Transcriptome and Metabolite Insights Into Domestication Process of Cultivated Barley in China

Yu Zhou

Huazhong Agricultural University

Guang Lu

Huazhong Agricultural University

Genlou Sun

Saint Mary's University

Daokun Sun

Huazhong Agricultural University

Xifeng Ren (✉ renxifeng@mail.hzau.edu.cn)

Huazhong Agricultural University

Research Article

Keywords: Domestication process, Barley, Gene flow, Selection, Alkaloid, Phenylpropanoid.

Posted Date: July 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-654119/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Plants on January 14th, 2022. See the published version at <https://doi.org/10.3390/plants11020209>.

Abstract

Background

Barley is one of the earliest domesticated crops and regarded as one of the founder of Neolithic transition in the Near East Fertile Crescent. Domestication process of cultivated barley (especially east-Asian cultivated barley) has been under debate because of the controversial origin centers of barley, which caused by widely dispersal of wild barley. What's more, no comprehensive study regarding alteration in metabolism during domestication has been delineated in barley so far.

Results

Transcriptomic and non-targeted metabolic analyses were performed for two wild barley populations (wild barley of Near East Fertile Crescent (Wb-NE), and wild barley of Tibetan Plateau (Wb-T)), and one cultivated barley population (cultivated barley of China (Cb-C)), the results revealed two stages of the domestication process of Cb-C, first from Wb-NE to Wb-T, and then from Wb-T to Cb-C. The Wb-T played an important intermediate role in the domestication from Wb-NE to Cb-C, and had made more genetic contribution than Wb-NE to Cb-C. Meanwhile, we found continuous gene flow, a large number of selective genes and metabolites during domestication. Divergent metabolites of alkaloids and phenylpropanoids were specific targeted in stages from Wb-NE to Wb-T and from Wb-T to Cb-C, respectively. The key missense SNPs in genes *HORVU6Hr1G027650* and *HORVU4Hr1G072150* might be related to the divergence of metabolites of alkaloids and phenylpropanoids during domestication.

Conclusions

Our results revealed that two stages of the domestication process of Cb-C, and distinct sets of metabolites were targeted by selection during the evolution from wild barley of the Near East Fertile Crescent to Tibetan wild barley to cultivated barley of China. Our findings not only provided genetic and metabolic insights into domestication process of barley but also highlighted the power of combining omics data for trait dissection.

Background

Barley is one of the earliest domesticated crops and regarded as one of the founder of Neolithic transition in the Near East Fertile Crescent [1]. Unlike wheat and other founder crops, the widely eastward dispersal of wild barley (*Hordeum vulgare* ssp. *spontaneum*), the progenitor of cultivated barley (*Hordeum vulgare* ssp. *vulgare*), extends the Near East Fertile Crescent into central Asia and even Tibetan Plateau [1–5]. Since the discovery of Tibetan wild barley (*Hordeum vulgare* ssp. *spontaneum* and *Hordeum agriocrithon* Åberg), increasing number of studies have argued that except a recognized commonly center of domestication in the Near East Fertile Crescent, Tibet was one of the centers of barley domestication [6–11]. Nevertheless, some research results didn't support this hypothesis that Tibet was a center of origin or domestication for barley

[12, 13]. What's more, some studies favored that barley has a monophyletic origin in the Fertile Crescent, but against multi-geographic origins for the crop [14–16].

Tibetan wild barleys do not exist as wild populations in the Qinghai Tibet Plateau but occur as weeds only at the edges of fields in the region [9–12]. Meanwhile, a large number of studies suggested that *Hordeum agriocrithon* Åberg has highly heterogeneous genetic structure, and originated from hybridization between six-rowed landraces and *Hordeum vulgare* ssp. *spontaneum* [12, 13, 16–19]. Thus the existence of Tibetan wild barley only provided weak evidence for the hypothesis of Tibet as one of the centers of barley origin or domestication. Tibetan cultivated barley (qingke) was derived from the Fertile Crescent, and was first introduced to southern Tibet most likely via north Pakistan, India, and Nepal between 3,500 and 4,500 years ago [12]. Thus *Hordeum agriocrithon* Åberg was likely imported to Tibet along with cultivated barley or other founder crops from the Near East Fertile Crescent. However, more evidence is needed to shed further light on introduction of *Hordeum vulgare* ssp. *spontaneum*.

