



Multi-omics identification of a key glycosyl hydrolase gene *FtGH1* involved in rutin hydrolysis in Tartary buckwheat (*Fagopyrum tataricum*)

Dili Lai^{1,2,†}, Kaixuan Zhang^{1,†}, Yuqi He^{1,†}, Yu Fan^{3,†}, Wei Li^{1,†}, Yaliang Shi^{1,†}, Yuanfen Gao^{1,†}, Xu Huang^{1,†}, Jiayue He¹, Hui Zhao¹, Xiang Lu¹, Yawen Xiao¹, Jianping Cheng², Jingjun Ruan², Milen I. Georgiev^{4,5} , Alisdair R. Fernie^{5,6}  and Meiliang Zhou^{1,*} 

¹State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

²College of Agriculture, Guizhou University, Guiyang, China

³School of Food and Biological Engineering, Chengdu University, Chengdu, China

⁴Laboratory of Metabolomics, Institute of Microbiology, Bulgarian Academy of Sciences, Plovdiv, Bulgaria

⁵Center of Plant Systems Biology and Biotechnology, Plovdiv, Bulgaria

⁶Department of Molecular Physiology, Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

Received 3 August 2023;

revised 16 October 2023;

accepted 20 November 2023.

*Correspondence (Tel/fax 10-82105476;

email zhoumeiliang@caas.cn)

[†]These authors contributed equally to this work.

Summary

Rutin, a flavonoid rich in buckwheat, is important for human health and plant resistance to external stresses. The hydrolysis of rutin to quercetin underlies the bitter taste of Tartary buckwheat. In order to identify rutin hydrolysis genes, a 200 genotypes mini-core Tartary buckwheat germplasm resource was re-sequenced with 30-fold coverage depth. By combining the content of the intermediate metabolites of rutin metabolism with genome resequencing data, metabolite genome-wide association analyses (GWAS) eventually identified a glycosyl hydrolase gene *FtGH1*, which could hydrolyse rutin to quercetin. This function was validated both in Tartary buckwheat overexpression hairy roots and in vitro enzyme activity assays. Mutation of the two key active sites, which were determined by molecular docking and experimentally verified via overexpression in hairy roots and transient expression in tobacco leaves, exhibited abnormal subcellular localization, suggesting functional changes. Sequence analysis revealed that mutation of the *FtGH1* promoter in accessions of two haplotypes might be necessary for enzymatic activity. Co-expression analysis and GWAS revealed that *FtbHLH165* not only repressed *FtGH1* expression, but also increased seed length. This work reveals a potential mechanism behind rutin metabolism, which should provide both theoretical support in the study of flavonoid metabolism and in the molecular breeding of Tartary buckwheat.

Keywords: Tartary buckwheat, Rutin, Hydrolysis, Re-sequencing, GWAS, Metabolome.

Introduction

In recent years, buckwheat is becoming a highly popular food due to its balanced amino acid composition and richness in resistant starch, vitamins, trace elements, and antioxidants (Huda *et al.*, 2021; Joshi *et al.*, 2020; Kreft *et al.*, 2020). Buckwheat is a unique pseudocereal crop, in that it contains large amounts of flavonoids including rutin, quercetin, isoquercitrin, and quercitrin (Li *et al.*, 2020; Zhang *et al.*, 2012), but does not yet represent an important calorific crop. Rutin, a flavonoid rich in buckwheat seeds, is widely used in pharmaceutical drugs because of its antibacterial and antioxidant properties, as well as its bioactivity in the reduction of blood pressure and with respect to its hypoglycaemic function (Calzada *et al.*, 2017; Kreft, 2016; Lee and Lee, 2010; Wang *et al.*, 2013; Yang *et al.*, 2008). The hydrolysis of rutin to quercetin is responsible for the strong bitterness exhibited by Tartary buckwheat (Suzuki *et al.*, 2021). Hence, studies focusing on the conversion of rutin to quercetin are of great importance. In this respect, the inhibition of rutin hydrolase could minimize quercetin formation, and thereby retain rutin contents, which would ultimately help in reducing the bitterness of buckwheat products.

As an important secondary metabolite, rutin is not only of great benefit to human health, but also has an important role for the buckwheat plant itself. For instance, rutin can protect the plant from solar UV radiation, dryness, cold, salinity, and a range of pests (Gaberscik *et al.*, 2002; Ismail *et al.*, 2015; Kreft *et al.*, 2002; Sharma *et al.*, 2021; Suzuki *et al.*, 2005; Yang *et al.*, 2022; Zhao *et al.*, 2021a). However, rutin in buckwheat is unstable and its accumulation is highly susceptible to environmental influences such as altitude, light, moisture, and other environmental factors (Hao *et al.*, 2016; Suzuki *et al.*, 2005; Tsurunaga *et al.*, 2013). Moreover, it is readily degraded by rutin hydrolase into quercetin (Jia *et al.*, 2019). Previous studies have made some progress in understanding rutin metabolism but it remains insufficient. Most of the current studies on rutin metabolism have focused on its upstream biosynthesis genes, including rhamnosyltransferases (RT), glycosyltransferases (GT), and flavonol synthases (FLS) (Zhang *et al.*, 2017). Currently, studies on rutin hydrolase activity are mainly focused on its enzyme activity for food processing applications (Lukšič *et al.*, 2016; Wu *et al.*, 2020b; Yoo *et al.*, 2012). However, there is a paucity of studies on the downstream genes involved in rutin hydrolysis, probably because the instability of rutin increases the difficulty in identification of the gene associated with its hydrolysis.

In order to identify the genes responsible for rutin hydrolysis, we have previously performed genome-wide association analyses (GWAS) by combining the 10-fold deep re-sequencing data and the contents of rutin and quercetin in a 480 germplasm resources of Tartary buckwheat (Zhang *et al.*, 2021). However, none of the genes in the associated regions encodes a rutin hydrolase. Subsequently, the metabolic genome-wide association study combining the 10-fold deep resequencing data and the contents of rutin metabolism intermediates in 200 microcore germplasm resources were further performed to identify rutin hydrolysis through the intermediate metabolites of rutin hydrolysis (Zhao *et al.*, 2023). However, still no rutin hydrolase was identified. We speculated that this might be due to the influence of the environment on rutin accumulation and the insufficient sequencing depth of the earlier studies. In the current study, the rutin content of the Tartary buckwheat planted in areas with different altitudes and the >30-fold deep re-sequencing data were used to further mine SNP variant loci in Tartary buckwheat germplasm resource. On combination with the content of different metabolites, GWAS was performed which identified a glycosyl hydrolase gene (*FtGH1*) responsible for the different rutin content in the Tartary buckwheat germplasm resources. The rutin hydrolytic activity of *FtGH1* was experimentally verified. Furthermore, the natural variation in an MYC transcription factor binding motif in the *FtGH1* promoter was identified. The regulatory effect of the upstream MYC transcription factor on the *FtGH1* promoter was further deciphered. This study thus not only provides clues for the identification of rutin hydrolase genes in buckwheat genetic resources, but also contributes to the exploration of key enzymatic genes involved in metabolic pathways of readily degradable metabolites.

Results

GWAS of rutin content at different altitudes with 10-fold depth resequencing data identified a candidate rutin hydrolase gene

As the accumulation of rutin in Tartary buckwheat seeds is easily influenced by environmental factors, we planted 480 Tartary buckwheat germplasm resources at five experimental sites situated at different altitudes. Only 226 were survived and produced seeds at all five experimental sites. The contents of rutin and quercetin in these 226 Tartary buckwheat accessions at five sites were subsequently investigated. The contents of rutin and quercetin were varied across the different altitudes (Figure 1a; Figure S1A; Data Set S1), further confirming that rutin accumulation is easily affected by environment, which may increase the difficulty in identification of genes related to rutin metabolism.

Subsequently, by combining the 10-fold deep whole-genome resequencing data (Zhang *et al.*, 2021) with the varied content of the rutin and quercetin of these 226 Tartary buckwheat accessions planted in different altitudes, GWAS of rutin-related metabolites was conducted to identify key genes involved in rutin hydrolysis (Figure 1). The results revealed that there was no significant locus in GWAS results of rutin content of seeds harvest at areas at altitudes of 1500 m and 3000 m (Figure 1c,f). However, there are significant loci associated with rutin content at 1000 m, 2000 m, and 2500 m altitudes, respectively (Figure 1b,d,e). Genes in these significant loci were further analysed but no genes were found annotated to be related to

rutin hydrolysis in the GWAS results of rutin content at 1000 m and 2000 m (Data Set S2 and S3). Interestingly, a significant locus on chromosome 2 [$-\log_{10}(P) > 5$] was identified using rutin content of Tartary buckwheat seeds obtained at 2500 m altitude (Figure 1e). Moreover, a gene encoding a beta-glucosidase type gene/glycosyl hydrolyse (*FtPinG0201454200*, *FtGH1*) was identified within this significant signal interval (Data Set S4). Similar to rutin, GWAS results of quercetin at five experimental sites with different altitudes produced significant results at all five test sites (Figure S1A–F). However, not all significant signalling intervals within have genes that are likely to be involved in rutin hydrolysis metabolism (Data Set S5–S8). Only one beta-glucosidase gene in the GWAS results of 2500 m was identified, which was the same gene identified by GWAS of rutin at 2500 m (Data Set S9). We therefore speculated that this candidate gene, encoding a beta-glucosidase, may be involved in the hydrolytic metabolism of rutin. These results partly confirm the difficulty of rutin hydrolase enzyme gene identification, but also endorse the implication that GWAS needs to be performed using phenotypes observed from different environmental conditions to explore the regulatory genes of metabolites that were greatly affected by the environment (Zhu *et al.*, 2022).

GWAS of the content of rutin-related flavonoids with >30-fold depth resequencing data further confirmed rutin hydrolase gene is necessary in rutin metabolism

The above results demonstrated that using 10-fold deep resequencing data to associate rutin and quercetin content at different altitudes only identified a single possible rutin hydrolase gene at 2500 m. This prompted us to use other methods to more accurately identify genes responsible for rutin hydrolase. Deep resequencing can improve the completeness of SNP mining in the whole genome, thus increasing the accuracy of key gene identification (Li *et al.*, 2021; Shao *et al.*, 2022; Zhao *et al.*, 2018). In order to find more accurate SNP signals across the whole genome, thus providing a stronger support for the identification of the rutin hydrolase genes, we re-sequenced a 200 microcore germplasm resource at 30-fold depth (Data Set S10). These 200 microcore germplasm resources have a wide range of genetic diversity and represent the largest population of Tartary buckwheat similar to 480 germplasm resources (Figure S2) (Zhang *et al.*, 2021; Zhao *et al.*, 2023). By combining the 10-fold depth resequencing data from 480 core germplasm resources (Zhang *et al.*, 2021) and the 30-fold depth resequencing data illustrated in this study, a higher depth (average depth > 40) resequencing data was used to analyse the SNPs associated with rutin hydrolysis (Data Set S10). A total of 1 363 970 SNPs were used for this associated study, exhibiting a significant increase in the number of SNPs compared to 10-fold (1 103 732 SNPs; Figure 2; Data Set S11). These increased SNP numbers were unevenly distributed across the eight Tartary buckwheat chromosomes. The most pronounced increase in SNP number appears to be around the approximately 6.9 Mb position on chromosome 7. This range does not appear to be found for any SNPs in the 10-fold depth resequencing (Figure 2a), whereas there are SNPs found in this range that were exploited in the 30-fold depth resequencing (Figure 2b). There were also increases in the number of SNPs on other chromosomes, and the exact statistics of their increases are presented in Data Set S11. In terms of different chromosomes, the highest total increase in SNPs was found on chromosome 1 (increasing 45 723 SNPs) and the lowest

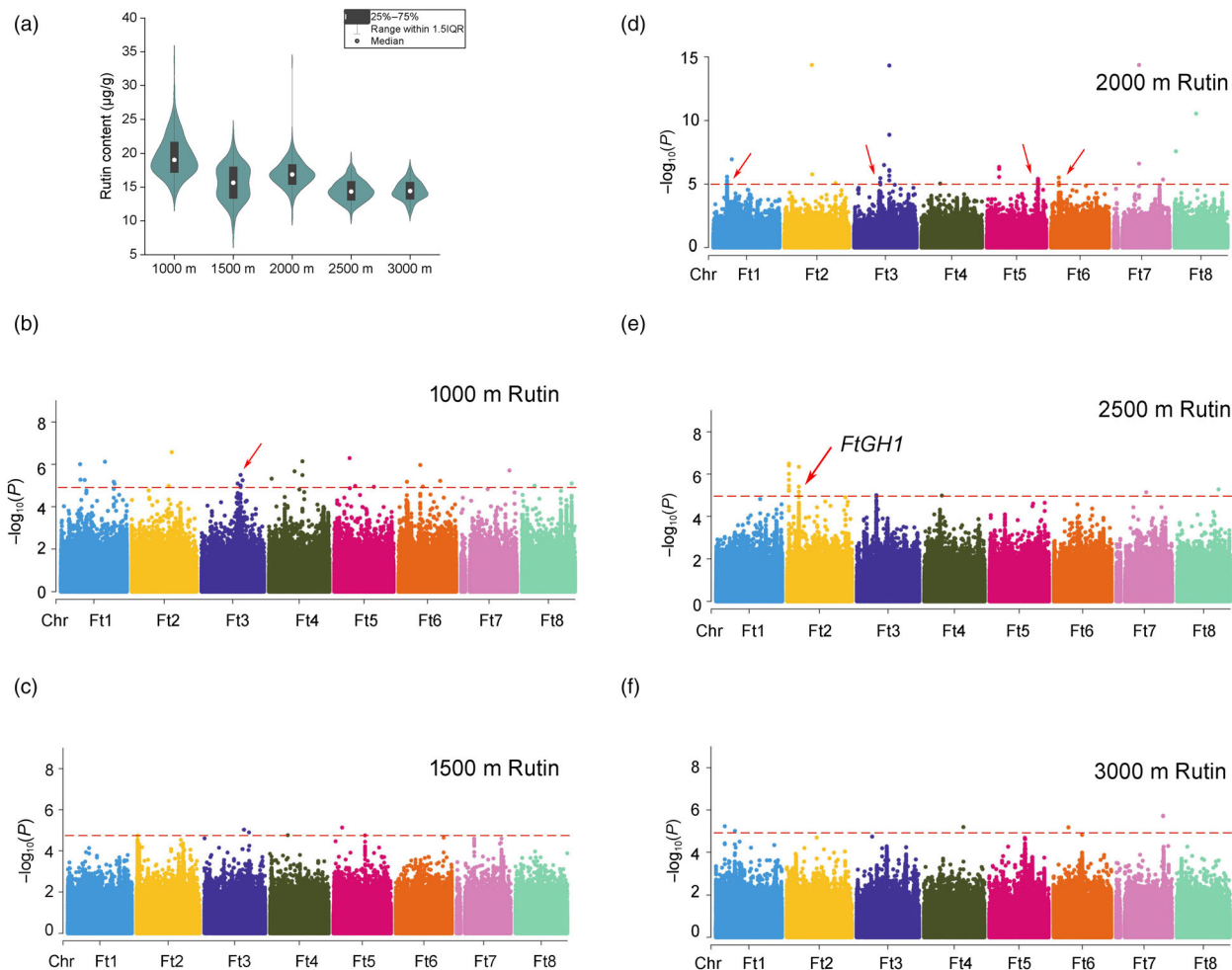


Figure 1 Genome-wide association analysis of rutin content at different altitudes. (a) Distribution of rutin content at different altitudes in 226 accessions of Tartary buckwheat. (b–f) Genome-wide association analysis results of rutin content with 10× depth whole-genome resequencing data in different environments. (b–f) The GWAS results of rutin content at altitudes of 1000 (b), 1500 (c), 2000 (d), 2500 (e), and 3000 (f) meters, respectively. The red arrow points to areas where GWAS have significant signals. The distribution of genes within the significant signal interval is shown in Data Set S2–S4.

