Medicinal Plant Biology

ARTICLE

Open Access

https://doi.org/10.48130/mpb-0026-0020

Medicinal Plant Biology **2026**, 5: e025

Whole-genome resequencing reveals population divergence and the domestication history of Tartary buckwheat

Peng Wang[1,2#], Qi Shen[1,3#], Yiling Chen[4#], Chi Song[5], Taoxiong Shi[6], Pengda Ma[7], Gangqiang Dong[4], Jianjun Chen[8], Qingfu Chen[6*], Wei Sun[1,9*] and Xiangxiao Meng[1*]

[1] *State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China*
[2] *The Engineering Research Center of Southwest Bio-Pharmaceutical Resources, Ministry of Education, Guizhou University, Guiyang 550025, China*
[3] *Institute of Medical Plant Physiology and Ecology, School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine, Guangzhou 510006, China*
[4] *Amway (China) Botanical R & D Centre, Wuxi 214115, China*
[5] *Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China*
[6] *Research Center of Buckwheat Industry Technology, Guizhou Normal University, Guiyang 550001, China*
[7] *College of Life Sciences, Northwest A & F University, Yangling 712100, China*
[8] *Hubei Key Laboratory of Quality and Safety of Traditional Chinese Medicine Health Food (Jing Brand Co., Ltd.), Daye 435111, China*
[9] *College of Pharmaceutical Science, Dali University, Dali 671003, China*
[#] Authors contributed equally: Peng Wang, Qi Shen, Yiling Chen
* Correspondence: cqf1966@163.com (Chen Q); wsun@icmm.ac.cn (Sun W); xxmeng@icmm.ac.cn (Meng X)

Abstract

Tartary buckwheat (*Fagopyrum tataricum*) is a pseudocereal valued for both its nutritional and medicinal properties, yet its evolutionary and domestication history remains debated. We analyzed whole-genome resequencing data from 118 wild and cultivated individuals of Tartary buckwheat and identified two genetically distinct groups of wild accessions distributed across Sichuan, Yunnan, and Xizang. The genetically diverse Wild I group was primarily distributed in northern Sichuan and was inferred to represent a possible ancestral wild population, whereas Wild II was mainly distributed in the eastern and central areas of Xizang. Our analysis suggests that cultivated Tartary buckwheat may have been domesticated from Wild II in eastern Xizang, and appears to have subsequently experienced a two-step process of domestication, an initial domestication phase from Wild II to landraces (Cultivar I), followed by a modern breeding and improvement phase (Cultivar II). We also estimated the genetic diversity and linkage disequilibrium of Tartary buckwheat groups, suggesting that the Wild I group retained high genetic diversity and experienced only a weaker genetic bottleneck, whereas the Wild II group underwent a much stronger domestication bottleneck before giving rise to cultivated populations. Selective sweep analysis uncovered genes linked to erect growth, seed development, and flavonoid biosynthesis—the pathway for medicinal compounds. Haplotype analysis revealed multiple genotypes in these genes, potentially explaining domestication traits and active metabolite accumulation. Our research broadens our understanding of the possible origins and divergence processes of Tartary buckwheat and provides a genetic foundation for future breeding efforts aimed at enhancing both agronomic performance and medicinal quality.

Keywords: Population structure, Evolutionary history, Whole-genome resequencing, Tartary buckwheat, Whole-genome resequencing

Citation: Wang P, Shen Q, Chen Y, Song C, Shi T, et al. 2026. Whole-genome resequencing reveals population divergence and the domestication history of Tartary buckwheat. *Medicinal Plant Biology* 5: e025 https://doi.org/10.48130/mpb-0026-0020

Introduction

Cultivated buckwheat species, Tartary buckwheat (*Fagopyrum tataricum*) and common buckwheat (*Fagopyrum esculentum*), which belong to Polygonaceae, are important economically cultivated grains and functional food crops with a huge range of cultivation from eastern Europe to eastern Asia[1–6]. Tartary buckwheat is generally accepted as a high nutraceutical crop and functional food because of its seeds, which are rich in essential amino acids and beneficial phytochemicals and contain abundant flavonoids, rutin, vitamins, and minerals[1,7]. Beyond its nutritional value, Tartary buckwheat has long been used in traditional medicine for treating hypertension, diabetes, and inflammatory conditions[4,8]. Modern pharmacological studies have confirmed that its bioactive flavonoids, particularly rutin and quercetin, exhibit potent antioxidant, anti-inflammatory, hypoglycemic, and cardiovascular-protective activities, making it a valuable medicinal plant resource[8,9]. Owing to these pharmacologically active metabolites, Tartary buckwheat has attracted increasing attention as an important medicinal plant as well as a functional food.

Tartary buckwheat is native to southern and western China, the Xizang plateau, and the high Himalayan hills of Nepal, India, and Pakistan[10,11]. Studies on the geographical distribution of allozymes in natural populations of wild Tartary buckwheat have suggested that *F. tataricum subsp. potanini* Batalin, the wild ancestor of cultivated Tartary buckwheat, was likely derived from the western part of Sichuan province and the northwestern part of Yunnan province in China[11]. Further research has identified the western part of Sichuan Province, the northwestern part of Yunnan province, and eastern areas of Xizang province as the key areas for the origin of cultivated Tartary buckwheat[10–12]. The domestication process from wild to cultivated Tartary buckwheat included significant phenotypic character changes, such as its shattering habit and seed dormancy. The release of the high-quality Tartary buckwheat genome has enabled whole-genome resequencing investigations of 510 Tartary buckwheat germplasms[13,14]. A previous whole-genome resequencing

© The Author(s)

https://doi.org/10.48130/mpb-0026-0020

study revealed two independent domestication events of Tartary buckwheat in southwestern and northern China, which contributed to the diversification of modern cultivated germplasm. Although previous population genomic studies have proposed domestication scenarios for Tartary buckwheat, the evolutionary relationships among geographically structured wild populations and their respective contributions to domestication remain unresolved. Given that domestication-related genetic changes may directly influence the biosynthesis of medicinal flavonoids, elucidating the evolutionary and domestication history of Tartary buckwheat is not only crucial for understanding its genetic diversity but also for the targeted improvement of its pharmaceutical quality.

Materials and methods

Plant material and re-sequencing

In total, 118 Tartary buckwheat accessions from worldwide distribution areas were collected. Information about each sample's variety name, original area of production, longitude, and latitude is listed in Supplementary Table S1. Samples were grown during the normal growing season in the experimental farm of the Research Center of Buckwheat Industry Technology, Guizhou Normal University. Leaf tissue was collected for each accession at the seedling stage, and genomic DNA was extracted for sequencing using the cetrimonium bromide (CTAB) method. Paired-end sequencing libraries with an insert size of 180 bp were constructed for each sample according to Illumina's standard instructions. Paired-end reads were obtained using the Illumina HiSeq 2000 platform. Sequencing was performed in multiple batches resulting in read lengths of 100 bp for the majority of samples, though a subset of 26 out of 51 cultivated accessions yielded 125-bp reads. All raw reads were uniformly quality-trimmed with identical filtering parameters prior to downstream analyses to ensure comparability for variant calling.

