Molecular Plant
Research article

CellPress
Partner Journal

Comparative and population genomics of buckwheat species reveal key determinants of flavor and fertility

Kaixuan Zhang[1,16], Yuqi He[1,16], Xiang Lu[1,16], Yaliang Shi[1,16], Hui Zhao[1,2,16], Xiaobo Li[3,16], Jinlong Li[1], Yang Liu[1], Yinan Ouyang[1], Yu Tang[1], Xue Ren[3], Xuemei Zhang[3], Weifei Yang[3], Zhaoxia Sun[4,5], Chunhua Zhang[6], Muriel Quinet[7], Zlata Luthar[8], Mateja Germ[8], Ivan Kreft[8,9], Dagmar Janovská[10], Vladimir Meglič[11], Barbara Pipan[11], Milen I. Georgiev[12,13], Bruno Studer[14], Mark A. Chapman[15] and Meiliang Zhou[1,*]

[1]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, National Crop Genebank Building, Zhongguancun South Street No. 12, Haidian District, Beijing 100081, China

[2]College of Agronomy, Sichuan Agricultural University, Chengdu 611130, China

[3]Annoroad Gene Technology (Beijing) Co., Ltd, Beijing 100176, China

[4]College of Agriculture, Institute of Agricultural Bioengineering, Shanxi Agricultural University, Taigu 030801, Shanxi, China

[5]Shanxi Key Laboratory of Minor Crops Germplasm Innovation and Molecular Breeding, Shanxi Agricultural University, Taiyuan 030031, Shanxi, China

[6]Tongliao Institute Agricultural and Animal Husbandry Sciences, Tongliao 028015, Inner Mongolia, China

[7]Groupe de Recherche en Physiologie Végétale (GRPV), Earth and Life Institute-Agronomy (ELI-A), Université Catholique de Louvain, Croix du Sud 4-5, boîte L7.07.13, B-1348, Louvain-la-Neuve, Belgium

[8]Biotechnical Faculty, University of Ljubljana, 1000 Ljubljana, Slovenia

[9]Nutrition Institute, Tržaška 40, 1000 Ljubljana, Slovenia

[10]Gene Bank, Crop Research Institute, Drnovská 507, Prague 6, Czech Republic

[11]Agricultural Institute of Slovenia, Hacquetova ulica, Ljubljana, Slovenia

[12]Laboratory of Metabolomics, Institute of Microbiology, Bulgarian Academy of Sciences, Plovdiv, Bulgaria

[13]Center of Plant Systems Biology and Biotechnology, Plovdiv, Bulgaria

[14]Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, Universitaetstrasse 2, 8092 Zurich, Switzerland

[15]Biological Sciences, University of Southampton, Life Sciences Building 85, Highfield Campus, Southampton SO17 1BJ, UK

[16]These authors contributed equally to this article.

***Correspondence: Meiliang Zhou (zhoumeiliang@caas.cn)**

https://doi.org/10.1016/j.molp.2023.08.013

ABSTRACT

Common buckwheat (*Fagopyrum esculentum*) is an ancient crop with a world-wide distribution. Due to its excellent nutritional quality and high economic and ecological value, common buckwheat is becoming increasingly important throughout the world. The availability of a high-quality reference genome sequence and population genomic data will accelerate the breeding of common buckwheat, but the high heterozygosity due to the outcrossing nature has greatly hindered the genome assembly. Here we report the assembly of a chromosome-scale high-quality reference genome of *F. esculentum* var. *homotropicum*, a homozygous self-pollinating variant of common buckwheat. Comparative genomics revealed that two cultivated buckwheat species, common buckwheat (*F. esculentum*) and Tartary buckwheat (*F. tataricum*), underwent metabolomic divergence and ecotype differentiation. The expansion of several gene families in common buckwheat, including *FhFAR* genes, is associated with its wider distribution than Tartary buckwheat. Copy number variation of genes involved in the metabolism of flavonoids is associated with the difference of rutin content between common and Tartary buckwheat. Furthermore, we present a comprehensive atlas of genomic variation based on whole-genome resequencing of 572 accessions of common buckwheat. Population and evolutionary genomics reveal genetic variation

Published by the Molecular Plant Shanghai Editorial Office in association with
Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

associated with environmental adaptability and floral development between Chinese and non-Chinese cultivated groups. Genome-wide association analyses of multi-year agronomic traits with the content of flavonoids revealed that *Fh05G014970* is a potential major regulator of flowering period, a key agronomic trait controlling the yield of outcrossing crops, and that *Fh06G015130* is a crucial gene underlying flavor-associated flavonoids. Intriguingly, we found that the gene translocation and sequence variation of *FhS-ELF3* contribute to the homomorphic self-compatibility of common buckwheat. Collectively, our results elucidate the genetic basis of speciation, ecological adaptation, fertility, and unique flavor of common buckwheat, and provide new resources for future genomics-assisted breeding of this economically important crop.

Key words: buckwheat, genomics, natural variation, adaptation, flavonoids, *Fagopyrum*

Zhang K., He Y., Lu X., Shi Y., Zhao H., Li X., Li J., Liu Y., Ouyang Y., Tang Y., Ren X., Zhang X., Yang W., Sun Z., Zhang C., Quinet M., Luthar Z., Germ M., Kreft I., Janovská D., Meglič V., Pipan B., Georgiev M.I., Studer B., Chapman M.A., and Zhou M. (2023). Comparative and population genomics of buckwheat species reveal key determinants of flavor and fertility. Mol. Plant. **16**, 1427–1444.

INTRODUCTION

Buckwheats are pseudocereal crops, belonging to the genus *Fagopyrum* (Polygonaceae). They have been cultivated for more than 4000 years in different parts of the world (Hunt et al., 2018). Common buckwheat (*Fagopyrum esculentum*) and Tartary buckwheat (*F. tataricum*) are two cultivated species with significant differences in their distribution, morphology and nutritional value (Lee et al., 2016; Joshi et al., 2020). Common buckwheat is widely distributed in temperate zones, including Eurasia, North America, South Africa, and Oceania, while Tartary buckwheat is mainly distributed in cooler climates, including high altitude. Tartary buckwheat is self-compatible, while common buckwheat is self-incompatible with two flower morphologies, i.e., short-styled and long-styled flowers (Matsui and Yasui, 2020). Flowers with different morphologies can cross-pollinate, while those with the same morphology cannot, resulting in a lower per unit yield of common buckwheat compared with Tartary buckwheat. The flower types of common buckwheat are determined by an *S*-locus containing an *S-ELF3* gene, whose intact sequence is only present in plants with short-styled flowers (Matsui et al., 2003; Yasui et al., 2012).

Due to the well-balanced amino acid composition, rich dietary fiber, and abundant bioactive flavonoids, buckwheat is regarded as an important element of a balanced diet. Rutin, which accounts for ca. 70% of the buckwheat flavonoids pool, is not reported in other crop species and gives pharmaceutical properties to buckwheat, such as maintenance of the normal elasticity of blood capillaries and/or prevention of hypertension and diabetes. It is worth noting that the content of several flavonoids, such as rutin, kaempferol, and quercetin, appears much lower in common buckwheat compared with Tartary buckwheat (Lee et al., 2016). Due to the observation that rutin generates strong bitterness when hydrolyzed, common buckwheat is generally regarded as a tastier alternative to Tartary buckwheat, further increasing its popularity in processed food, for instance, soba in Japan, cold noodles in Korea, galettes in France, blini in Russia, and pancake mixes and muffins in North America (Lee et al., 2016; Kreft et al., 2020). The less bitterness of common buckwheat probably contributes to its wider distribution than Tartary buckwheat, owing to the consumer's preference. The high nutritional value, gluten-free property, and broad geographic distribution, along with the diverse uses make common buckwheat an economically important crop throughout the world.

Reference genomes are valuable resources for investigations of molecular genetics, molecular breeding, evolution, and domestication. The sequencing and assembly of the Tartary buckwheat genome (489.3 Mb) has facilitated the identification of genes involved in secondary metabolism and stress tolerance (Zhang et al., 2017). Based on this reference genome, a comprehensive database of genomic variation was constructed from whole-genome resequencing of 510 Tartary buckwheat germplasms (Zhang et al., 2021). Subsequently, multiple domestication events and key loci associated with agronomic traits were identified (Zhang et al., 2021). Recently, we assembled a high-quality genome (1.08 Gbp) of golden buckwheat (*F. cymosum*) showing a one-to-one syntenic relationship with chromosomes of Tartary buckwheat (He et al., 2022). A comparison of these two genomes revealed the genetic basis of metabolomic divergence and ecotype differentiation (He et al., 2022).

Homo (*F. esculentum* var. *homotropicum*) is a self-pollinated variant of common buckwheat. It has a similar morphology to common buckwheat and is regarded as an ideal model for common buckwheat genome research (Yasui et al., 2004). Here we report a high-quality reference genome of *F. esculentum* var. *homotropicum*. The use of it as a representative of common buckwheat, comparative genomic analysis with Tartary buckwheat, enables us to systematically explore the genetic basis of wider environmental adaptability and more pleasant flavor of common buckwheat than Tartary buckwheat. We also present genome resequencing data from 572 common buckwheat germplasm accessions, and population genomic analysis was used to identify genes likely responsible for environmental adaptability between native Chinese and non-Chinese cultivated groups. Based on genome-wide association study (GWAS) of multi-year agronomic traits and flavor-associated metabolites, a series of candidate genes likely responsible for yield, fertility, and flavor were identified. This study provides insights into the self-incompatible and unique flavor of common buckwheat and will eventually help accelerating the breeding process of common buckwheat.

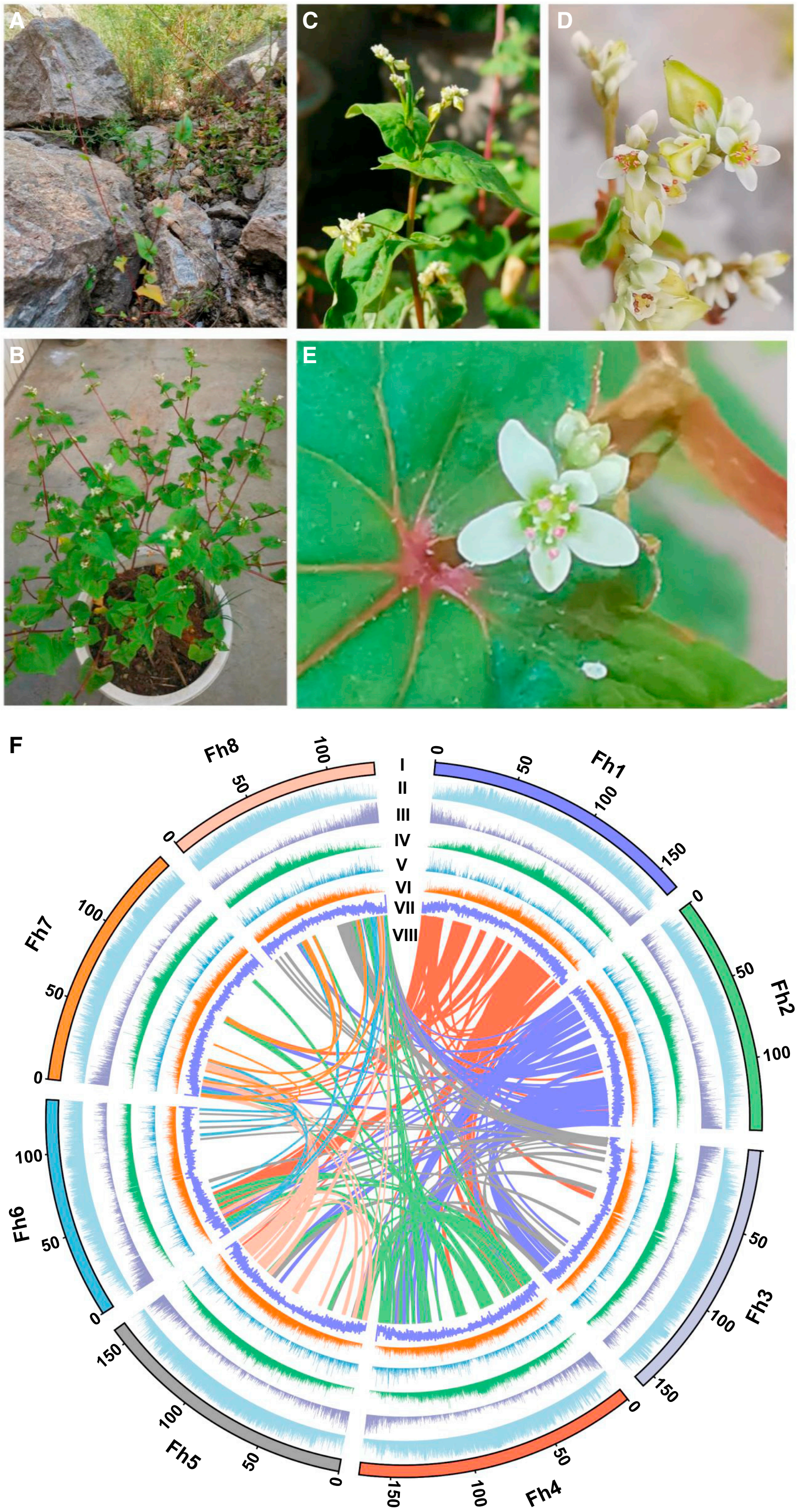

Figure 1. Morphological and genomic features of *F. esculentum* var. *homotropicum*
(A–E) Plants grown in the wild **(A)** and greenhouse **(B–E)**. The whole plant **(A and B)**, leaves **(C)**, and flowers **(D and E)** are shown.
(F) Circos plot of genomic features in *F. esculentum* var. *homotropicum*. The genomic feature in the concentric circles indicates the eight pseudochromosomes (I); the density of TEs, genes, Gypsy elements, Copia, elements, and LTRs, respectively (II-VI); GC content (VII); and intra-genome collinear relationships of the pseudochromosomes (VIII).

