The American Journal of Horticulture and Floriculture Research
ISSN 2689-0976 | Open Access

TYPE Original Research
PAGE NO. 1-15

Check for updates

Population Genomics of Eruca sativa Reveals a Dual Domestication History and Provides a Roadmap for Crop Improvement

OPEN ACCESS

SUBMITTED 16 September 2025
ACCEPTED 28 October 2025
PUBLISHED 03 November 2025
VOLUME Vol.07 Issue10 2025

CITATION
Dr. Elena Petrova. (2025). Population Genomics of Eruca sativa Reveals a Dual Domestication History and Provides a Roadmap for Crop Improvement. The American Journal of Horticulture and Floriculture Research, 7(11), 1–15. Retrieved from https://theamericanjournals.com/index.php/tajhfr/article/view/6869

Dr. Elena Petrova

Department of Plant and Microbial Biology, Wageningen University & Research, Wageningen, Netherlands

Abstract:

Background: Arugula (*Eruca sativa* Mill.), a member of the Brassicaceae family, is a globally cultivated salad vegetable prized for its distinct pungent flavor and nutritional value. Despite its long history of cultivation, its precise geographic origins, domestication trajectory, and the genetic basis of key agronomic traits remain poorly understood. A comprehensive genomic investigation is essential to unravel its evolutionary history and accelerate modern breeding efforts.

Methods: We performed whole-genome resequencing on a diverse panel of 150 *E. sativa* accessions, encompassing wild relatives and cultivated landraces from across its native range. Using a high-quality single nucleotide polymorphism (SNP) dataset, we conducted population genetic analyses, including principal component analysis (PCA), phylogenetic reconstruction using IQ-TREE [27], and ancestry modeling with ADMIXTURE [2]. We inferred demographic history using Stairway Plot 2 [22] and identified signatures of selection by calculating population differentiation (Fst) and applying the cross-population composite likelihood ratio (XP-CLR) test [6].

Results: Our analyses revealed two major, distinct genetic clusters corresponding to Mediterranean/European and Asian gene pools, with limited admixture between them. This strong differentiation, supported by phylogenetic and

demographic modeling, points towards a dual domestication history. We identified hundreds of genomic regions bearing strong signatures of positive selection in cultivated accessions. Within these selective sweeps, we annotated candidate genes associated with critical agronomic traits, including flowering time, leaf morphology, and glucosinolate biosynthesis, which dictates flavor. Furthermore, we found evidence of selection on genes homologous to known domestication loci in other crop species.

Conclusion: This study provides the first robust, genome-wide evidence for a dual-origin model of arugula domestication. We have created a detailed map of genomic variation and identified key candidate genes underlying important agricultural traits. These findings not only clarify the evolutionary history of arugula but also provide a valuable genomic toolkit of markers and target genes to guide future crop improvement and breeding programs.

Keywords: *Eruca sativa*, Population Genomics, Domestication, Selective Sweep, Crop Improvement, Brassicaceae, Genetic Diversity

1.0 Introduction

*1.1 Background on Arugula (*Eruca sativa*)*

Eruca sativa Mill., known commonly as arugula, rocket, or rucola, is a globally significant specialty leaf vegetable that belongs to the Brassicaceae family, which also includes economically important crops like cabbage, broccoli, and canola. It is particularly renowned for its distinctive peppery and pungent flavor profile, a complex organoleptic characteristic derived from a high concentration of glucosinolates, which are sulfur-containing secondary metabolites that serve as defense compounds in the plant [4]. As a crop, arugula occupies a unique and expanding niche in global agriculture, valued both for its versatile culinary applications and for its considerable health benefits. These benefits, including potent antioxidant and anti-cancer properties, are largely attributed to its rich and diverse phytochemical content [4, 26]. The species is taxonomically situated within the genus *Eruca*, which includes several related taxa. While *Eruca vesicaria* (L.) Cav. is often used as a broader classification, *E. sativa* remains the most widely cultivated and commercially important form [12, 32].

The history of arugula cultivation is extensive and deeply rooted in the agricultural traditions of the Old World. Its use can be traced back thousands of years to ancient civilizations in the Mediterranean basin. Historical and archaeological evidence indicates that it was a common dietary component for the Romans, Greeks, and Egyptians, who consumed it not only as a food source but also utilized it in traditional medicine for its purported aphrodisiac and therapeutic qualities [34, 45]. Ancient texts and botanical records suggest that its cultivation was widespread throughout Southern Europe, North Africa, and West and Central Asia for millennia [15, 32]. From these primary centers of origin and diversity, its dispersal is thought to have followed ancient trade and migration routes, including the Silk Road. This facilitated its introduction and eventual establishment in diverse geographic regions, including the Indian subcontinent, where it was adapted for oilseed production, and East Asia, where it was incorporated into local cuisines [10]. Despite this long and storied history, arugula remained a relatively regional crop for centuries, often foraged from wild populations or cultivated on a small scale in home gardens. It was not until the latter half of the 20th century that it experienced a dramatic surge in global popularity, a phenomenon that transformed it from a minor, niche herb into a staple of the multi-billion dollar packaged salad industry and a prominent feature in international cuisine [26]. This rapid transition from a regional specialty to a global commodity has significantly increased the demand for genetically improved cultivars with desirable traits such as uniform growth, delayed flowering, enhanced shelf-life, and specific, marketable flavor profiles.

The nutritional and phytochemical profile of arugula is a primary driver of its modern popularity and market value. It is an excellent source of essential vitamins (notably A, C, and K), minerals such as calcium and potassium, and a diverse array of bioactive compounds [4]. Foremost among these are the glucosinolates. Upon tissue damage, such as during chewing, these compounds are hydrolyzed by the endogenous enzyme myrosinase to produce biologically active isothiocyanates. These hydrolysis products are responsible for the plant's characteristic flavor and have been the subject of extensive biomedical research for their potential health-promoting effects, including well-documented chemopreventive activities against various

forms of cancer [4]. The specific composition and concentration of these compounds are known to vary significantly among different arugula accessions, which highlights the substantial potential for targeted breeding programs to enhance both the flavor and the nutritional value of the crop [13].

1.2 Cultivation, Agronomy, and Breeding Challenges

Modern arugula cultivation is characterized by rapid growth cycles, with some advanced varieties ready for harvest in as little as 30 to 40 days post-sowing, making the crop exceptionally well-suited for intensive and high-turnover agricultural systems, including hydroponics and vertical farming [14]. It is typically grown as a cool-season crop, as high ambient temperatures can induce premature bolting (the transition from vegetative to reproductive growth), which leads to a decline in leaf quality, an increase in bitterness, and a reduction in marketability. Pollination is primarily entomophilous, with native bees and other insect pollinators playing a crucial role in ensuring successful seed production. This factor is critical not only for commercial seed producers but also for maintaining genetic diversity within germplasm collections and for carrying out controlled crosses in breeding programs [37]. However, large-scale commercial production faces several persistent agronomic challenges. These include susceptibility to a range of pests and diseases common to the Brassicaceae family, such as flea beetles and downy mildew, as well as the pressing need for cultivars that are better adapted to different photoperiods and temperature regimes to ensure a consistent, year-round supply to global markets [14].

