphanus sativus* root phototropism protein 3
6:14855547:T:C	9.88E-10	-6.02	0.14	0.88	Rr0R6c028045	-2906	PREDICTED: *Raphanus sativus* receptor-like protein kinase HSL1
6:15931253:G:A	1.26E-09	-4.89	0.16	0.72	Rr0R6c028197	801	PREDICTED: *Raphanus sativus* proline-rich receptor-like protein kinase PERK10
6:22479046:T:C	3.79E-09	-5.79	0.16	0.88	Rr0R6c029369	1455	PREDICTED: *Raphanus sativus* WAT1-related protein At1g09380-like
7:17743948:C:A	3.05E-09	3.45	0.49	0.52	Rr0R7c030956	2500	PREDICTED: *Raphanus sativus* uncharacterized LOC130499928 (LOC130499928)
7:28100740:A:G	2.33E-09	-6.55	0.12	0.98	Rr0R7c031977	0	PREDICTED: *Raphanus sativus* S-alkyl-thiohydroximate lyase SUR1
7:2989270:A:G	8.33E-09	-4.9	0.14	0.77	Rr0R7c030214	-14459	Unknown
8:10664996:T:A	1.06E-09	-4.48	0.17	0.65	Rr0R8c034553	16945	PREDICTED: *Raphanus sativus* WAT1-related protein At1g44800

8:15155304:G:C	9.42E-09	-6.5	0.12	1.02	Rr0R8c034892	15927	Unknown
8:21030482:G:C	1.97E-11	-5.83	0.15	0.75	Rr0R8c034966	40392	PREDICTED: *Brassica rapa* uncharacterized LOC117132821
8:22053223:A:G	5.32E-09	-5.26	0.14	0.81	Rr0R8c035010	-5266	Unknown
8:22589864:A:T	5.14E-10	-4.2	0.21	0.6	Rr0R8c035059	7386	PREDICTED: *Raphanus sativus* uncharacterized LOC108822786 (LOC108822786), transcript variant X1, mRNA
8:23103665:G:A	6.41E-10	-6.44	0.14	0.92	Rr0R8c035098	5577	Unknown
8:23433730:C:T	1.35E-10	-4.22	0.21	0.58	Rr0R8c035129	-8649	PREDICTED: *Raphanus sativus* secretory carrier-associated membrane protein 4
8:25435199:C:T	1.06E-09	-4.21	0.18	0.61	Rr0R8c035324	-839	PREDICTED: *Raphanus sativus* vegetative cell wall protein gp1
8:29079034:T:A	8.66E-09	-5.36	0.13	0.84	Rr0R8c035687	2037	PREDICTED: *Raphanus sativus* eukaryotic translation initiation factor 3 subunit K
8:39082169:A:C	2.03E-09	-6.16	0.14	0.92	Rr0R8c037314	0	PREDICTED*: Raphanus sativus* high mobility group B protein 15-like
9:15697853:A:G	9.96E-10	-5.16	0.15	0.75	Rr0R9c039508	-10663	PREDICTED: *Brassica napus* protein DETOXIFICATION 8
9:21292065:G:T	5.38E-10	-6.65	0.13	0.95	Rr0R9c040039	11132	PREDICTED*: Brassica napus* cytochrome P450 72A15-like
9:46840239:C:T	3.70E-09	-4.21	0.18	0.64	Rr0R9c040779	52656	PREDICTED: *Raphanus sativus* NADPH:adrenodoxin oxidoreductase, mitochondrial
9:51123460:G:A	8.73E-09	-3.77	0.22	0.59	Rr0R9c041181	-951	Unknown
9:54205187:A:G	6.21E-10	-6.98	0.11	1	Rr0R9c041599	-2171	PREDICTED: *Raphanus sativus* receptor homology region, transmembrane domain- and RING domain-containing protein 2
9:57948695:C:T	7.66E-09	-5.46	0.15	0.85	Rr0R9c042043	0	PREDICTED*: Raphanus sativus* probable N-acetyltransferase HLS1