on chromosome 5 (increasing 19 647 SNPs). Interestingly, the increase in the number of SNPs was also exhibited in the exons: the most on chromosome 1 and the least on chromosome 5.

As the GWAS using the 10-fold resequencing data identified a significant locus associated with the rutin content at 2500 m altitude, we analysed the changes in the number of SNPs in this region within roughly 120 kb compared with that after 30-fold resequencing data. It was found that the total number of SNPs in this region increased by 258, which is 27.71% higher than that of the previous study (Data Set S12). The increase was mainly concentrated in upstream, downstream, intergenic, intronic, and exonic regions. The number of SNPs in UTR5 and UTR3 remained unchanged. However, the most interesting thing is that the number of SNPs in this exonic region is decreased after 30-fold genome resequencing. The above results showed that 30-fold genome resequencing significantly improved the accuracy of SNP mining, which can reduce the number of false negatives for subsequent GWAS results and thus enhance the accuracy of key genes identification for rutin metabolism.

To further identify genes responsible for rutin metabolism, GWAS was further performed by combining the 30-fold deep

whole-genome resequencing data of 200 minicore germplasm resources obtained in this study and the content of rutin, quercetin, and their related flavonoids obtained from metabolomic profiling of a 200 minicore germplasm resource (Zhao *et al.*, 2023) (Figure S3, Data Set S13). There was, however, no significant signal found in rutin (Figure S4A). That said, there was a significant signal on chromosome 6 of quercetin GWAS (Figure S4B); in-depth analysis of this signal interval did not find any potential genes that might be involved in the hydrolytic metabolism of rutin (Data Set S14). For this reason, we broadened the substances analysed by association study, including isoquercitrin, quercitrin, hyperin, isohyperoside, and quercetin-7-O-glucoside, which are metabolites related to rutin and quercetin (Figure S3). Interestingly, all five of these metabolites exhibited associations in the same region of chromosome II (Figure 3a–e; Data Set S15–S19). A significant SNP is located on *FtGH1* (Figure 3f,g). In-depth analysis of the genes in this region revealed that FtPinG0201454200 was still present in it. This gene belongs to glycosyl hydrolase family 1, and evolutionary tree clustering analysis also revealed a high degree of similarity to several known enzymes (rutinosidase, rutin hydrolase,

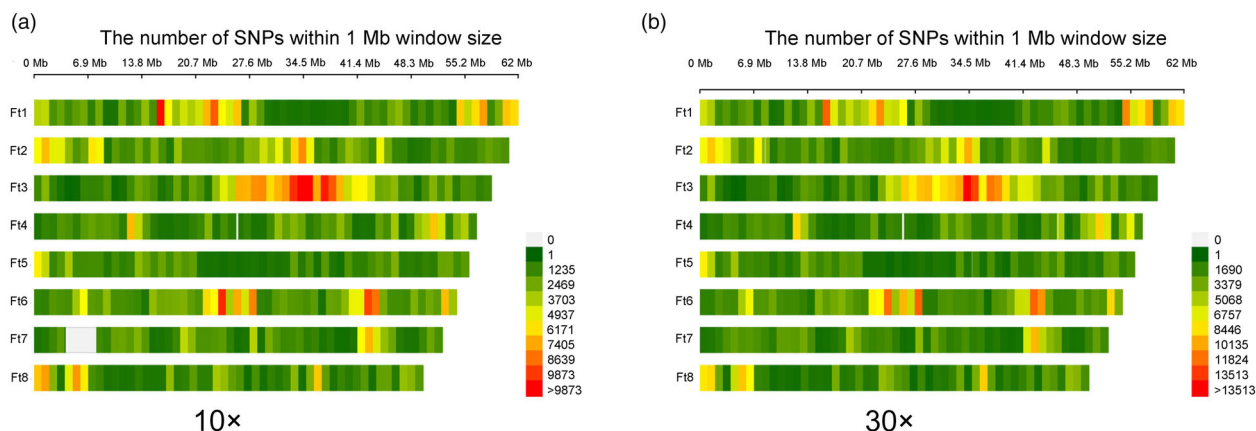


Figure 2 High depth whole-genome resequencing of Tartary buckwheat and the distribution of SNPs in chromosomes. (a) and (b) indicate the distribution profiles of different chromosomal SNPs after whole-genome resequencing of Tartary buckwheat with 10× and 30× depth, respectively.

and β -glucosidase) (Figure S5). *FtGH1* consists mainly of a glycosyl hydrolase domain, which is 477 amino acids long, in addition to a transmembrane region at the N terminus (Figure 3g). The Pearson correlation analysis showed that all four substances were significantly correlated with rutin except for quercitrin, further confirming these metabolites were closely associated with rutin metabolism (Figure 3h; Data Set S20). This is further evidence that the above five metabolites identified to the genes are more accurate. These results furthermore strongly suggest that this gene may be involved in the rutin catabolic pathway.

As the metabolic pathway of rutin-related flavonoids is extremely complex, and there are mutual transformation relationships between different substances. In order to exclude the influence of genes responsible for other reactions, GWAS with the ratio of the content of five key metabolites associated with rutin or quercetin were conducted. GWAS of the ratios of five key metabolites to rutin showed that the ratios of three of them (isoquercitrin, quercitrin, hyperin) to rutin were able to re-identify *FtGH1* (Figure S6A–E; Data Set S21). The above results imply that *FtGH1* is very important in the metabolic pathway of rutin. In addition, the same method was used for the GWAS of the ratios of the five key metabolites to quercetin content, and surprisingly, all their ratios could be identified to *FtGH1* (Figure S6F–J; Data Set S21). This result suggests that *FtGH1* is essential for the interconversion between the five key metabolites and quercetin.

Functional analysis of Tartary buckwheat rutin hydrolase gene *FtGH1*

To gain further insight into the function(s) of *FtGH1* in plants, we overexpressed *FtGH1* in Tartary buckwheat hairy roots and analysed the flavonoid metabolite content in different overexpression lines (Figure S7). The content of rutin was significantly reduced in overexpression lines compared with the EV (empty vector) control, and the content of quercetin was significantly increased (Figure 4a,b), confirming the hydrolytic effect of *FtGH1* on rutin. Furthermore, *FtGH1* was identified in the regions associated with the contents of isoquercitrin, quercitrin, hyperin, isohyperoside, and quercetin-7-O-glucoside (Figure 3a–e). And these substances are all derived from quercetin and have a glycosyl group added to quercetin. This suggests that *FtGH1*, as

a glucosidase, may hydrolyse these metabolites to produce quercetin. Therefore, detecting the content of these metabolites in the hairy roots is equally important. In addition, other flavonoids also showed significant changes in the hairy root overexpression lines. For example, the content of isoquercitrin was significantly reduced (Figure S8A). These substances are closely related to key compounds such as rutin and quercetin, and the substances associated with *FtGH1* according to GWAS. These results, therefore, indicate that our GWAS results are reliable and the associated genes can significantly affect the associated substances. In addition, *FtGH1* was transiently overexpressed into tobacco (*Nicotiana benthamiana*) leaves, and its protein was able to be expressed to detectable levels in tobacco leaves (Figure 4c,d; Figure S8). Then, the content of flavonoid metabolism in tobacco leaves was examined using LC–MS, with the results revealing that the content of rutin in tobacco leaves injected with *FtGH1* transient overexpression bacterial solution was significantly decreased compared to both the empty vector and non-injected control groups. By contrast, the quercetin content was significantly increased. This result is similar to that of Tartary buckwheat overexpressing hairy roots, further supporting the hydrolytic activity of *FtGH1* on rutin. Further analysis of the changes in other substances in tobacco showed that the levels of hyperin and quercetin-7-O-glucoside were also significantly reduced in transiently overexpressed leaves (Figure S8B,C). For which substrate is catalysed by *FtGH1*, we performed an in vitro enzyme activity assay of *FtGH1*. The results showed that *FtGH1* was able to catalyse the generation of quercetin from rutin in vitro (Figure 4e, Figure S9). However, we have not found the function of *FtGH1* on other compounds for the moment, and perhaps more future efforts need to be devoted to the functional expansion of *FtGH1*. The above results confirm the key role of *FtGH1* in rutin metabolism.

Ser382 and Ser472 was responsible for hydrolysis activity of *FtGH1*

To better understand the function of *FtGH1*, we used a homologous protein of *FtGH1* as a model for molecularly docking studies using different compounds. This was used to simulate their binding and in attempt to find the key active sites. Rutin is able to enter the active pocket of *FtGH1* protein (Figure 4f).

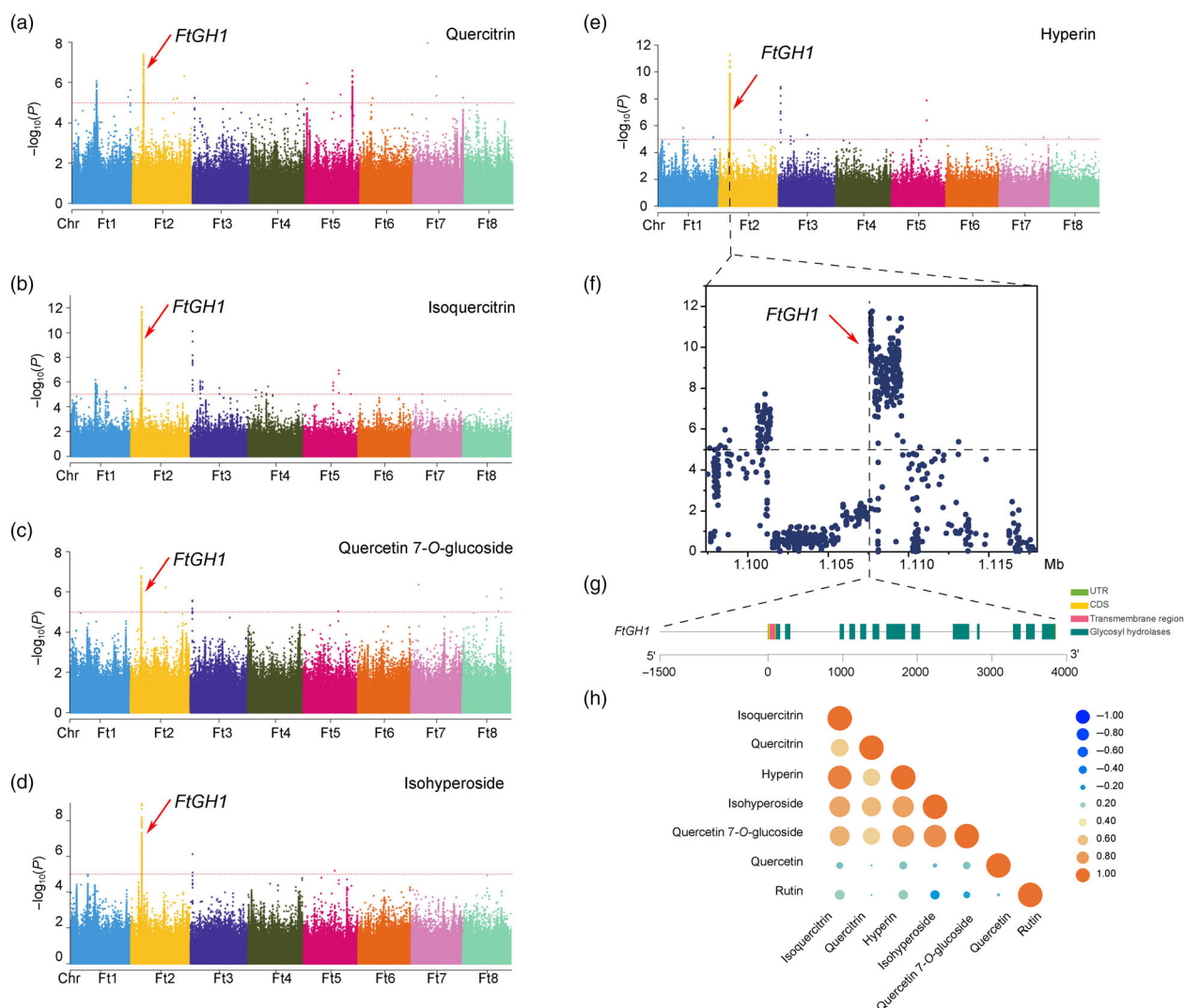


Figure 3 GWAS of 30-fold buckwheat whole-genome resequencing data with rutin metabolism-related substances. (a–e) Manhattan plots of genome-wide association analysis of rutin metabolism-related substances using 30× depth whole-genome resequencing data. The GWAS were performed using the content of quercitrin (a), isoquercitrin (b), quercetin-7-O-glucoside (c), isohyperoside (d), and hyperin (e) in 200 mini-core germplasm resources, respectively. The red arrow points to the significant GWAS signals where *FtGH1* located. (f) Distribution of loci information on the significant signal of GWAS in chromosome 2 for hypericin. (f) Gene structure information of *FtGH1*. (h) Pearson correlation analysis of rutin, quercetin, quercitrin, isoquercitrin, quercetin-7-O-glucoside, isohyperoside, hyperin, quercetin, and rutin. The small dots in the upper right corner represent the correlation coefficients.