Read mapping and variant calling

The raw paired-end reads of 118 accessions were trimmed to remove adapters and low-quality bases using fastp (version: 0.20.0)[15]. The reads were filtered with a sliding window of 4 bp, with an average Phred score scale of 20 within the window. The trimmed reads were mapped to the Tartary buckwheat cultivar 'Pinku1' reference genome using BWA-MEM (version: 0.7.17) with the default parameters[14]. After alignment, Picard (version: 2.18.17, http://broadinstitute.github.io/picard) was used to remove polymerase chain reaction (PCR) duplicates according to the mapping coordinates.

Variation detection followed the best practice workflow recommended by GATK (version 3.8.1). In brief, variants were called for each accession by the GATK HaplotypeCaller. A joint genotyping step for comprehensive variation union was performed on the gVCF files. In the filtering step, the single-nucleotide polymorphism (SNP) filter was set at Quality by Depth (QD) < 2.0, Mapping Quality (MQ) < 40.0, Fisher Strand (FS) > 60.0, Strand Odds Ratio (SOR) > 3.0, Mapping Quality Rank Sum Test (MQRankSum) < −2.5 and Read Position Rank Sum Test (ReadPosRankSum) < −8.0, and the insertion–deletion (indel) filter was set at QD < 2.0, FS > 200.0, MQ < 40.0, SOR > 10.0 and ReadPosRankSum < −20.0. Four subsets of Tartary buckwheat SNPs were defined using the following filtering criteria. (1) A basic set of 3,542,659 SNPs was created by excluding nonbiallelic loci, sites with > 20% missing calls, and Minor Allele Frequency (MAF) < 5%. These filtering thresholds follow standard practices in population genomic studies of crop species, including buckwheat[13] and other selfing crops[16,17]. (2) A core SNP set of 254,956 SNPs was derived from the basic SNP set using a two-step pruning procedure with PLINK (version: 1.9) with the parameters '--indep-pairwise 50 5 0.5' to minimize linkage disequilibrium (LD) among markers for population structure analyses.

SNP and indel annotations were performed according to the Tartary buckwheat 'Pinku1' reference genome using SnpEff (version: 4.3t). The coverage of each accession against each chromosome of the Tartary buckwheat genome was counted on the basis of the aligned BAM file using Qualimap (version: 2.2.1)[18]. SNP and indel densities across each chromosome were counted with 100-kb sliding windows using VCFtools (version: 0.1.17).

Phylogenetic and population analyses

A phylogenetic tree was constructed from the core SNP set by using the neighbor-joining method in PHYLIP (version: 3.697, https://phylipweb.github.io/phylip)[19], and Identity By State (IBS) distance matrices were calculated using PLINK. The R package ggtree was used to color the phylogenetic tree.

Principal component analysis (PCA) was performed with EIGENSOFT (version: 6.1.4) with the core SNP set, and the first four eigenvectors were plotted[20]. Population structure was analyzed using ADMIXTURE (version: 1.3) with a block-relaxation algorithm with the core SNP set. To explore the convergence of individuals, we predefined the number of genetic clusters (K) from two to nine and ran the crossvalidation (CV) error procedure. ADMIXTURE was then run again on the whole core SNP set 10 times per K value using a common set of 10randomly generated seeds[21]; the Q-matrices were aligned using Pong and clustered on the basis of similarity[22]. Then the matrices belonging to the largest cluster were averaged to produce the final matrix of admixture proportions.

LD was calculated using PopLDdecay (version: 3.34, https://github.com/BGI-shenzhen) with the default settings. The pairwise r^2 values within different chromosomes were calculated. The LD for each group was calculated using SNP pairs only from the corresponding group with PopLDdecay default filtering.

Genome scanning for selective sweep signals

We used a cross-approach based on genetic differentiation (F_{ST}) and nucleotide diversity (θ_π) to investigate selection signals across the whole genome. A 200-kb sliding window with a 10-kb step approach was applied to quantify F_{ST} and θ_π by using VCFtools (version: 0.1.17)[23]. For each pairwise comparison, the θ_π ratio was calculated as θ_π of the first group / θ_π of the second group (e.g., θ_π Wild I / θ_π cultivar, θ_π Wild II / θ_π cultivar, and θ_π Wild I / θ_π Wild II). The candidates that were in the top 5% of both values were selected as the selective signals. The 5% empirical cutoff was chosen to identify regions with the strongest selection signals while maintaining a manageable number of candidate genes for downstream analysis, following established practices in crop domestication studies[24,25].

Functional enrichment analysis

Candidate genes identified within selective sweep regions were subjected to Gene Ontology (GO, https://geneontology.org) and Kyoto Encyclopedia of Genes and Genomes (KEGG, www.kegg.jp) pathway enrichment analyses, with the complete reference genome annotation of Tartary buckwheat 'Pinku1' used as the background

gene set. Benjamini–Hochberg false discovery rate (FDR) correction was applied to adjust for multiple testing, and GO terms and KEGG pathways with an adjusted *p*-value (p_{adj}) < 0.05 were considered to be significantly enriched.

Results

Genomic variation

In this study, we conducted whole-genome resequencing of 118 Tartary buckwheat accessions, including 51 wild and 67 cultivated accessions from China, Nepal, and Germany, with a focus on wild samples from Sichuan province (Supplementary Table S1). The sequencing generated 422.921 gigabases (Gb) of high-quality clean data, with an average of 3.584 Gb per accession, representing approximately 7.8-fold coverage of the haploid Tartary buckwheat genome (Supplementary Tables S2 and S3). Individual BioSample accessions are provided in Supplementary Table S2. Using the Tartary buckwheat 'Pinku1' reference genome[14], we identified 3,542,659 SNPs and 448,860 small indels (Supplementary Tables S4 and S5). The identified SNPs and indels were mapped to the eight chromosomes and six super-scaffolds, spanning a total genomic length of 453.12 Mb. Of the identified SNPs, 2.69%, 5.24%, and 65.59% were located in the exons, introns, and intergenic regions, respectively (Supplementary Table S6). For the identified indels, the corresponding proportions were 1.04% (exons), 7.82% (introns) and 58.67% (intergenic regions) (Supplementary Table S7).