During domestication, selection process driven the improvement of adaptation for plants and animals to be more suitable for cultivation or rearing [20]. The untangling domestication process of barley is fundamental to understanding the origins and early diffusion of agrarian culture [21]. So far, the undisputed research results about cultivated barley in east-Asia are: (1) cultivars and landraces of barley in east-Asia exhibited significant divergence in genetic compositions between those in West [12, 14, 21, 22]; (2) Tibetan wild barley showed larger genomic contribution for cultivated barley of east-Asia than other wild barley populations. Nevertheless, these evidences are far from complete in depicting the domesticated landscape of cultivated barley of east-Asia.

In view of the large genome of barley and the wide genetic diversity of wild barley [23–25], RNA-Seq has become a very effective and powerful technology in generating comprehensive transcriptome profiles and in inferring phylogenetic relationship [26]. Compared with wild barley, cultivated barley underwent dramatic morphological transformations during domestication, such as the rachis, seed dormancy, seed size and number, and flowering time [27, 28]. In domestication process of crops, growing evidence showed that morphological changes were frequently accompanied by coordinated metabolite alterations [29–33]. So, to address the knowledge gap about domestication process for cultivated barley of China, We analyzed transcriptomic and metabolic data from three barley populations contained 29 accessions (9 wild barley from the Near East Fertile Crescent, 10 wild barley from Tibetan Plateau and 10 cultivated barley from China) to depict domestication process of cultivated barley of China. We showed that domestication process cultivated barley of China can be divided into two stages, the first is from wild barley of the Near East Fertile Crescent to Tibetan wild barley, and the second is from Tibetan wild barley to cultivated barley of China. Furthermore, we utilized an evolutionary metabolomics method to comprehensively assess metabolite divergence in seedlings of the two domesticated stages. We revealed that distinct sets of metabolites were targeted by selection during the evolution from wild barley of the Near East Fertile Crescent to Tibetan wild barley to cultivated barley of China. Our findings not only provided genetic and metabolic insights into domestication process of barley but also highlighted the power of combining omics data for trait dissection.

Results

One Way ANOVA analysis of eight morphological traits of grain

Morphology of grains and spikes in the three types of barley were shown in Fig. 1B and Fig. 1C, respectively. Both Wb-NE and Wb-T had brittle rachis while Cb-C had non-brittle rachis (Fig. 1C). One way ANOVA analysis of eight morphological traits of grain showed that Wb-NE had the highest grain perimeter (27.56 mm), the highest length-width ratio (4.05), the highest grain length (12.04 mm), the lowest grain width (3.00 mm), the lowest roundness of grain (0.25) and the lowest density factor (0.90 mg/mm²), compared with these traits in other two populations (Table 1). Grain area and grain diameter didn't showed significant difference among the three populations (Table 1).

Table 1
Eight morphological traits of grain in three barley populations.

Populations	Grain area (mm ²)	Grain perimeter (mm)	length-width ratio	Grain length (mm)	Grain width (mm)	Roundness of grain	Density factor (mg/mm ²)	Grain diameter (mm)
Wb-NE	26.20 ^a	27.56 ^a	4.05 ^a	12.04 ^a	3.00 ^b	0.25 ^c	0.90 ^b	5.74 ^a
Wb-T	24.70 ^a	24.66 ^b	3.20 ^b	10.51 ^b	3.30 ^a	0.31 ^b	1.10 ^a	5.57 ^a
Cb-C	22.62 ^a	22.05 ^b	2.78 ^b	9.20 ^c	3.34 ^a	0.35 ^a	1.12 ^a	5.35 ^a
Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China.								

Genetic diversity and evolutionary relationships of the three populations

RNA sequencing of 29 barley accessions generated a total of 185.60 gigabase (Gb) of high-quality cleaned sequences, with an average of 6.19 Gb per accession (Supplementary Data 1) and revealed 1,182,394 raw SNPs. After stringent quality filtering, we obtained 176,741 high-quality SNPs for subsequent analyses.