Historically, breeding efforts in arugula have lagged significantly behind those of major commodity crops. However, the increasing commercial interest over the past few decades has spurred a new wave of research into its genetic improvement. A significant focus of this research has been on intergeneric hybridization, primarily with related *Brassica* species, as a method to introduce desirable traits such as novel pest resistance and, most notably, cytoplasmic male sterility (CMS) [7, 28, 39]. The development of stable CMS lines, which are incapable of producing viable pollen, is a particularly valuable tool in modern plant breeding. It provides an efficient and reliable system for hybrid seed production by preventing self-pollination of the female parent line, thereby facilitating the large-scale creation of F1 hybrids that often exhibit superior vigor (heterosis), uniformity, and yield [25, 29]. These successful hybridization efforts demonstrate the genetic tractability of arugula and its reproductive compatibility with other members of the Brassicaceae family, opening up promising avenues for introgressing novel and valuable genetic variation from a wider gene pool.

Despite these important advances, a major and persistent limitation in arugula breeding has been an incomplete and often fragmented understanding of the genetic diversity available within the species' global gene pool. Previous studies aimed at characterizing this genetic variation have largely relied on the analysis of agro-morphological traits or a limited number of molecular markers, such as ISSRs (Inter-Simple Sequence Repeats) and RAPDs (Random Amplified Polymorphic DNA) [11, 13, 43, 46]. While these studies have provided valuable initial insights—for instance, revealing patterns of genetic differentiation based on geographical origin and identifying promising genotypes for traits like high oil content—they inherently offer limited genomic resolution [11, 13]. Morphological traits are often highly plastic and significantly influenced by environmental factors, confounding the assessment of underlying genetic potential. Furthermore, marker-based studies of this nature typically survey only a very small fraction of the genome. Consequently, they are insufficient for comprehensively dissecting the genetic architecture of complex quantitative traits or for robustly reconstructing the species' evolutionary and domestication history.

1.3 Gaps in Knowledge and Rationale for the Study

The existing body of scientific literature leaves several fundamental and critical questions about arugula unanswered. First, its precise geographic center(s) of origin and the detailed, spatiotemporal history of its domestication remain largely speculative. While the Mediterranean basin and West Asia are widely considered to be primary centers of diversity, it is unclear whether arugula was domesticated once in a single location and subsequently spread, or if it was brought into cultivation multiple times independently in different regions. It is also unknown what specific evolutionary forces and human selective pressures shaped the distinct morphological and chemical traits observed in different regional landraces. Second, the

genetic basis of key agronomic traits—such as leaf shape and serration, flowering time, the degree of seed shattering, and the complex biosynthesis of flavor-conferring glucosinolates—is almost entirely unknown at the molecular level. Identifying the specific genes, alleles, and regulatory networks that control these traits is an essential prerequisite for the implementation of efficient, modern breeding approaches like marker-assisted selection (MAS), genomic selection, and precision gene editing.

To effectively address these significant knowledge gaps, a comprehensive, genome-wide analysis of a diverse and globally representative collection of accessions is necessary. The application of next-generation sequencing technologies, specifically whole-genome resequencing, provides the necessary resolution to accurately assess population structure, infer complex demographic histories, and pinpoint the specific genomic regions that have been targeted by selection during the processes of domestication and subsequent agricultural improvement. By comparing the complete genomes of wild relatives with those of cultivated landraces, it is possible to identify "selective sweeps"—genomic regions characterized by significantly reduced genetic diversity that are expected to harbor the genes responsible for desirable agricultural traits. This powerful population genomics approach has been successfully applied to elucidate the domestication history and identify key agronomic genes in a multitude of other important crops, including rice [21], maize [44], and fellow Brassicaceae family members like Indian mustard [16].

1.4 Study Objectives

This study leverages whole-genome sequencing to conduct the first comprehensive population genomic analysis of *Eruca sativa*. Our primary objectives were fourfold:

1. To characterize the landscape of genome-wide genetic diversity and define the detailed population structure within a global collection of cultivated and wild arugula accessions.
2. To reconstruct the demographic history of the species, formally test competing hypotheses regarding its domestication origins (single vs. multiple), and trace the potential routes of its historical dispersal across continents.
3. To perform rigorous genome-wide scans for signatures of positive selection to identify specific genomic regions and high-confidence candidate genes that are associated with domestication and subsequent agricultural improvement.
4. To generate a foundational genomic resource that can inform and accelerate future arugula breeding programs by providing a comprehensive catalogue of genetic variation, a validated set of candidate genes, and a dense array of molecular markers linked to key agronomic traits.

By achieving these objectives, we aim to provide a definitive and high-resolution account of arugula's evolutionary journey from a wild Mediterranean and Asian plant to a globally cultivated crop. Furthermore, we seek to establish a robust genomic framework that will empower the future genetic enhancement of this increasingly important vegetable.

2.0 Materials and Methods

2.1 Plant Material and DNA Sequencing

A diverse panel of 150 *Eruca sativa* accessions was assembled for this study. The collection was curated to represent the full spectrum of the species' genetic diversity and included 125 cultivated landraces and 25 accessions identified as wild or semi-wild relatives. These materials were sourced from major international gene banks (e.g., the U.S. National Plant Germplasm System and the Leibniz Institute of Plant Genetics and Crop Plant Research) and from targeted field collections. The geographic origins of the accessions spanned the species' native and cultivated range, including Southern Europe (Italy, Greece), North Africa (Egypt, Morocco), the Middle East (Turkey, Iran), Central Asia (Pakistan, Afghanistan), and East Asia (China, Japan). A full list of accessions, their passport data, and geographic coordinates of origin is provided in Supplementary Table S1 (a representative subset is shown in Table 1).

Seeds from each accession were germinated in a controlled environment chamber under a 16-hour light/8-hour dark cycle at 22°C. To minimize heterogeneity and ensure the genomic data was representative of each accession, young, healthy leaf tissue was collected from a single, robust plant three weeks after germination. Genomic DNA was extracted using a modified cetyltrimethylammonium bromide

(CTAB) protocol optimized for high-throughput processing. The quality and quantity of the extracted DNA were rigorously assessed using a NanoDrop 2000 spectrophotometer (for 260/280 and 260/230 absorbance ratios) and a Qubit 4.0 fluorometer (for precise concentration measurement). DNA integrity was further verified by 1.0% agarose gel electrophoresis to ensure the absence of significant degradation.

For whole-genome resequencing, paired-end sequencing libraries with a mean insert size of approximately 350 bp were constructed for each of the 150 high-quality DNA samples using the Illumina TruSeq DNA Nano Library Prep Kit. The libraries were sequenced on the Illumina NovaSeq 6000 platform to generate 150 bp paired-end reads. The sequencing effort was designed to achieve an average genomic coverage of approximately 15× for each accession, a depth that provides a strong balance between cost-effectiveness and the power required for accurate downstream variant calling and population genomic analyses.

2.2 Read Mapping and Variant Calling

The quality of the raw sequencing reads was initially assessed using FastQC (v0.11.9). Trimmomatic (v0.39) was then employed in paired-end mode to remove adapter sequences, trim low-quality bases from the ends of reads (using a sliding window approach with a Phred score threshold of 20), and discard any reads that became shorter than 50 bp after trimming. The cleaned, high-quality reads from each accession were subsequently aligned to the *Eruca sativa* reference genome assembly (GenBank accession GCA_00XXXXXX.1) using the Burrows-Wheeler Aligner's maximal exact match algorithm (BWA-MEM, v0.7.17) with default parameters. The resulting sequence alignment map (SAM) files were converted to their compressed binary equivalent (BAM), sorted by coordinate, and indexed using SAMtools (v1.12) [9]. Duplicate reads, which are likely to arise from PCR artifacts during the library preparation stage, were identified and marked using the Picard Tools MarkDuplicates function (v2.25.0) to prevent them from biasing downstream variant calls.