RESULTS

Genome assembly and annotation

The genome of *F. esculentum* var. *homotropicum* (Homo; Figure 1A–1E) was *de novo* assembled using 45.83 Gb of PacBio HiFi long reads, 151.98 Gb of Illumina short reads, and 200.82 Gb of Hi-C data (Supplemental Table 1). Based on *k-mer* analysis, the estimated genome size of Homo was 1.33 Gb with a low level of heterozygosity (0.05%) and high repeat content (60.32%; Supplemental Figure 1 and Supplemental Table 2). Assembly of the PacBio long reads alone generated an initial genome of 1.24 Gb with a contig N50 of 102.67 Mb (Supplemental Table 3). Using Hi-C crosslinking, 97.23% (1.20/1.24 Gb) of the genome was anchored to eight pseudomolecules spanning 1.20 Gb, which were visualized and corrected by 3D-DNA, with contig and scaffold N50 values of 53.43 Mb and 158.30 Mb, respectively (Figure 1F, Supplemental Figure 2, and Supplemental Tables 5–7). The final assembly showed a size similar to the estimated genome size based on flow cytometry (~1.20 Gb) and a previous study of the common buckwheat genome (He et al., 2022). Evaluation of the genome assembly with Benchmarking Universal Single-Copy Orthologs (BUSCO) demonstrated 96% of conserved orthologs were complete and only 4% were fragmented or missing, testifying to the high quality and completeness of the reference assembly of the Homo genome (Supplemental Table 7).

We annotated 38 704 protein-encoding genes, with an average gene and coding sequence (CDS) length of 3232 base pairs (bp) and 1248 bp, respectively (Supplemental Tables 8 and 9). The predicted genes comprised an average of 4.68 exons of 326 bp length. In addition, we identified noncoding RNAs (ncRNAs), including 121 microRNAs, 2606 tRNAs, 752 rRNAs, and 63 small nuclear RNAs (Supplemental Table 10). Repetitive sequences accounted for a large proportion

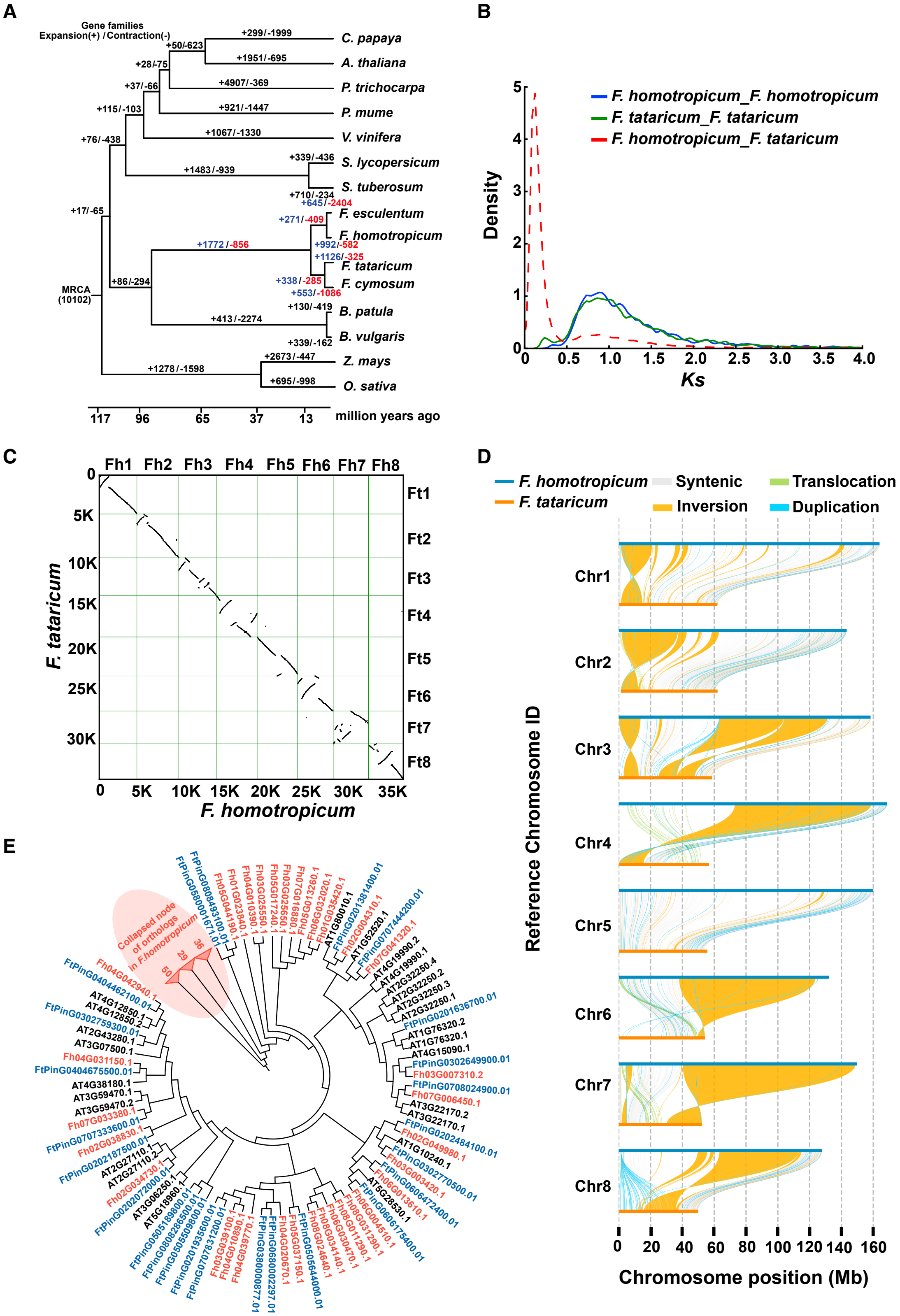

Figure 2. Comparative analyses of Homo and Tartary buckwheat genomes

(A) Phylogenetic tree showing the evolutionary relationship of *F. esculentum* var. *homotropicum* to 14 other plant species, with their divergence time on the x axis. The numbers on branches separated by slashes indicate expansions/contractions of gene families.

(legend continued on next page)

of the Homo buckwheat genome; in total 997.14 Mb of the genome (80.53%) was identified as repeats (Supplemental Table 11). Of these, long terminal repeat retrotransposons (LTR-RTs) are the most abundant repeat type, and the largest LTR-RT superfamilies were *Gypsy* (685.38 Mb, 55.35%) and *Copia* (75.61 Mb, 6.11%). The LTRs were primarily located in intergenic regions; however, 12.52% of the LTRs were found close to genes or in exons/introns (Supplemental Figure 3A). A recent burst of LTR-RT activity about 0.22 Mya was detected in the Homo genome, which was not detected in the Tartary buckwheat genome, where most insertions took place ca. 1 MYA (Supplemental Figure 3B). This recent burst of insertions contributes to the greater genome size of Homo relative to Tartary buckwheat, which has only 189.33 Mb of LTR-RTs.

Divergence between common buckwheat and Tartary buckwheat

To investigate the evolution of the buckwheat genome, we identified 39 422 orthogroups from two monocots and 13 eudicots, including four *Fagopyrum* species (Supplemental Table 12), and performed phylogenetic analysis based on 889 single-copy orthologs (Figure 2A). The phylogenetic tree revealed that Tartary buckwheat and golden buckwheat are more closely related to each other than they are to common buckwheat (Homo and *F. esculentum*; Figure 2A), which is consistent with previous studies (Ohsako and Li, 2020). Tartary buckwheat shared a common ancestor with common buckwheat ca. 11.3 million years ago (Mya), and the divergence of Homo and *F. esculentum* occurred ca. 2.9 Mya (Figure 2A). A total of 33 362 Homo genes were clustered into 17 144 orthogroups, including 403 Homo-specific orthogroups (Supplemental Table 13). We found 590 expanded and 346 contracted orthogroups that show statistical significance ($p < 0.05$) in the Homo genome (Supplemental Table 14). These expanded and contracted orthogroups were significantly enriched in Kyoto Encyclopedia of Genes and Genomes (KEGG) terms of photosynthesis, terpenoid biosynthesis, ABC transporter, and amino acid synthesis (Supplemental Figure 4), which might be related to the evolution of cultivated buckwheat species.

To further explore the divergence of cultivated buckwheat species, we analyzed whole-genome duplications (WGDs). The distribution of Ks (synonymous substitutions per synonymous site) for paralogous gene pairs within Homo and Tartary buckwheat showed a common peak at Ks = 0.88 (Figure 2B and Supplemental Tables 15 and 16), suggesting a common WGD occurring ca. 67.7 Mya. Based on the Ks peak (Ks = 0.12) for orthologous gene pairs between Homo and Tartary buckwheat, we estimated that the divergence of the two species occurred ca. 9.2 Mya (Figure 2B and Supplemental Table 17), close to the divergence time based on the MCMCTREE method (Figure 2A).

We further aligned the genomes of Homo and Tartary buckwheat and identified 86 collinear blocks (19 425 collinear gene pairs; Supplemental Table 18), and several structural variants (SVs) (inversions and duplications) that were present on all chromosomes (Figure 2C and 2D, Supplemental Figures 5 and 6). Given the shared WGD before Homo and Tartary buckwheat diverged and the lack of evidence of large-scale genome duplication, the analysis of synteny further indicated that a non-WGD burst of LTR duplications caused the large size of the Homo genome. The ω value (Ka/Ks) of collinear gene pairs between Homo and Tartary buckwheat was used to detect evidence for positive selection. We found that 82 genes appeared under positive selection (Supplemental Table 19) and were mainly enriched in cellular process, metabolic process, and response to stimulus (Supplemental Figure 7). Seven large inversions (i.e., with more than 1000 genes in each) were identified and genes in these inversions were enriched in gene ontology (GO) terms, including response to metal ions, cellular response to stimulus, toxin metabolic process, and secondary metabolite catabolic process (Supplemental Figure 8 and Supplemental Table 20), which could play a role in the divergent environmental adaptation between the two cultivated species.

We then compared transcription factor (TF) families in Homo and Tartary buckwheat (Supplemental Table 21) and found that the *FAR-RED IMPAIRED RESPONSE 1* (*FAR1*) family has an apparent expansion in Homo (n = 148 copies) relative to Tartary buckwheat (n = 25; Figure 2E). *FAR1* genes play a critical role in plant light response (Tang et al., 2013; Xie et al., 2020), carbon starvation responses (Ma and Li, 2021), and biotic and abiotic stress response (Fernández-Calvo et al., 2020; Liu et al., 2021; Wang et al., 2021), so we speculate that the six-fold difference in the number of *FAR1* genes could be a factor underlying the greater environmental tolerances and global distribution of common buckwheat than Tartary buckwheat.