The results of genetic diversity analysis for the three populations were summarized in Table 2. The highest nucleotide diversity (π = 0.23036), highest Watterson's estimator (θ_W = 0.21255), highest minor allele frequency (MAF = 0.1745) and greatest polymorphism information content (PIC = 0.2176) were observed in Wb-NE, followed by the Wb-T population with π = 0.18906, θ_W = 0.15944, MAF = 0.1453 and PIC = 0.1796 (Table 2). The lowest in those indexes were found in Cb-C (Table 2).

Table 2
Estimate of genetic diversity per base pair.

Populations	π	θ_w	MAF	PIC
Wb-NE	0.23036	0.21255	0.1745	0.2176
Wb-T	0.18906	0.15944	0.1453	0.1796
Cb-C	0.11034	0.10183	0.112	0.1048
π , nucleotide diversity. θ_w , Watterson's estimator. MAF, Minor allele frequency. PIC, Polymorphism information content. Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China.				

An unrooted phylogenetic tree, PCA scatter plot and population structure ($K = 2 \sim 4$) were shown in Fig. 2. The unrooted phylogenetic tree showed an obvious separation of wild barleys (*Hordeum vulgare* ssp. *spontaneum*) and cultivated barleys (*Hordeum vulgare* ssp. *vulgare*) (Fig. 2A). The wild barley accessions were further divided into two clusters, one Wb-NE cluster and one Wb-T cluster (Fig. 2A). Two accessions (HS30 and HS31) of Wb-NE and one accession (Huangqingke) of Cb-C exhibited closer genetic relationships with Wb-T, suggesting that the diffusion process of barley were accompanied by gene flow. PCA scatter plot also displayed that cultivated accessions were clustered together and obviously separated from wild accessions (Fig. 2B). Furthermore, the highest genetic diversity was observed in Wb-NE because of the largest hollow ellipse, followed by Wb-T and Cb-C (Fig. 2B). An optimal population division appeared when the K values was 2 (cv error = 0.82356) (Supplementary Data 2). All accessions were separated into wild accessions and cultivated accessions. The results were consistent with the results of PCA and unrooted Neighbor-joining tree (Fig. 2A-B). When the K values was 3, Wb-T was separated from Wb-NE (Fig. 2C). The emergence of Wb-T might be due to the spread of Wb-NE accompanied by domesticated barley accessions from west to east. Then continuous gene flows took place between the adrift wild barley and cultivated barley during the process of diffusion or after settlement in Tibet.

In order to test evolutionary relationships of the three populations, we constructed rooted phylogenetic trees based 165,848 high-quality SNPs using the bulbous barley (*Hordeum bulbosum*) as an outgroup that was differentiated from barley about 5.80 MYA (million years) ago (<http://www.timetree.org/>). As a result of the bootstrap consensus tree, Wb-NE was located at the basal position on the rooted tree (node 1) (Fig. 3A), and Wb-T was derived from Wb-NE (node 2) (Fig. 3A), supporting that Wb-NE was ancestral to Wb-T. Cb-C was differentiated with Wb-T in node 3, which suggested Wb-T was ancestral to Cb-C (Fig. 3A). Evolutionary relationships revealed by the phylogenetic tree suggested that Wb-T might be evolved from Wb-NE and Cb-C evolved from Wb-T (Fig. 3B). The taxa of Wb-T are mixed with two Wb-NE accessions (HS30 and HS31) and a Cb-C accession (Huangqingke), suggested continuous gene flow accompanied by evolution process of barley.

Search of Identical-by-descent (IBD) blocks and test of gene flow

We tested the gene flow among populations. When K was 3, one Wb-T accession HS18 showed more genetic composition of Cb-C, while one Cb-C accession Huangqingke showed more genetic composition of Wb-T