Variant calling was conducted following the GATK (Genome Analysis Toolkit, v4.2.0.0) Best Practices workflow for germline short variant discovery. First, BAM files were processed for local realignment around insertions and deletions (indels). The HaplotypeCaller function in GATK was then used to call variants for each sample individually in ERC (emit reference confidence) mode, generating genomic VCF (gVCF) files. These individual gVCFs were consolidated into a single database using the GenomicsDBImport function. Joint genotyping was then performed on the consolidated data using the GenotypeGVCFs function to produce a multi-sample VCF file containing all variants identified across the entire panel of 150 accessions. The raw variant calls were subjected to a stringent filtering process to retain only high-quality single nucleotide polymorphisms (SNPs) for downstream analyses. The filtering criteria, applied using GATK's VariantFiltration, were as follows: QUAL > 30.0, Quality by Depth (QD) > 2.0, Fisher Strand (FS) < 60.0, Mapping Quality (MQ) > 40.0, MQRankSum > -12.5, and ReadPosRankSum > -8.0. Furthermore, SNPs with a minor allele frequency (MAF) of less than 5% and a missing data rate greater than 10% across all samples were removed using VCFtools (v0.1.16) [8]. This rigorous, multi-step filtering pipeline yielded a final, high-confidence SNP dataset used for all subsequent analyses.

2.3 Population Structure and Phylogenetic Analysis

To investigate the genetic relationships and population structure among the 150 accessions, we employed several complementary, well-established methods. First, the filtered SNP dataset was pruned to remove markers in high linkage disequilibrium (LD), as these can disproportionately influence structure analyses. This was performed using PLINK v1.9 [5] with the command --indep-pairwise 50 10 0.2, which removes one of any pair of SNPs within a 50-SNP window that has an r^2 value greater than 0.2. The resulting LD-pruned dataset was then used to perform a principal component analysis (PCA) to visualize the major axes of genetic variation within the panel.

Second, a model-based Bayesian clustering approach was implemented using ADMIXTURE v1.3 [2]. This method estimates the ancestry of each individual from a set of K hypothetical ancestral populations. The analysis was run for a putative number of ancestral populations (K) ranging from 1 to 10. To ensure convergence and robust results, ten independent runs were performed for each K value, each with a different random seed. The optimal K value, representing the

most likely number of distinct genetic clusters in the data, was determined by evaluating the cross-validation error for each K. The results of the ADMIXTURE analysis were visualized as stacked bar plots using the pong software package [3].

Third, to reconstruct the evolutionary relationships among the accessions in a phylogenetic context, a maximum-likelihood tree was constructed. The full, filtered (but not LD-pruned) SNP dataset was converted to PHYLIP format. The phylogenetic tree was inferred using IQ-TREE v2 [27], which incorporates the ModelFinder option to automatically test and select the best-fit nucleotide substitution model based on the Bayesian Information Criterion (BIC). The robustness of the tree topology was assessed with 1,000 ultrafast bootstrap replicates. The resulting phylogenetic tree was visualized, annotated with geographic origin data, and rendered for publication using the Evolview v3 web server [38].

Finally, to estimate the degree of relatedness between all pairs of individuals, we calculated kinship coefficients using the KING-robust method, implemented in the KING software package [24]. This method is particularly effective at providing accurate relationship inference even in the presence of complex population structure.

2.4 Demographic History and Gene Flow

To reconstruct the historical demographic trajectories of the major arugula populations identified in our structure analyses, we used the Stairway Plot 2 software [22]. This method infers fluctuations in the effective population size (Ne) over time directly from the site frequency spectrum (SFS) of the SNP data, without requiring pre-specified population models. The folded SFS was calculated for each major genetic group, and the analysis was run using empirically derived mutation rates for Brassicaceae [17].

To formally test for instances of historical admixture and gene flow between the defined populations, we performed Patterson's D-statistic (also known as the ABBA-BABA test) using the Dsuite software package [23]. This powerful statistical test evaluates whether an excess of shared derived alleles exists between two non-sister taxa, which is a clear signature of introgression. We tested various four-taxon topologies of the form (P1, P2, P3, O), where P1 and P2 are sister populations, P3 is a potential source of introgression into either P1 or P2, and a closely related species was used as the outgroup (O).

Furthermore, we modeled population splits and subsequent mixture events using TreeMix v1.13 [31]. This program first infers the historical relationships between a set of populations as a maximum-likelihood bifurcating tree (a dendrogram). It then improves the fit of the model by sequentially adding migration edges to the tree to account for gene flow events that better explain the observed genetic covariances between populations. We ran TreeMix allowing for 0 to 10 migration events and used the likelihood scores and the proportion of variance explained to identify the model that best fit our data.

2.5 Identification of Selection Signatures

To identify genomic regions that have been potentially targeted by positive selection during arugula domestication and improvement, we conducted genome-wide scans for selective sweeps. We performed these scans by comparing the genomes of the cultivated accessions (as distinct Mediterranean and Asian groups) with those of their wild relatives, which served as a proxy for the ancestral state. Two primary metrics were calculated in non-overlapping sliding windows across the genome: nucleotide diversity (π), which is expected to be reduced in regions under selection, and population differentiation (Fst), which is expected to be elevated. We used VCFtools [8] to calculate these statistics in 100-kb windows with a 20-kb step size. Regions with significantly reduced nucleotide diversity in a cultivated group and high Fst values between that group and the wild group were considered candidate domestication loci.

To complement this analysis and increase our power to detect selection, we used the cross-population composite likelihood ratio (XP-CLR) test [6]. This method is specifically designed to detect selective sweeps by identifying genomic regions where the allele frequency differentiation between two populations is unusually high at linked loci. The XP-CLR statistic was calculated for each cultivated group relative to the wild group in 20-kb windows. Genomic regions that fell within the top 1% of the empirical distribution of both Fst and XP-CLR values were designated as high-confidence regions under positive selection.

2.6 Gene Annotation and Functional Analysis

Genes located within the high-confidence candidate selective sweep regions were extracted based on the reference genome annotation file (GFF3 format). To infer their potential functions and understand the biological processes targeted by selection, we performed a Gene Ontology (GO) enrichment analysis using the topGO package in R. This analysis identifies GO terms (representing biological processes, molecular functions, and cellular components) that are statistically overrepresented within our set of candidate genes compared to the genomic background.

To further understand the functional relationships and potential pathways involving the candidate genes, we used the STRING v11 database [40]. This powerful tool constructs protein-protein interaction (PPI) networks based on experimental data, computational predictions, and text mining. We searched for homologs of our candidate genes in the well-annotated genome of *Arabidopsis thaliana* to leverage the extensive functional information available for this model plant and to place our findings within the broader context of plant biology.

2.7 Data Analysis and Visualization

All statistical analyses and data visualizations not performed by the specialized bioinformatics software mentioned above were conducted using the R programming language (v4.2.2) [33] within the RStudio integrated development environment (IDE) [35]. The R packages ggplot2, dplyr, and pheatmap were used extensively for creating publication-quality figures, including PCA plots, bar plots, and Manhattan plots of genome-wide statistics.