Flavonoid divergence between common buckwheat and Tartary buckwheat

The different content of flavonoids, especially rutin, is the main reason for differences in bitterness between common and Tartary buckwheat. To quantify metabolic differences between the two cultivated species, we carried out metabolomic analysis of Homo seeds (HSE) and Tartary buckwheat seeds (TSE). A total of 1048 annotated metabolites were characterized, of which the abundance of 503 metabolites was found significantly different between HSE and TSE (Supplemental Table 22), 229 metabolites were significantly more abundant in TS and 274 in HSE (Figure 3A and Supplemental Table 22). Based on KEGG analysis, these differential metabolites were enriched in 82 KEGG terms including these for flavonoids, flavone and flavanol biosynthesis, and ABC transporters (Figure 3B, Supplemental Figure 9, and Supplemental Table 23). Of the differentially abundant flavonoids, 58 were significantly more abundant in TSE, including quercetin, kaempferol, isorhamnetin, quercetin-3-*O*-rutinoside (rutin), kaempferol-3-*O*-rutinoside (nicotiflorin), isorhamnetin-3-*O*-rutinoside (narcissin), naringenin-7-*O*-rutinoside (narirutin), and

(B) Distribution of synonymous nucleotide substitutions inter- or intraspecies.

(C and D) Syntenic analysis of the *F. esculentum* var. *homotropicum* and *F. tataricum* genomes.

(E) Neighbor-joining tree of expanded *FhFARs*. The gene IDs of *F. esculentum* var. *homotropicum*, *F. tataricum*, and *Arabidopsis* are shown in red, blue, and black, respectively.

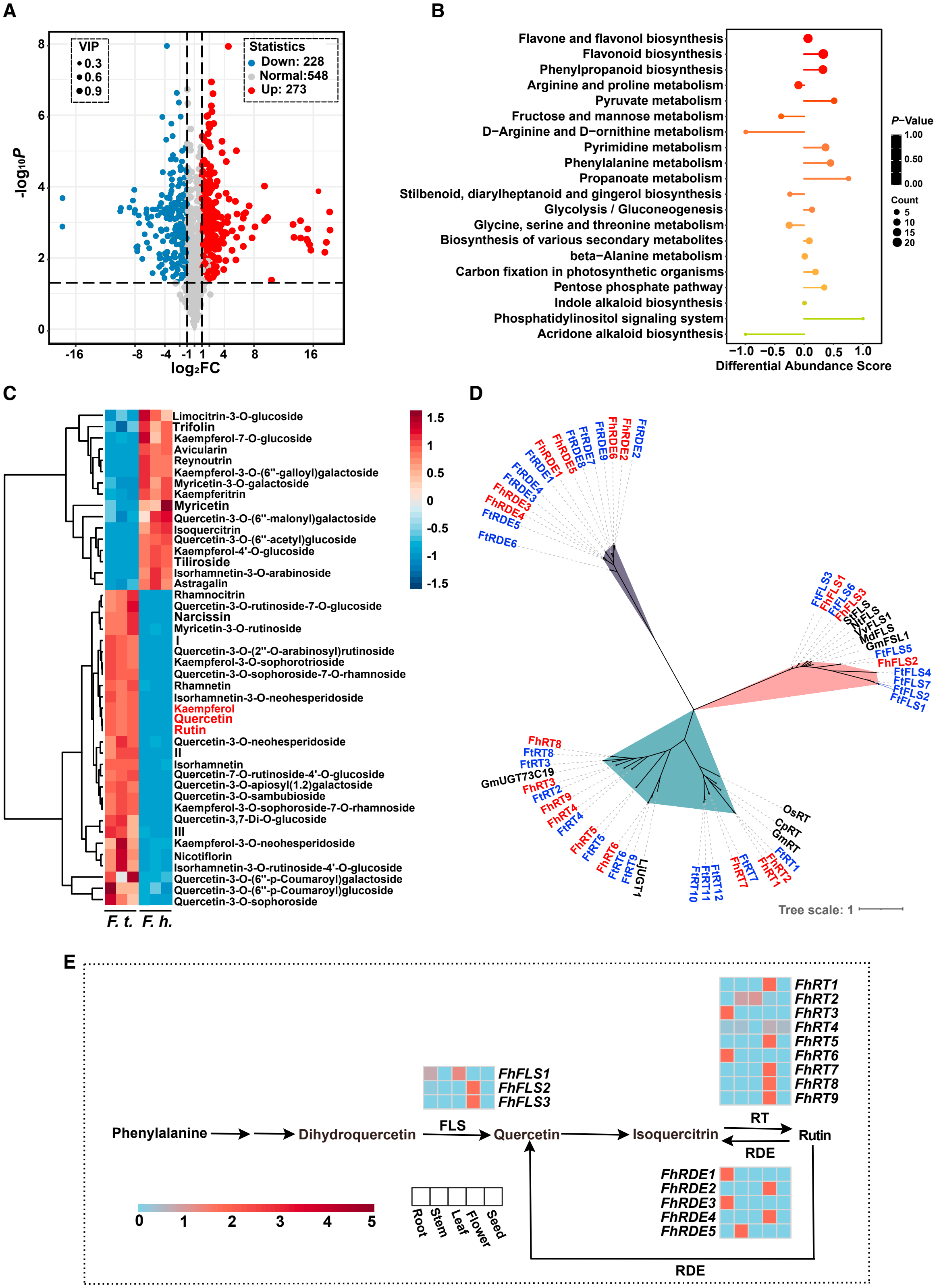

Figure 3. Metabolomic analysis and genes involved in flavonoid divergence between common and Tartary buckwheat

(A) Volcano plot of the metabolic differences between *F. esculentum* var. *homotropicum* and *F. tataricum*. The red and blue dots represent differential metabolites, and the gray dots represent metabolites with nonsignificant differences. The $\log_2$ fold change, *p* value, and VIP thresholds were set to ±1, 0.05, and 1, respectively.

(legend continued on next page)

myricetin-3-*O*-rutinoside (Figure 3C and Supplemental Table 22). Fifty-six flavonoids were significantly more abundant in HS, most of which were glucoside and galactoside containing ones, such as quercetin-3-*O*-glucoside (isoquercitrin), kaempferol-7-*O*-glucoside, kaempferol-4′-*O*-glucoside, kaempferol-3-*O*-glucoside (astragalin), kaempferol-3-*O*-galactoside (trifolin), naringenin-7-*O*-glucoside (prunin), myricetin-3-*O*-galactoside, apigenin-8-C-Glucoside (vitexin), and apigenin-6-*C*-glucoside (isovitexin; Figure 3C and Supplemental Table 22). These results agree with previous reports, that these compounds contribute to the bitter flavor of Tartary buckwheat (Joshi et al., 2020; Lee et al., 2016).

To illuminate the genetic basis of the lower rutinoside content in Homo, we investigated the UDP-glycosyltransferases (UGTs), which are involved in the last two steps of the rutin biosynthetic pathway. The number of *UGT* genes was similar between Homo (171) and Tartary buckwheat (169), and these UGTs were clustered into 20 groups by alignment with *Arabidopsis* UGTs (Supplemental Figure 10). We found nine and 12 rhamnosyltransferase genes (*RTs*), encoding the enzymes in the last step of rutin synthesis, in Homo and Tartary buckwheat, respectively (Figure 3D and Supplemental Table 24). Flavonol synthase (FLS) enzymes catalyze the formation of flavonols from dihydroflavonols, such as dihydrokaempferol to kaempferol and dihydroquercetin to produce quercetin. Here, we found three *FLS* genes in Homo and seven in Tartary buckwheat (Figure 3D and Supplemental Table 24). In addition, β-glucosidase (BGLU) is responsible for the degradation of rutinosides, generating the bitter flavor when buckwheat is processed. The number of rutin degrading enzyme genes (*RDEs*), encoding BGLU is six in Homo and nine in Tartary buckwheat (Figure 3D and Supplemental Table 24). Therefore, the small sizes of rutin synthesis gene families (*FLS* and *RT*) in Homo could play a role in the reduced accumulation of rutin in Homo, and less rutin hydrolysis gene family (*RDE*) could underlie the less bitter flavor of common buckwheat, compared with Tartary buckwheat. Flavonoids are more abundant in flowers and leaves than in roots and seeds in buckwheat, which probably results from the differential expression of flavonoid biosynthesis genes (Li et al., 2010). Here, from transcriptome sequences of different Homo tissues, we found that the expression of *FhFLSs*, *FhRTs,* and *FhRDEs* genes exhibited tissue specificity (Figure 3E). Most *FhFLS* (2/3) and *FhRT* (6/9) genes were highly expressed in flowers (Figure 3E), contributing to the highest accumulation of flavonoids in buckwheat flower.

The population structure of common buckwheat represents environmental adaptation

Cultivated buckwheat originated and was domesticated in China, with dissemination from north China to other buckwheat-producing countries (Zhang et al., 2017). To investigate the intraspecific evolution of common buckwheat, we collected 572 common buckwheat accessions from 33 countries, including 433 Chinese accessions (Figure 4A and Supplemental Table 25). The phenotyping of 433 Chinese accessions was performed in three locations for 2 years, namely in Kangbao (Hebei province, 2020), Yuanmou (Yunnan province, 2021), and Danzhou (Hainan province, 2021). Common buckwheat is a cross-pollinated crop and distributed in a wide range of latitude and longitude, variation in the flowering and maturity time will greatly affect their pollination and yield. Therefore, we focused on full bloom date (FB; 50% of plants are flowering), maturity date (MD), and main stem number (MSN) determining plant height (Supplemental Table 26). The same phenotypic traits and scoring criteria were implemented across these populations. As rutin is among the most important quality trait of buckwheat, we further quantified the rutin content (RC) of seeds, which ranged from 0.022 to 1.184 mg/g (Supplemental Table 26). Moreover, we recorded the flower morphology (FM) of the single plant whose leaves were used for DNA extraction and genome resequencing (Supplemental Table 26).

We resequenced 572 accessions to an average depth of 10× (Supplemental Table 25), and 24 847 225 single-nucleotide polymorphisms (SNPs) were identified by alignment to the Homo reference genome. We employed phylogenetic analysis, principal-component analysis (PCA), and model-based clustering to examine the population structure in common buckwheat. The phylogenetic analysis divided all accessions into two major clades: group I included all 169 samples that are from outside China as well as a few Chinese accessions, and group II included 403 Chinese accessions (Supplemental Figure 11). The PCA (Figure 4B) and model-based clustering analysis (K = 2; Figure 4C) further supported the population structure, primarily dividing the accessions into Chinese and non-Chinese accessions. Therefore, it appears that the accessions from outside China are derived from a small sub-population of Chinese accessions in group I and not from group II, with the latter providing possible untapped breeding material. The cross-validation error exhibited the optimum K = 2 (Supplemental Figure 12), and the group I accessions remained predominantly as two clusters when K = 3 or 4 (Figure 4C). This may have resulted from cross-pollination of common buckwheat and the frequent germplasm exchange within China.

The accessions used in this study are distributed across a wide range of environments, which may account for some of the population structure. We estimated population differentiation (F_{ST}) between group I and group II and identified 339 regions, containing 1637 genes, in the top 1% of the F_{ST} distribution (Figure 4D and Supplemental Table 27). GO enrichment of these genes included flower development, regulation of metabolic process, embryo development, embryo development ending in seed dormancy, and post-embryonic development (Supplemental Figure 13). We assume that a subset of these high F_{ST} regions

(B) KEGG pathway enrichment analysis and differential abundance (DA) scores of differential metabolites (DMs) in *F.t.* vs. *F.h.*.

(C) Heatmap and clustering of differentially expressed flavonoids in *F. esculentum* var. *homotropicum* and *F. tataricum*.

(D) Neighbor-joining tree of *RTs*, *FLSs,* and *RDEs* genes of *F. esculentum* var. *homotropicum* and *F. tataricum*. Genes from *F. esculentum* var. *homotropicum* are marked with red and *F. tataricum* with blue.