Population structure

In total, 254,956 (7.20%) nontightly linked SNP loci were selected for phylogenetic analysis, revealing that the 118 Tartary buckwheat accessions could be divided into four main groups (Fig. 1a). Wild I and Wild II consisted of 26 and 25 Tartary buckwheat accessions, respectively, whereas Cultivar I and Cultivar II contained 48 and 19 Tartary buckwheat accessions, respectively. Wild I, which exhibited the highest genetic diversity and was mainly distributed in northern Sichuan (e.g., accessions CATCM009 from Maoxian; CATCM010–016 from Maerkang; CATCM018, CATCM019, and CATCM024 from Wenchuan; CATCM020 and CATCM021 from Jiuzhaigou; and CATCM042–047 from Lixian), eastern Xizang (e.g., accessions CATCM030, CATCM031, CATCM033, and CATCM034 from Mangkang and CATCM056 from Jiacha), and northern Yunnan (e.g., accessions CATCM025 and CATCM026 from Deqin), which is inferred to represent a possible ancestral wild population of Tartary buckwheat. Wild II mainly included samples distributed in the eastern and central areas of Xizang province, such as Mangkang (e.g., CATCM032, CATCM035–040), Jiacha (e.g., CATCM049, CATCM050, CATCM053–055, and CATCM057–060), and Gongbujiangda (e.g., CATCM027, CATCM028, and CATCM029) counties. These results are consistent with eastern Xizang representing an important region associated with the domestication of cultivated Tartary buckwheat. Cultivar I comprised cultivars and landraces from multiple provinces of China, as well as from Nepal, Italy, and Germany, representing the major cultivated genetic group. Cultivar II mainly consisted of modern breeding lines and improved cultivars from different regions of China. These phylogenetic groupings were further supported by the analysis of 483 cultivars[13]. The population

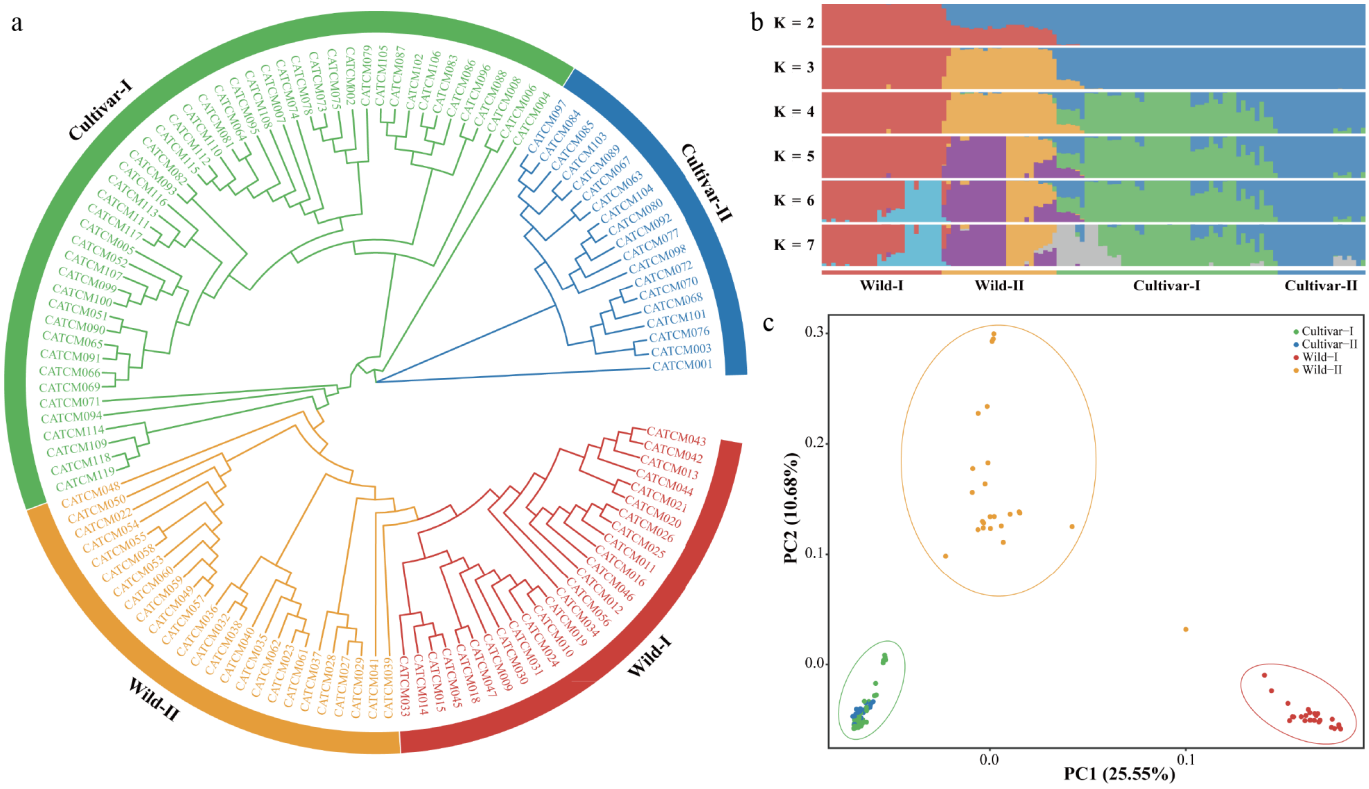


Fig. 1 Population structure and genetic diversity of Tartary buckwheat accessions. (a) Neighbor-joining clustering of 118 Tartary buckwheat accessions. Branch colors indicate the subgroup Wild I (red), the subgroup Wild II (orange), the subgroup Cultivar I (green), and the subgroup Cultivar II (blue). (b) Individual ancestry coefficients from K = 2 to K = 7. (c) Principal component analysis showing genetic differences among these two groups of wild Tartary buckwheat and cultivars.

https://doi.org/10.48130/mpb-0026-0020

admixture graph and structure analysis supported the distinction between the major Tartary buckwheat categories, with the wild and cultivated populations clearly separating into different groups (Fig. 1b). In the structure analysis, the wild and cultivated populations were separated when K = 2, but the two wild populations were separated when K = 3, and the two cultivated groups were separated when K = 4. At K = 5 and K = 6, additional substructure was detected within the Wild I and Wild II groups, respectively. Although K = 7 showed the minimal CV error, the improvement from K = 4 to K = 7 was extremely marginal (Δ = 0.019). Notably, K = 4 exhibited outstanding run stability with a tiny CV standard deviation (SD = 0.003) across replicates, demonstrating robust and consistent convergence (Supplementary Table S8 and Supplementary Fig. S1). Therefore, K = 4 was selected as the biologically interpretable solution that resolves the four major genetic lineages (Wild I, Wild II, Cultivar I, and Cultivar II), whereas K = 7 represents the numerically optimal solution with the minimum CV error, though it captures additional substructure within the wild groups that is not the focus of the present study. The separation of cultivated accessions into two genetic clusters in our study is broadly consistent with the differentiation patterns observed in the phylogenetic reconstruction of 483 cultivars[13], although the grouping criteria differ between the two studies. PCA supported the existence of three subpopulations corresponding to Wild I, most of Wild II and cultivar accessions (Fig. 1c, Supplementary Fig. S2), suggesting significant genetic differences between them, whereas the two cultivated groups exhibited significant overlap.