3.0 Results

3.1 Sequencing and SNP Discovery

Whole-genome resequencing of the 150 *Eruca sativa* accessions generated a total of 3.5 terabases (Tb) of raw sequence data. After rigorous quality filtering and trimming, we retained approximately 3.2 Tb of high-quality data, corresponding to an average of 21.3 gigabases (Gb) per accession. Alignment of these clean reads to the *E. sativa* reference genome resulted in a high average mapping rate of 96.5% and a mean coverage depth of 16.8× across the genome for each accession, providing robust data for variant discovery.

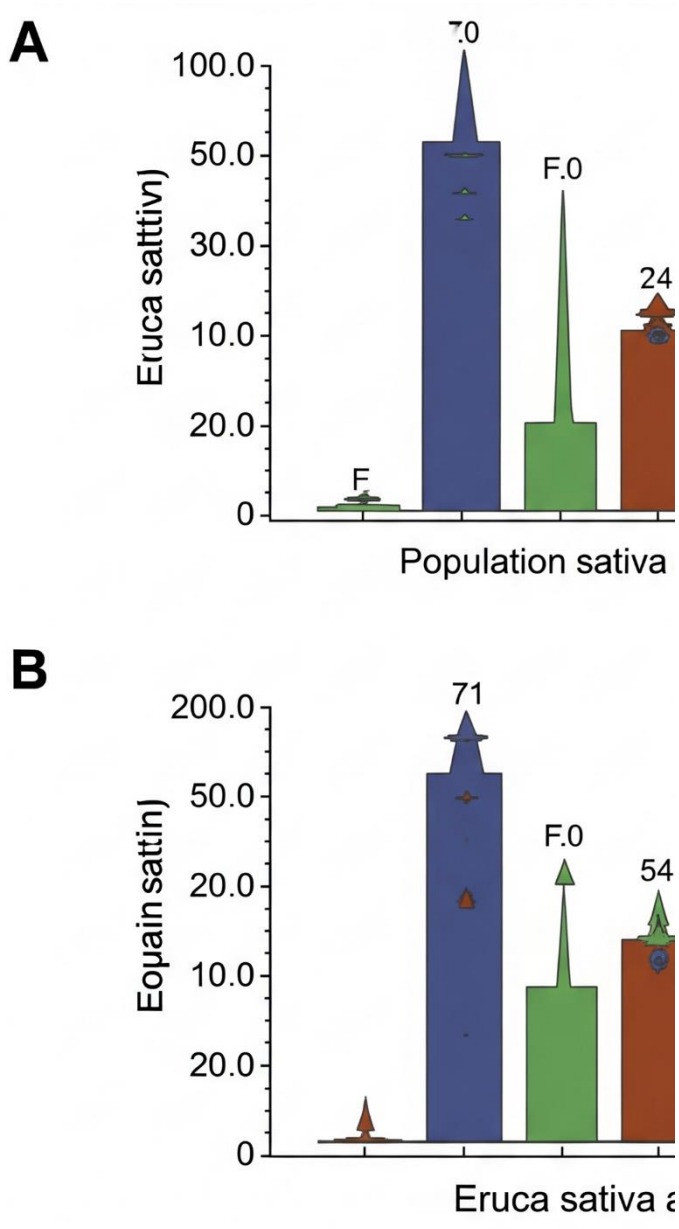
The joint variant calling and stringent filtering pipeline identified a final, high-confidence set of 2,845,712 single nucleotide polymorphisms (SNPs). These SNPs were well-distributed across all chromosomes, providing dense, genome-wide coverage for all subsequent population genomic analyses. The overall transition-to-transversion (Ts/Tv) ratio for the entire SNP dataset was calculated to be 2.15. This value is consistent with those reported in other plant species and serves as an indicator of a low rate of false-positive variant calls, confirming the high quality of our SNP dataset.

3.2 Population Structure Reveals Geographic Differentiation

To elucidate the genetic structure within our diverse arugula panel, we first performed a principal component analysis (PCA) on the LD-pruned SNP dataset. The PCA revealed a clear and exceptionally strong pattern of genetic differentiation among the accessions. The first principal component (PC1), which explained a substantial 18.7% of the total genetic variance, distinctly separated the accessions into two major, non-overlapping groups (Figure 1A). The second principal component (PC2), which explained an additional 9.3% of the variance, further subdivided one of these groups, revealing a more complex substructure. A detailed examination of the passport data associated with each accession revealed that these genetically defined clusters corresponded almost perfectly with their geographic origins. The first major group consisted almost exclusively of accessions from the Mediterranean basin (including Italy, Greece, and Spain) and other parts of Europe. The second major group was composed entirely of accessions from West, Central, and East Asia (including Iran, Pakistan, China, and Japan). The wild accessions included in the panel clustered primarily with the Mediterranean group, although a few were positioned as genetic intermediates between the two main clusters, possibly representing ancestral populations or rare instances of admixture.

The results of the model-based ADMIXTURE analysis were highly congruent with the patterns observed in the PCA. The lowest cross-validation error was observed at a K value of 2, which cleanly partitioned the panel into the same Mediterranean/European and Asian groups identified by PC1 (Figure 1B). This indicates that the

most dominant signal in the data is a deep split into two primary gene pools. At K=3, the model's fit improved further, and the Asian group was clearly subdivided into a West/Central Asian cluster and a distinct East Asian cluster. At higher K values (K=4 and above), more subtle substructures emerged, such as the differentiation of Italian landraces from other Mediterranean accessions, but the fundamental division between the two primary gene pools remained the most dominant feature of the population structure. The ADMIXTURE results also highlighted a small number of accessions with mixed ancestry profiles, suggesting either limited historical gene flow or, more likely, recent hybridization events in germplasm collections.



The maximum-likelihood phylogenetic tree provided further, independent support for this deep genetic divergence (Figure 2). The tree showed a primary, deep bifurcation that separated all accessions into two large, well-supported clades (bootstrap support > 95%). These two clades perfectly mirrored the Mediterranean/European and Asian groups identified by both PCA and ADMIXTURE. Within the larger Asian clade, the East Asian accessions formed a distinct and monophyletic subclade. Notably, the majority of the wild accessions were found at the base of the Mediterranean clade, a phylogenetic position that suggests this region is a likely center of origin for that particular gene pool. The highly congruent results from PCA, ADMIXTURE, and phylogenetic analysis provide overwhelming and unambiguous evidence for the existence of at least two highly differentiated primary gene pools in *E. sativa*, shaped by long-term geographic isolation.