(E) A simplified representation of the biosynthetic pathway of rutin metabolism. The expression of each gene in different tissues (root, stem, leaf, flower, and seed) of *F. esculentum* var. *homotropicum* is presented in terms of FPKM.

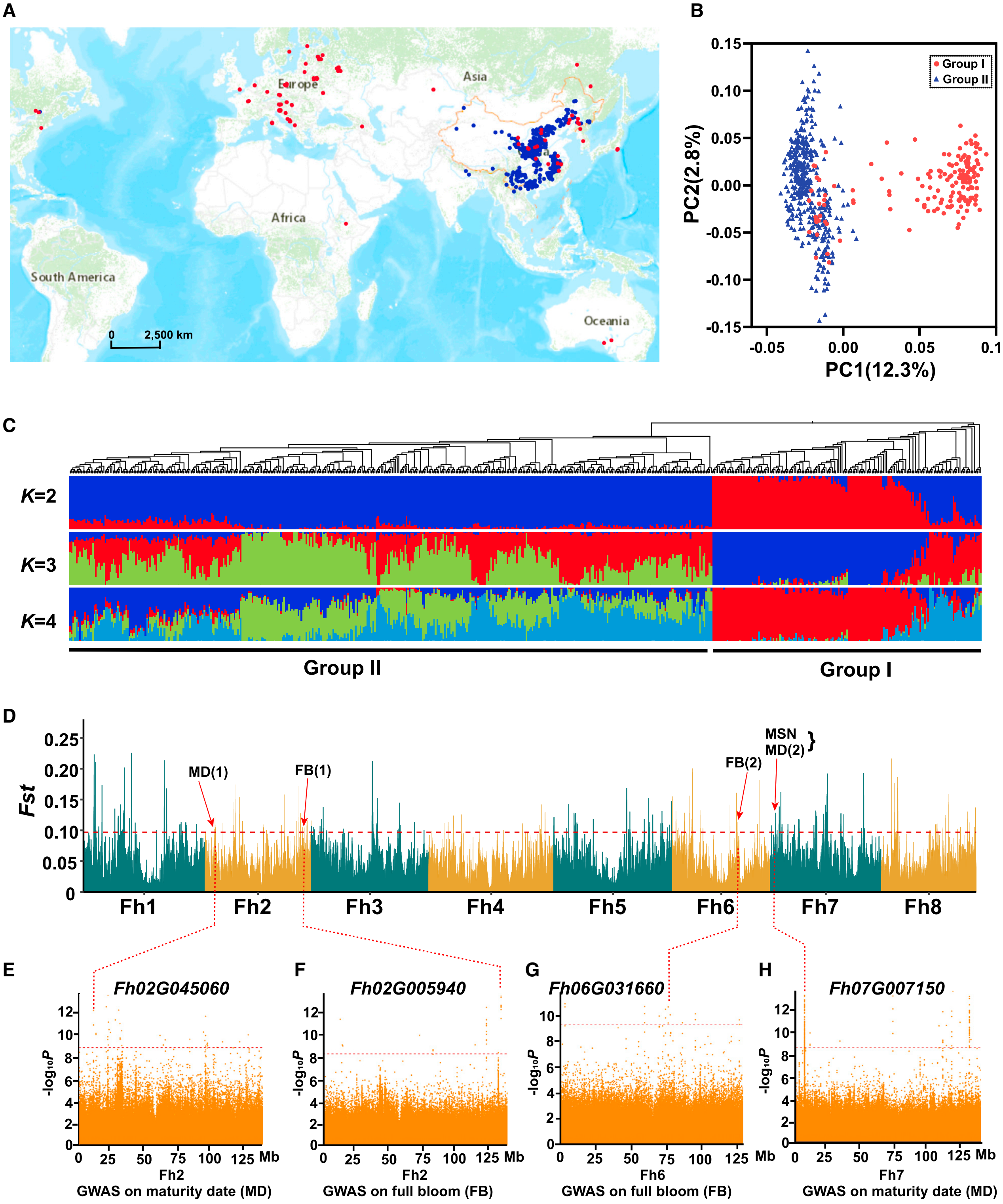

Figure 4. Population feature and divergence of common buckwheat

(A) Geographic distribution of common buckwheat accessions used in the population genomic analysis. Accessions found in groups I and II are shown in red and blue, respectively.

(B) Principal-component analysis of common buckwheat accessions, showing the first two components. Colors correspond to the group I and group II.

(C) Population structure analysis with different numbers of clusters (K = 2, 3, and 4) with the phylogenetic tree superimposed on top.

(D) Highly divergent genomic regions between group I and group II. The horizontal dashed line indicates the top 1% of F_{ST}. Red vertical lines indicate high-divergence regions that overlapped GWAS signals.

(E–H) Local Manhattan plots of GWAS signals overlapping with high-divergence genomic regions for maturity date **(E and H)** and full bloom **(F and G)**.

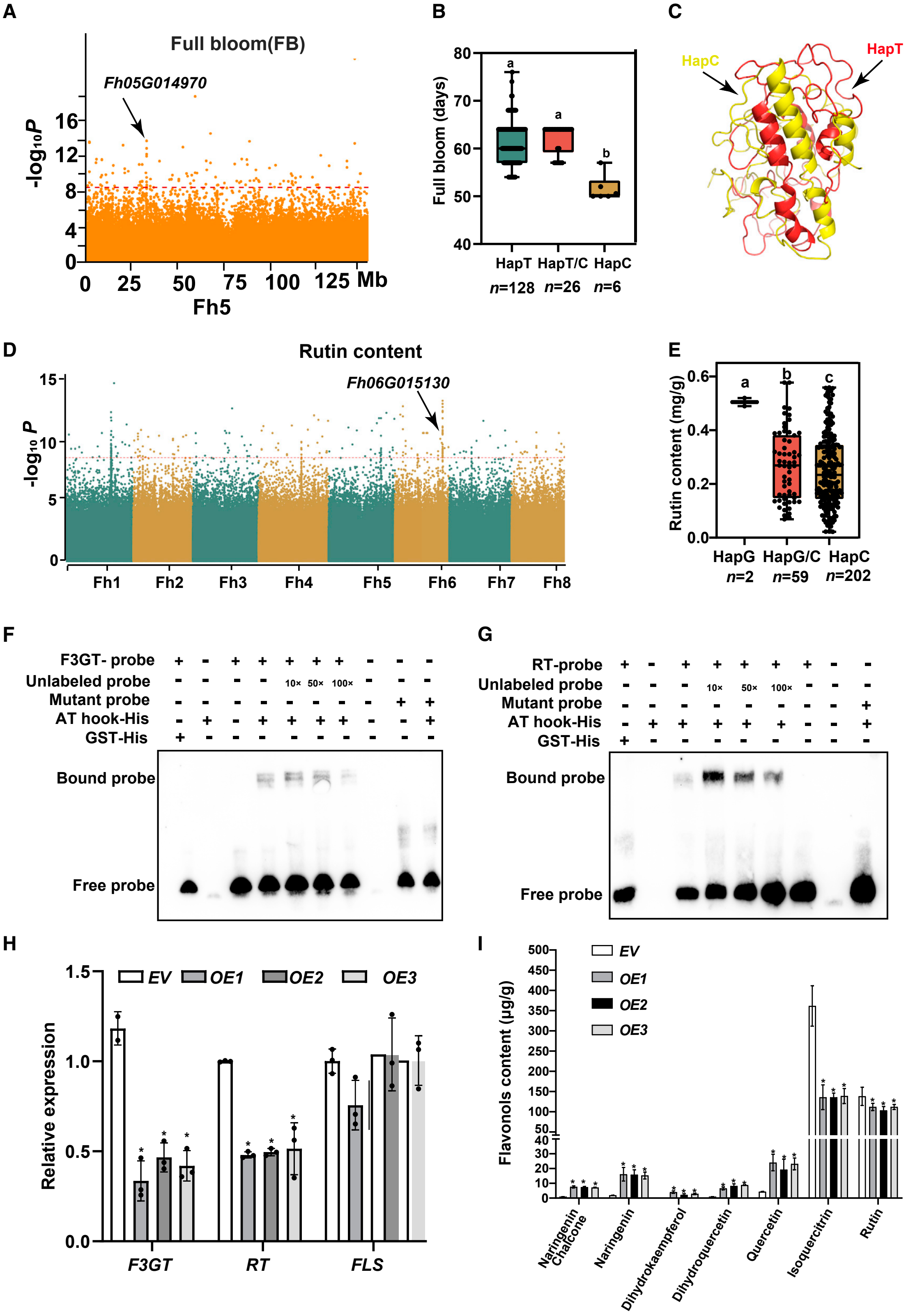

Figure 5. Identification of genes associated with agronomic traits of buckwheat

(A) Manhattan plot for GWAS association with full bloom (FB, days). The dashed line indicates the threshold −logP = 8.5. The black arrow indicates the SNP in *Fh05G014970*.

(legend continued on next page)

underlie divergence between the two groups. Two TF genes in the high F_{ST} regions, *Fh06G000090* (*FhNF-Y*) encoding a nuclear factor Y (NF-Y) and *Fh02G044180* (*FhB3P*) encoding a B3 domain protein, were also found to be under positive selection between the two cultivated buckwheats (Supplemental Table 19). NF-Y plays multiple essential roles in plant growth, flower development, and drought tolerance (Hou et al., 2014; Hwang et al., 2019). The B3 domain-containing protein is critical for accelerating flowering in response to prolonged cold treatment and promotes flowering by directly activating flowering activation genes *SOC1* and *FT* (King et al., 2013; Yu et al., 2020). There is a possibility that the divergence of flower time and stress responses occurred in intraspecific evolution of common buckwheat.

Identification of genes associated with environmental adaptation and flavonoid metabolism in common buckwheat

To identify the genetic loci related to important buckwheat traits, GWAS for MSN, MD, FB, FM, and RC was carried out (Supplemental Figure 14). Five GWAS signals were associated with MSN, MD, and FB (Figure 4D–4H), including two genes for MD, *Fh02G045060* and *Fh07G007150*, and two genes for FB, *Fh02G005940* and *Fh06G031660* (Supplemental Table 28). Functions of their orthologs in *Arabidopsis* (Motose et al., 2004; Sugiyama et al., 2006; Vij and Tyagi, 2006; Wohlbach et al., 2008; Solanke et al., 2009; Meinke, 2020) suggest that these genes are related to plant growth, development, and resistance to stresses.

The GWAS for FB in Hebei (2020) identified in total 38 significant genomic regions comprising 191 genes (Supplemental Figure 14E–14G and Supplemental Table 29). The gene *Fh05G014970*, encoding a protein phosphatase inhibitor, was also identified in GWAS for FB in Yunnan (2021; Figure 5A). Two alleles were identified in *Fh05G014970*, plants with alleles T and T/C (n = 154) have significantly longer FB than with allele C (n = 6; Figure 5B and Supplemental Figure 15). This non-synonymous SNP (T/C) resulted in the amino acid change from serine (S) to asparagine (N) in the C-domain (Figure 5C) that is the typical phosphorylation site, and therefore could affect protein function (Yasui et al., 2012). All the above results indicate that *Fh05G014970* may play a crucial role in controlling flowering time and thus the crop yield of common buckwheat.

The GWAS for RC identified 68 candidate genes in 17 significantly associated genome regions (Figure 5D and Supplemental Table 30). A gene on chromosome 6 (*Fh06G015130*), encoding an AT-hook motif nuclear-localized protein 20 TF, was identified in the selective sweep between groups I and II (Supplemental Figure 16 and Supplemental Table 31). Two alleles (G, C) and their heterozygote (G/C) were resolved based on an SNP in the exon (Supplemental Figure 17). RC was greater with allele G than allele C and with the heterozygote intermediate (Figure 5E). This non-synonymous SNP leads to the amino acid change from aspartate (D) to histidine (H) at position 236 in the C-terminal, which possibly affects protein-protein interactions. The *Arabidopsis* ortholog of *Fh06G015130* can directly bind to the jasmonate-responsive element (JRE, CAAT[A/G]AA[A/T]) of its target genes in the regulation of plant secondary metabolism (Vom et al., 2007; Zhou and Memelink, 2016). *Fh04G007350* and *Fh03G002630* genes encoding two key enzymes (F3GT and RT) of the last two steps of rutin biosynthesis, possess one and two JREs in their promoters (Supplemental Figure 18), respectively, suggesting possible binding of *Fh06G015130* to their promoters. By employing electrophoretic mobility shift assay (EMSA), the binding of *Fh06G015130* to the promoters of *F3GT* and *RT* was further confirmed (Figure 5F and 5G). Moreover, we overexpressed *Fh06G015130* in common buckwheat hairy roots (Supplemental Figures 19 and 20) and found that the rutin and isoquercitrin content were both significantly decreased, while the content of their upstream precursors, such as quercetin, dihydroquercetin, and naringenin, were increased compared with the wild-type (Figure 5I), indicating a mechanistic link between *Fh06G015130* and rutin biosynthesis. The expression of *F3GT* and *RT* decreased in the *Fh06G015130*, overexpressing hairy root lines (Figure 5H).