Genetic diversity and linkage disequilibrium

To assess genetic diversity and population differentiation among Tartary buckwheat groups, we calculated the nucleotide diversity (π) and pairwise population differentiation (F_{ST}) for the four genetic groups. The π value in the Wild I group (1.28×10^{-3}) and Wild II group (1.07×10^{-3}) was substantially higher than that Cultivar I (5.20×10^{-4}) and Cultivar II (4.40×10^{-4}) (Fig. 2a), which is consistent with the general pattern of wild populations harboring higher genetic diversity than domesticated populations. The ratio of diversity ($\pi_{\text{Wild I}}/\pi_{\text{Wild II}}$) in wild Tartary buckwheat (1.19) was comparable with that reported for castor bean (*Ricinus communis*) (1.14)[26], suggesting that Wild II experienced a weak genetic bottleneck. In contrast, the ratio of diversity $\pi_{\text{Wild II}}/\pi_{\text{Cultivar I}}$ was 2.05, which was significantly higher than that of rice (*Oryza sativa*) (1.25) and soybean (*Glycine max*) (1.58), similar to that of cucumber (*Cucumis sativus*) (1.96) and lower than tomato (*Solanum lycopersicum*) (2.63)[16,17,27], suggesting that Wild II and cultivar groups underwent a higher domestication bottleneck overall. The pairwise F_{ST} between the two wild groups and the two cultivar groups revealed differentiation patterns: Wild II showed low intergroup F_{ST} values with Cultivar I and II (0.209 and 0.173, respectively), suggesting a close genetic relationship and supporting Wild II as the most likely progenitor lineage among the sampled wild groups. In contrast, Wild I exhibited remarkably high *Fst* values with Cultivar I and II (0.701 and 0.695, respectively) (Fig. 2a), suggesting substantial genetic divergence and suggesting that this lineage contributed little to the domestication of cultivated Tartary buckwheat. Together, these results support a pronounced domestication bottleneck during the inferred divergence between Wild II and cultivated Tartary buckwheat, consistent with a putative domestication transition from a Wild II-like ancestral population. We next examined LD decay across the four groups to further characterize their population genetic patterns (Fig. 2b). The LD decay to half its maximum value was seen at 3.3 and 19.8 kb in Wild I and Wild II, respectively, compared with 79 and 158 kb in Cultivar I and Cultivar II, respectively. The LD decay to half distances in the wild groups was shorter than that in the cultivar groups, suggesting that artificial domestication and selection increased LD, a common pattern during crop domestication.

Identification of selective sweep regions and genes

To identify genomic regions subjected to selection during Tartary buckwheat groups, a 200-kb sliding window analysis with a 10-kb step was carried out for θ_π ratios ($\theta_{\pi\ \text{Wild}}/\ \theta_{\pi\ \text{Cultivar}}$) and F_{ST} values. Between wild and cultivated groups, most selective signals in the cultivated group ($\log_2\theta_\pi$ ratio > 6.495 and $F_{ST} > 0.478$) (Fig. 2c). Similar patterns were also shown between Wild I and the cultivar groups ($\log_2\theta_\pi$ ratio > 5.158 and $F_{ST} > 0.884$) and between Wild II and the cultivar groups ($\log_2\theta_\pi$ ratio > 4.77 and $F_{ST} > 0.43$). However, between the two wild groups, selective signals were comparably distributed in Wild I ($\log_2\theta_\pi$ ratio < −0.969 and $F_{ST} > 0.796$) and Wild II ($\log_2\theta_\pi$ ratio > 1.558 and $F_{ST} > 0.796$) (Supplementary Fig. S3). These results suggest that divergence between the two wild groups was primarily shaped by natural evolutionary processes, whereas domestication imposed much stronger positive selection on cultivated populations. In total, 55 cultivated–wild selective sweeps in cultivated groups (15.04 Mb or 3.31% of the genome, containing 1,199 genes) were identified, which corresponded to 35 cultivated–Wild I selective sweeps in the cultivated groups (8.98 Mb or 1.98% of the genome, containing 766 genes) and 36 cultivated–Wild II selective sweeps in the cultivated groups (12.18 Mb or 2.68% of the genome, containing 853 genes) (Supplementary Table S9). Subsequent analyses focused on the cultivated–wild comparison, in which 783 candidate genes were identified within selective sweep regions (Supplementary Table S10). GO enrichment analysis showed that these candidate genes were significantly enriched in cellular and metabolic processes, catalytic activity, and binding functions, whereas the KEGG analysis highlighted translation- and amino acid metabolism-related pathways (Supplementary Fig. S4). The protein content and composition in Tartary buckwheat showed a significant change during the domestication process, which was consistent with the selective sweep. For example, the top 30% of selected genes were found to be involved in branched-chain amino acid metabolic processes, according to the GO enrichment analysis (Fig. 2d). Moreover, the KEGG enrichment analysis indicated that the top 30% of selected genes were involved in ribosome biogenesis in eukaryotes (Fig. 2e).

During domestication and improvement, several key agronomic traits have been modified through artificial selection[24,25,28]. In Tartary buckwheat, traits such as prostrate growth habit, seed shattering, and seed dormancy were selected against during the transition from wild to cultivated forms, whereas seed size, weight, and protein and starch content were also altered over the course of domestication[29]. Selected genes were identified between wild and cultivar Tartary buckwheat populations. Interestingly, a number of genes, particularly those encoding transcription factors involved in shattering, erect or dwarf plant growth, flowering period, inflorescence architecture, seed size, and resistance to pathogens and environmental stresses, were identified during the Tartary buckwheat domestication process (Fig. 3a, b). For instance, zinc finger CCHC-type protein (FtPinG0505481600.01) has been reported to regulate seed size in cereals and the model legume species *Medicago truncatula*[30]. Haplotype networks showed two SNP variations that existed in Wild I and Wild II, and the variants were shared by the Wild II and cultivated groups of Tartary buckwheat (Fig. 3c).