(Fig. 2C; Fig. 4A). These phenomena suggested ancient admixture in domestication process of barley. In order to gain further insights into the evolutionary relationships among the three populations and assess the admixture state, we searched for identical-by-descent (IBD) blocks. In the paired accessions within populations, as many as 30 of 45 pairs shared IBD blocks within Wb-T were found, followed by Cb-C with 16 of 45 pairs shared IBD blocks (Fig. 4B; Table 3), only 3 of 36 pairs shared IBD blocks within Wb-NE (Fig. 4B; Table 3). In the paired accessions between populations, 19 of 100 pairs shared IBD blocks were found between Wb-T and Cb-C, followed by 9 of 90 pairs shared IBD blocks between Wb-T and Wb-NE (Fig. 4B; Table 3). Only 1 of 90 pairs shared IBD blocks between Cb-C and Wb-NE (Fig. 4B; Table 3). These results suggested that Wb-T accessions were the most admixture due to gene flow within or between populations. Furthermore, the average length of shared IBD block between Cb-C and Wb-T (9.67 Mb) were longer than that between Wb-T and Wb-NE (8.04 Mb), suggesting that gene flow occurred earlier between Wb-NE and Wb-T than between Wb-T and Cb-C (Supplement Data 3).

Table 3
Accession pairs shared IBD blocks within and among populations.

Population	Cb-C	Wb-T	Wb-NE
Cb-C	30 (45)		
Wb-T	19 (100)	30(45)	
Wb-NE	1(90)	9(90)	3(36)
Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China.			
The numbers in brackets denote all accession pairs within or between populations.			

Although the IBD analysis did not reveal conclusive evidence for gene flow (without statistics to participate in judgment), we used the results to guide further analyses. Specifically, we used the presence of between-populations IBD blocks (Fig. 4B) to identify probable cases of gene flow and then further examined these using the ABBA-BABA test [34]. We used HS18, a wild barley of Near East Fertile Crescent at the basal position in the rooted phylogenetic trees (Fig. 2A-B), as the outgroup. Three tree topological models of trios were simulated. They are (HS18,(Wb-NE,(Wb-T,Cb-C))), (HS18,(Wb-T,(Wb-NE,Cb-C))) and (HS18,(Cb-C,(Wb-NE,Wb-T))), respectively. As a result, the most suitable topological model of (HS18,(Wb-NE,(Wb-T,Cb-C))) was inferred (Fig. 4C), which indicated gene flow between Wb-NE and Cb-C because the D statistic (0.167869) was greater than 0 and the Z score (6.0212) was greater than 3 (Fig. 4D). The result suggests that Wb-NE accessions are possible genome donor of Cb-C. We also found gene flow between Wb-T and Cb-C from two other topological models (Fig. 4D), suggesting that Wb-T made genetic contributions to the genome of Cb-C.

Wb-T has made more genetic contribution than Wb-NE to Cb-C

Based on the results of population structure inference (Fig. 2C) and rooted phylogenetic tree (Fig. 3A), we deleted 4 accessions (including Huangqingke, HS30, HS31 and HS108) to minimize the effect of genetic drift between populations. The retained 25 accessions (9 accessions in Cb-C, 9 accessions in Wb-T and 7

accessions in Wb-NE, respectively) represented the three pure populations for further analysis. In order to clarify that Wb-T or Wb-NE has higher genetic contributions to Cb-C, the three pure populations were used to perform genomic similarity analysis. We used 150-kb windows and 75-kb overlapping slide windows to detect the potential similarity of genomic regions, resulting in the best coverage along the reference genome of barley. As a result, total 1155 windows with high similarity detected fell into the selection criteria, which account for ~ 3.58% of barley reference genome. The 1155 windows were showed in Circos diagrams (Fig. 5A-B). The numbers of windows with high similarity was 560 between Wb-T and Cb-C (Fig. 5A-B), and was 281 between Wb-NE and Cb-C (Fig. 5B). More windows (656) with high similarity between Wb-NE and Wb-T were detected (Fig. 5A), indicating a close genetic relationship between the two wild barley population. Compared with the numbers of windows between Wb-NE and Cb-C (281) (Fig. 5B), more windows were detected between Wb-T and Cb-C (560) (Fig. 5A-B), indicating that Wb-T has made more genetic contributions than Wb-NE to Cb-C.