In the above GWAS for RC, we also identified an ABC transporter G family gene (*Fh03g007120*; Supplemental Table 30) that is putatively involved in the transport of a large spectrum of structurally different compounds (Grafe and Schmitt, 2021). As transporters play a vital role in the biosynthesis of rutin (Li et al., 2019), we speculate that *Fh03g007120* might be involved in rutin metabolism; however, the function of *Fh03g007120* needs to be further studied in detail. In summary, all these candidate genes and genetic loci can possibly be used to improve the quality and yield of common buckwheat.

(B) Boxplots for full bloom (FB) of common buckwheat based on the three haplotypes of *Fh05G014970*. The central lines indicate the median, and the box limits represent the upper and lower quartiles. Whiskers extend to encompass data that fall within 1.5 times the interquartile range, and dots represent outliers. Different letters above the box and whiskers indicate significant differences ($p < 0.05$, two-tailed Student's *t*-test).

(C) Protein structure of *Fh05G014970* in different haplotypes prediction with AlphaFold. Structural alignment of the HapC (yellow) and HapT (red) protein structure prediction with AlphaFold.

(D) Manhattan plot for GWAS on rutin content. The dashed line indicates the threshold $-\log P = 8.5$. The black arrow indicates the SNP in *Fh06G015130*.

(E) Boxplots showing rutin content for the three haplotypes (Hap). Different letters above the box and whiskers indicate significant differences ($p < 0.05$, two-tailed Student's *t*-test).

(F and G) Binding of *F3GT* **(F)** and RT **(G)** promoter fragments with His-AT-hook with or without various concentrations of competitor probes and the mutant probes. Upper arrow indicates the shift and lower arrow indicates the free probes.

(H) Expression of *FhF3GT* and *FhRT* genes in AT-hook overexpressed hairy root lines evaluated by qRT-PCR. The data are shown as mean ± SEM; $n = 3$ biological replicates; two-tailed Student's *t*-test; $^{*}p < 0.05$. **(I)** the corresponding flavonoid contents in different overexpressed hairy root lines. EV, overexpression lines with A4 *Agrobacterium* with empty vector infection; OE1-3, overexpression lines with A4 *Agrobacterium* harboring pCambia1307-Fh06G015130 recombinant plasmid. The asterisks indicate significant differences between each transgenic line and EV plants (two-tailed Student's *t*-test; $^{*}p < 0.05$).

Determinants of fertility in buckwheat species

Common buckwheat accessions are self-incompatible and heterostylous (Figure 6A), which results in low seed set and per unit yield. Homo, however, is self-compatible and homostylous (Figure 6A), which is ideal for buckwheat breeding and has been used to generate self-compatible common buckwheat lines (e.g., Kyukei SC2 and Kyushu PL4; Matsui and Yasui, 2020). To identify the genomic region determining the heteromorphic self-incompatibility, we carried out GWAS analysis for FM and found a strong signal on chromosome 2 (Figure 6B). This genomic region (49.68–65.17 Mb) overlapped with the candidate *S*-locus region of common buckwheat (He et al., 2023) and was named as S^h locus (Figure 6C).

A supergene *S LOCUS EARLY FLOWERING3* (*S-ELF3*; GenBank: AB668590.1) in the *S* locus has been reported to determine the presence of short-styled flowers (Yasui et al., 2012). Interestingly, there is an absence of homo *S-ELF3* gene (*FhS-ELF3*) in the S^h locus, while it was identified by the GWAS signal (SNP: chr4_108890352) on chromosome 4 (Figure 6B and 6D and Supplemental Table 32). The high coverage of HiFi reads in the Homo genome assembly ruled out the possibility that the location of *FhS-ELF3* gene is caused by incomplete or inaccurate genome assembly (Supplemental Figure 21). High genetic differentiation (F_{ST}) between long and short style individuals extended ~40 kb flanking the homo *S-ELF3* gene (Figure 6E). We then compared the *FhS-ELF3* sequences with *FeS-ELF3* and found the sequence variation in the *FhS-ELF3* exon and the C-terminal FhS-ELF3 protein (Figure 6F and Supplemental Figure 22), which might determine the homomorphic flower. These results further demonstrate that alleles at *S-ELF3* are crucial to control the short flower morph and will help develop molecular markers for identification of flower morphology in buckwheat.

To investigate the phylogenetic relationship of the style morphology in the genus *Fagopyrum*, we amplified the *S-ELF3* homologous genes from different species, including *F*. *cymosum*, *F*. *pugense*, and *F*. *longistylum*. The phylogenetic tree of *S-ELF3* genes clearly divided these buckwheat species into two groups (Cymosum and Urophyllum) as expected (Figure 6G). However, unlike the well-known closer relationship between *F*. *dibotrys* and *F*. *tataricum*, *FdS-ELF3* showed a closer relationship with *FeS-ELF3* and *FhS-ELF3*. This may reflect that Tartary buckwheat has undergone more selection pressure in terms of style morphology and fertility, resulting in a conflict between the topology of the gene tree and that of the species tree.

DISCUSSION

Common buckwheat is among the two main cultivated species of buckwheat, with continuously increasing global popularity. Here, we present a chromosome-scale genome assembly of *F*. *esculentum* var. *homotropicum*, an inbred and self-compatible variant of common buckwheat, which would be able to promote further genetic and genomic studies of *Fagopyrum* species and common buckwheat populations.

For cultivated buckwheat, the geographic distribution as well as phenotypic variations for traits such as flavonoid content and floral morphology varied markedly. The comparative genomics of Homo and Tartary buckwheat identified an expansion of the *FAR1* family in Homo. As *FAR1* genes in other plants are involved in regulating responses to stresses (Fernández-Calvo et al., 2020; Liu et al., 2021; Wang et al., 2021), we assume that the greater number of *FAR1* in common buckwheat (148 vs. 25) is one of the reasons why this species has enhanced environmental adaptability and hence geographic distribution than Tartary buckwheat. We then collected and sequenced 572 common buckwheat germplasm from throughout the world to further enrich the understanding of genetic basis of wide distribution and agronomic traits of common buckwheat. Based on the comprehensive data of genomic and phenotypic variation, we identified multiple genetic loci and genes associated with the flowering period and rutin biosynthesis (Figure 5). The population genomic analysis demonstrated the germplasm clearly separated into two groups, group I and group II. By comparing the genomes from groups I and II, we identified highly divergent regions and selective sweeps, and several genes in these regions are involved in flower development, metabolic process, and stress resistance. Two transcription factors of these genes (*FhNF-Y* and *FhB3P*) that putatively regulate floral development, flowering period, and cold tolerance (King et al., 2013; Hou et al., 2014; Hwang et al., 2019; Yu et al., 2020) have undergone positive selection between Homo and Tartary buckwheat, suggesting their important function for both interspecific and intraspecific evolution. Meantime, we found not only the metabolism divergence between common and Tartary buckwheat, but also a regulatory gene (*Fh06G015130*) responsible for rutin biosynthesis in the selective sweep between group I and II. These results revealed that the environmental adaptation of common buckwheat is accompanied by the variation in their secondary metabolites. In addition, it appears that after the two groups diverged, expansion out of China was from only group I, therefore material from group II could be used in breeding material for outside China that is currently untapped.

Engineering a self-compatible variety has been considered as one of the aims of common buckwheat breeding. The flower types of common buckwheat are determined by an *S*-locus containing an *S-ELF3* gene, whose intact sequence is only present in plants with short-styled flowers (Matsui et al., 2003; Yasui et al., 2012). We locate a candidate S^h locus on chromosome 2 by GWAS of long- and short-styled individuals (Figure 6B), and this genomic region exhibited collinearity with the *S*-locus of common buckwheat (Figure 6C). Several genes in the S^h locus show high identity to genes of common buckwheat (Supplemental Table 32), which might contribute to flower development; however, the key dominant *S-ELF3* gene was absent in the S^h locus. Surprisingly, the *FhS-ELF3* gene is present in the genetic loci (10.87 Mb–10.89 Mb) of chromosome 4, associated with flower morphology. The alignment of the S^h locus and common buckwheat *S* locus identified a structural variation, which might lead to the significant GWAS signal on chromosome 2. A DNA transposon and an LTR insertion existed in the first intron and downstream of the *FhS-ELF3* gene, respectively (Figure 6E), which may contribute to the translocation of this *S-ELF3* gene in different chromosomal locations. The gene sequence variation between *FhS-ELF3* and *FeS-ELF3* may result in their different functions

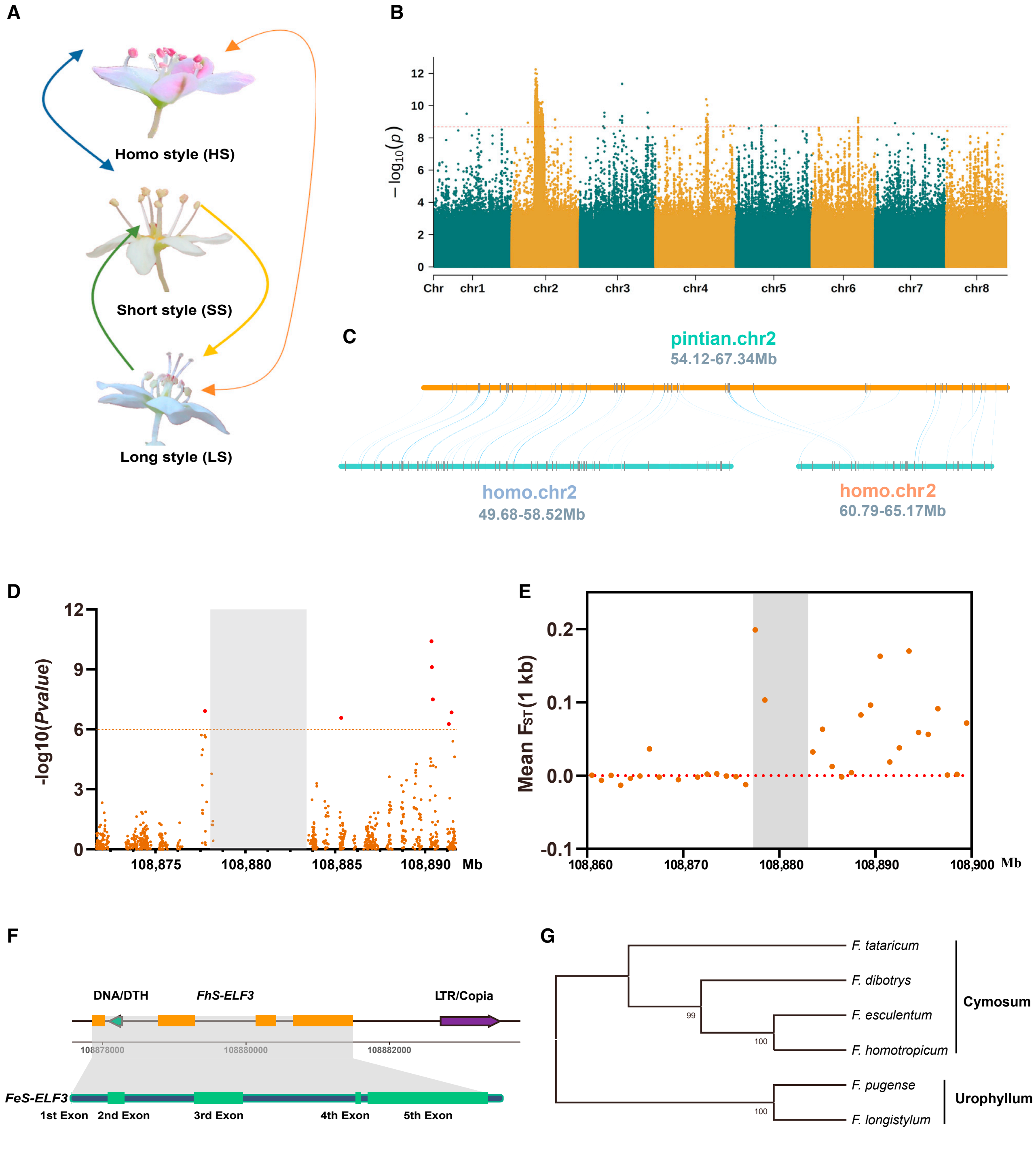

Figure 6. Identification of genes associated with floral morphology of buckwheat

(A) Common buckwheat individuals with short style (SS) and long style (LS), and Homo individual with homo style (HS).