Moreover, amino acid residues of the *FtGH1* protein are able to form hydrogen bonds, π -bonds, and hydrophobic interactions with rutin. The implicated residues include: Glu-206, Phe-207, Thr-342, Phe-344, Leu-367, and Ser-382. Among which, the residues that bind to rutin are Glu-206, Phe-207, and Ser-382. In addition, we predicted the docking of other substrates (hypericin, quercetin, isoquercitrin, quercetin-7-O-glucoside) with the *FtGH1* protein. The results showed that certain residues (Phe-207, Ser-382, Trp-384, and Ser-472) appeared more frequently in these compounds (Figure S10A–D, Data Set S22), and thus were selected for site-directed mutagenesis to investigate the critical active sites of the *FtGH1* protein. We obtained overexpression materials of these mutant sites in Tatar buckwheat hairy roots (Figure S11A) and detected the corresponding flavonoid contents. The results showed that the rutin content in the hairy roots of the S382A and S472A mutation sites was significantly

up-regulated compared with that of the original *FtGH1*, and its up-regulation level was similar to that of the empty vector (Figure 4g). In contrast, the rutin content in the hairy roots of the F207A and S384A mutation sites was not significantly different compared with that of the original *FtGH1*. Their mutations did not affect the function of *FtGH1*. It is also interesting that the quercetin content in the hairy roots of these mutant sites showed a correspondence with the rutin content (Figure S11B). The wild-type and mutated *FtGH1* were then independently transiently transformed into tobacco leaves. Western blots were then performed to detect the expression of all forms of *FtGH1* protein, with the results demonstrating that expression was detected for all proteins (Figure S11C). Furthermore, the rutin content of these leaves was measured, and compared to the wild-type *FtGH1*, with the rutin content being significantly increased in S382A and S472A mutants

(Figure 4h). These results suggest that the enzymatic activity was disrupted after mutation at these two positions, identifying them as critical sites for FtGH1 activity.

As the subcellular localization alteration of protein often occurs alongside protein function changes (Barberis *et al.*, 2021; Lundberg and Börner, 2019), exploring the subcellular localization of FtGH1 is essential for studying its function. We next constructed

fusion expression vectors of GFP with FtGH1 or its mutant form, and co-transformed it with the cell membrane maker fused with mCherry protein into tobacco leaves. The results showed that the signal of FtGH1 overlaps with the cell membrane maker signal (Figure 4i). Thus, FtGH1 is cell membrane localized and it functions at the cell membrane. For the four mutations in FtGH1, where the subcellular localization of F207A and S384A was not

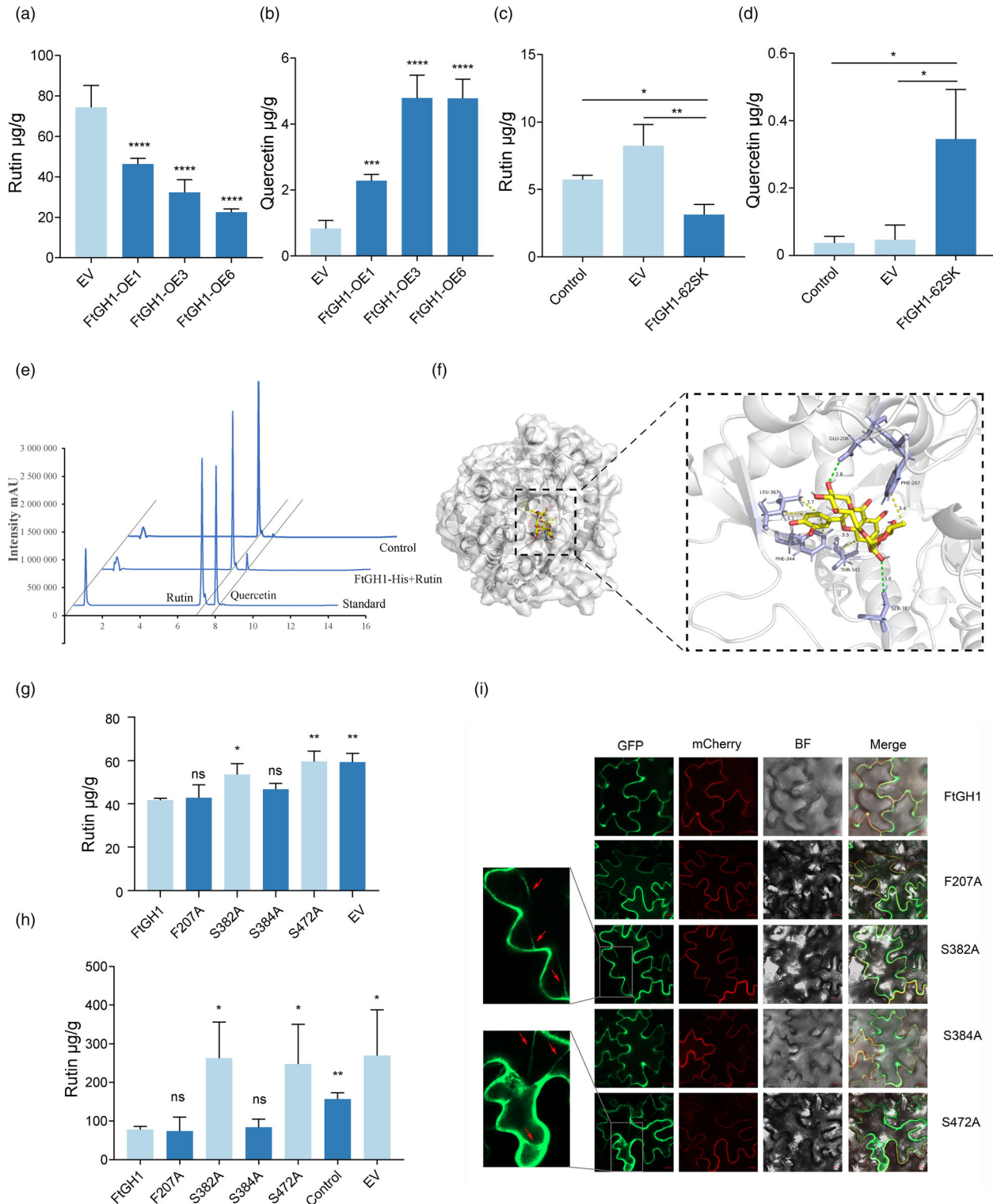


Figure 4 Functional analysis of *FtGH1* involvement in rutin metabolic pathway. (a,b) The content of rutin (a) and quercetin (b) in *FtGH1* overexpressed buckwheat hairy roots. Data are shown as mean \pm SD from three biological replicates ($n = 5$). Analyse the significance of differences between EV and others. Asterisk indicates significant difference between EV and others. $P < 0.001$ and $P < 0.0001$ are designated by *** and ****, respectively (one-way ANOVA, Tukey's post-test). (b,c) The rutin (c) and quercetin (d) content in *FtGH1* transient overexpressed tobacco leaves. Data are shown as mean \pm SD from three biological replicates ($n = 3$). $*P < 0.05$ and $**P < 0.01$ (one-way ANOVA, Tukey's post-test). (e) *FtGH1* in vitro enzyme activity assay. HPLC showed different reaction curves. The pD2P empty vector extract was used as a negative control. (f) Molecular docking of rutin with *FtGH1* homologous protein. The green dashed line represents hydrogen bonding, the yellow dashed line represents hydrophobic interaction, and the blue dashed line represents π -bonding. (g) Rutin content in Tartary buckwheat hairy roots after stable expression of *FtGH1* and its mutants. Data are expressed as mean \pm SD of three biological replicates ($n = 3$). Significant differences between *FtGH1* and other genes were analysed. $*P < 0.05$ and $**P < 0.01$ (one-way ANOVA, Tukey's post-test). (h) Rutin content of in tobacco leaves after transient expression of *FtGH1* and its mutant version in tobacco leaves. Data are shown as mean \pm SD from three biological replicates ($n = 3$). *FtGH1* was analysed for significance of difference with others. $*P < 0.05$ and $**P < 0.01$, as calculated using two-tailed Student's *t*-test. (i) Subcellular localization of *FtGH1* and its mutant protein after transient expression in tobacco leaves.

altered. Interestingly, there is a slight change in the localization of S382A and S472A, whose GFP signals seem to be located in the cellular cytoskeleton as well. This seems to imply the existence of abnormal localization in the cell, and this particular situation seems to interfere with their normal function.

Natural variation in *FtGH1* promoter was responsible for *FtGH1* expression regulation

By analysing the sequence of *FtGH1*, a T/C mutation was found at the -172 bp of the *FtGH1* promoter (Figure 5a). Haplotype analysis revealed the content of isoquercitrin, hyperin, isohyperoside, and quercetin-7-*O*-glucoside in accessions of Hap. T was significantly lower than in accessions of Hap. C (Figure S12A,C–E), while the content of quercitrin exhibited no significant difference (Figure S12B). Rutin hydrolase activity analysis illustrated that accessions in Hap. T has higher hydrolase activity compared to accessions in Hap. C (Figure 5b; Figure S13). Moreover, the expression level of *FtGH1* in seeds germinated for 4 days is much higher in Hap. T accessions than that of Hap. C accessions (Figure 5c), suggesting this T/C mutation may be the key factor affecting the activity of rutin hydrolase. To investigate whether the T/C mutation could change the activity of the promoters, they were cloned into pGreenII-0800-LUC to detect the activity of LUC. The luciferase reporter system showed that the luciferase activity of Hap. T was indeed significantly higher than that of Hap. C (Figure 5d). The above results suggest that these two haplotypes affect the expression of *FtGH1* by influencing the difference in *FtGH1* promoter activity.

The different binding activities of *FtbHLH165* on *FtGH1* promoters of the two haplotypes results in varied *FtGH1* expression

As the T/C mutation on *FtGH1* promoter altered *FtGH1* expression, we next analysed the *cis*-acting elements in the 1.5-kb promoter region of *FtGH1* for two haplotypes (Data Set S23). Interestingly, the T/C mutation was located on the MYC *cis*-acting elements (Figure S14). For this reason, the upstream MYC transcription factor was further screened. Previous studies, using transcriptome and proteomics, have demonstrated that flavonoids are gradually accumulated during seed development (Deng et al., 2022; Gao et al., 2017). So transcriptome data of Tartary buckwheat seed development stages were used for co-expression clustering analysis (Liu et al., 2018), and the expression pattern of *FtGH1* showed an increase with the extension of seed development time (Figure S15). All MYC transcription factors identified in Tartary buckwheat transcriptomes were used for expression pattern identification, and seven MYC transcription factors

showed decreasing expression patterns with the extension of seed development time (Figure S15). This situation implies that the upstream MYC transcription factor may negatively regulate *FtGH1* expression. As JA signalling plays positive roles in flavonoid metabolism (Ding et al., 2022; Zhou et al., 2017), and MYCs and JAMs are important members of the JA signalling pathway (Ruan et al., 2019). To screen for the most likely MYC transcription factors regulating *FtGH1*, a phylogenetic tree was constructed using the known positive regulatory transcription factors of JA signalling in *Arabidopsis*, MYC2, MYC3, and MYC4 (Cheng et al., 2011; Fernández-Calvo et al., 2011; Niu et al., 2011), and negative regulatory transcription factors of JA signalling, JAM1, JAM2, and JAM3 (Nakata et al., 2013; Sasaki-Sekimoto et al., 2013). The results revealed that *FtbHLH165* (*FtPinG0201231100*) is most closely related to JAM1, JAM2, and JAM3 based on the phylogenetic tree (Figure 5e). Therefore, *FtbHLH165* is the most likely candidate to regulate the expression of *FtGH1*, and as such was used for subsequent functional validation.

First, yeast one-hybrid was used to validate the binding of *FtbHLH165* to the promoter of *FtGH1*. The results revealed that *FtbHLH165* binds to both haplotype promoters, and they were able to grow normally on triple-deficient medium containing 3-AT (Figure 5f; Figure S16). At 6 mM/L 3-AT, the growth status of Hap. C yeast was better than that of Hap. T, which may indicate that *FtbHLH165* has a stronger ability to bind to Hap. C than Hap. T. Secondly, a luciferase reporter assay was used to validate the inhibitory effect of *FtbHLH165* on the *FtGH1* promoter. *FtbHLH165* was able to produce a repressive effect on both haplotype promoters (Figure 5g). Furthermore, considering the LUC/REN ratio data, *FtbHLH165* showed significantly stronger inhibition on Hap. C compared to Hap. T. This result corresponds with the difference in rutin hydrolase activity and the gene expression level between the two haplotypes. Therefore, the inhibition of *FtbHLH165* on the two haplotype promoters is also one of the key factors leading to the differential rutin hydrolase activity in different accessions.