There were only 122 unique genomic regions between Wb-NE and Cb-C (Fig. 5C). In contrast, 383 and 461 unique genomic regions were detected between Wb-T and Cb-C, and between Wb-T and Wb-NE, respectively (Fig. 5C). Meanwhile, unique genomic regions on each chromosome were summarized in Fig. 5D. Compared with the numbers of unique windows between Cb-C and Wb-NE, the numbers of unique windows between Cb-C and Wb-T exhibited overwhelming advantages on chromosome 1H (481.8%), 2H (260.0%), 5H (400.0%) and 7H (485.7%) (Fig. 5D).

Based on the results of unrooted and rooted phylogenetic trees, PCA, population structure, IBD blocks, the ABBA-BABA test, and genomic similarity, we proposed two evolutionary stage of the Cb-C domestication process, first from Wb-NE to Wb-T, and then from Wb-T to Cb-C. Wb-T played an important intermedia role in the domestication from Wb-NE to Cb-C, and has made more genetic contributions than Wb-NE to Cb-C. Furthermore, continuous gene flow in the domestication process has occurred.

Positive selection signals from Wb-NE to Wb-T and from Wb-T to Cb-C

The three pure barley populations described in genomic similarity analysis were used to calculate sliding windowed fixation index (F_{st}) between populations. The average F_{st} values was calculated in the genomic region from 48Mbs to 49Mbs on the 3H chromosome, which contains the two genes (*Btr1* and *Btr2*) regulating brittle rachis. As a result, the lowest average F_{st} value of 0.1115 was observed between Wb-NE and Wb-T, suggesting weaker signals of selective sweeps in the region between the two wild barley populations. By contrary, strong signals of selective sweeps were detected in comparison between Wb-NE and Cb-C ($F_{st} = 0.573$) and between Wb-T and Cb-C ($F_{st} = 0.456$) in the same genomic region (Fig. 6).

Genomic regions with the $F_{st} > 0.5$ threshold values, were used to conduct selective sweep analysis and locate genes influenced by natural selection or artificial selection. As a result, 1090 windows under strong selective sweeps were detected in comparison between Wb-NE and Wb-T and contained 686 high-confidence genes (Supplementary Data 4). More windows (6546) and more high-confidence genes (3529) were detected between Wb-T and Cb-C (Supplementary Data 5). In comparison between Wb-T and Cb-C, distribution of selective genes on chromosome 1H ~ 7H were 479, 504, 615, 629, 506, 253 and 459, respectively, much less

than those between Wb-NE and Wb-T (Table 4). 261 and 1390 high-confidence genes were KEGG annotated in comparison between Wb-NE and Wb-T, and between Wb-T and Cb-C, respectively (Supplementary Data 4; Supplementary Data 5).

Table 4
Distribution of selected genes on each chromosome.

Group	1H	2H	3H	4H	5H	6H	7H	Unknow	Total
Wb-NE vs Wb-T	122	60	64	85	60	18	244	33	686
Wb-T vs Cb-C	479	504	615	629	506	253	459	84	3529
Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China.									

Three populations exhibited significant divergence in their metabolomes

To determine the difference in the metabolomes among the three populations, the seedling metabolomes of three populations including 25 accessions described in genomic similarity analysis were quantified using a non-targeted metabolomics approach (see methods for detail). In total, 8828 non-redundant metabolites were detected (Supplementary Data 6), including 328 structure-annotated metabolites (Supplementary Data 7). These annotated metabolites covered 12 metabolite classes, including 114 amino acids and derivatives, 54 sugar and derivatives, 37 carboxylic acids, 24 lipides, 21 glycosides, 17 nucleic acids and derivative, 15 amines, 11 vitamins, 9 hormones, 6 phenylpropanoides, 3 alkaloid and 17 unclassified metabolites (Fig. 7A; Supplementary Data 7).

Principal component analysis (PCA) and hierarchical cluster were performed based on converted contents of all metabolites. Interestingly, the 25 barley accessions were also separated into three clusters (Fig. 7B). Moreover, Wb-T resided between Wb-NE and Cb-C, indicating that metabolic divergence might have occurred during the evolution from Wb-NE to Wb-T and from Wb-T to Cb-C. Similar results of PCA were obtained by the hierarchical cluster tree (Fig. 7C).