(B) Manhattan plot for GWAS association with flower morphology (FM). The dashed line indicates the threshold −logP = 8.5. The black arrow indicates the peak SNP on Chr2 (chr2_49742739) and Chr4 (chr4_108890352).

(C) Local synteny of candidate *S*-locus in *F. esculentum* (pintian.chr2) and *F. esculentum* var. *homotropicum* (homo.chr2). Lines link the collinearity gene of two chromosomes.

(D) Manhattan plot of the local region neighboring the *FhS-ELF3* gene.

(E) Genetic differentiation (F_{ST}) in 1kb-windows around 20 Mb of the *FhS-ELF3* gene.

(F) Position and schematic diagram of *FhS-ELF3* gene structure (orange) and the comparison with *FeS-ELF3* (green).

(G) The phylogenetic tree of *S-ELF3* in *Fagopyrum* species.

on the styled determination. Further molecular experiments are strongly recommended to investigate the gene function of *FhS-ELF3* and *FeS-ELF3* and other floral genes involved in the *S*-locus. In the meantime, searching for excellent haplotypes would be beneficial for the molecular breeding for buckwheat.

METHODS

Plant materials and phenotyping

Homo buckwheat (*F. esculentum* var. *homotropicum*; collected from Ninglang, Yunnan province) was grown in the greenhouse and young leaves of 4-week-old plants were used for DNA extraction and sequencing. For genome resequencing, 572 common buckwheat accessions were collected from 33 countries (Figure 4A and Supplemental Table 24), which covered most areas where common buckwheat is currently cultivated. For phenotyping, 433 Chinese common buckwheat accessions were planted in Beijing (39°58′ N, 116°20′ E) in 2019, Kangbao (Hebei province, 41°51′ N, 114°36′ E) in 2020, Yuanmou (Yunnan province, 25°42′ N, 101°52′ E), and Danzhou (Hainan province, 19°31′ N, 109°34′ E) in 2021. Five plants from each accession were used for the measurement of the agronomic traits. One of the five plants in Hebei were used to record the flower morphology and its leaves were used for DNA extraction. Seeds were used for the measurement of RC.

Whole-genome sequencing

Genomic DNA was extracted using the CTAB method (Allen et al., 2006). The long-read library was size-selected by Sage ELF for molecules about 20 kb. After primer annealing and the binding of SMRT bell templates to polymerases with the DNA/Polymerase Binding Kit, sequencing was carried out on PacBio Sequel II platform. Three cells were obtained in HiFi mode.

Genome survey

In order to estimate the genome size, heterozygosity, and repeat content of *F. esculentum* var. *homotropicum*, jellyfish (v2.1.4; Marcais and Kingsford, 2011) was used to generate a 17–35 K-mer counts table for the high-quality reads. GenomeScope (Vurture et al., 2017) was used to estimate the general characteristics of the genome.

Hi-C library construction and sequencing

The Hi-C libraries were prepared using fresh immature shoots of *F. esculentum* var. *homotropicum* according to standard procedures. Nuclear DNA was first cross-linked *in situ*, extracted, and then digested by the restriction enzyme *Mbo*I. The biotinylated DNA fragments were enriched to construct a sequencing library, and sequencing on an MGI DNBSEQ™ platform; 200Gb data for the sample was obtained.

Genome assembly and quality assessment

After running the pbccs read consensus tool (v6.0.0, smrtlink v10.1.0.119588) with the parameter (–hifi-kinetics –chunk 22/30 -j 30 –min-passes 3 –min-rq 0.99), we recovered 45.83 Gb of sequence in HiFi reads. The primary assembly was performed with PacBio long reads using HiCanu (v2.1) (Koren et al., 2017) with default parameters. After assembly, the following approaches were employed to evaluate the quality of genome: (1) BUSCO (Simao et al., 2015) was applied to evaluate the genome completeness using the embryophyta_odb10 as query; and (2) short reads were mapped against the genome to evaluate the mapping rate and coverage of Illumina reads to genome.

For the chromosome-level genome assembly, Hi-C reads were first mapped to contigs with default parameters (-m haploid -t 15 000 -s 2) using Juicer alignment pipeline (v1.5.6) (Durand et al., 2016). Contigs were then clustered into pseudochromosomes using the 3D-DNA pipeline (Dudchenko et al., 2017) with default parameters. The default parameters were used in contig anchoring steps. As original anchored results always have disordered and misoriented contigs, manual correction and validation were performed to obtain the final nuclear genome sequences using Juicerbox (v2.13.07, https://github.com/aidenlab/Juicebox).

Gene annotation

Three strategies (homology, transcriptome, and *ab initio*) were used to predict the protein-coding genes of the *F. esculentum* var. *homotropicum* genome. Homologies from three species (*Arabidopsis thaliana*: https://arabidopsis.org/download_files/Genes/TAIR10_genome_release/TAIR10_chromosome_files/TAIR10_chr_all.fas.gz, *F. cymosum*: (He et al., 2022), and *F. tataricum*: https://ncbi.nlm.nih.gov/genome/38383) were used for blast searches against the genome and the output was used to predicted gene structure using GeneWise (https://www.ebi.ac.uk/~birney/wise2/) with default parameters. The RNA-seq reads from multiple tissues were mapped to the genome using HISAT2 (v2.1.0; Kim et al., 2015) to generate gene structures and train a model for AUGUSTUS (http://augustus.gobics.de/; Stanke et al., 2004). *Ab initio* gene prediction was performed with AUGUSTUS with the trained model. The Geta pipeline (https://github.com/chenlianfu/geta) was used to integrate annotation from all homology-based, transcriptome-based and *ab initio* predictions to generate a comprehensive protein-coding gene set. The tRNAs were predicted using tRNA scan-SE (v1.3.1; Lowe and Eddy, 1997), and other ncRNA sequences were searched against Rfam (http://rfam.xfam.org/). For functional annotation, predicted proteins were used in a blast search against various functional databases including NT (http://ncbi.nlm.nih.gov/nucleotide/), Swissprot (http://web.expasy.org/docs/swiss-prot_guideline.html), NR (https://ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz), eggNOG (https://eggnogdb.embl.de/; Powell et al., 2012), GO (http://geneontology.org/; Ashburner et al., 2000), and KEGG (https://www.genome.jp/kegg/; Kanehisa et al., 2012) and HMMER (v3.1; Finn et al., 2011) search against PFAM (http://xfam.org/).

Repeat annotation and estimation of LTR insertion times

Tandem repeats and transposable elements were identified for the *F. esculentum* var. *homotropicum* genome. We used Tandem Repeat Finder (v4.09; Benson, 1999) to predict tandem repeats with the following parameters: Match = 2, Mismatch = 7, Delta = 7, PM = 80, PI = 10, Minscore = 50, and MaxPeriod = 500’. A combination of *de novo* and homology-based approach was used to identify transposable elements. RepeatMasker (v4.0.7; https://www.repeatmasker.org/) and RepeatProteinMask were used to search for known repeat sequences; LTR_retriever, LTR_FINDER (v1.05; https://github.com/xzhub/LTR_Finder) and RepeatModeler (v1.0.11; https://www.repeatmasker.org/RepeatModeler/) were then used to search the *de novo* repeats. The LTR libraries and consensus repetitive libraries identified by RepeatModeler were combined as the input data for RepeatMasker. The predicted LTRs were further classified into intact and non-intact LTRs, and the insertion time was calculated using the formula T = K/2r (K is the divergence rate, and r is the neutral mutation rate, we used 6.5×10^{-9} in LTR_retriever) using the scripts implemented in the LTR_retriever package (Ou and Jiang, 2018).

Phylogenetic analysis and gene family analysis

Fourteen species were selected to construct a phylogenetic tree with *F. esculentum* var. *homotropicum*, including 12 eudicots (*Beta vulgaris*, *F. tataricum*, *F. esculentum*, *F. cymosum*, *Populus trichocarpa*, *Solanum tuberosum*, *Vitis vinifera*, *Solanum lycopersicum*, *Carica papaya*, *Beta patula*, *Arabidopsis thaliana,* and *Prunus mume*) and two monocots (*Oryza sativa* and *Zea mays*) downloaded from NCBI. Protein sequences were filtered by removing short sequences (less than 50 amino acids) and choosing the longest isoform to represent each protein-coding gene. All-versus-all blastp (v2.2.26) was performed with an E-value cutoff of 1e-5 for all species proteins. Then we used OrthoMCL (v2.09; Li et al., 2003) to cluster genes into families with the MCL inflation parameter of 1.5. Multiple sequence alignment was performed with MUSCLE (https://www.drive5.com/muscle/; Edgar, 2004). Four-fold degenerate sites were extracted from sequence of each single-copy gene families and concatenated to a supergene for each species. PhyML (v3.0; Guindon et al., 2010) was used to construct a phylogenetic tree using four-fold degenerate site with maximum likelihood method. MCMCTREE in the package of PAML (http://abacus.gene.ucl.ac.uk/software/paml.html; Yang, 2007) was used to estimate divergence time using soft fossil calibrations collected from the TimeTree website (https://timetree.org/; Sanderson, 2003). We calibrated the model using divergence time between *F. tataricum* and *A. thaliana* (111-131 Mya), *S. lycopersicum* and *B. vulgaris* (107-131 Mya), and *S. lycopersicum* and *S. tuberosum* (5.23–9.40 Mya). The CodeML utility in the PAML (Yang, 2007) software package was used to calculate the Ka and Ks rates of orthologous genes between Homo and Tartary buckwheat.

To model gene family expansion and contraction across the phylogeny, CAFÉ (v4.1; https://sourceforge.net/projects/cafehahnlab/) with parameters "-p 0.05 -t 1 -r 10 000" was used. We employed probabilistic graphical models to estimate the size of each gene family at each ancestral node of the phylogenetic tree topology using the orthologous genes inferred from OrthoMCL, and to obtain a family-wise *p* value (based on a Monte Carlo re-sampling procedure) to indicate whether it has a significant expansion or contraction in each gene family across species. Transcriptions factors were predicted in PlantTFDB (http://planttfdb.gao-lab.org/) with iTAK.

Genome synteny and WGD

Syntenic blocks were identified using JCVI (v0.84; https://github.com/tanghaibao/jcvi) with parameters "-g -3 -e 1e-05 -u 10 000." The syntenic regions of the target species and the comparison species were shown graphically. The MCScanX (Wang et al., 2012) was first employed to identify syntenic blocks of gene pairs. To assess the history of WGD in *F. esculentum* var. *homotropicum*, a whole-paranome approach was used to obtain an initial Ks distribution where genes were first clustered, followed by pairwise comparison and Ks estimation within clusters. Whole-paranome Ks estimation and subsequent mixture modeling were performed with the WGD package using the commands ksd and mix (Dong et al., 2019).

Chromosome structural variation analysis

Taking a query and referenced genome *F. esculentum* var. *homotropicum* and *F. tataricum*, genome comparation was conducted by nucmerv4.0 with the default parameters (-1 -i 90 -L 100) and delta-filter for filtering. Assemblytics (Nattestad and Schatz, 2016) was used to identify structured variation that contained insertions (INS), deletions (DEL), and duplications (DUP) with the parameter "unique_length_required: 500, min_size: 50, max_size: 1 000 000." The obtained results are converted into VCF files using SURVIVOR software. The inversion (INV) and Translocation (TRA) are detected using Syri (Goel et al., 2019) and combined with final SV results, then ANNOVAR (Wang et al., 2010) was used to annotate the variation.