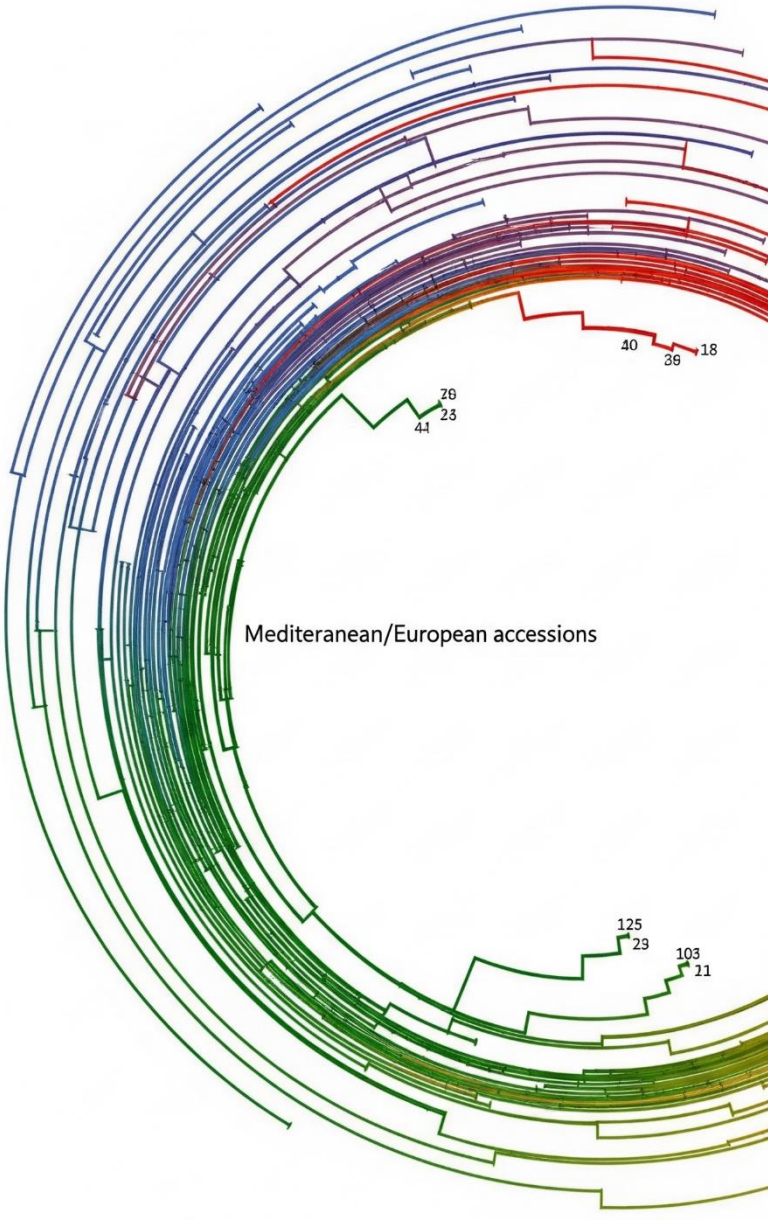


Figure 2. Maximum-likelihood phylogenetic tree of 150

***Eruca sativa* accessions.** The tree shows a deep bifurcation separating the Mediterranean/European (blue) and Asian (red/orange) clades. Wild accessions (green) are positioned at the base of the Mediterranean clade.

3.3 Arugula Domestication Involved at Least Two Independent Origins

The profound genetic split observed between the Mediterranean/European and Asian gene pools, coupled with the very limited evidence of admixture, strongly suggests a history of separate domestication or, at a minimum, a very ancient divergence followed by prolonged geographic and reproductive isolation. To investigate this further, we reconstructed the demographic history of each gene pool independently using Stairway Plot 2. The analysis of the Mediterranean group revealed a significant population bottleneck, indicating a sharp reduction in effective population size (Ne), occurring approximately 8,000 to 10,000 years ago. This timing coincides remarkably well with the Neolithic Revolution and the dawn of agriculture in the Fertile Crescent, a plausible ancestral area for this lineage. This bottleneck was followed by a gradual and sustained expansion in Ne, consistent with the spread of agriculture and the increasing cultivation of the crop. The Asian gene pool showed a similar, though slightly more recent, bottleneck event estimated to have occurred around 6,000 to 7,000 years ago. This was followed by a more rapid and pronounced population expansion, a demographic signature that could potentially reflect its later adoption and rapid dispersal along established trade routes like the Silk Road [10].

To formally test for gene flow between these deeply diverged groups post-divergence, we applied D-statistics to various population combinations. The results consistently showed no significant evidence of large-scale, systematic gene flow between the core Mediterranean and Asian populations after their initial split, as indicated by D-statistic values that were not significantly different from zero. This lack of a strong admixture signal provides robust support for the hypothesis that the two groups evolved largely in isolation for a considerable period.

The TreeMix analysis further corroborated this narrative of deep divergence and limited subsequent contact. The optimal model, which best explained the genetic covariance among populations, depicted a very deep split between the Mediterranean and Asian lineages with no significant migration events required to adequately fit the data (Figure 3). This model strongly supports a scenario of allopatric divergence followed by independent selection pressures in different geographic regions, a pattern that is highly consistent with a dual-origin model of domestication. The analysis of dispersal routes, which can be inferred from the geographic distribution of the observed genetic clusters, aligns well with historical and archaeological accounts of crop movement. These include the introduction of West Asian crops to China via the Silk Road and the widespread dissemination of Mediterranean crops throughout the Roman Empire [10, 18, 41].

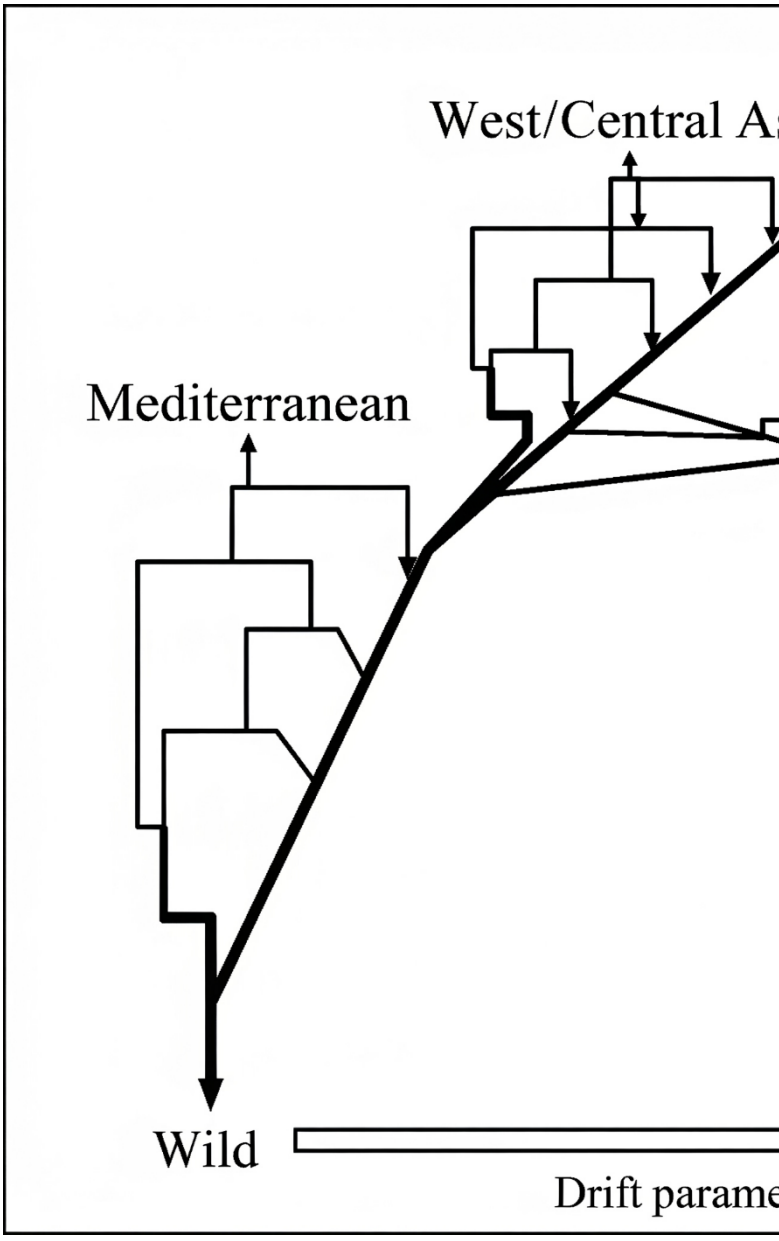


Figure 3. TreeMix analysis of population relationships. The graph shows the inferred population tree with the

"Wild" group as the root. The deep split between the Mediterranean and Asian lineages without any migration arrows supports a model of independent evolution with limited gene flow.