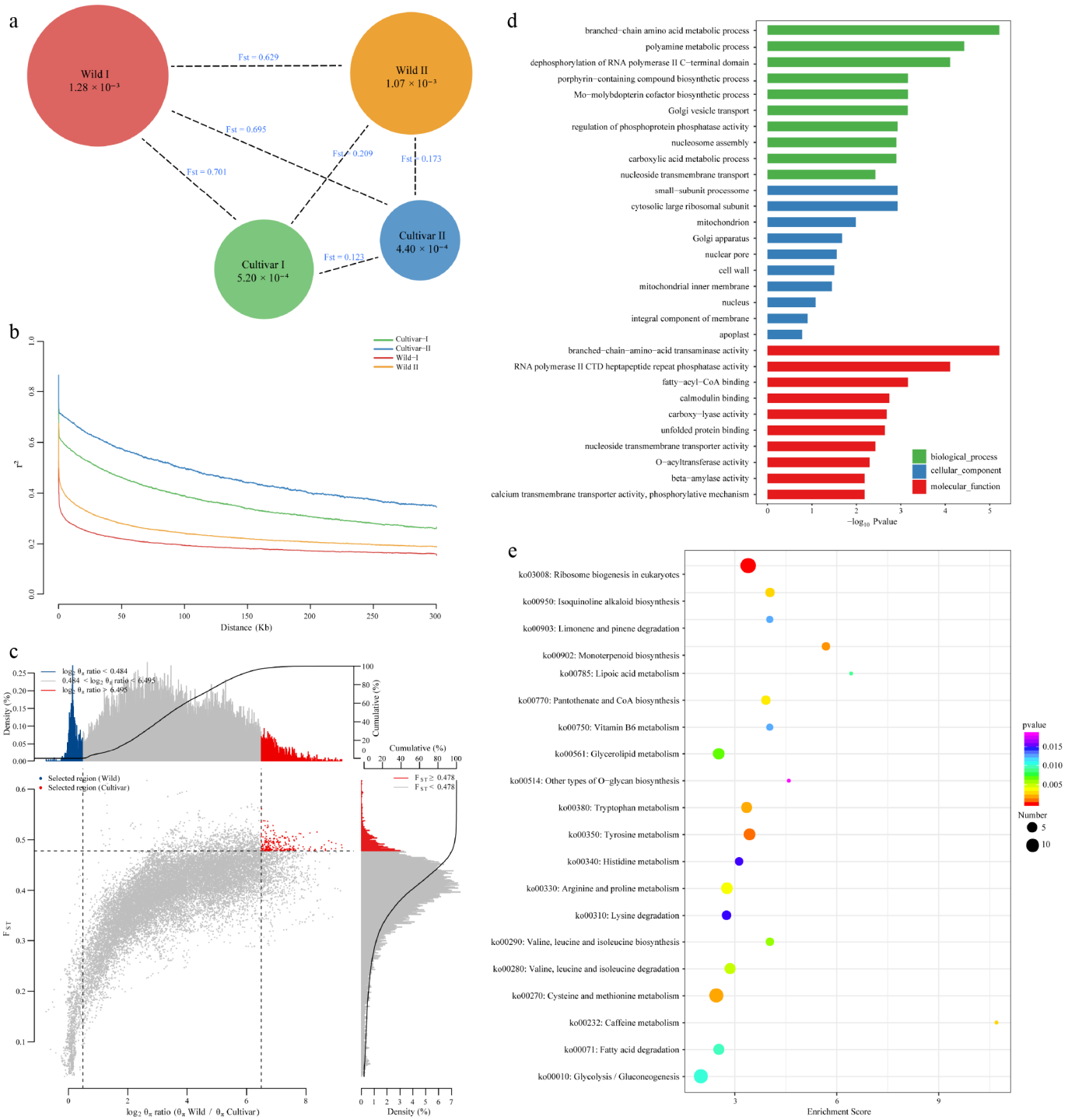


Fig. 2 Genetic diversity, LD, and selection signals. (a) Nucleotide diversity (π) and pairwise population fixation index (F_{ST}). (b) Decay of LD. The colors represent groups: Wild I (red), Wild II (orange), Cultivar I (green), and Cultivar II (blue). (c) Genomic regions with strong selective sweep signals in wild and cultivar groups of Tartary buckwheat. Distributions of θ_π ratios and F_{ST} values were calculated in 200-kb sliding windows and 10-kb steps. Data points located to the left and right of the left and right vertical dashed lines, respectively (corresponding to the 5% left and right tails of the empirical θ_π ratio distribution), and above the horizontal dashed line (the 5% right tail of the empirical F_{ST} distribution) were identified as selected regions. (d) GO enrichment analysis of the genes in selective sweep regions. (e) KEGG enrichment analysis of the genes in selective sweep regions.

The leucine-rich repeat receptor-like kinase (LRR-RLK) (FtPinG0201347300.01) is known as a regulator of seed size in *Arabidopsis thaliana*[31]. Eighteen SNPs existed in Wild I and Wild II. There were two variations in Wild II, and the change was also observed in the cultivated groups (Fig. 3d). The WD40 domain protein (FtPinG0505979100.01) was significantly higher in accessions of Tartary buckwheat with higher grain weight, which was also previously reported[13]. Haplotype networks showed that four SNP variations were shared between Wild II and the cultivated groups in our research (Fig. 3e). Genes for the BTB/POZ domain-containing protein