To gain a deeper understanding of the regulation of *FtGH1* by *FtbHLH165* in genetics, we overexpressed *FtbHLH165* into Tartary buckwheat hairy roots (Figure S17). This overexpression material was also used for transcriptome sequencing to verify whether the overexpression of *FtbHLH165* suppressed the expression of *FtGH1* and the effect on some other genes involved in flavonoid metabolism (Data Set S24). In *FtbHLH165* overexpressing lines, *FtbHLH165* itself was significantly up-regulated, but *FtGH1* was significantly repressed (Figure 5h, Figure S18). A Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis revealed that many differentially expressed genes

are involved in secondary metabolic biosynthesis and the phenylpropanoid biosynthesis pathway (Figure S19). This provides further evidence that *FtbHLH165* is involved in the flavonoid

synthesis pathway in Tartary buckwheat. In addition, some other flavonoid metabolism genes were significantly up- or down-regulated, such as flavonoid synthase (*FtFLS*), a gene known to

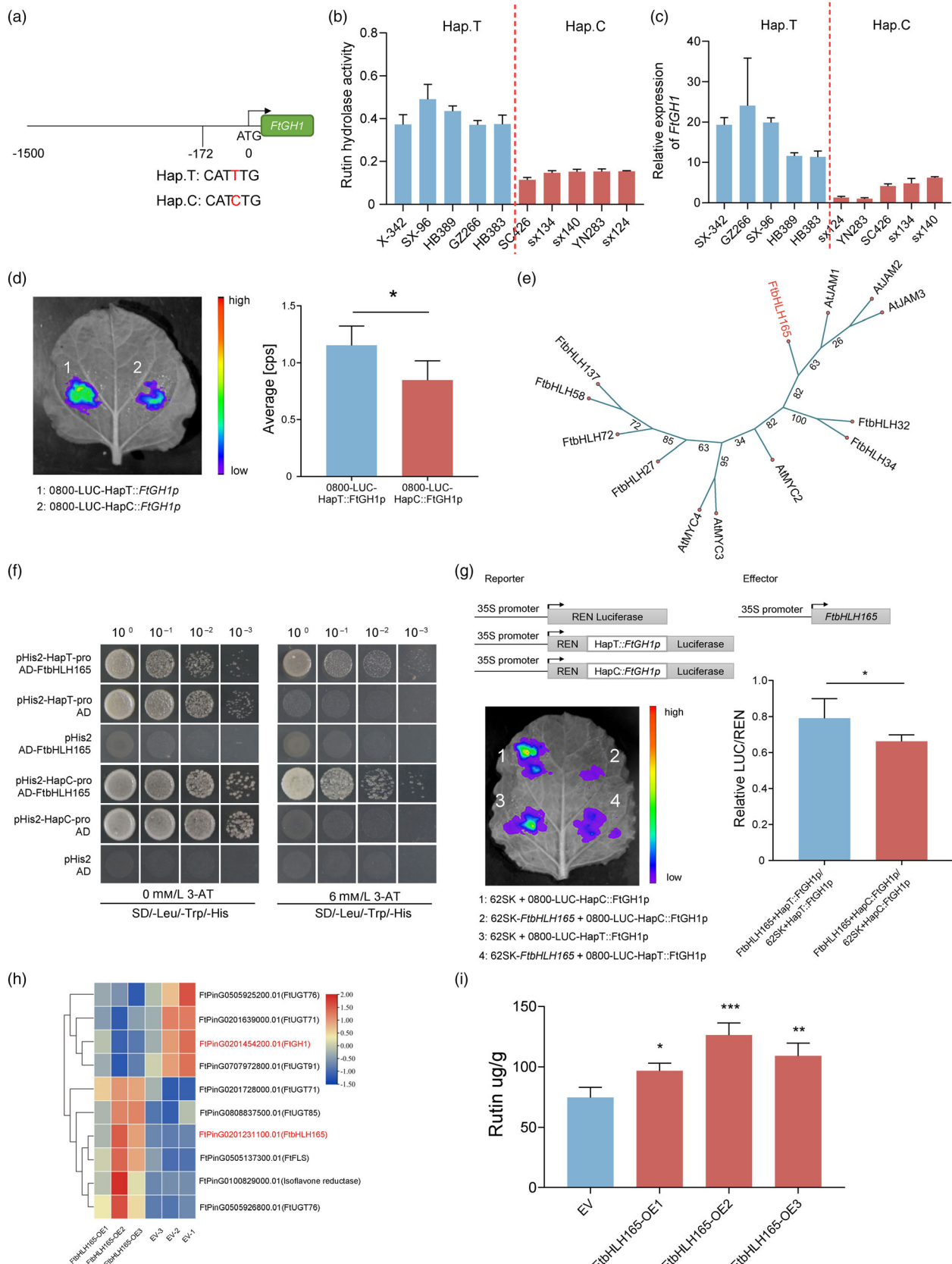


Figure 5 Genetic diversity analysis of *FtGH1* and regulation of *FtGH1* by *FtbHLH165*. (a) The T/C mutation of on *FtGH1* promoter in two haplotypes. (b,c) Rutin hydrolase activity (b) and relative expression (c) level of the *FtGH1* gene in two haplotypes of *FtGH1* accessions. (d) Identification of the activity of the Hap. T and Hap. C promoters of two haplotypic materials of *FtGH1*. Data are shown as mean \pm SD from three biological replicates ($n = 6$). * $P < 0.05$, as calculated using two-tailed Student's *t*-test. (e) Evolutionary tree construction of seven buckwheat MYC transcription factors with three positively regulated *bHLH* (MYC2/3/4) and three negatively regulated *bHLH* (JAM1/2/3) in *Arabidopsis thaliana*. Multiple sequence proteins are aligned using the MUSCLE algorithm. IQ-TREE was used to construct a phylogenetic tree. The optimal model for constructing the maximum likelihood (ML) unrooted evolutionary tree was JTT + F + G4. (f) Yeast one-hybrid assay of transcription factor *FtbHLH165* with the promoter of *FtGH1*. (g) Luciferase assay to validate the binding of transcription factor *FtbHLH165* to the *FtGH1* promoter and repress LUC expression. Data are shown as mean \pm SD from three biological replicates ($n = 5$). * $P < 0.05$, as calculated using two-tailed Student's *t*-test. (h) Analysis of differentially expressed genes for flavonoid synthesis in the transcriptome of *FtbHLH165* overexpressing buckwheat hairy roots (log2 fold changes >2 or <0.5 , q value <0.05). (i) Rutin content in *FtbHLH165* overexpressed buckwheat hairy roots. Data are shown as mean \pm SD from three biological replicates ($n = 3$). Analyse the significance of differences between EV and others. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$ (one-way ANOVA, Tukey's post-test).

be involved in the conversion of dihydroquercetin to quercetin (Li et al., 2013). A large number of glycosyltransferase genes were found to be significantly affected, with three glycosyltransferase genes significantly down- or up-regulated, respectively. In addition, the content of rutin in *FtbHLH165* overexpressing hairy roots was significantly elevated (Figure 5i), further confirming *FtbHLH165* is an important regulator in rutin metabolism.

We also investigated the co-linear genes of *FtGH1* and *FtbHLH165* in representative plants of the genus *Fagopyrum*, as well as in other monocotyledonous (rice, maize, and sorghum) and dicotyledonous plants (soybean, *Arabidopsis thaliana*, and tobacco). In order to explore whether they are conserved in this regulatory relationship. Interestingly, *FtGH1* has a single copy of the co-linear gene only with golden buckwheat (*Fagopyrum dibotrys*) and common buckwheat (*Fagopyrum esculentum*). There are no co-linear genes for *FtGH1* in any of the above monocotyledonous and dicotyledonous plants. *FtbHLH165* has two co-linear genes in both golden buckwheat and common buckwheat. *FtbHLH165* has no co-linear genes in any of the above monocotyledonous and dicotyledonous plants except for three co-linear genes in soybeans (Figure S20, Data Set S25). We also analysed the promoter sequences of genes that are co-linear with *FtGH1* in golden buckwheat and common buckwheat. The results showed that in common buckwheat, the promoter sequence of the *FtGH1* co-linear gene contained very few MYC cis-acting elements (only two). Moreover, at the position of the key MYC element in the *FtGH1* promoter, the MYC element is absent in common buckwheat at its same position (Figure S21, Data Set S26).

***FtbHLH165* is a multi-functional gene that regulates the balance between the yield and quality in Tartary buckwheat**

In order to further understand the gene function of *FtbHLH165*, we searched all GWAS results accumulated by our group in the previous period. We found *FtbHLH165* in the GWAS results for seed perimeter, an agronomic trait (Data Set S27). The results of this experiment revealed that the significant locus interval on chromosome 2 contained *FtbHLH165* (Figure S22). Within the 200 kb range of the significant signal, *FtbHLH165* is tightly linked to the lead SNP (Figure 6a; Data Set S28). This suggests that *FtbHLH165* may be involved in regulating seed perimeter. Therefore, seeds of *FtbHLH165* overexpressing *Arabidopsis* were used to observe the phenotype (Figure 6b, Figure S23). The seed perimeter, length, and width of all three *FtbHLH165* overexpressing *Arabidopsis* lines were further measured. The perimeter of the seeds overexpressing *FtbHLH165* in *Arabidopsis* is

significantly larger than that of wild-type *Arabidopsis* seeds (Figure 6c). Moreover, it was found that the seed lengths of all three *FtbHLH165* overexpressing *Arabidopsis* were longer than wild-type *Arabidopsis* seeds (Figure 6d). However, the seed width did not change significantly (Figure 6e). The above results showed that the main factor contributing to the longer perimeter was the significant increase in their seed length. In addition, the thousand seed weight of *Arabidopsis* seeds overexpressing *FtbHLH165* was significantly increased (Figure 6f). The above results suggest that in addition to regulating the rutin content, *FtbHLH165* is able to regulate seed length.

Previous studies have shown that the accumulation of flavonoids in Tartary buckwheat can increase plant resistance to *Rhizoctonia solani* (He et al., 2023). As the overexpression of *FtbHLH165* could increase the content of flavonoids in Tartary buckwheat hairy roots (Figure 5i), the disease resistance of *FtbHLH165* was further investigated. The leaves of transgenic plants with heterologous overexpression of *FtbHLH165* in *A. thaliana* were used to identify the resistance of *R. solani*. Overexpression of *FtbHLH165* in *A. thaliana* significantly enhanced disease resistance compared with the wild type (Figure 6g). In addition, *FtbHLH165* was transiently overexpressed in tobacco leaves and inoculated with *R. solani*. The results also showed that overexpression of *FtbHLH165* significantly increased the disease resistance of leaves compared with the empty vector (Figure S24). We additionally detected flavonoids in *A. thaliana* plants overexpressing *FtbHLH165*. Our results revealed that some flavonoids including nicotiflorin, afzelin, and rutin accumulated in *FtbHLH165* overexpression lines (Figure 6h). As these metabolites were previously shown to enhance plant disease resistance (He et al., 2023; Yang et al., 2016), we speculate that *FtbHLH165* might enhance plant resistance to *R. solani* through increasing flavonoid biosynthesis.

Discussion

More and more methods have been used for the identification of key genes in recent years, such as transcriptomics, QTL (quantitative trait locus) mapping, GWAS, and many others. However, each of these methods has its own advantages and disadvantages, and sometimes it is necessary to combine various techniques to achieve our objectives. The use of QTL analyses has made great progress in gene identification, for example, identifying cotton fibre development-related genes (Wang et al., 2020), barley flowering time genes (Afsharyan et al., 2020), and wheat yield-related trait genes (Yang et al., 2021). However, the difficult and laborious construction of QTL mapping

populations, somewhat limited number of detections, and low accuracy have limited the development of this method. Transcriptome analysis, a method that has flourished in recent years, also plays an important role in selecting candidate genes. However, transcriptome analysis is more focused on detecting genes related to plant growth and development and stress responses, such as *Arabidopsis* response to intense light stress (Huang *et al.*, 2019), wheat inflorescence development

(Feng *et al.*, 2017), rice growth and development (Zhao *et al.*, 2020), and moso bamboo underground shoot growth (Wei *et al.*, 2017). In addition, transcriptome data are generally used as co-expression analysis to screen for genes that are co-expressed with the target gene (Chang *et al.*, 2019; Xiao *et al.*, 2023). Transcriptome analysis seems to be limited for the screening of key genes of plant secondary metabolites, so it is essential to select suitable methods to identify secondary