Different metabolic sets were targeted during the evolution from Wb-NE to Wb-T and from Wb-T to Cb-C

We used a Q_{st} - F_{st} comparison strategy (see methods for detail) to identify the specific metabolites that were targeted by selection or by neutral processes in domesticated stages from Wb-NE to Wb-T and from Wb-T to Cb-C, respectively. Q_{st} and F_{st} were used to estimate the divergence for metabolic level and for molecular level between populations, respectively. At the molecular level, the average F_{st} from Wb-NE to Wb-T (0.23733) was much higher than that from Wb-T to Cb-C (0.14329), suggesting that greater genomic divergence appeared in the stage from Wb-T to Cb-C due to natural or artificial selection (Fig. 8A). The SNPs were considered to be selective with $F_{st} > 0.326$ in the stage from Wb-NE to Wb-T, and so did those in the stage from Wb-NE to Wb-T with $F_{st} > 0.347$ (Fig. 8B). Those SNPs that deviated from neutrality were excluded, and two sets of neutral SNPs including 51097 and 96208 SNPs was generated in domesticated stages from Wb-

NE to Wb-T and from Wb-T to Cb-C, respectively (data not shown). At the metabolic level, 99% confidence intervals (CIs) of Q_{st} at each metabolite was compared with F_{st} of neutral SNPs to determine divergent metabolites driven by selection. As a result, of the 8828 non-redundant metabolites (Supplementary Data 6), total 3851 metabolites showed significant differences in metabolite contents due to selection in the two domesticated stages (Fig. 8C; Fig. 8D), including 1850 metabolites from Wb-NE to Wb-T and 2766 metabolites from Wb-T to Cb-C (Fig. 8D; Supplementary Data 8; Supplementary Data 9). Among the 8828 metabolites that showed evidence of selection, only 765 (8.67%) metabolites were common targeted in both the two domesticated stages (Fig. 8C; Fig. 8D), indicating that different sets of metabolites were targeted during the domesticated stages from Wb-NE to Wb-T and from Wb-T to Cb-C.

Among the 328 structure-annotated metabolites (Supplementary Data 7), 132 metabolites (126 classified and 6 unclassified metabolites) showed divergence in the two domesticated stages (Supplementary Data 10; Supplementary Fig. 1), including 55 metabolites between Wb-NE and Wb-T, and 108 metabolites between Wb-T and Cb-C (Supplementary Data 11; Supplementary Data 12). To determine which class of metabolites was involved in specific domesticated stage, we analyzed the distribution of 11 metabolite classes from 126 classified metabolites. As a result, these 11 metabolite classes exhibited three types of evolutionary features (Fig. 8E).

Type I of metabolite class only included alkaloids with 2 metabolites that were specially targeted in evolutionary process from Wb-NE to Wb-T (Fig. 8E). Two alkaloid metabolites, N-methyltyramine (M152T230) and Hordenine (M166T177), exhibited selective status between Wb-NE and Wb-T because their CIs of Q_{st} were above 0.326, while exhibited neutral status between Wb-T and Cb-C (Fig. 8F). Type II of metabolite class was targeted in both two domesticated stages, including lipids, hormones, nucleic and derivate, amino acid and derivate, amines, carboxylic acids and sugar and derivate, with metabolite numbers of 17, 4, 12, 42, 8, 8 and 15, respectively (Fig. 8E). Type III metabolite class were mainly targeted in domesticated stage from Wb-T to Cb-C, including phenylpropanoids, glycosides, and vitamins (Fig. 8E). All of 5 metabolites (3 compounds) in phenylpropanoids class, exhibited selective status between Wb-T to Cb-C because their CIs of Q_{st} were above 0.347 (Fig. 8G). In phenylpropanoids class, only one metabolite (M147T299, 4-Hydroxycinnamic) exhibited selective status in evolutionary stage from Wb-NE to Wb-T, but the other four metabolite showed neutral status in the stage (Fig. 8G).