https://doi.org/10.48130/mpb-0026-0020

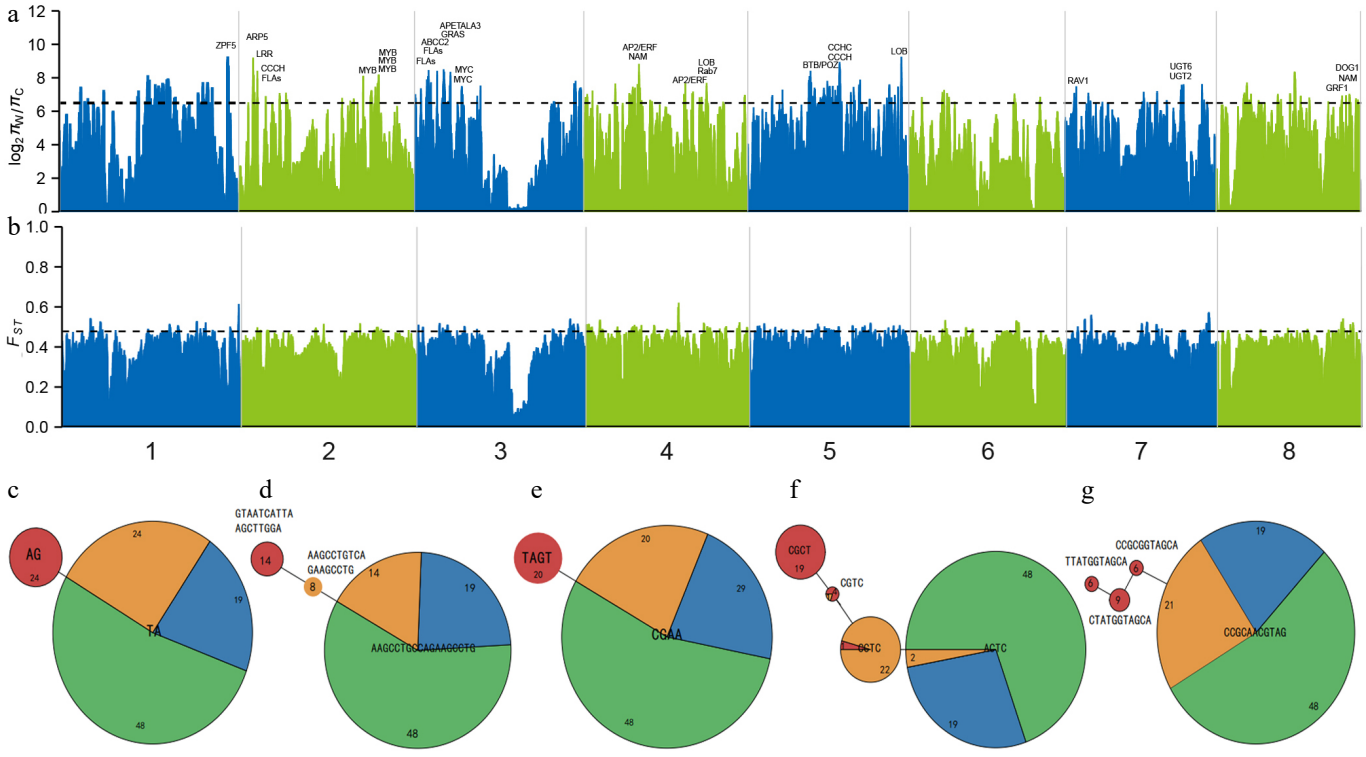


Fig. 3 Identification of selective sweep signals between wild and cultivar groups of Tartary buckwheat. (a) Selective sweep signals identified by θ_π ratios. (b) Selective sweep signals identified by relative divergence (F_{ST}). Haplotype networks of typical selected genes are shown for (c) zinc finger CCHC-type protein, (d) LRR receptor-like serine kinase, (e) WD40 domain protein, (f) and (g) UDP-glucosyl transferase. The colors in the pie chart indicate Wild I (red), Wild II (orange), Cultivar I (green), and Cultivar II (blue). Circle size is proportional to haplotype frequency, and branch length indicates the number of mutational steps.

(FtPinG0505497800.01) and LOB domain-containing protein (FtPinG0404479200.01 and FtPinG0506019300.01) determining plant architecture were identified in Tartary buckwheat[32,33]. Three APETALA2 (AP2) domain-containing genes, the APETALA2/ethylene response factor (AP2/ERF) (FtPinG0403956300.01 and FtPinG0404513600.01) and the RAV1 transcription factor (FtPinG0707223800.01), which were reported as important genes controlling seed shattering in the domestication of rice and wheat (*Triticum aestivum*), were more strongly selected during the domestication process[34,35]. Notably, we also observed candidate genes involved in flavonoid-related pathways, including four MYB genes, two MYC genes, and two UGT family members. Given the known roles of these gene families in plants' secondary metabolism[36–40], they are reasonable targets for functional testing. Therefore, the genes we screened may also be involved in the transcriptional regulation and biosynthesis of flavonoid metabolites in Tartary buckwheat (Supplementary Table S11). The annotations of the candidate genes and their selection statistics can be found in Supplementary Table S11. These genes are worthy of further functional validation. Collectively, these candidate genes provide insights into the genetic basis of domestication-associated phenotypic evolution in Tartary buckwheat. To investigate genetic divergence between the two wild lineages, we further analyzed selective signals distinguishing Wild I and Wild II. In total, 133 and 284 genes were identified in Wild I and Wild II, respectively (Supplementary Tables S12 and S13). Genes under selection in Wild II were enriched in glycerone kinase activity and phenylpropanoid biosynthesis, whereas those in Wild I were primarily associated with vesicle-mediated transport, polygalacturonase activity, and plant–pathogen interactions (Supplementary Fig. S5). Haplotype network analysis showed that the majority of sequence variation can be observed between the two wild lineages, whereas relatively few additional variants distinguish the cultivated groups from Wild II (Fig. 3f, g, Supplementary Table S11).

Discussion

Wild Tartary buckwheat is distributed across a wide area of temperate Asia, from southern China to the Himalayan foothills and then to northern Pakistan[11]. Although previous studies based on molecular markers suggested the existence of two or three wild genetic groups[11,12], our genome-wide resequencing data provide much higher resolution and consistently resolve the wild accessions into two genetically distinct lineages, Wild I and Wild II. Wild I exhibited the highest nucleotide diversity (π) and the greatest genetic differentiation from the cultivated groups, as reflected by its markedly higher pairwise F_{ST} values. Within Wild I, the highly polymorphic wild accessions from northern Sichuan (e.g., accessions CATCM009 from Maoxian; CATCM010–016 from Maerkang; CATCM018, CATCM019, and CATCM024 from Wenchuan; CATCM020 and CATCM021 from Jiuzhaigou; and CATCM042–047 from Lixian) were the most likely candidates for the probable center of origin of wild Tartary buckwheat[11]. Wild II is primarily distributed in the eastern and central areas of Xizang, regions that have previously been proposed as important areas for the diversification of wild Tartary buckwheat and the domestication of cultivated forms[13,41]. This hypothesis is supported by both the geographic distribution and genetic relationships of the wild and cultivated accessions[21]. Further support comes from a recent population genomic analysis

of 567 Tartary buckwheat accessions[41], in which several wild accessions from the Himalayan Wild group collected in Xizang appear to correspond to the Wild II lineage identified in our study despite the differences in nomenclature. Consistent with these findings, our genome-wide population structure analyses provide independent evidence supporting eastern Xizang as a candidate region associated with Tartary buckwheat's domestication[41]. Our findings are broadly consistent with, yet distinct from, the two major genome-wide studies of Tartary buckwheat domestication published previously. Zhang et al.[13] first established the three-group population structure (Himalayan Wild, Southwest Landraces, North Landraces) on the basis of 510 accessions and proposed that Tartary buckwheat experienced independent domestication events in southwestern and northern China. He et al.[41] subsequently expanded the sampling to 567 accessions, added critical wild accessions from the Himalayan region, and refined the origin hypothesis by confirming the Himalayan region as the center of origin on the basis of higher nucleotide diversity and faster LD decay in the HW group. Our study complements these findings by providing a more densely sampled view of wild diversity in the Himalayan foothills, identifying two distinct wild lineages (Wild I and Wild II) that correspond to previously recognized but unresolved genetic clusters. Several limitations of this study should be acknowledged. First, the present sampling is geographically imbalanced and cannot identify a unique domestication center or prove a direct progenitor–descendant relationship. In particular, 14 of 26 Wild I accessions were from Sichuan, whereas 20 of 25 Wild II accessions were from Xizang, so fine-scale phylogeographic inference may be affected by the sampling design. Second, the candidate genes identified under selective sweeps lack direct phenotypic, metabolite, or functional validation. Despite these limitations, we believe that the combination of enriched wild sampling and multiple complementary population genomic analyses provides a solid framework for understanding the domestication history of Tartary buckwheat.

Tartary buckwheat is a self-fertilizing species. The LD of cultivated Tartary buckwheat groups was similar to that of soybean cultivars (133 kb) and cultivated rice (123 and 167 kb in *O. sativa* subspecies *indica* and *O. sativa* subspecies *japonica*, respectively) but much higher than cultivated maize (*Zea mays*) (30 kb) and *Brassica oleracea* (36.8 kb)[17,27,42]. The slower LD decay observed in cultivated populations relative to wild ones is consistent with the expected consequences of domestication, including population bottlenecks that reduce the effective population size and intense artificial selection that extends haplotype homozygosity around the target loci. Conversely, the more rapid LD decay in Wild I may reflect large, stable populations with high recombination rates.