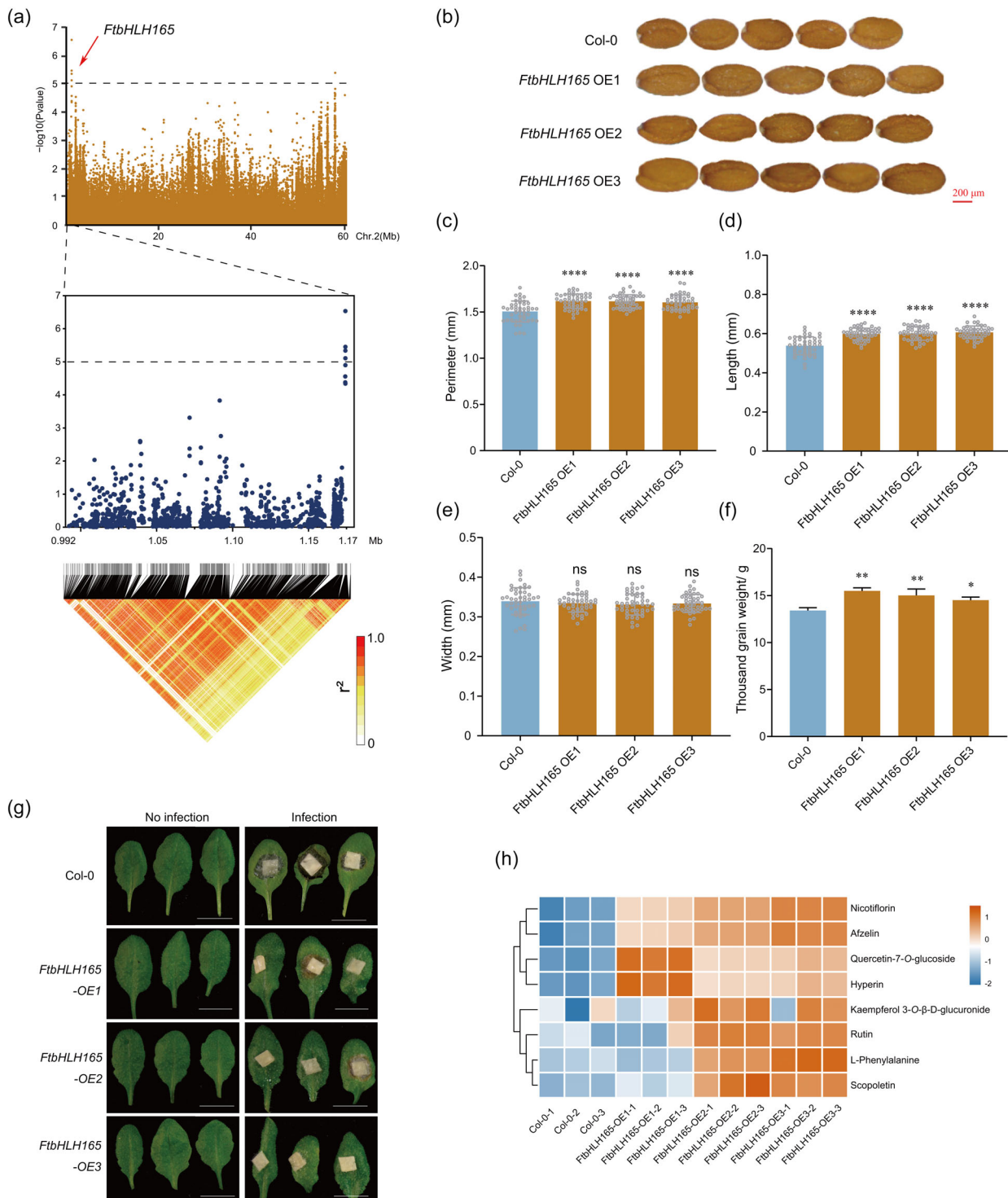


Figure 6 GWAS analysis of *FtbHLH165* and its regulation of seed size and disease resistance in *Arabidopsis thaliana*. (a) GWAS results in seed circumference with 30-fold whole gene resequencing data on chromosome 2 identified *FtbHLH165*. GWAS significant signal 200 kb range gene linkage analysis. (b) Phenotypic characterization of *Arabidopsis* seeds overexpressing *FtbHLH165*. (c–f) Seed perimeter (c), seed length (d), and seed width (e), and thousand seed weight (f) of *FtbHLH165* overexpressing *Arabidopsis* seeds were determined. Seed perimeter, seed length, and seed width data are shown as the mean \pm SD of three biological replicates ($n = 15$). **** $P < 0.0001$ (one-way ANOVA, Tukey's post-test). Seed thousand weight data are shown as the mean \pm SD of three biological replicates ($n = 3$). * $P < 0.05$ and ** $P < 0.01$ (one-way ANOVA, Tukey's post-test). (g) Phenotypes of *Arabidopsis* lines heterologously expressing *FtbHLH165* and infected with *R. solani* AG4-HGI 3. Two-week-old *Arabidopsis* leaves in vitro were inoculated with subcultured *R. solani* for 24 h. The leaves of wild-type (Col-0) were used as negative control to study the morphological characteristics of *Arabidopsis* leaves. These phenotypes were observed in the leaves of 3 *Arabidopsis* seedlings ($n = 3$). Different batches of *Arabidopsis* leaves were tested three times. A photo shows typical experimental results. Scale bar, 1 cm. (h) Heat map analysis of differential flavonoid metabolites in *FtbHLH165* overexpressing *Arabidopsis* compared to Col-0.

metabolite genes. The rapid development of bioinformatic technologies, alongside the torrent of genome sequencing and resequencing data GWAS, has been widely applied and proven highly effective in plants (Fang and Luo, 2019), and it is widely used for the localization of key genes. For example, it was highly useful in identifying key genes controlling oil content in oilseed rape (Tang et al., 2021) and the genes related to powdery mildew resistance in cotton (Zhao et al., 2021b). In addition, GWAS has made great progress in defining secondary metabolism related genes. For example, many genes regulating glucosinolate content in *Arabidopsis* (Francisco et al., 2016; Wang et al., 2022), flavonoid and steroidal glycoalkaloid content in tomato (Alseekh et al., 2015), and vitamin contents in maize (Albert et al., 2022; Diepenbrock et al., 2021) were defined by this approach. Furthermore, *GmMPK1*, a gene related to isoflavone content, was identified by combining GWAS and linkage analysis (Wu et al., 2020a). *GmMYB29* was identified involved in the soybean isoflavone anabolic pathway by GWAS (Chu et al., 2017). Moreover, many genes involved in flavonoid modification were also identified in wheat seeds by mGWAS (Chen et al., 2020). Increased resequencing depth can result in a significant increase in genome-wide coverage of sequencing reads from individual samples and an increased potential for sequencing reads from different samples to be aligned to the same location in the reference genome, and the above results lead to an increase in the number of SNPs detected in population samples. The increased depth of whole-genome resequencing enables the mining of more SNPs, providing a contribution to the identification of complex genes (Li et al., 2021; Shao et al., 2022; Zhao et al., 2018). Our association analysis based on high-depth whole-genome resequencing combined with key metabolite content resulted in significant signals detected in both the key metabolites (rutin, quercetin) and their upstream or downstream multiple substances (isoquercitrin, quercetin, hyperin, isohyperoside, and quercetin-7-O-glucoside). Thus, this also demonstrated GWAS-based identification of key genes of secondary metabolites is also effective for Tartary buckwheat. This is a reference source for similar studies in the future—particularly those associated with readily degradable metabolites.

The identification of key genes for secondary metabolites is often extremely difficult because of the minimal metabolite content, the influence of the external environment, the complexity of their metabolic pathways, and other factors. First, the minimal content of secondary metabolites may also lead to difficulties in the identification of key genes. Tartary buckwheat contains relatively high levels of rutin compared to tomato (Verhoeven et al., 2002), *Sophora japonica* flowers (Chen and Hsieh, 2010), amaranth (*Amaranthus* spp.) (Kalinova and Dadakova, 2009), olea europaea leaf (Lee and Lee, 2010), asparagus (*Asparagus*

officinalis) (Yi et al., 2019), and other rutin-containing plants. However, in Tartary buckwheat, the rutin content is minimal compared to nutrients such as starch. Secondly, the rutin content is affected by the environment. Previous studies have demonstrated that flavonoids in Tartary buckwheat are generated as protective substances against UV-B radiation (Huang et al., 2016; Özbolt et al., 2008; Tsurunaga et al., 2013). Overall, there is a decreasing trend in the content of rutin with increasing altitude and UV intensity (Figure 1a). This indicates that not just rutin hydrolase, but environmental factors such as UV radiation are important factors influencing the content of rutin. This also explains why the content of rutin and quercetin varies greatly at different altitudes, with significant signals identified by rutin and quercetin GWAS only at an altitude of 2500 m. It could be that environmental factors such as UV light in different environments greatly affect the content of flavonoids such as rutin and quercetin (Figure 1a, Figure S1A). Interestingly, observation of the metabolites isoquercitrin, quercitrin, hyperin, isohyperoside, quercetin-7-O-glucoside, and rutin revealed that they all possess a glycosyl. The rutin hydrolase is a glucoside hydrolase, which implies that it may be involved in the degradation pathway of multiple metabolites and has the ability to hydrolyse the glycosidic bonds of multiple metabolites (Kalinová et al., 2018). Finally, rutin metabolism is a complexity pathway. Rutin is produced through the phenylalanine synthesis pathway which is highly regulated at multiple key upstream enzyme genes including RT, GT, and FLS (Zhang et al., 2017). It is known that it can be hydrolysed to isoquercitrin (de Araujo et al., 2013) or quercetin (Jia et al., 2019), which in turn can be glycosylated to produce several other compounds including quercitrin, hyperin, and isohyperoside. The complexity of the metabolic pathway of rutin leads to difficulties in the identification of its key genes; however, this explains why it took us multiple efforts to locate the rutin hydrolase gene. Namely, the initial identification of this gene by rutin and quercetin content at 2500 m altitude, as well as high-depth resequencing combined with rutin-related metabolites (isoquercitrin, quercitrin, hyperin, isohyperoside, and quercetin-7-O-glucoside) re-identified this gene.

Plants often face a wide variety of adversities such as UV, drought, disease, and other stresses in their natural environment. However, rutin, as a flavonoid produced by secondary metabolism, is of great importance to both the plant itself (He et al., 2023) and to humans (Ola et al., 2015; Xu et al., 2014). As an important factor for balancing rutin metabolism, rutin hydrolase is not agriculturally favoured. Indeed, in agriculture, people hope to breed as many high-rutin Tartary buckwheat varieties as possible since this would not only reduce the bitterness of Tartary buckwheat products but also maximize their benefits in the human diet. At this point, two strategies are being considered for the breeding of high-rutin Tartary buckwheat

varieties. One is to synthesize as much rutin as possible, and the other is to minimize rutin hydrolysis. Research on rutin synthesis is already quite extensive (Ding *et al.*, 2021; Li *et al.*, 2019; Lin *et al.*, 2023; Zhang *et al.*, 2018), but research on rutin metabolism is relatively scarce. For this reason, we identified the rutin hydrolase gene *FtGH1* in this study. Moreover, molecular docking predicted possible key sites of enzymatic activity of *FtGH1*, and mutations in these key sites resulted in abnormal subcellular localization of the protein (Figure 4i), which may lead to loss of enzymatic activity (Figure 4g,h). In addition, the repression of *FtGH1* by upstream transcription factors could also serve the purpose of reducing rutin hydrolysis. To this end, the transcriptome was used to explore possible transcription factors with *FtbHLH165* being shown to repress *FtGH1* expression (Figure 5f,g), thereby increasing the content of rutin (Figure 5h). Interestingly, the transcriptional regulatory relationship of *FtbHLH165* to *FtGH1* may be conserved in *Fagopyrum*. This regulatory relationship may not exist in other plants because no collinear genes for *FtGH1* have been identified in other plants. In addition, for common buckwheat, the absence of the MYC element in the promoter of the *FtGH1* collinear gene seems to explain why the rutin content of common buckwheat is lower than that of Tartary buckwheat and golden buckwheat (Jiang *et al.*, 2007; Kreft *et al.*, 2006). The underlying molecular mechanism may be the lack of transcriptional repression of the related rutin hydrolase by a bHLH-like transcription factor. Of course, its reality needs to be verified in depth by future molecular experiments. Overall, our study can provide theoretical guidance for future molecular breeding of high rutin containing crops. We suggest that mutations in the key active sites S382 and S472 of *FtGH1* are advantageous for cultivating high-rutin Tartary buckwheat varieties. Moreover, the Hap. C type promoter, which binds more strongly to the transcriptional repressor *FtbHLH165*, can likewise provide assistance in the development of high-rutin varieties. That said, challenges to achieving these goals exist, such as the limited current availability of genetic transformation systems in Tartary buckwheat. Although successful overexpression or knockout of genes in Tartary buckwheat has been reported and stable genetically transformed plants have been obtained (Pinski and Betekhtin, 2023). However, this process is very time-consuming and requires a high degree of patience from collaborators.

In the current study, GWAS was performed by combining high depth whole-genome resequencing with the analysis of flavonoid secondary metabolites. Following this route, the key gene of rutin hydrolysis metabolism *FtGH1* was identified, and the downstream pathway of rutin catabolism was elucidated with the transcription factors regulating this gene being identified. These research results represent a reference for the molecular design of rutin breeding in Tartary buckwheat, and can furthermore provide a theoretical basis for the cloning of complex genes and even the analysis of catabolic pathways of readily degradable metabolites.

Methods

Plant materials

The 226 Tartary buckwheat materials were planted at five different elevations: Sichuan Province (Hagan Township, Zhaojue County, 103.18° E, 28.10° N, 1006.4 m; Lamujue Township, Meigu County, 103.10° E, 28.13° N, 1520.5 m; Xincheng Township, Zhaojue County, 102.84° E, 27.98° N, 2010.7 m; Saladipo Township, Zhaojue County, 102.61° E, 27.92° N, 2530.2 m; and Riha Township, Zhaojue County, 103.08° E, 27.96° N, 3065.6 m).

Detailed information on these Tartary buckwheat materials, and how they were grown, managed, and harvested, has been presented previously (Zhang *et al.*, 2021). The method of constructing 200 micro-core germplasm resources of Tartary buckwheat and the cultivation method refer to previous research (Zhao *et al.*, 2023).

Flavonoid content analysis

The 226 Tartary buckwheat seeds were dried at 105 °C for 30 min, and keeping constant weight at 65 °C, the seeds were smashed and filtered through a 40-mesh sieve. 0.2 g of the sample was dissolved in 20 mL of 80% methanol and then extracted at 50 °C for 45 min under ultrasound at 40 KHz. The obtained solution was filtered through a 0.22 µm organic microporous membrane and analysed by high-performance liquid chromatography (LC–MS, Agilent G6500 series HPLC–QQQ). The chromatographic column was HSS T3 (2.1 mm × 100 mm × 1.8 µm), and the mobile phases were (A) water/formic acid (99.9/0.1, v/v) and (B) acetonitrile/formic acid (99.9/0.1, v/v). The elution gradient was referenced to the previous method (Zhang *et al.*, 2021). Three biological replicates were measured for each sample. The growth of positive hairy roots of Tartary buckwheat refer to previous studies (Zhang *et al.*, 2021). Transiently expressing tobacco leaves and *Arabidopsis* leaves were freeze-dried at –40 °C. The flavonoid content of hairy roots and *Arabidopsis* leaves was assayed with reference to the LC–MS method described above.