Identification of key positive selective genes and SNPs that influence the metabolic divergence

In our study, we found that two alkaloid compounds (N-Methyltyramine and Hordenine) exhibited specific divergence in evolutionary stage from Wb-NE to Wb-T (Fig. 8F; Fig. 9A). Further, 5 of 261 positive selective genes (the F_{st} value > 0.5 of the window where the gene is located) in the evolutionary process were in the same regulatory pathway with the 2 divergent compounds (Table 5). The 5 selective genes regulate alcohol dehydrogenase, acylpyruvate hydrolase (FAHD1), aspartate aminotransferase (GOT1), aromatic-L-amino-acid/L-tryptophan decarboxylase (DDC) and 4-hydroxyphenylpyruvate dioxygenase (HPD), respectively. Among the 5 genes, *HORVU1Hr1G010120*, *HORVU3Hr1G073220* and *HORVU6Hr1G027650* were closer to N-Methyltyramine and Hordenine (Fig. 9A).

Table 5

Positive selective genes related to alkaloid compounds in evolutionary stage from Wb-NE to Wb-T

Gene ID	KEGG ID	Function
<i>HORVU1Hr1G010130</i>	K00001	alcohol dehydrogenase
<i>HORVU4Hr1G013370</i>	K01557	FAHD1, acylpyruvate hydrolase
<i>HORVU3Hr1G073220</i>	K14454	GOT1, aspartate aminotransferase
<i>HORVU1Hr1G010120</i>	K01593	DDC, aromatic-L-amino-acid/L-tryptophan decarboxylase
<i>HORVU6Hr1G027650</i>	K00457	HPD, 4-hydroxyphenylpyruvate dioxygenase

Since there are multiple SNPs in gene structure of *HORVU6Hr1G027650* (Fig. 9B), it was used to further analyze the relationship with the two compounds. We found 19 SNPs in the structure of *HORVU6Hr1G027650* (Fig. 9B), of which 13 SNPs were located in the coding region (8 synonymous mutations and 5 missense mutations) (Fig. 9C). In addition, a SNP named *6H:109670128* causes stop lost (Fig. 9C). Five missense mutations in the coding region of *HORVU6Hr1G027650* detected in 16 accessions from Wb-NE and Wb-T produced four haplotypes (Fig. 9C-D). Hap1 to hap4 were detected in Wb-NE, while only hap4 was found in Wb-T (Fig. 9D). Moreover, hap1 contained the most accessions (accession number = 4) of Wb-NE, while hap4 contained all accessions (9 accessions) of Wb-T (Fig. 9D). The “t test” between the 4 Wb-NE accessions with hap1 and 9 Wb-T accessions with hap4, showed that both N-Methyltyramine and Hordenine in Wb-T exhibited significant more content than in Wb-NE (Fig. 9E-F).

We also found that three phenylpropanoids compounds (4-Hydroxycimamic, L-Phenylalanine and Apiin) exhibited specific divergence in domesticated stage from Wb-T to Cb-C (Fig. 8G; Fig. 10A). Further, 41 of 1390 positive selective genes (the F_{st} value > 0.5 of the window where the gene is located) in the domestication stage were related to the 3 divergent compounds (Supplementary Data13). There are four genes that directly regulate these three divergent compounds (especially 4-Hydroxycimamic and L-Phenylalanine). Among the four genes, both *HORVU2Hr1G038110* and *HORVU2Hr1G038120* regulate phenylalanine ammonia-lyase (PAL), *HORVU3Hr1G080830* and *HORVU4Hr1G072150* regulate trans-cinnamate 4-monooxygenase (C4H) and 4-coumarate-CoA ligase (4CL), respectively (Table 6).

Table 6

The key positive selective genes involved in phenylpropanoid metabolites

Gene ID	KEGG ID	Function
<i>HORVU2Hr1G038110</i>	K10775	PAL, phenylalanine ammonia-lyase
<i>HORVU2Hr1G038120</i>	K10775	PAL, phenylalanine ammonia-lyase
<i>HORVU3Hr1G080830</i>	K00487	C4H, CYP73A, trans-cinnamate 4-monooxygenase
<i>HORVU4Hr1G072150</i>	K01904	4CL, 4-coumarate-CoA ligase