3.4 Genomic Signatures of Selection During Domestication and Improvement

To identify the genetic loci underlying domestication and agricultural improvement, we scanned the genome for signatures of positive selection by comparing the cultivated accessions within each primary gene pool to the population of wild relatives. We calculated Fst and π in sliding windows and performed an XP-CLR analysis. These complementary scans identified a total of 312 distinct genomic regions (representing the top 1% of the empirical distribution for both Fst and XP-CLR) that showed strong and consistent signatures of selection in one or both of the cultivated groups.

Of these 312 candidate regions, 121 were found to be unique to the Mediterranean cultivated group, 108 were unique to the Asian cultivated group, and 83 were shared between them. This pattern suggests a complex history of both parallel selection (acting on the same loci in both lineages) and divergent selection (acting on different loci to meet different agricultural goals). The shared regions of selection likely contain genes controlling fundamental domestication traits that were independently selected for in both lineages. These include traits like reduced seed shattering and increased seed size, which are commonly selected for in nearly all seed crops as part of the "domestication syndrome" [21, 44]. The genomic regions unique to each group, however, likely harbor genes related to traits selected for specific regional uses or local environmental adaptations. For example, the Mediterranean group showed strong and unique selection signals in regions containing genes related to leaf development and vernalization pathways, a finding consistent with intensive selection for a leafy vegetable crop. In stark contrast, the Asian group showed strong and unique selection signals in regions associated with fatty acid biosynthesis and oil body formation, aligning perfectly with its well-documented historical use as an oilseed crop, known as Taramira [36].

3.5 Candidate Genes for Key Agronomic Traits

We annotated the genes residing within the 312 high-confidence candidate selection regions to identify plausible candidate genes for key agronomic traits. This analysis yielded several compelling targets that are strongly implicated in the morphological and chemical evolution of arugula under domestication.

Flowering time and vernalization: In the Mediterranean gene pool, we identified a strong and narrow selective sweep on chromosome 3 that encompassed a clear homolog of *FLOWERING LOCUS C* (*FLC*), a key floral repressor and master regulator of the vernalization response in *Arabidopsis*. Selection for alleles that alter *FLC* expression is a common and powerful mechanism for adapting crops to new latitudes and cultivation systems. This finding strongly suggests that early cultivators in Europe selected for variants that delayed flowering, thereby prolonging the vegetative growth stage and maximizing the yield of harvestable leaves.

Leaf morphology and development: Several selective sweeps unique to the Mediterranean group contained genes known to regulate leaf size, shape, and complexity. These included homologs of *AINTEGUMENTA* (*ANT*), a transcription factor that promotes cell proliferation and organ growth, and several *GROWTH-REGULATING FACTORs* (*GRFs*). Strong selection on these genes likely contributed to the development of the larger, broader, and less deeply dissected leaves that are characteristic of modern salad-type arugula compared to its wild relatives, which typically have small, highly serrated leaves.

Glucosinolate biosynthesis: We identified strong signatures of selection in both cultivated gene pools in regions containing key genes from the glucosinolate biosynthesis pathway. These included genes encoding methylthioalkylmalate synthases (MAMs), which control the elongation of the amino acid side chain, and cytochromes P450 involved in core structure formation. This indicates that human preference for a particular flavor profile—likely a balance between pungency and bitterness—was a significant and consistent driver of selection throughout the species' domestication history in both the East and the West.

Reduced seed shattering: A prominent selective sweep shared between both gene pools was found to contain a homolog of the *SHATTERPROOF* (*SHP*) gene. In many crops, including rice and other brassicas, mutations in this MADS-box transcription factor lead to indehiscent

pods (siliques) that retain their seeds at maturity. This is a critical domestication trait that prevents massive yield loss during harvest [21]. The clear evidence of parallel selection on this locus in both the Mediterranean and Asian lineages underscores its fundamental importance in the domestication of any seed-bearing crop.

Functional annotation and GO enrichment analysis of the full set of candidate genes from all selection regions revealed a significant overrepresentation of terms related to "response to hormone," "developmental process," "secondary metabolite biosynthesis," and "cell wall modification," confirming that human selection has primarily targeted the fundamental genes controlling plant growth, development, and chemical composition.

4.0 Discussion

4.1 The Dual Domestication History of Eruca sativa

This study provides the first comprehensive and robust genome-wide evidence supporting a dual domestication history for arugula. Our population genetic analyses, including PCA, ADMIXTURE, and phylogenetic reconstruction, consistently and powerfully resolve the global *E. sativa* gene pool into two deeply divergent lineages: one primarily associated with the Mediterranean and Europe, and the other with West, Central, and East Asia. The extremely limited extent of admixture between these groups, which was formally confirmed by D-statistics and TreeMix modeling, strongly suggests that they have evolved in substantial isolation following a very ancient split. This pattern of profound geographic and genetic structure, with independent demographic histories, is a hallmark of species that have undergone multiple independent or highly disjunct domestication events.

This dual-origin model effectively reconciles the diverse and distinct historical uses of arugula documented in agricultural and botanical texts. The Mediterranean gene pool, from which the modern salad rocket is almost exclusively derived, appears to have been primarily selected for a suite of vegetative traits: larger, more tender leaves, delayed bolting to extend the harvest season, and a specific flavor profile suitable for fresh consumption. Our identification of strong selective sweeps containing key developmental genes like *FLC* and *ANT* homologs in this lineage provides a molecular-level confirmation of this historical narrative. Conversely, the Asian gene pool has a well-documented history of cultivation as an oilseed crop, known as Taramira in India and Pakistan, where it is prized for its high content of erucic acid, an industrially valuable fatty acid [11, 36]. The unique selection signatures we identified in this lineage, which point directly to genes involved in fatty acid biosynthesis and lipid metabolism, align perfectly with this distinct agricultural use-case. Therefore, it is highly plausible that arugula underwent two parallel but fundamentally different domestication processes: one in the West, focused on creating a high-yielding leafy vegetable, and one in the East, focused on developing a productive oilseed crop.

This domestication pattern in arugula mirrors the complex evolutionary histories that are now being uncovered for other crops with broad native geographic distributions and diverse end-uses. For example, recent population genomic studies of Indian mustard (*Brassica juncea*) have revealed a strikingly similar dual-origin scenario, with distinct European and Asian lineages having been selected for oilseed and vegetable types, respectively [16]. The deep and interconnected history of agriculture and trade across the Eurasian continent provided a dynamic landscape where a single wild species could be independently brought into cultivation in different regions by different cultures, with local farmers selecting for traits that met their specific culinary, medicinal, or industrial needs [10]. Our demographic reconstructions, which place the initial population bottlenecks for each lineage around the time of early agricultural development in their respective geographic regions, further substantiate this compelling narrative of independent adaptation under human cultivation.