DNA preparation and 30-fold depth re-sequencing

Young Tartary buckwheat leaves during the nutritional growth period were used to extract genomic DNA with reference to the CTAB method (Murray and Thompson, 1980). The concentration and quality of whole genomic DNA were detected by Qubit 3.0, and genomic DNA was sent to ANOPOAD GENOME Technology (Beijing) Co. for whole-genome resequencing. At least 1 µg genomic DNA was used to construct the library. The Burrows-Wheeler Aligner program (Li, 2013) with default parameters was used to map all sequencing reads to the assembled genome (Zhang *et al.*, 2017). The detailed SNP calling method is based on previous research (Zhang *et al.*, 2021).

SNP calling

Tartary buckwheat 10× depth resequencing data of 226 germplasm were taken from a previous study (Zhang *et al.*, 2021; Zhao *et al.*, 2023). Besides, 30× highly depth whole-genome resequencing of the 200 micro-core accessions was completed in this study. All the clean reads were mapped to the reference genome (Zhang *et al.*, 2017) using the BWA-MEM program of the BWA software (Li, 2013) with the default parameter. Then, the aligned reads were sorted by samtools software (Li, 2011; Li *et al.*, 2009) and removed duplicates by Picard (<https://broadinstitute.github.io/picard/>). Final alignment file was used for the SNP detection through Haplotyper function with the GATK pipeline (McKenna *et al.*, 2010). The population variants were filtered with the following criteria: (a) QUAL >30.0; (b) QD >5.0; (c) FS < 60.0; (d) MQ0 ≥ 4 and ((MQ0/(1.0*DP)) > 0.1); (e) DP >5. All variants were remapped to the reference genome and annotated in the genomic position using ANNOVAR software (Wang *et al.*, 2010a).

Genome-wide association study (GWAS)

Population SNPs with minor allele frequency (MAF) ≥ 0.05 and missing rate ≤0.1 were used for GWAS, which generated

1 103 732 and 1 363 970 SNPs in 10-fold and 30-fold resequencing data sets, respectively. The Efficient Mixed-Model Association eXpedited program (EMMAX) (Kang *et al.*, 2010) was used for association analysis. After evaluated by Linkage disequilibrium and effective number of SNPs, the significance threshold was estimated to be approximately $P = 10^{-5}$. The 100 kb flanking region of peak SNPs were considered as candidate region associated with target phenotype. Based on the significant loci determined by GWAS, the candidate regions were further analysed and annotated for all related genes.

Construction of overexpressed materials

Total RNA was extracted using the RNAPure Pure Factory Plus Kit (DP441, Beijing, China) and HiScript III RT SuperMix for qPCR (R323-01, Vazyme, Nanjing, China) was used for reverse transcription. The CDSs of *FtGH1* and *FtbHLH165* were inserted into the pCambia 1307 vector and then transformed into *Agrobacterium* A4 to obtain transgenic hairy roots, respectively, according to the method described previously (Li *et al.*, 2019; Zhou *et al.*, 2017). The methods of obtaining of hairy roots were referred to the previous method of the group (Zhang *et al.*, 2021). The pCambia 1307-*FtbHLH165* vector plasmid was transformed into *Agrobacterium tumefaciens* GV3101 to obtain overexpressing *Arabidopsis* seeds (Clough and Bent, 1998). Subsequently harvested seeds were screened on MS solid medium containing 30 µg/mL hygromycin B, and overexpression of *FtbHLH165* was verified in combination with PCR as well as qRT-PCR. PCR was performed to amplify whether *FtbHLH165* was inserted into the *Arabidopsis* genome using DNA from different *Arabidopsis* lines of *FtbHLH165* as a template. RNA from different *Arabidopsis* lines of *FtbHLH165* was used as a template for qRT-PCR to detect *FtbHLH165* expression. qRT-PCR was performed using ChamQ Universal SYBR qPCR Master Mix (Q711, Vazyme, Nanjing, China). Primer sequences are given in Data Set S29.

FtGH1 transient transformation of tobacco leaves and protein expression assay

The pGreen-62SK is a transient overexpression vector which is capable of transient overexpression in tobacco leaves. *FtGH1* fusion his tag was constructed together into the pGreen-62SK vector. Successfully transformed *Agrobacterium* GV3101 was used to inject 4-week-old tobacco leaves of uniform size, with a full area of each leaf injected. Three replicates were incubated in dark for 24 h. Samples were taken at 70 h after injection and freeze-dried at low temperature for 24 h. Meanwhile the tobacco leaves were used to extract *FtGH1* protein, and then western blotting was performed to detect the expression of *FtGH1*-his. The same mass was weighed for flavonoid extraction and then assayed by LC-MS (QQQ).

Recombinant protein purification and enzyme assay

The full-length cDNA of *FtGH1* was amplified and constructed into the pD2P_1.08e-8His expression vector. Positive cloned DNA templates were added directly to ProteinFactory RXN to a final template concentration of 20 ng/µL. Protein production was accomplished by the reaction of 5 mL of ProteinFactory RXN reaction solution in a 12-well plate container at 28 °C on a 150 rpm shaker for 6 h. The ProteinFactory RXN reaction solution was then centrifuged at 1986 g for 3 min at 4 °C and the supernatant was collected. His-Monster Beads were used to purify recombinant proteins. Immunoblotting of *FtGH1*-His was performed with

anti-His (1 : 3000; CW0082A, CWBIO, Taizhou, Jiangsu, China) and anti-mouse IgG (1 : 8000; CW0102, CWBIO) antibodies.

For *FtGH1* activity assay, 0.1 µg purified protein was added to the reaction buffer (20 mM ammonium acetate buffer, pH 5.0, 0.5 mM rutin (SR8250; Solarbio, Beijing, China), 1 mM ATP) and incubated at 37 °C for 30 min. After terminated by freeze dryer at −40 °C, the dried reaction products were re-dissolved in 80% methanol for analysed by HPLC-MS (Nexera XR UHPLC/HPLC System).

Molecular docking

Molecular docking studies were performed with the AutoDock4 program (Bitencourt-Ferreira *et al.*, 2019). We docked rutin, hyperin, isohyperoside, quercitrin, isoquercitrin, and quercetin-7-O-glucoside into the binding pocket of *FtGH1* homologous protein. The best docking mode was chosen based on the lowest binding free energy calculated by AutoDock4, and the resulting complexes of these compounds bound to *FtGH1* were obtained.

Subcellular localization of *FtGH1*

The full-length CDS of *FtGH1* was constructed into pCambia 1300 vector and the successful vector was transformed into *Agrobacterium tumefaciens* GV3101. pJIT-mCherry-Mem was injected into *N. benthamiana* leaves as a cell membrane maker together with *FtGH1::GFP*. A laser confocal microscope (Zeiss LSM900) was used to observe the subcellular localization at 561 (excitation)/590–640 nm (emission) for mCherry and 488 (excitation)/500–530 nm (emission) for GFP. Primer sequences are listed in Data Set S26.

Cloning of *FtGH1* promoter and dual luciferase reporter assay

DNA of Tartary buckwheat material sx140 (Hap. C) and SX-342 (Hap. T) was used as a template to clone the promoter of *FtGH1*. The promoter analysis was performed using PlantCARE (Lescot *et al.*, 2002). The promoters of both haplotypes were constructed into the pGreenII 0800-LUC vector, respectively. In addition, *FtbHLH165* was constructed into pGreen-62SK for detecting the regulation of *FtGH1* promoter. All the above vectors were transformed into *Agrobacterium tumefaciens* GV3101 and infiltration assays were performed on two mature leaves of 4-week-old tobacco plants using a needleless syringe. Luc fluorescence imaging was detected with a plant live imager (Night SHADE LB985). A multifunctional enzyme marker (Biotek Synergy Mx) was used to detect LUC/REN activity levels.

Phylogenetic tree construction and determination of rutin hydrolase activity

Phylogenetic trees of *FtGH1* and *FtbHLH165* were performed by MEGA X based on the neighbour-joining method and the maximum-likelihood method, respectively (Kumar *et al.*, 2018; Saitou and Nei, 1987; Tamura *et al.*, 2011). Forty buckwheat seeds were used for the rutin hydrolase activity assay. Defatted Tartary buckwheat powder (2 g) was extracted in 30 mL 200 mM acetate buffer (pH 4.0) at 4 °C for 3 h. The rutin hydrolase activity of different materials of Tartary buckwheat was measured by the method of Chen and Gu (2011).

Yeast one-hybrid assays

Yeast one-hybrid using the pHIS2 and pGADT7 systems. The target *cis*-element in *FtGH1* was inserted into the pHIS2 vector as a reporter, and *FtbHLH165* was inserted into the pGADT7 vector containing the GAL4 transcriptional activation structural domain

as an effector. The effector and reporter were co-transformed into the Y1H Gold strain. The transformants were plated on minimal synthetic definition (SD)-deficient Leu(−L) and Trp(−T) glucose medium, and then on three-deficient medium containing 3-AT (SD-LTH) to verify whether the transcription factors bind *cis*-acting elements. The Y187 yeast strain (Cat.NO. ZC1603) was purchased from Beijing Zoman Biotechnology Co., Ltd. Y1H experiments were performed according to the method of the manufacturer (<http://www.zomanbio.com/>).

Transcriptome assays

FtbHLH165-OE1 overexpressing hairy roots and empty vector hairy roots were used for transcriptome analysis. The above hairy roots were transferred to MS liquid medium containing 100 mg/mL cefotaxime and grown in the dark at 22 °C, 160 r/min for 2 weeks with shaking. Three independent biological replicates were performed for each sample. The transcriptome sequencing of hairy roots was performed by Shanghai OE Biotech. Differential expression analysis was conducted using DEGseq (Wang *et al.*, 2010b) and DESeq2 (Love *et al.*, 2014). Adjusted *P*-values ≤ 0.05 for genes (Benjamini–Hochberg method) with fold change < 0.5 or > 2 (*FtbHLH165* compared to EV). The functional enrichment of these differentially expressed genes was analysed using the KEGG database. We performed a co-clustering analysis of gene expression patterns using R package TCseq on the transcriptome of buckwheat seeds with different development (Liu *et al.*, 2018).

Phenotypic observation of *Arabidopsis* seeds

Mature dry *Arabidopsis* seeds were photographed under a Zeiss stereomicroscope (Carl Zeiss Microscopy GmbH). Fifteen seeds of each line were randomly selected to be measured and three independent replicates were performed. For seed weight, 100 seeds were weighed with an analytical balance. Seed phenotype images are measured in seed circumference, seed length, and seed width using ImageJ (Schneider *et al.*, 2012).

Evaluation of disease resistance of *FtbHLH165* overexpression lines

For *R. solani* infected *A. thaliana*, isolated leaves of two-week-old *Arabidopsis* seedlings were inoculated with sub-cultured mycelial disks and incubated with 28 °C for 24 h according to methods described previously (He *et al.*, 2023). For *R. solani* infected tobacco leaves, *FtbHLH165* was transiently overexpressed in tobacco leaves for 48 h. And the isolated leaves were inoculated with *R. solani* according to methods described previously (He *et al.*, 2023). The phenotypes were observed after 24 h. Three biological replicates were performed and the experiment was carried out three times.

Statistical analysis

Histograms were plotted using Origin 2019b. Pearson correlation was analysed using SigmaPlot 12.0 ($P < 0.05$). Significant differences were assessed using two-tailed Student's *t*-tests and one-way analysis of variance (ANOVA, $P < 0.05$) performed by GraphPad Prism 8 software. Statistical data are provided as Data Set S30.

Acknowledgements

We thank the public laboratory of the Biotechnology Research Institute, Chinese Academy of Agricultural Sciences for use of the HPLC and triple-quadrupole MS/MS instrument, and for providing

technical assistance. This research was supported by National Natural Science Foundation of China (32241042, 32161143005, and 32111540258), The Youth Innovation Program of Chinese Academy of Agricultural Sciences (Y2022QC02), European Union Horizon 2020 project Planta SYST (SGA-CSA No. 739582 under FPA No. 664620), and Science and Technology Project of Hebei Education Department (BJ2019022).

Authors contributions

M.Z. designed and managed the project. M.Z. and K.Z. organized the funding for this research. M.Z. and K.Z. provided the genetic materials. D.L., Y.S., Y.G., X.H., Y.H., and W.L. performed data analysis and figure design. D.L., Y.F., W.L., X.H., J.H., H.Z., X.L., and Y.X. performed most of the experiments. D.L., Y.H., M.I.G., A.R.F., J.C., J.R., K.Z., and M.Z. wrote the manuscript. All authors read and approved the manuscript.