Sequencing and variant calling

Total genomic DNA was extracted from a single plant for each accession using a modified CTAB method (Marcais and Kingsford, 2011). Libraries with 500 bp insert size were constructed and sequenced on Illumina HiSeq X Ten platform (Illumina, San Diego, CA) with 150 bp read length of paired-end (PE) reads by Annoroad Gene Technology (Beijing, China). The raw PE reads were filtered using fqtools (v0.1.8) with the following criteria: (1) reads containing more than five adapter-polluted bases, (2) reads with low quality (Phred quality value < 19) accounting for more than 50% of the total bases, and (3) reads with more than 5% N bases. The clean reads of each accession were then mapped onto the genome using BWA (Li and Durbin, 2009) with default parameters. The mapped reads were sorted using SAMtools (v0.1.19), and duplicate reads were removed using Picard (v1.13; https://broadinstitute.github.io/picard/) MarkDuplicates. Reads mapping more than two places were also filtered out.

The Genome Analysis Toolkit (McKenna et al., 2010) from sentieon (https://www.sentieon.com/) was applied for variant calling. HaplotypeCaller model was used for SNPs and Indels calling, and the GVCFs of each accession were merged using GVCFtyper. The SNPs and Indels were filtered with GATK VariantFiltration (QD < 10, FS > 10.0, DP < 4, QUAL <30). ANNOVAR (Wang et al., 2010) was employed to annotate for all the qualified variants based on the GFF file.

Population genomic analysis

All SNPs were filtered to remove loci with minor allele frequency (MAF) <0.02, missing rate of SNPs site >0.1 and missing rate of all samples >0.3, and the remaining 24 237 975 SNPs were used for population analyses. SNPs in high linkage disequilibrium (LD) were discarded with PLINK (–indep-pairwise 50 5 0.5), and the filtered data (11,744,764 SNPs) were supplied for PCA, and phylogenetic and population structure analysis.

A phylogenetic tree was constructed with treebest (v1.9.2) (www.treesoft.sourceforge.net/treebest.shtml) with 100 bootstrap replicates based on the NJ model. The tree was visualized and modified by iTOL (https://itol.embl.de/). The PCA was conducted using EIGENSOFT (v6.0.1; Price et al., 2006) (https://github.com/DReichLab/EIG). The ADMIXTURE (Alexander et al., 2009) analysis was also used for estimating the population structure. To identify the most likely population group number, the initial burn-in period was set to 50 000 followed by 100 000 Markov chain Monte Carlo iterations, and *K* (the tested number of populations) was set from 2 to 8. Finally, delta K was calculated to predict the best K value according to the previous methods (Evanno et al., 2005).

Pairwise population differentiation (F_{ST}) was estimated using VCFtools with a 50-kb sliding window and 25-kb steps, and the top 1% values were considered as the candidate high-divergence regions. The cross-population composite likelihood ratio test XP-CLR v1.0 (Chen et al., 2010) was performed in 50-kb sliding windows with a step size of 25 kb. The highest XP-CLR values, accounting for 5% of the genome, were considered as selected regions.

Metabolome analysis

One gram of dried seed samples of *F. tataricum* cv. Pinku and *F. esculentum* var. *homotropicum* (three replicates per group) were weighed and frozen in liquid nitrogen. Sample preparation for the metabolomics and data analysis were performed at Wuhan Metware Biotechnology Co., Ltd (Wuhan, China) using standard procedures, and three biological replicates of each treatment were analyzed. Ultra-performance liquid chromatography (UHPLC; SHIMADZU Nexera X2, https://www.shimadzu.com.cn/) in tandem with mass spectrometry (MS/MS; Applied Biosystems 4500 QTRAP, https://www.appliedbiosystems.com.cn/) were used for data acquisition. Linear Ion Trap (LIT) and triple quadrupole (QQQ) scans were obtained on a triple quadrupole linear ion trap mass spectrometer (QTRAP) in AB4500 Q TRAP UPLC/MS/MS system. By using the self-build database MWDB (Metware database), substances were qualitatively analyzed according to the secondary spectrum information, and the multiple reaction monitoring mode (MRM) of triple quadrupole mass spectrometer was used to quantify metabolites (Fraga et al., 2010). Metabolite annotation is based on the accurate mass of metabolites, MS2 fragments, MS2 fragment isotope distribution, and retention time (RT). Through the self-developed intelligent secondary spectrum matching method, the secondary spectrum and RT of the metabolites in the samples are compared intelligently with the database MWDB one by one. The MS tolerance and MS2 tolerance are set to 2 ppm and 5 ppm, respectively. In the MRM mode, the quadrupole first filters the precursor ions (precursor ions) of the target substance, and excludes ions corresponding to other molecular weight substances to initially eliminate interference. The precursor ions are fragmented after the induced ionization of the collision chamber. A characteristic fragment ion was selected by triple quadrupole filtering to eliminate the interference of non-target ions. Through a PCA, an orthogonal partial least squares discriminant analysis model and DAMs were screened with fold change ≥ 2, $p \leq 0.05$, and VIP (variable importance in project) ≥ 1. Finally, the KEGG database was used for the pathway enrichment analysis of DAMs.

Measurement of flavonoids content

Seeds were dried at 105°C for 30 min and at 65°C until reaching constant weight, and then ground and filtered using an 80-mesh sieve; 0.1 g powder was used for flavonoid extraction in 10 ml 80% methanol (v/v). Transgenic hairy roots were dried to a constant weight and 0.1 g ground powder was used for flavonoid extraction in 5 ml of 80% methanol (v/v). Ultrasonic extraction was carried out at 50°C and 80 kHz for 30 min. Crude extracts were passed through a 0.22-μm organic microporous filter and then analyzed using a UPLC-QQQ/MS (Agilent UPLC 1290II-G6400 QQQ MS, Agilent, Santa Clara, CA, USA). The mobile phase consisted of solvent A, pure water with 0.1% formic acid, and solvent B, acetonitrile with 0.1% formic acid. Sample measurements were performed with a gradient program that employed the starting conditions of 98% A, 2% B kept for 2 min. From 2 to 4 min, a linear gradient to 90% A, 10% B was programmed, followed by a linear gradient to 20% A, 80% B from 4 to 11 min. Subsequently, a composition of 2% A, 98% B was adjusted within 0.10 min and kept for 1.9 min. Finally, a composition of 98% A, 2% B was adjusted within 0.10 min and kept for 1.9 min. The column oven was set to 40°C and the injection volume was 2 μL. The effluent was alternatively connected to an ESI-triple-quadrupole LIT (Q TRAP)-MS. Flavonoid content was determined by comparing the peak area with authentic standards (Sigma-Aldrich, USA).

Genome-wide association study

After filtering the SNP data for the samples used in the phenotypic characterization (MAF $\leq$0.05 and miss $\leq$0.02), 24 847 225 SNPs were used in the following analyses. First, LD analysis was carried out with PopLDdecay (v1.29; https://github.com/BGI-shenzhen/PopLDdecay). Subsequently, Genome-wide Efficient Mixed Model Association algorithm (GEMMA; Zhou and Matthew, 2012) with the LMM model was used for GWAS analysis on the agronomic traits including full bloom (FB), maturity date (MD), main stem number (MSN), rutin content (RC), and flower morphology (FM). The effective SNP numbers were calculated using gec v0.2 (Li et al., 2012), and a stringent Bonferroni correction (0.05/the effective SNP number) was used to collect the association signals based on p value (effective SNP number = 16 024 279, threshold $\approx$8.5). Manhattan and qq-plot plots were visualized by CMplot packages (Yin et al., 2021).

Gene clone and transgenic hairy roots generation

The CDS fragments of *Fh06G015130* were amplified using PCR with the primers illustrated in Supplemental Table 29 and inserted into the pCambia1307 vector. Recombinant plasmids were transformed into *Agrobacterium* A4 to generate transgenic hairy roots using previously described methods (Zhang et al., 2021). Transgenic lines identified by PCR were then moved to MS (HRY0519304A, PhytoTech, USA) liquid medium including 100 mg/ml cefotaxime with shaking (120 rpm) in the dark at 22°C. Hairy roots were collected after 2 weeks and dried for determination of the rutin content.

Electrophoretic mobility shift assays

The CDSs of *Fh06G015130* were inserted into pET28a containing an His tag, which were expressed in *Escherichia coli* BL21 (DE3 Strain). After an 8-h induction by 0.5 mM isopropyl β-D-thiogalactoside at 22°C, 100 rpm, the recombinant proteins was purified by an Ni-NTA Agarose column according to the previous method (Dahro et al., 2022). Quadruple JRE elements were synthesized according to the sequences of *FhF3GT* and *FhRT* promoters and labeled with biotin at 5′-start to be used as probes. The EMSAs were conducted as the LightShift Chemiluminescent EMSA Kit (Number, 20148, Thermo Fisher Scientific, USA) according to the manufacturer's instructions.

Quantitative RT-PCR analyses

Total RNA was extracted using an RNA Easy Plant Tissue Kit (DP452, Tiangen, Beijing, China). Reverse transcription was performed using Hi-Script III RT SuperMix for qPCR (+gDNA wiper; R323, v21.1, Vazyme, Nanjing, China) and the manufacturer's protocol. The qRT-PCR was carried out using Taq Pro Universal SYBR qPCR Master Mix (Q712, v20.1, Vazyme, Nanjing, China) according to the protocol. The statistical analysis was performed by the $2^{-\triangle\triangle CT}$ method. Primers are illustrated in Supplementary Table 33.

RNA library construction and sequencing analysis

We collected roots, stems, leaves, and flowers of Homo to perform RNA-seq, and total RNA was isolated using TRIzol Reagent (Invitrogen, Carlsbad, California, USA). RNA concentration was measured using Qubit RNA Assay Kit in Qubit 3.0 Flurometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bio-analyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). Libraries were generated using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations. Total RNA (4 μg) with RNA integrity number (RIN) >7.5 was used as input material for library construction. The RNA-seq library was commercially performed using the DNBSEQ-T7 (DNBSEQ-T7, RRID:SCR_017981) with a PE read length of 150 bp at Annoroad Gene Technology Co. Ltd.

Statistical analysis

Two-tailed Student's *t*-tests were carried out for all statistical analyses to determine significance. Significance levels were defined as $^*p < 0.05$; $^{**}p < 0.01$; and $^{***}p < 0.001$.

DATA AND CODE AVAILABILITY

All genomic sequence data for genome assembly, genome resequencing of 572 buckwheat accessions, and raw transcriptome and metabolome data are available from Genome Sequence Archive (GSA) database in CNCB (https://ngdc.cncb.ac.cn/) under the following projects accession PRJCA010349.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This research was supported by the National Key R&D Program of China (2022YFE0140800), the European Union Horizon 2020 project ECOBREED (771367), the Youth Innovation Program of Chinese Academy of Agricultural Sciences (No. Y2022QC02), Project of Sanya Yazhou Bay Science and Technology City (SCKJ-JYRC-2022-22), National Natural Science Foundation of China (32161143005, 31911530772, 32111540258), PlantaSYST (SGA No 739582 under FPA No. 664620), and the BG05M2OP001-1.003-001-C01 project, financed by the European Regional Development Fund through the "Science and Education for Smart Growth" Operational Programme and Slovenian Research Agency, program P4-0077 "Genetics and Modern Technologies of Crops".

AUTHOR CONTRIBUTIONS

M.Z., D.J., V.M., B.S., M.A.C., and M.I.G. designed and managed the project. Y.T., Z.S., C.Z., M.Q., Z.L., M.G., I.K., D.J., V.M., B.P., and M.Z. contributed to genetic material. X.L., X.R., X.Z., and W.Y. contributed to generation of genome assembly, genomic comparison, and whole-genome resequencing data. K.Z., J.L., Z.S., and C.Z. performed phenotyping and quality trait measurement. K.Z., Y.H., X.L., H.Z., Y.S., and M.Z. performed data analysis and/or figure design. K.Z., J.L., Y.L., Y.O., and H.Z extracted DNA and RNA, and performed gene functional analysis.