The cultivated species was domesticated, involving major phenotypic transformations[29]. In the context of the domestication process of Tartary buckwheat, researchers have identified a greater number of genes and transcription factors that are involved in various plant growth characteristics, such as shattering, erect or dwarf growth, flowering period, inflorescence architecture, seed size, and resistance to pathogens and environmental stresses. Through genome-wide association studies, Zhang et al. located the association of the *FtUFGT3* (FtPinG0100123400) gene with flavonoid metabolism and found that this gene has been demonstrated to be involved in flavonoid glycosylation through enzymatic assays, and found that the AP2 transcription factor (FtPinG0404616900) was key for grain weight, although more Uridine diphosphate (UDP)-glucosyl transferase and AP2 domain-containing genes were located in selective signals in the group[13]. Additionally, the LRR receptor-like serine kinase plays crucial roles in development and stress responses, and the WD40 domain protein (FtPinG0505979100.01) was found to be associated with higher grain weight[13]. The flavonoid biosynthesis pathway, which produces pharmacologically active compounds such as rutin and quercetin, was among the most prominently enriched pathways in our selective sweep analysis. Candidate genes in this pathway, including *FtUFGT3* and several AP2/ERF transcription factors, showed haplotype structures differentiating the wild and cultivated groups[13], suggesting that selection for agronomic traits may have indirectly influenced the accumulation of medicinal metabolites. Both reported genes were also identified in our research. These selected sites and genes can provide insights into the possible functional divergence and molecular mechanisms between wild and cultivated Tartary buckwheat.

Conclusions

This study provides a genome-wide perspective on the population divergence and possible domestication history of Tartary buckwheat based on whole-genome resequencing of 118 wild and cultivated accessions. Our analyses identified two genetically distinct wild lineages, Wild I and Wild II, distributed across Sichuan, Yunnan, and Xizang. Wild I, mainly distributed in northern Sichuan, was inferred to represent a possible ancestral wild population, whereas Wild II, mainly distributed in the eastern and central areas of Xizang, appears to be a possible progenitor lineage among the sampled wild groups. The results are consistent with a two-step domestication scenario, in which cultivated Tartary buckwheat may have been domesticated from Wild II in eastern Xizang, first giving rise to landraces (Cultivar I), followed by a modern breeding and improvement phase (Cultivar II). Genetic diversity and LD analyses suggest that Wild I retained high diversity and experienced only a weaker bottleneck, whereas a much stronger domestication bottleneck occurred during the divergence of Wild II from cultivated groups. Selective sweep analysis identified candidate genes potentially associated with erect growth, seed development, and flavonoid biosynthesis (the pathway for medicinal compounds), and haplotype analysis revealed multiple genotypes in these genes, potentially explaining domestication traits and active metabolite accumulation. These findings broaden our understanding of the possible origins and divergence processes of Tartary buckwheat and provide a genetic foundation for future breeding efforts. Although broader geographic sampling and functional validation of the candidate genes are needed to confirm the domestication scenario and the roles of these loci, the putative candidate genes associated with flavonoid biosynthesis and seed-related traits identified here could potentially serve as targets for marker-assisted breeding to improve both medicinal quality and agronomic performance in Tartary buckwheat.

Author contributions

The authors confirm their contributions to the paper as follows: conceptualization, software, validation: Wang P; methodology, investigation: Wang P, Shi T; data curation: Wang P, Shen Q, Song C; writing – original draft: Wang P, Shen Q; writing – review and editing: Wang P, Chen Y, Song C, Sun W, Meng X; supervision: Chen Y, Dong G, Chen J, Chen Q, Meng X; resources: Chen Y, Shi T, Dong G, Chen J, Chen Q, Meng X; writing – review: Ma P; project administration: Sun W, Meng X; funding acquisition: Meng X. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation (GSA: CRA012229). Individual BioSample accessions, file names, and coverage statistics are provided in Supplementary Table S2. All data are publicly accessible at http://bigd.big.ac.cn/gsa without restriction.

Acknowledgments

We gratefully acknowledge the contributions of Yuying Yan and Zhigui Bao to this work. This work was supported by the China Academy of Chinese Medical Sciences (CACMS) Innovation Fund (CI2025E00-17, CI2025E00-09, CI2026E002 and CI2024E003), the Fundamental Research Funds for the Central Public Welfare Research Institutes (ZZ15-YQ-039 and ZXKT22051), the Guizhou Provincial Basic Research Programme (Natural Science) (Qiankehe Foundation MS [2026]074), and the Natural Science Foundation of Hubei Province (2024AFD019). We are grateful for the instrument support provided by the large-scale instrument sharing platform of Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences. We are grateful for the instrument support provided by the large-scale instrument sharing platform of Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences.

Conflict of interest

Although the authors Yiling Chen and Gangqiang Dong are employees of Amway (China) Botanical R&D Centre, and Jianjun Chen is an employee of the Hubei Key Laboratory of Quality and Safety of Traditional Chinese Medicine Health Food (Jing Brand Co., Ltd.), the work presented herein is independent academic research and is not related to the commercial interests of these companies. The authors declare that no financial or other contractual agreements between the company and the authors or their institutions influenced the design, results, or reporting of this study. The other authors declare that they have no conflict of interest.

Supplementary information accompanies this paper online at: https://doi.org/10.48130/mpb-0026-0020.

Dates

Received 30 June 2026; Revised 18 September 2026; Accepted 22 September 2026; Published online 27 October 2026

References

[1] Jiang P, Burczynski F, Campbell C, Pierce G, Austria JA, et al. 2007. Rutin and flavonoid contents in three buckwheat species *Fagopyrum esculentum, F. tataricum*, and *F. homotropicum* and their protective effects against lipid peroxidation. *Food Research International* 40:356–364

[2] Liu T, Wang P, Chen Y, Sun B, Li Q, et al. 2024. LC–MS and MALDI–MSI-based metabolomic approaches provide insights into the spatial–temporal metabolite profiles of Tartary buckwheat achene development. *Food Chemistry* 449:139183

[3] Ma KH, Kim NS, Lee GA, Lee SY, Lee JK, et al. 2009. Development of SSR markers for studies of diversity in the genus *Fagopyrum*. *Theoretical and Applied Genetics* 119:1247–1254

[4] Wang P, Li Q, Wei J, Zeng S, Sun B, et al. 2023. Germplasm resources and metabolite marker screening of high-flavonoid Tartary buckwheat (*Fagopyrum tataricum*). *Journal of Agricultural and Food Chemistry* 71:20131–20145