References

- Afsharyan, N.P., Sannemann, W., Leon, J. and Ballvora, A. (2020) Effect of epistasis and environment on flowering time in barley reveals a novel flowering-delaying QTL allele. *J. Exp. Bot.* **71**, 893–906.
- Albert, E., Kim, S., Magallanes-Lundback, M., Bao, Y., Deason, N., Danilo, B., Wu, D. *et al.* (2022) Genome-wide association identifies a missing hydrolase for tocopherol synthesis in plants. *Proc. Natl. Acad. Sci. USA* **119**, e2113488119.
- Alseekh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., Kleessen, S. *et al.* (2015) Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell* **27**, 485–512.
- de Araujo, M.E., Moreira Franco, Y.E., Alberto, T.G., Sobreiro, M.A., Conrado, M.A., Priolli, D.G., Frankland Sawaya, A.C. *et al.* (2013) Enzymatic deglycosylation of rutin improves its antioxidant and antiproliferative activities. *Food Chem.* **141**, 266–273.
- Barberis, E., Marengo, E. and Manfredi, M. (2021) Protein subcellular localization prediction. *Methods Mol. Biol.* **2361**, 197–212.
- Bitencourt-Ferreira, G., Pinto, V. and de Azevedo, W. (2019) Docking with AutoDock4. *Methods Mol. Biol.* **2053**, 125–148.
- Calzada, F., Solares-Pascasio, J.I., Ordóñez-Razo, R.M., Velázquez, C., Barbosa, E., García-Hernández, N., Méndez-Luna, D. *et al.* (2017) Antihyperglycemic activity of the leaves from *Annona cherimola* Miller and Rutin on Alloxan-induced diabetic rats. *Pharm. Res.* **9**, 1–6.
- Chang, Y.M., Lin, H.H., Liu, W.Y., Yu, C.P., Chen, H.J., Wartini, P.P., Kao, Y.Y. *et al.* (2019) Comparative transcriptomics method to infer gene coexpression networks and its applications to maize and rice leaf transcriptomes. *Proc. Natl. Acad. Sci. USA* **116**, 3091–3099.
- Chen, P. and Gu, J. (2011) A rapid measurement of rutin-degrading enzyme activity in extract of tartary buckwheat seeds. *Food Bioprod. Process.* **89**, 81–85.
- Chen, H.N. and Hsieh, C.L. (2010) Effects of *Sophora japonica* flowers (Huaihua) on cerebral infarction. *Chinese Med.* **5**, 1–4.
- Chen, J., Hu, X., Shi, T., Yin, H., Sun, D., Hao, Y., Xia, X. *et al.* (2020) Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. *Plant Biotechnol. J.* **18**, 1722–1735.
- Cheng, Z., Sun, L., Qi, T., Zhang, B., Peng, W., Liu, Y. and Xie, D. (2011) The bHLH transcription factor MYC3 interacts with the Jasmonate ZIM-domain proteins to mediate jasmonate response in *Arabidopsis*. *Mol. Plant* **4**, 279–288.
- Chu, S., Wang, J., Zhu, Y., Liu, S., Zhou, X., Zhang, H., Wang, C.E. *et al.* (2017) An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean. *PLoS Genet.* **13**, e1006770.
- Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.

- Deng, J., Zhao, J., Huang, J., Damaris, R.N., Li, H., Shi, T., Zhu, L. et al. (2022) Comparative proteomic analyses of Tartary buckwheat (*Fagopyrum tataricum*) seeds at three stages of development. *Funct. Integr. Genomics* **22**, 1449–1458.
- Diepenbrock, C.H., Ilut, D.C., Magallanes-Lundback, M., Kandianis, C.B., Lipka, A.E., Bradbury, P.J., Holland, J.B. et al. (2021) Eleven biosynthetic genes explain the majority of natural variation in carotenoid levels in maize grain. *Plant Cell* **33**, 882–900.
- Ding, M., Zhang, K., He, Y., Zuo, Q., Zhao, H., He, M., Georgiev, M.I. et al. (2021) FtBPM3 modulates the orchestration of FtMYB11-mediated flavonoids biosynthesis in Tartary buckwheat. *Plant Biotechnol. J.* **19**, 1285–1287.
- Ding, M., He, Y., Zhang, K., Li, J., Shi, Y., Zhao, M., Meng, Y. et al. (2022) JA-induced FtBPM3 accumulation promotes FTERF-EAR3 degradation and rutin biosynthesis in Tartary buckwheat. *Plant J.* **111**, 323–334.
- Fang, C. and Luo, J. (2019) Metabolic GWAS-based dissection of genetic bases underlying the diversity of plant metabolism. *Plant J.* **97**, 91–100.
- Feng, N., Song, G., Guan, J., Chen, K., Jia, M., Huang, D., Wu, J. et al. (2017) Transcriptome profiling of wheat inflorescence development from spikelet initiation to floral patterning identified stage-specific regulatory genes. *Plant Physiol.* **174**, 1779–1794.
- Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J.M., Gimenez-Ibanez, S., Geerinck, J., Eeckhout, D. et al. (2011) The Arabidopsis bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell* **23**, 701–715.
- Francisco, M., Joseph, B., Caligagan, H., Li, B., Corwin, J.A., Lin, C., Kerwin, R.E. et al. (2016) Genome wide association mapping in *Arabidopsis thaliana* identifies novel genes involved in linking allyl glucosinolate to altered biomass and defense. *Front. Plant Sci.* **7**, 1010.
- Gaberscik, A., Voncina, M., Trost, T., Germ, M. and Olof Björn, L. (2002) Growth and production of buckwheat (*Fagopyrum esculentum*) treated with reduced, ambient and enhanced UV-B radiation. *J. Photochem. Photobiol. B.* **66**, 30–36.
- Gao, J., Wang, T., Liu, M., Liu, J. and Zhang, Z. (2017) Transcriptome analysis of filling stage seeds among three buckwheat species with emphasis on rutin accumulation. *PLoS One* **12**, e0189672.
- Hao, J., Wu, T., Li, H., Wang, W. and Liu, H. (2016) Dual effects of slightly acidic electrolyzed water (SAEW) treatment on the accumulation of gamma-aminobutyric acid (GABA) and rutin in germinated buckwheat. *Food Chem.* **201**, 87–93.
- He, Y., Zhang, K., Li, S., Lu, X., Zhao, H., Guan, C., Huang, X. et al. (2023) Multi-omics analysis reveals the molecular mechanisms underlying virulence in Rhizoctonia and jasmonic acid-mediated resistance in Tartary buckwheat (*Fagopyrum tataricum*). *Plant Cell*, 2773–2798.
- Huang, X., Yao, J., Zhao, Y., Xie, D., Jiang, X. and Xu, Z. (2016) Efficient rutin and quercetin biosynthesis through flavonoids-related gene expression in *Fagopyrum tataricum* Gaertn. hairy root cultures with UV-B irradiation. *Front. Plant Sci.* **7**, 63.
- Huang, J., Zhao, X. and Chory, J. (2019) The Arabidopsis transcriptome responds specifically and dynamically to high light stress. *Cell Rep.* **29**, 4186–4199.e3.
- Huda, M.N., Lu, S., Jahan, T., Ding, M., Jha, R., Zhang, K., Zhang, W. et al. (2021) Treasure from garden: bioactive compounds of buckwheat. *Food Chem.* **335**, 127653.
- Ismail, H., Dragisic Maksimovic, J., Maksimovic, V., Shabala, L., Zivanovic, B.D., Tian, Y., Jacobsen, S.E. et al. (2015) Rutin, a flavonoid with antioxidant activity, improves plant salinity tolerance by regulating K⁺ retention and Na⁺ exclusion from leaf mesophyll in quinoa and broad beans. *Funct. Plant Biol.* **43**, 75–86.
- Jia, P., Wang, Y., Niu, Y., Han, X., Zhu, Y., Xu, Q., Li, Y. et al. (2019) Cloning and molecular characterization of rutin degrading enzyme from tartary buckwheat (*Fagopyrum tataricum* Gaertn.). *Plant Physiol. Biochem.* **143**, 61–71.
- Jiang, P., Burczynski, F., Campbell, C., Pierce, G., Austria, J.A. and Briggs, C.J. (2007) Rutin and flavonoid contents in three buckwheat species *Fagopyrum esculentum*, *F. tataricum*, and *F. homotropicum* and their protective effects against lipid peroxidation. *Food Res. Int.* **40**, 356–364.
- Joshi, D.C., Zhang, K., Wang, C., Chandora, R., Khurshid, M., Li, J., He, M. et al. (2020) Strategic enhancement of genetic gain for nutraceutical development in buckwheat: a genomics-driven perspective. *Biotechnol. Adv.* **39**, 107479.
- Kalinova, J. and Dadakova, E. (2009) Rutin and total quercetin content in amaranth (*Amaranthus* spp.). *Plant Foods Hum. Nutr.* **64**, 68–74.
- Kalinová, J.P., Vrchotová, N. and Tríska, J. (2018) Contribution to the study of rutin stability in the achenes of Tartary buckwheat (*Fagopyrum tataricum*). *Food Chem.* **258**, 314–320.
- Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C. et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354.
- Kreft, M. (2016) Buckwheat phenolic metabolites in health and disease. *Nutr. Res. Rev.* **29**, 30–39.
- Kreft, S., Strukelj, B., Gaberscik, A. and Kreft, I. (2002) Rutin in buckwheat herbs grown at different UV-B radiation levels: comparison of two UV spectrophotometric and an HPLC method. *J. Exp. Bot.* **53**, 1801–1804.
- Kreft, I., Fabjan, N. and Yasumoto, K. (2006) Rutin content in buckwheat (*Fagopyrum esculentum* Moench) food materials and products. *Food Chem.* **98**, 508–512.
- 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.
- Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549.
- Lee, O.H. and Lee, B.Y. (2010) Antioxidant and antimicrobial activities of individual and combined phenolics in *Olea europaea* leaf extract. *Bioresour. Technol.* **101**, 3751–3754.
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P. et al. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30**, 325–327.
- Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993.
- Li, H. (2013) *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. arXiv:1303.3997v2.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Li, X., Kim, Y.B., Kim, Y., Zhao, S., Kim, H.H., Chung, E., Lee, J.H. et al. (2013) Differential stress-response expression of two flavonol synthase genes and accumulation of flavonols in tartary buckwheat. *J. Plant Physiol.* **170**, 1630–1636.
- Li, J., Zhang, K., Meng, Y., Li, Q., Ding, M. and Zhou, M. (2019) FtMYB16 interacts with Ftimportin- α 1 to regulate rutin biosynthesis in tartary buckwheat. *Plant Biotechnol. J.* **17**, 1479–1481.
- Li, J., Hossain, M.S., Ma, H., Yang, Q., Gong, X., Yang, P. and Feng, B. (2020) Comparative metabolomics reveals differences in flavonoid metabolites among different coloured buckwheat flowers. *J. Food Compos. Anal.* **85**:103335.
- Li, C., Wang, G., Li, H., Wang, G., Ma, J., Zhao, X., Huo, L. et al. (2021) High-depth resequencing of 312 accessions reveals the local adaptation of foxtail millet. *Theor. Appl. Genet.* **134**, 1303–1317.
- Lin, H., Yao, Y., Sun, P., Feng, L., Wang, S., Ren, Y., Yu, X. et al. (2023) Haplotype-resolved genomes of two buckwheat crops provide insights into their contrasted rutin concentrations and reproductive systems. *BMC Biol.* **21**, 87.
- Liu, M., Ma, Z., Zheng, T., Sun, W., Zhang, Y., Jin, W., Zhan, J. et al. (2018) Insights into the correlation between Physiological changes in and seed development of tartary buckwheat (*Fagopyrum tataricum* Gaertn.). *BMC Genomics* **19**, 648.
- Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Lukić, L., Bonafaccia, G., Timoracka, M., Vollmannova, A., Trček, J., Nyambe, T.K., Melini, V. et al. (2016) Rutin and quercetin transformation during preparation of buckwheat sourdough bread. *J. Cereal Sci.* **69**, 71–76.