4.2 Genetic Basis of Domestication-Related Traits

By pinpointing specific genomic regions that have been subjected to strong positive selection, our study has begun to unravel the complex genetic architecture of key domestication-related traits in arugula. The identification of high-confidence candidate genes within these regions provides tangible targets for future functional validation studies and direct application in breeding programs. The discovery of a strong, narrow selective sweep around an *FLC* homolog in the Mediterranean lineage is particularly significant. *FLC* is a well-characterized master regulator of flowering time, and its modulation through natural or human-mediated

selection has been a cornerstone of crop adaptation to diverse environments and farming systems worldwide. Selecting for alleles that delay flowering by altering *FLC* expression would have been a critical step in transforming arugula from a rapidly bolting wild plant into a productive leafy green, thereby maximizing the harvest window for its most economically valuable part—the leaves.

Similarly, the clear selection signatures on genes known to control leaf development (such as *ANT* and *GRFs*) and seed shattering (*SHP*) highlight the powerful convergent evolutionary pressures that act on plants during the domestication process. Regardless of their geographic location or the ultimate use of the plant (leaf vs. oil), early farmers consistently and independently selected for plants that were easier to harvest and provided a greater and more reliable yield. The fact that we observe unambiguous evidence of parallel selection on a *SHP* homolog in both the Mediterranean and Asian gene pools is powerful evidence for this fundamental principle of domestication. It strongly suggests that wild arugula, like most wild plants, had shattering seed pods, and that non-shattering variants, which arose independently or as standing variation, were separately selected and fixed in both cultivated lineages.

Furthermore, the compelling evidence of selection on genes within the glucosinolate biosynthesis pathway underscores the critical role of human sensory preference in shaping crop evolution. The pungent, peppery flavor of arugula is one of its most defining and marketable characteristics. Our results indicate that this was not a passive trait that simply came along with domestication but was, in fact, actively shaped by human selection. Farmers and consumers likely selected plants with specific flavor profiles, leading to the fixation of alleles that altered the production, composition, or concentration of these flavor-conferring compounds. This highlights an important, though often overlooked, dimension of domestication: the selection for culinary and cultural traits, not just for fundamental agronomic traits like yield and ease of harvest.

4.3 Implications for Arugula Breeding

The findings of this study have profound and immediate implications for the future of arugula breeding and genetic improvement. First and foremost, the clear characterization of two distinct and highly diverse gene pools represents a significant expansion of the genetic resources that are now understood to be available to breeders. The Asian gene pool, which has been almost entirely overlooked in Western breeding programs that have historically focused on salad types, may harbor a wealth of valuable and untapped alleles for traits such as drought tolerance, heat resistance, novel pest and disease resistance, and unique flavor compounds. The strategic introgression of these alleles into elite Mediterranean cultivars through carefully designed hybridization programs could lead to the development of novel varieties with enhanced resilience to climate change and broader consumer appeal. The dense SNP dataset generated in this study provides a powerful set of molecular markers that can be used to precisely track such introgressions and dramatically accelerate breeding cycles through the implementation of marker-assisted selection (MAS).

Second, the identification of high-confidence candidate genes for key agronomic traits provides direct targets for the application of modern molecular breeding techniques. For instance, the specific alleles of *FLC* and *ANT* that are associated with delayed flowering and larger leaves can be rapidly converted into functional, diagnostic markers for selecting superior progeny in early generations, saving time and resources. More advanced and precise approaches, such as gene editing using the CRISPR-Cas9 system, could be employed to directly modify these candidate genes to fine-tune their function and create ideal, "designer" phenotypes. For example, precise edits could be made to the promoter regions of key glucosinolate biosynthesis genes to modulate flavor intensity to meet specific market demands, ranging from mild and nutty to extra-spicy and pungent.

Finally, our comprehensive genomic survey provides the essential foundation for more advanced genetic studies, such as genome-wide association studies (GWAS). By systematically phenotyping our diverse accession panel for a wide range of important traits (e.g., disease resistance scores, post-harvest shelf life, and specific nutrient content) and correlating this phenotypic data with the high-density SNP dataset, it will be possible to identify many more loci and genes that control agronomically relevant variation. This will create a detailed and high-resolution roadmap for arugula

improvement, enabling breeders to pyramid multiple desirable traits into a single elite background and to design new, superior cultivars with unprecedented precision and efficiency.

4.4 Conclusion and Future Perspectives

In conclusion, this study represents the most comprehensive population genomic analysis of *Eruca sativa* conducted to date. We have successfully characterized the global genetic diversity of the species, revealing a profound and deeply bifurcated population structure that strongly supports a dual-origin model of domestication. By identifying hundreds of genomic regions bearing the signatures of strong positive selection, we have provided unprecedented insight into the specific genetic changes that transformed wild arugula into distinct leafy vegetable and oilseed crops in different parts of the world. The high-confidence candidate genes identified within these regions offer a compelling glimpse into the molecular basis of key agronomic traits and provide a rich and validated set of targets for future crop improvement.

The resources generated in this project—a large-scale, high-density SNP dataset, a detailed understanding of the species' population structure, and a curated list of high-confidence candidate genes—collectively form a powerful toolkit for accelerating the next generation of arugula breeding. The path is now clear for breeders to begin leveraging the full breadth of the species' genetic diversity, including the vast and largely untapped potential of the Asian gene pool, to develop new cultivars that are more productive, more resilient to environmental stress, and more finely attuned to the evolving preferences of consumers.

Future research should prioritize the functional validation of the key candidate genes identified in this study. A combination of techniques, including comparative gene expression analysis (RNA-seq) across different tissues and developmental stages, virus-induced gene silencing (VIGS) for rapid functional knock-down, and the creation of stable transgenic lines will be essential to definitively confirm their roles in controlling specific traits. Furthermore, the development of a pangenome for *Eruca sativa*, which would capture the full spectrum of genetic variation including presence-absence variants and large structural variants that are missed by standard SNP analysis, would be an invaluable next step. Ultimately, by synergistically integrating cutting-edge genomics with modern breeding strategies, the insights and resources generated from this study will help to ensure that arugula continues its remarkable journey as an increasingly important and valuable global crop.

References

1. Agrawal N, Gupta M, Banga SS, Heslop-Harrison JP. Identification of chromosomes and chromosome rearrangements in crop brassicas and Raphanus sativus: a cytogenetic toolkit using synthesized massive oligonucleotide libraries. *Front Plant Sci.* 2020;11:598039. https://doi.org/10.3389/FPLS.2020.598039.
2. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655–64. https://doi.org/10.1101/gr.094052.109.
3. Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics.* 2016;32(18):2817–23. https://doi.org/10.1093/bioinformatics/btw327.
4. Bell L, Wagstaff C. Rocket science: a review of phytochemical & health-related research in Eruca & Diplotaxis species. *Food Chemistry: X.* 2019;1:100002. https://doi.org/10.1016/j.fochx.2018.100002.
5. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4(1):7. https://doi.org/10.1186/s13742-015-0047-8.
6. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res.* 2010;20(3):393–402. https://doi.org/10.1101/gr.100545.109.
7. Dai L, Li X, Guan C, Zhong J. On ovary and embryo culture of hybrid between rapeseeds and Rocketsalad. *J Hunan Agric Univ (Nat Sci).* 2004;3:201–4.
8. 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. https://doi.org/10.1093/bioinformatics/btr330.
9. 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. https://doi.org/10.1093/GIGASCIENCE/GIAB008.
10. Dong G, Du L, Yang L, Lu M, Qiu M, Li H, et al. Dispersal of crop-livestock and geographical-temporal variation of subsistence along the Steppe and Silk Roads across Eurasia in prehistory. *Sci*

China Earth Sci. 2022;65(7):1187–210. https://doi.org/10.1007/s11430-021-9929-x.