[5] Yang W, Qian G, Chen Y, Liu T, Wan H, et al. 2022. Profiling of polyphenols for in-depth understanding of Tartary buckwheat sprouts: correlation between cultivars and active components, dynamic changes and the effects of ultraviolet B stress. *Food Chemistry: X* 14:100295

[6] Tang Z, Huang L, Gou J, Chen H, Han X. 2014. Genetic relationships among buckwheat (*Fagopyrum*) species from southwest China based on chloroplast and nuclear SSR markers. *Journal of Genetics* 93:849–853

[7] Bai CZ, Feng ML, Hao XL, Zhong QM, Tong LG, et al. 2015. Rutin, quercetin, and free amino acid analysis in buckwheat (*Fagopyrum*) seeds from different locations. *Genetics and Molecular Research* 14:19040–19048

[8] Zhu F. 2016. Chemical composition and health effects of Tartary buckwheat. *Food Chemistry* 203:231–245

[9] Wu Z, Li R, Sun M, Hu X, Xiao M, et al. 2024. Current advances of *Carthamus tinctorius* L.: a review of its application and molecular regulation of flavonoid biosynthesis. *Medicinal Plant Biology* 3(1):1–11

[10] Tsuji K, Ohnishi O. 2001. Phylogenetic position of east Tibetan natural populations in Tartary buckwheat (*Fagopyrum tataricum* Gaert.) revealed by RAPD analyses. *Genetic Resources and Crop Evolution* 48:63–67

[11] Ohnishi O. 1998. Search for the wild ancestor of buckwheat III. The wild ancestor of cultivated common buckwheat, and of Tatary buckwheat. *Economic Botany* 52:123–133

[12] Tsuji K, Ohnishi O. 2001. Phylogenetic relationships among wild and cultivated Tartary buckwheat (*Fagopyrum tataricum* Gaert.) populations revealed by AFLP analyses. *Genes & Genetic Systems* 76:47–52

[13] Zhang K, He M, Fan Y, Zhao H, Gao B, et al. 2021. Resequencing of global Tartary buckwheat accessions reveals multiple domestication events and key loci associated with agronomic traits. *Genome Biology* 22(1):23

[14] Zhang L, Li X, Ma B, Gao Q, Du H, et al. 2017. The Tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. *Molecular Plant* 10:1224–1237

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

[16] Choi JY, Zaidem M, Gutaker R, Dorph K, Singh RK, et al. 2019. The complex geography of domestication of the African rice Oryza glaberrima. *PLoS Genetics* 15:e1007414

[17] Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, et al. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology* 33:408–414

[18] García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, et al. 2012. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 28:2678–2679

[19] Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8:28–36

[20] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38:904–909

[21] Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19:1655–1664

[22] Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. 2016. Pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* 32:2817–2823

[23] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158

[24] Cao K, Zhou Z, Wang Q, Guo J, Zhao P, et al. 2016. Genome-wide association study of 12 agronomic traits in peach. *Nature Communications* 7:13246

[25] Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, et al. 2019. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nature Genetics* 51:857–864

[26] Jackson AJ, Hershey DM, Chesnut T, Xu M, Peters RJ. 2014. Biochemical characterization of the *Castor* bean ent -kaurene synthase(-like) family supports quantum chemical view of diterpene cyclization. *Phytochemistry* 103:13–21

[27] Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, et al. 2012. Comparative population genomics of maize domestication and improvement. *Nature Genetics* 44:808–811

[28] Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, et al. 2017. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nature Genetics* 49:1082–1088

[29] Fuller DQ, Allaby RG. 2009. Seed dispersal and crop domestication: shattering, germination and seasonality in evolution under cultivation. In *Fruit Development and Seed Dispersal*. ed. Østergaard L. Oxford: Wiley-Blackwell. pp. 238–295 doi: 10.1002/9781444314557.ch7

[30] Radkova M, Revalska M, Kertikova D, Iantcheva A. 2019. Zinc finger CCHC-type protein related with seed size in model legume species *Medicago truncatula*. *Biotechnology & Biotechnological Equipment* 33:278–285

[31] Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A. 2005. *MINISEED3 (MINI3)*, a *WRKY* family gene, and *HAIKU2 (IKU2)*, a leucine-rich repeat (*LRR*) *KINASE* gene, are regulators of seed size in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 102(48):17531–17536

[32] Dong Z, Li W, Unger-Wallace E, Yang J, Vollbrecht E, et al. 2017. Ideal crop plant architecture is mediated by *tassels replace upper ears1*, a BTB/POZ ankyrin repeat gene directly targeted by TEOSINTE BRANCHED1. *Proceedings of the National Academy of Sciences of the United States of America* 114(41):E8656–E8664

[33] Xu C, Luo F, Hochholdinger F. 2016. LOB domain proteins: beyond lateral organ boundaries. *Trends in Plant Science* 21:159–167

[34] Xie W, Ding C, Hu H, Dong G, Zhang G, et al. 2022. Molecular events of rice AP2/ERF transcription factors. *International Journal of Molecular Sciences* 23(19):12013

[35] Gao F, Chen JM, Xiong AS, Peng RH, Liu JG, et al. 2009. Isolation and characterization of a novel AP2/EREBP-type transcription factor OsAP211 in *Oryza sativa*. *Biologia Plantarum* 53(4):643–649

[36] Liu W, Mu H, Yuan L, Li Y, Li Y, et al. 2023. VvBBX44 and VvMYBA1 form a regulatory feedback loop to balance anthocyanin biosynthesis in grape. *Horticulture Research* 10(10):uhad176

[37] Yang Y, Wang J, Han F, Zhang J, Gao M, et al. 2024. Characterization of UGT71, a major glycosyltransferase family for triterpenoids, flavonoids and phytohormones-biosynthetic in plants. *Forestry Research* 4:e035

[38] Ye Y, Liu RY, Li X, Zheng XQ, Lu JL, et al. 2024. CsMYB67 participates in the flavonoid biosynthesis of summer tea leaves. *Horticulture Research* 11:uhad231

[39] Thakur N, Kanwar R, Karchuli AS, Negi S. 2026. A coordinated network of MYB regulators orchestrates anthocyanin biosynthesis in banana. *Horticulture Research* 13:uhaf361

[40] Yong Y, Bi H, Li M, Zhu Y, Zhou Q, et al. 2026. LIR3MYB-mediated flavonoid biosynthesis confers cold stress tolerance in *Lilium lancifolium* through the LIDREB-LICHS2 regulatory cascade. *Horticulture Research* 13:uhag065

[41] He Y, Zhang K, Shi Y, Lin H, Huang X, et al. 2024. Genomic insight into the origin, domestication, dispersal, diversification and human selection of Tartary buckwheat. *Genome Biology* 25(1):61

[42] Cheng F, Sun R, Hou X, Zheng H, Zhang F, et al. 2016. Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nature Genetics* 48(10):1218–1224

Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit https://creativecommons.org/licenses/by/4.0/.