- Lundberg, E. and Borner, G.H.H. (2019) Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K. et al. (2010) The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.
- Murray, M.G. and Thompson, W.F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**, 4321–4326.
- Nakata, M., Mitsuda, N., Herde, M., Koo, A.J., Moreno, J.E., Suzuki, K., Howe, G.A. et al. (2013) A bHLH-type transcription factor, ABA-INDUCIBLE BHLH-TYPE TRANSCRIPTION FACTOR/JA-ASSOCIATED MYC2-LIKE1, acts as a repressor to negatively regulate jasmonate signaling in arabidopsis. *Plant Cell* **25**, 1641–1656.
- Niu, Y., Figueroa, P. and Browse, J. (2011) Characterization of JAZ-interacting BHLH transcription factors that regulate jasmonate responses in Arabidopsis. *J. Exp. Bot.* **62**, 2143–2154.
- Ola, M.S., Ahmed, M.M., Ahmad, R., Abuhashish, H.M., Al-Rejaie, S.S. and Alhomida, A.S. (2015) Neuroprotective effects of rutin in streptozotocin-induced diabetic rat retina. *J. Mol. Neurosci.* **56**, 440–448.
- Özbolt, L., Kreft, S., Kreft, I., Germ, M. and Stibilj, V. (2008) Distribution of selenium and phenolics in buckwheat plants grown from seeds soaked in Se solution and under different levels of UV-B radiation. *Food Chem.* **110**, 691–696.
- Pinski, A. and Betekhtin, A. (2023) Efficient Agrobacterium-mediated transformation and genome editing of *Fagopyrum tataricum*. *Front. Plant Sci.* **14**, 1270150.
- Ruan, J., Zhou, Y., Zhou, M., Yan, J., Khurshid, M., Weng, W., Cheng, J. et al. (2019) Jasmonic acid signaling pathway in plants. *Int. J. Mol. Sci.* **20**, 2479.
- Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Sasaki-Sekimoto, Y., Jikumaru, Y., Obayashi, T., Saito, H., Masuda, S., Kamiya, Y., Ohta, H. et al. (2013) Basic helix-loop-helix transcription factors JASMONATE-ASSOCIATED MYC2-LIKE1 (JAM1), JAM2, and JAM3 are negative regulators of jasmonate responses in Arabidopsis. *Plant Physiol.* **163**, 291–304.
- Schneider, C., Rasband, W. and Eliceiri, K. (2012) NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675.
- Shao, Z., Shao, J., Huo, X., Li, W., Kong, Y., Du, H., Li, X. et al. (2022) Identification of closely associated SNPs and candidate genes with seed size and shape via deep re-sequencing GWAS in soybean. *Theor. Appl. Genet.* **135**, 2341–2351.
- Sharma, A.D., Kaur, J. and Chand, T.S.P. (2021) Spectral fingerprinting revealed modulation of plant secondary metabolites in providing abiotic stress tolerance to invasive alien plants *Lantana camara* (L.), *Parthenium hysterophorus* (L.), *Ricinus communis* (L.), and *Ageratum conyzoides* (L.) (plant metabolites in stress tolerance to invasive plants). *BioTechnologia (Pozn)* **102**, 307–319.
- Suzuki, T., Honda, Y. and Mukasa, Y. (2005) Effects of UV-B radiation, cold and desiccation stress on rutin concentration and rutin glucosidase activity in tartary buckwheat (*Fagopyrum tataricum*) leaves. *Plant Sci.* **168**, 1303–1307.
- Suzuki, T., Morishita, T., Noda, T., Ishiguro, K., Otsuka, S. and Katsu, K. (2021) Breeding of buckwheat to reduce bitterness and rutin hydrolysis. *Plants (Basel, Switzerland)* **10**, 791.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.
- Tang, S., Zhao, H., Lu, S., Yu, L., Zhang, G., Zhang, Y., Yang, Q.Y. et al. (2021) Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. *Mol. Plant* **14**, 470–487.
- Tsurunaga, Y., Takahashi, T., Katsube, T., Kudo, A., Kuramitsu, O., Ishiwata, M. and Matsumoto, S. (2013) Effects of UV-B irradiation on the levels of anthocyanin, rutin and radical scavenging activity of buckwheat sprouts. *Food Chem.* **141**, 552–556.
- Verhoeven, M., Bovy, A., Collins, G., Muir, S., Robinson, S., de Vos, C.H.R. and Colliver, S. (2002) Increasing antioxidant levels in tomatoes through modification of the flavonoid biosynthetic pathway. *J. Exp. Bot.* **53**, 2099–2106.
- Wang, K., Li, M. and Hakonarson, H. (2010a) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164.
- Wang, L., Feng, Z., Wang, X., Wang, X. and Zhang, X. (2010b) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138.
- Wang, L., Yang, X., Qin, P., Shan, F. and Ren, G. (2013) Flavonoid composition, antibacterial and antioxidant properties of tartary buckwheat bran extract. *Ind. Crops Prod.* **49**, 312–317.
- Wang, F., Zhang, J., Chen, Y., Zhang, C., Gong, J., Song, Z., Zhou, J. et al. (2020) Identification of candidate genes for key fibre-related QTLs and derivation of favourable alleles in *Gossypium hirsutum* recombinant inbred lines with *G. barbadense* introgressions. *Plant Biotechnol. J.* **18**, 707–720.
- Wang, P., Clark, N.M., Nolan, T.M., Song, G., Bartz, P.M., Liao, C.Y., Montes-Serey, C. et al. (2022) Integrated omics reveal novel functions and underlying mechanisms of the receptor kinase FERONIA in *Arabidopsis thaliana*. *Plant Cell* **34**, 2594–2614.
- Wei, Q., Jiao, C., Guo, L., Ding, Y., Cao, J., Feng, J., Dong, X. et al. (2017) Exploring key cellular processes and candidate genes regulating the primary thickening growth of Moso underground shoots. *New Phytol.* **214**, 81–96.
- Wu, D., Li, D., Zhao, X., Zhan, Y., Teng, W., Qiu, L., Zheng, H. et al. (2020a) Identification of a candidate gene associated with isoflavone content in soybean seeds using genome-wide association and linkage mapping. *Plant J.* **104**, 950–963.
- Wu, X., Fu, G., Li, R., Li, Y., Dong, B. and Liu, C. (2020b) Effect of thermal processing for rutin preservation on the properties of phenolics & starch in Tartary buckwheat achenes. *Int. J. Biol. Macromol.* **164**, 1275–1283.
- Xiao, X., Zhu, M., Liu, Y., Zheng, J., Cui, Y., Xiong, C., Liu, J. et al. (2023) Phenotypic and gene co-expression network analyses of seed shattering in divergent sorghum (*Sorghum* spp.). *Crop J.* **11**, 478–489.
- Xu, P.X., Wang, S.W., Yu, X.L., Su, Y.J., Wang, T., Zhou, W.W., Zhang, H. et al. (2014) Rutin improves spatial memory in Alzheimer's disease transgenic mice by reducing Abeta oligomer level and attenuating oxidative stress and neuroinflammation. *Behav. Brain Res.* **264**, 173–180.
- Yang, J., Guo, J. and Yuan, J. (2008) In vitro antioxidant properties of rutin. *LWT – Food Sci. Technol.* **41**, 1060–1066.
- Yang, W., Xu, X., Li, Y., Wang, Y., Li, M., Wang, Y., Ding, X. et al. (2016) Rutin-mediated priming of plant resistance to three bacterial pathogens initiating the early SA signal pathway. *PLoS One* **11**, e0146910.
- Yang, Y., Amo, A., Wei, D., Chai, Y., Zheng, J., Qiao, P., Cui, C. et al. (2021) Large-scale integration of meta-QTL and genome-wide association study discovers the genomic regions and candidate genes for yield and yield-related traits in bread wheat. *Theor. Appl. Genet.* **134**, 3083–3109.
- Yang, L., Kang, Y., Liu, J., Li, N., Sun, H., Ao, T. and Chen, W. (2022) Foliar spray with rutin improves cadmium remediation efficiency excellently by enhancing antioxidant and phytochelatin detoxification of *Amaranthus hypochondriacus*. *Int. J. Phytoremediation* **24**, 1060–1070.
- Yi, T.G., Yeoung, Y.R., Choi, I.Y. and Park, N.I. (2019) Transcriptome analysis of *Asparagus officinalis* reveals genes involved in the biosynthesis of rutin and protodioscin. *PLoS One* **14**, e0219973.
- Yoo, J., Kim, Y., Yoo, S.-H., Inglett, G.E. and Lee, S. (2012) Reduction of rutin loss in buckwheat noodles and their physicochemical characterisation. *Food Chem.* **132**, 2107–2111.
- Zhang, Z.-L., Zhou, M.-L., Tang, Y., Li, F.-L., Tang, Y.-X., Shao, J.-R., Xue, W.-T. et al. (2012) Bioactive compounds in functional buckwheat food. *Food Res. Int.* **49**, 389–395.
- Zhang, L., Li, X., Ma, B., Gao, Q., Du, H., Han, Y., Li, Y. et al. (2017) The tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. *Mol. Plant* **10**, 1224–1237.
- Zhang, K., Logacheva, M.D., Meng, Y., Hu, J., Wan, D., Li, L., Janovska, D. et al. (2018) Jasmonate-responsive MYB factors spatially repress rutin biosynthesis in *Fagopyrum tataricum*. *J. Exp. Bot.* **69**, 1955–1966.
- Zhang, K., He, M., Fan, Y., Zhao, H., Gao, B., Yang, K., Li, F. et al. (2021a) Resequencing of global Tartary buckwheat accessions reveals multiple

- domestication events and key loci associated with agronomic traits. *Genome Biol.* **22**, 23.
- Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q. et al. (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* **50**, 278–284.
- Zhao, Z., Zhang, Z., Ding, Z., Meng, H., Shen, R., Tang, H., Liu, Y.G. et al. (2020) Public-transcriptome-database-assisted selection and validation of reliable reference genes for qRT-PCR in rice. *Sci. China Life Sci.* **63**, 92–101.
- Zhao, D., Zhang, Y., Lu, Y., Fan, L., Zhang, Z., Chai, M. and Zheng, J. (2021a) Genome sequence and transcriptome of *Sorbus pohuashanensis* provide insights into population evolution and leaf sunburn response. *J. Genet. Genomics* **49**, 547–558.
- Zhao, Y., Chen, W., Cui, Y., Sang, X., Lu, J., Jing, H., Wang, W. et al. (2021b) Detection of candidate genes and development of KASP markers for Verticillium wilt resistance by combining genome-wide association study, QTL-seq and transcriptome sequencing in cotton. *Theor. Appl. Genet.* **134**, 1063–1081.
- Zhao, H., He, Y., Zhang, K., Li, S., Chen, Y., He, M., He, F. et al. (2023) Rewiring of the seed metabolome during Tartary buckwheat domestication. *Plant Biotechnol. J.* **21**, 150–164.
- Zhou, M., Sun, Z., Ding, M., Logacheva, M.D., Kreft, I., Wang, D., Yan, M. et al. (2017) FTSAD2 and FtJAZ1 regulate activity of the FtMYB11 transcription repressor of the phenylpropanoid pathway in *Fagopyrum tataricum*. *New Phytol.* **216**, 814–828.
- Zhu, F., Micha, W.A., Yavir, B., Saleh, A., Agustin, Z. and Alisdair, R.F. (2022) Bringing more players into play: leveraging stress in genome wide association studies. *J. Plant Physiol.* **271**, 153657.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

- Figure S1** Genome-wide association analysis of quercetin content at different altitudes.
- Figure S2** Principal component analysis of Tatar buckwheat varieties, showing the first two components.
- Figure S3** Schematic diagram of the synthesis of flavonoids in Tartary buckwheat.
- Figure S4** GWAS results of rutin and quercetin with 30-fold genomic resequencing data.
- Figure S5** Evolutionary tree analysis of FtGH1 and its homologous proteins.
- Figure S6** GWAS of the ratio of five key metabolites to rutin or quercetin re-identified *FtGH1*.
- Figure S7** Identification and expression assay of buckwheat overexpressing hairy root positive lines of *FtGH1*.
- Figure S8** Determination of isoquercitrin content after overexpression of *FtGH1* in hairy roots and hyperin and quercetin 7-O glucoside content after overexpression of *FtGH1* in tobacco leaves.
- Figure S9** Western blot assay was performed to detect the FtGH1 protein.
- Figure S10** Molecular docking of rutin-related metabolites with FtGH1 protein.
- Figure S11** Identification of positive hairy roots of *FtGH1* mutant sites and corresponding quercetin content and protein expression of *FtGH1* of mutant sites in tobacco leaves.
- Figure S12** Haplotype analysis of different metabolites.
- Figure S13** Investigation of rutin hydrolase activity in 40 buckwheat germplasm resources.
- Figure S14** Overview of the distribution of E-box and G-box-like *cis*-acting elements in the promoter of *FtGH1* in two haplotype materials.

Figure S15 Analysis of the expression patterns of *FtGH1* and MYC transcription factors in the transcriptome of buckwheat seed development.

Figure S16 Schematic diagram of the two haplotypes of FtGH1 promoter used in Y1H for the experimental compartments.

Figure S17 Identification and expression assay of buckwheat overexpressing hairy root positive lines of *FtbHLH165*.

Figure S18 Expression of *FtGH1* in *FtbHLH165* overexpressing hairy roots.

Figure S19 The most significantly enriched KEGG metabolic pathway in buckwheat with *FtbHLH165* overexpressing hairy roots compared to empty vector hairy roots differentially expressing genes.

Figure S20 Synteny analysis of *FtGH1* and *FtbHLH165* with other representative plant species.

Figure S21 Distributional profiles of MYC *cis*-acting elements in the promoters of FtGH1 and its homologous genes in golden buckwheat (*Fagopyrum dibotrys*) and common buckwheat (*Fagopyrum esculentum*).

Figure S22 GWAS results of buckwheat seed circumference with 30-fold whole-genome resequencing.

Figure S23 Identification and expression assay of positive lines overexpressing *FtbHLH165* in *Arabidopsis thaliana*.

Figure S24 Heterologous transient overexpression of *FtbHLH165* and infection of *Rhizoctonia solani* AG4-HGI 3.

Data Set S1 The contents of rutin and quercetin in Tartary buckwheat at different altitudes.

Data Set S2 Candidate genes determined by GWAS analysis of rutin content at 1000 m altitude using 10× resequencing data.

Data Set S3 Candidate genes determined by GWAS analysis of rutin content at 2000 m altitude using 10× resequencing data.

Data Set S4 Candidate genes determined by GWAS analysis of rutin content at 2500 m altitude using 10× resequencing data.

Data Set S5 Candidate genes determined by GWAS analysis of quercetin content at 1000 m altitude using 10× resequencing data.

Data Set S6 Candidate genes determined by GWAS analysis of quercetin content at 1500 m altitude using 10× resequencing data.

Data Set S7 Candidate genes determined by GWAS analysis of quercetin content at 2000 m altitude using 10× resequencing data.

Data Set S8 Candidate genes identified by GWAS for quercetin content at 3000 m environment with 10-fold resequencing data.

Data Set S9 Candidate genes determined by GWAS analysis of quercetin content at 2500 m altitude using 10× resequencing data.

Data Set S10 Details of 30-fold whole gene resequencing of 200 microcore Tartary buckwheat germplasm resources.

Data Set S11 Number of SNPs on Buckwheat chromosome under 10-fold and 30-fold resequencing.

Data Set S12 Number of SNPs counted under 10-fold and 30-fold resequencing in the approximately 120-kb range of the region where *FtGH1* is located on chromosome II.

Data Set S13 Content of rutin-related metabolites in 200 core germplasm of buckwheat.

Data Set S14 Candidate genes identified by GWAS for quercetin content with 30-fold resequencing data.

Data Set S15 Candidate genes identified by GWAS for quercitrin content with 30-fold resequencing data.

Data Set S16 Candidate genes identified by GWAS for isoquercitrin content with 30-fold resequencing data.

Data Set S17 Candidate genes identified by GWAS for quercetin 7-O-glucoside content with 30-fold resequencing data.

Data Set S18 Candidate genes identified by GWAS for isohyperoside content with 30-fold resequencing data.

Data Set S19 Candidate genes identified by GWAS for hyperin content with 30-fold resequencing data.

Data Set S20 Correlation analysis of seven flavonoids in Tartary buckwheat.

Data Set S21 GWAS of the ratios of five key metabolites to rutin or quercetin revealed significant locus information.

Data Set S22 Docking scores of different substrates with FtGH1 protein.

Data Set S23 Distribution of the *FtGH1* promoter in two haplotype materials.

Data Set S24 Analysis of differentially expressed genes in the *FtbHLH165* overexpression hairy root transcriptome.

Data Set S25 One-to-one orthologous relationships between Tartary buckwheat and other plants.

Data Set S26 Distribution of cis-acting elements in the promoter sequences of FtGH1 and its homologous genes in golden buckwheat and common buckwheat.

Data Set S27 Phenotypic data of seed circumference of 200 Tartary buckwheat germplasm resources.

Data Set S28 Candidate genes for seed perimeter identified by GWAS using 30-fold resequencing data.

Data Set S29 Primers used in this study.

Data Set S30 Summary of statistical analyses.