K.Z., Y.H., X.L., B.S., M.A.C., and M.Z. wrote and finalized the manuscript. All authors read and approved the paper.

ACKNOWLEDGMENTS

We thank Lunhao Shen, Kui Ren, and Ayi Shen for their efforts on the seed harvest and phenotyping. We thank Feng He, Di Yang, Yanrong Hao, and Qian Zuo for their efforts on the preparation of plant samples and molecular experiments. We thank Zelin Yi, Xiaomei Fang, Baili Feng, and the National Crop Genebank of China for providing most of the common buckwheat germplasm resources. No conflict of interest declared.

Received: August 24, 2023
Revised: August 28, 2023
Accepted: August 28, 2023
Published: August 30, 2023

REFERENCES

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. **19**:1655–1664.

Allen, G.C., Flores-Vergara, M.A., Krasynanski, S., Kumar, S., and Thompson, W.F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat. Protoc. **1**:2320–2325.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. Nat. Genet. **25**:25–29.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27**:573–580.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Res. **20**:393–402.

Dahro, B., Wang, Y., Khan, M., Zhang, Y., Fang, T., Ming, R., Li, C., and Liu, J.H. (2022). Two AT-Hook proteins regulate A/NINV7 expression to modulate sucrose catabolism for cold tolerance in *Poncirus trifoliata*. New Phytol. **235**:2331–2349.

Dong, S., Xiao, Y., Kong, H., Feng, C., Harris, A.J., Yan, Y., and Kang, M. (2019). Nuclear loci developed from multiple transcriptomes yield high resolution in phylogeny of scaly tree ferns (*Cyatheaceae*) from China and Vietnam. Mol. Phylogenet. Evol. **139**, 106567.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science **356**:92–95.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click System for analyzing loop-resolution Hi-C experiments. Cell Syst. **3**:95–98.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.

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

Fernández-Calvo, P., Iñigo, S., Glauser, G., Vanden Bossche, R., Tang, M., Li, B., De Clercq, R., Nagels Durand, A., Eeckhout, D., Gevaert, K., et al. (2020). FRS7 and FRS12 recruit NINJA to regulate expression of glucosinolate biosynthesis genes. New Phytol. **227**:1124–1137.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. **39**:W29–W37.

Fraga, C.G., Clowers, B.H., Moore, R.J., and Zink, E.M. (2010). Signature-discovery approach for sample matching of a nerve-agent precursor using liquid chromatography-mass spectrometry, XCMS, and chemometrics. Anal. Chem. **82**:4165–4173.

Goel, M., Sun, H., Jiao, W.B., and Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. **20**:277.

Gräfe, K., and Schmitt, L. (2021). The ABC transporter G subfamily in Arabidopsis thaliana. J. Exp. Bot. **72**:92–106.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59**:307–321.

He, M., He, Y., Zhang, K., Lu, X., Zhang, X., Gao, B., Fan, Y., Zhao, H., Jha, R., Huda, M.N., et al. (2022). Comparison of buckwheat genomes reveals the genetic basis of metabolomic divergence and ecotype differentiation. New Phytol. **235**:1927–1943.

He, Q., Ma, D., Li, W., Xing, L., Zhang, H., Wang, Y., Du, C., Li, X., Jia, Z., Li, X., et al. (2023). High-quality *Fagopyrum esculentum* genome provides insights into the flavonoid accumulation among different tissues and self-incompatibility. J. Integr. Plant Biol. **65**:1423–1441.

Hou, X., Zhou, J., Liu, C., Liu, L., Shen, L., and Yu, H. (2014). Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. Nat. Commun. **5**:4601.

Hunt, H.V., Shang, X., and Jones, M.K. (2018). Buckwheat: a crop from outside the major Chinese domestication centres? A review of the archaeobotanical, palynological and genetic evidence. Veg. Hist. Archaeobotany **27**:493–506.

Hwang, K., Susila, H., Nasim, Z., Jung, J.Y., and Ahn, J.H. (2019). Arabidopsis ABF3 and ABF4 Transcription Factors Act with the NF-YC Complex to Regulate SOC1 Expression and Mediate Drought-Accelerated Flowering. Mol. Plant **12**:489–505.

Joshi, D.C., Zhang, K., Wang, C., Chandora, R., Khurshid, M., Li, J., He, M., Georgiev, M.I., and Zhou, M. (2020). Strategic enhancement of genetic gain for nutraceutical development in buckwheat: A genomics-driven perspective. Biotechnol. Adv. **39**, 107479.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. **40**:D109–D114.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**:357–360.

King, G.J., Chanson, A.H., McCallum, E.J., Ohme-Takagi, M., Byriel, K., Hill, J.M., Martin, J.L., and Mylne, J.S. (2013). The Arabidopsis B3 domain protein VERNALIZATION1 (VRN1) is involved in processes essential for development, with structural and mutational studies revealing its DNA-binding surface. J. Biol. Chem. **288**:3198–3207.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. **27**:722–736.

Kreft, I., Zhou, M., Golob, A., Germ, M., Likar, M., Dziedzic, K., and Luthar, Z. (2020). Breeding buckwheat for nutritional quality. Breed Sci. **70**:67–73.

Lee, D.G., Woo, S., and Choi, J.S. (2016). Biochemical Properties of Common and Tartary Buckwheat: Centered with Buckwheat Proteomics in: Molecular Breeding and Nutritional Aspects of Buckwheat (ELSEVIER United Kingdom), pp. 239–259.

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

Li, J., Zhang, K., Meng, Y., Li, Q., Ding, M., and Zhou, M. (2019). FtMYB16 interacts with Ftimportin-alpha1 to regulate rutin biosynthesis in tartary buckwheat. Plant Biotechnol. J. **17**:1479–1481.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**:2178–2189.

Li, M.X., Yeung, J.M.Y., Cherny, S.S., and Sham, P.C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum. Genet. **131**:747–756.

Li, X., Park, N.I., Xu, H., Woo, S.H., Park, C.H., and Park, S.U. (2010). Differential expression of flavonoid biosynthesis genes and accumulation of phenolic compounds in common buckwheat (*Fagopyrum esculentum*). J. Agric. Food Chem. **58**:12176–12181.

Liu, Z., An, C., Zhao, Y., Xiao, Y., Bao, L., Gong, C., and Gao, Y. (2021). Genome-wide identification and characterization of the CsFHY3/FAR1 gene family and expression analysis under biotic and abiotic stresses in tea plants (*Camellia sinensis*). Plants **10**:570.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.

Ma, L., and Li, G. (2021). Arabidopsis FAR-RED ELONGATED HYPOCOTYL3 negatively regulates carbon starvation responses. Plant Cell Environ. **44**:1816–1829.

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics **27**:764–770.

Matsui, K., and Yasui, Y. (2020). Genetic and genomic research for the development of an efficient breeding system in heterostylous self-incompatible common buckwheat (*Fagopyrum esculentum*). Theor. Appl. Genet. **133**:1641–1653.

Matsui, K., Tetsuka, T., Nishio, T., and Hara, T. (2003). Heteromorphic incompatibility retained in self-compatible plants produced by a cross between common and wild buckwheat. New Phytol. **159**:701–708.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**:1297–1303.

Meinke, D.W. (2020). Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. New Phytol. **226**:306–325.

Motose, H., Sugiyama, M., and Fukuda, H. (2004). A proteoglycan mediates inductive interaction during plant vascular development. Nature **429**:873–878.

Nattestad, M., and Schatz, M.C. (2016). Assemblytics: a web analytics tool for the detection of variants from an assembly. Bioinformatics **32**:3021–3023.

Ohsako, T., and Li, C. (2020). Classification and systematics of the *Fagopyrum* species. Breed Sci. **70**:93–100.

Ou, S., and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. **176**:1410–1422.

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., et al. (2012). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res. **40**:D284–D289.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **38**:904–909.

Sanderson, M.J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics **19**:301–302.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210–3212.

Solanke, A.U., Sharma, M.K., Tyagi, A.K., and Sharma, A.K. (2009). Characterization and phylogenetic analysis of environmental stress-responsive SAP gene family encoding A20/AN1 zinc finger proteins in tomato. Mol. Genet. Genom. **282**:153–164.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. **32**:W309–W312.

Sugiyama, A., Shitan, N., Sato, S., Nakamura, Y., Tabata, S., and Yazaki, K. (2006). Genome-wide analysis of ATP-binding cassette (ABC) proteins in a model legume plant, *Lotus japonicus*: comparison with Arabidopsis ABC protein family. DNA Res. **13**:205–228.

Tang, W., Ji, Q., Huang, Y., Jiang, Z., Bao, M., Wang, H., and Lin, R. (2013). FAR-RED ELONGATED HYPOCOTYL3 and FAR-RED IMPAIRED RESPONSE1 transcription factors integrate light and abscisic acid signaling in Arabidopsis. Plant Physiol. **163**:857–866.

Vij, S., and Tyagi, A.K. (2006). Genome-wide analysis of the stress associated protein (SAP) gene family containing A20/AN1 zinc-finger(s) in rice and their phylogenetic relationship with Arabidopsis. Mol. Genet. Genom. **276**:565–575.

Vom Endt, D., Soares e Silva, M., Kijne, J.W., Pasquali, G., and Memelink, J. (2007). Identification of a bipartite jasmonate-responsive promoter element in the Catharanthus roseus ORCA3 transcription factor gene that interacts specifically with AT-Hook DNA-binding proteins. Plant Physiol. **144**:1680–1689.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics **33**:2202–2204.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. **38**:e164.

Wang, T., Ren, L., Li, C., Zhang, D., Zhang, X., Zhou, G., Gao, D., Chen, R., Chen, Y., Wang, Z., et al. (2021). The genome of a wild Medicago species provides insights into the tolerant mechanisms of legume forage to environmental stress. BMC Biol. **19**:96.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. **40**:e49.

Wohlbach, D.J., Quirino, B.F., and Sussman, M.R. (2008). Analysis of the Arabidopsis histidine kinase ATHK1 reveals a connection between vegetative osmotic stress sensing and seed maturation. Plant Cell **20**:1101–1117.

Xie, Y., Zhou, Q., Zhao, Y., Li, Q., Liu, Y., Ma, M., Wang, B., Shen, R., Zheng, Z., and Wang, H. (2020). FHY3 and FAR1 integrate light signals with the miR156-SPL module-mediated aging pathway to regulate Arabidopsis flowering. Mol. Plant **13**:483–498.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586–1591.

Yasui, Y., Mori, M., Aii, J., Abe, T., Matsumoto, D., Sato, S., Hayashi, Y., Ohnishi, O., and Ota, T. (2012). S-LOCUS EARLY FLOWERING 3 is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. PLoS One **7**, e31264.

Yasui, Y., Wang, Y., Ohnishi, O., and Campbell, C.G. (2004). Amplified fragment length polymorphism linkage analysis of common buckwheat (*Fagopyrum esculentum*) and its wild self-pollinated relative *Fagopyrum homotropicum*. Genome **47**:345–351.

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., and Liu, X. (2021). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Dev. Reprod. Biol. **19**:619–628.

Yu, Y., Qiao, L., Chen, J., Rong, Y., Zhao, Y., Cui, X., Xu, J., Hou, X., and Dong, C.H. (2020). Arabidopsis REM16 acts as a B3 domain transcription factor to promote flowering time via directly binding to the promoters of SOC1 and FT. Plant J. **103**:1386–1398.

Zhang, K., He, M., Fan, Y., Zhao, H., Gao, B., Yang, K., Li, F., Tang, Y., Gao, Q., Lin, T., et al. (2021). Resequencing of global Tartary buckwheat accessions reveals multiple domestication events and key loci associated with agronomic traits. Genome Biol. **22**:23.

Zhang, L., Li, X., Ma, B., Gao, Q., Du, H., Han, Y., Li, Y., Cao, Y., Qi, M., Zhu, Y., et al. (2017). The Tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. Mol. Plant **10**:1224–1237.

Zhou, M., and Memelink, J. (2016). Jasmonate-responsive transcription factors regulating plant secondary metabolism. Biotechnol. Adv. **34**:441–449.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. **44**:821–824.