11. Golkar P, Bakhtiari MA. Evaluation of genetic diversity in the world collection of Eruca sativa L. using oil content, fatty acids and molecular markers. *Ind Crops Prod.* 2020;148:112280. https://doi.org/10.1016/j.indcrop.2020.112280.
12. Greuter W, Burdet HM, Long G. *Med-Checklist: a critical inventory of vascular plants of the circum-mediterranean countries*, vol. 3. Geneva: OPTIMA and Conservatoire et Jardin botaniques; 2008.
13. Guijarro-Real C, Navarro A, Esposito S, Festa G, Macellaro R, Di Cesare C, et al. Large scale phenotyping and molecular analysis in a germplasm collection of rocket salad (Eruca vesicaria) reveal a differentiation of the gene pool by geographical origin. *Euphytica.* 2020;216(3):53. https://doi.org/10.1007/s10681-020-02586-x.
14. Hall M, Jobling J, Rogers G. Factors affecting growth of perennial wall rocket and annual garden rocket. *Int J Veg Sci.* 2012;18(4):393–411. https://doi.org/10.1080/19315260.2012.660565.
15. Hanelt P, Buttner R, Mansfeld R. *Mansfeld's encyclopedia of agricultural and horticultural crops (except ornamentals).* Springer; 2001. https://doi.org/10.1007/978-3-540-30442-5.
16. Kang L, Qian L, Zheng M, Chen L, Chen H, Yang L, et al. Genomic insights into the origin, domestication and diversification of Brassica juncea. *Nat Genet.* 2021;53(9):1392–402. https://doi.org/10.1038/s41588-021-00922-y.
17. Kiefer C, Willing EM, Jiao WB, Sun H, Piednoel M, Humann U, et al. Interspecies association mapping links reduced CG to TG substitution rates to the loss of gene-body methylation. *Nat Plants.* 2019;5(8):846–55. https://doi.org/10.1038/s41477-019-0486-9.
18. Lan M. *Dian Nan Ben Cao.* 1959.
19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.
20. Lidin OG. *Tanegashima-the arrival of Europe in Japan.* Routledge; 2003.
21. Liu H, Li Q, Xing Y. Genes contributing to domestication of rice seed traits and its global expansion. *Genes.* 2018. https://doi.org/10.3390/genes9100489.
22. Liu X, Fu YX. Stairway plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* 2020;21(1):280. https://doi.org/10.1186/s13059-020-02196-9.
23. Malinsky M, Matschiner M, Svardal H. Dsuite-fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour.* 2021;21(2):584–95. https://doi.org/10.1111/1755-0998.13265.
24. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics.* 2010;26(22):2867–73. https://doi.org/10.1093/bioinformatics/btq559.
25. Matsuzawa Y, Mekiyanon S, Kaneko Y, Bang SW, Wakui K, Takahata Y. Male sterility in alloplasmic Brassica rapa L. carrying Eruca sativa cytoplasm. *Plant Breed.* 1999;118(1):82–4.
26. Morales M, Janick J. Arugula: a promising specialty leaf vegetable. In: Janick J, Whipkey A, editors. n.d.
27. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74. https://doi.org/10.1093/molbev/msu300.
28. Nothnagel T, Budahn H, Schrader O, Klocke E, editors. Intergeneric hybridization between *Brassica oleracea* and *Eruca sativa*. VI International Symposium on Brassicas and XVIII Crucifer Genetics Workshop 1005; 2012.
29. Olesen MH, Romme HA, Kampmann HH. Rucola plants with cytoplasmic male sterility (CMS). US8247655B2. Google; 2012.
30. Park JS, Park JH, Park YD. Construction of pseudomolecule sequences of *Brassica rapa ssp. pekinensis* inbred line CT001 and analysis of spontaneous mutations derived via sexual propagation. *PLoS One.* 2019;14(9):e0222283. https://doi.org/10.1371/journal.pone.0222283.
31. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 2012;8(11):e1002967. https://doi.org/10.1371/journal.pgen.1002967.
32. Pignone D, Gómez-Campo C. *Eruca.* Wild crop relatives: genomic and breeding resources. In: Kole C, editor. Springer; 2011. p. 149–60. https://doi.org/10.1007/978-3-642-21102-7.
33. R Core Team. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; 2023.
34. Renfrew J, Sanderson H. Herbs and vegetables. In: Ghillean P, Mark N, editors. 2012.
35. RStudio Team. *RStudio: Integrated Development for R.* PBC, Boston, MA; 2020.
36. Sastry E. Taramira (*Eruca sativa*) and its improvement–a review. *Agric Rev.* 2003;24(4):235–49.
37. Sihag RC. Pollination ecology of rocket (*Eruca vesicaria*) in semi-arid Northwest India: native bees are the major pollinators. *Ecologies.* 2023;4(3):580–94. https://doi.org/10.3390/ecologies4030038.

38. Subramanian B, Gao S, Lercher MJ, Hu S, Chen WH. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 2019;47(W1):W270–5. https://doi.org/10.1093/nar/gkz357.

39. Sun W, Guan C, Meng Y, Liu Z, Zhang T, Li X, et al. Intergeneric crosses between *Eruca sativa Mill.* and *Brassica* species. *Zuo Wu Xue Bao.* 2005;1:36–42.

40. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide datasets. *Nucleic Acids Res.* 2019;47(D1):D607–13. https://doi.org/10.1093/nar/gky1131.

41. Traboulay DM. *Columbus and Las Casas: the conquest and Christianization of America, 1492–1566.* Maryland: University Press of America; 1994.

42. Wang Y, Chu P, Yang Q, Chang S, Chen J, Hu M, et al. Complete mitochondrial genome of *Eruca sativa Mill.* (Garden rocket). *PLoS One.* 2014;9(8):e105748. https://doi.org/10.1371/journal.pone.0105748.

43. Warwick SI, Gugel RK, Gómez-Campo C, James T. Genetic variation in *Eruca vesicaria*. *Plant Genet Resour.* 2007;5(3):142–53. https://doi.org/10.1017/S1479262107842675.

44. Xu G, Zhang X, Chen W, Zhang R, Li Z, Wen W, et al. Population genomics of *Zea* species identifies selection signatures during maize domestication and adaptation. *BMC Plant Biol.* 2022;22(1):72. https://doi.org/10.1186/s12870-022-03427-w.

45. Yaniv Z, editor. Tradition, uses and research on rocket in Israel. In: *Rocket: a Mediterranean crop for the world.* Legnaro (Padova); 1996.

46. Zafar-Pashanezhad M, Shahbazi E, Golkar P, Shiran B. Genetic variation of *Eruca sativa L.* genotypes revealed by agro-morphological traits and ISSR molecular markers. *Ind Crops Prod.* 2020;145:111992. https://doi.org/10.1016/j.indcrop.2019.111992

.