Since there were multiple SNPs in gene structure of *HORVU4Hr1G072150* (Fig. 10B), it was used to further analyze the relationship with L-Phenylalanine and 4-Hydroxycimamic. We found 6 SNPs in the structure of *HORVU4Hr1G072150* (Fig. 10B), and all of them were located in the coding region. Among the six SNPs, four SNPs were missense mutations, while two were synonymous mutations (Fig. 10B). The four missense SNPs can be further from 16 haplotypes (Fig. 10C), but only three haplotypes (hap1, hap15 and hap16) were found in our 18 accessions containing Wb-T and Cb-C (Fig. 10D). Hap1 contained the all accessions (accession number = 9) of Cb-C and few accessions (accession number = 3) of Wb-T, while hap15 contained most accessions (accession number = 5) of Wb-T (Fig. 10D). Hap16 only contained 1 accession of Wb-T (Fig. 10D). For the 2 divergent compounds (L-Phenylalanine and 4-Hydroxycimamic), The “t test” between the 9 Cb-C accessions with hap1 and 5 Wb-T accessions with hap15 indicated that both L-Phenylalanine and 4-Hydroxycimamic exhibited more content in Wb-T than in Cb-C (Fig. 10E-F).

Discussion

Origin and domestication process of cultivated barley of China

The domestication of crops and the expansion of agriculture fundamentally reshape the historical process of human beings [36–36]. The Near East Fertile Crescent is the primary habitat of wild barley (*Hordeum vulgare* ssp. *spontaneum*) [37]. However, its isolated populations have spread as far as to North African and European shores of the Mediterranean Basin and East Asia [1]. It has been reported that the majority of genetic diversity of barley cultivars in Europe and America were contributed by wild barley from the Near East Fertile Crescent, while the cultivars in Central Asia and the Far East were contributed genetically by barley wild accessions of Central Asia [21]. So far, the domestication process cultivated barley of China are still under debate [38]. Two presuppositions have been proposed. One suggested the cultivated barley in China was introduced from the Near East Fertile Crescent [15, 39, 40], and another replenished that cultivated barley in China might have locally derived from two-rowed or six-rowed wild barley from Qinghai Tibet Plateau [41, 42]. We used transcriptomic data and metabolic data to infer the origin and domestication process of cultivated barley of China. Among the three populations, Wb-NE has the highest in genetic diversity, followed by Wb-T, and Cb-C (Table 2). The results favored the hypothesis of barley originated from the Near East Fertile Crescent [12, 14]. An unrooted phylogenetic tree, PCA scatter plot and inference of population structure showed a separation of wild barley and cultivated barley (Fig. 2A-C), which was consistent with the results of previous research [12]. Furthermore, the results of PCA and unrooted phylogenetic tree showed that Wb-T was an intermediate subgroup between Wb-NE and Cb-C (Fig. 2A-B). The Wb-NE and Wb-T were divided into one group when K was 2, but divided into 2 groups when K was 3 (Fig. 2C). The rooted phylogenetic trees showed Wb-NE was differentiated from bulbous firstly, then Wb-T was differentiated from Wb-NE, and Cb-C was differentiated from WB-T lastly (Fig. 3A-B). So we suggested that both Wb-T and Cb-C originated from the Near East Fertile Crescent [12]. The highest genomic similarity regions were found in comparison between Wb-NE and Wb-T, followed by between Wb-T and Cb-C, and finally between Wb-NE and Cb-C (Fig. 5A-B). The results showed that Wb-T has made more genetic contributions than Wb-NE to Cb-C, which is

consistent with results of Dai's study [43]. With the above-mentioned results, we proposed the two stages of domestication process of cultivated barley in China, first from Wb-NE to Wb-T, and then from Wb-T to Cb-C. Moreover, the emergence of Wb-T was accompanied by the introduction of domesticated barley accessions from the Near East Fertile Crescent.

Domestication process of barley was accompanied with gene flow

In domestication process of barley, gene flow is not an isolated event, but a continuous process. Gene flow events made domestication process of barley is more complex, because they had probably existed in the early stages of domestication [24]. In our study, we speculated that some highly heterogeneous accessions were arisen by continuous gene flow events during domestication process (Fig. 2A-C; Fig. 3A-B). This conjecture was confirmed by our results of IBD search, ABBA-BABA test and previous results (Fig. 4B-D) [14, 44]. By using ABBA-BABA statistics, we verified that both Wb-NE and Wb-T showed gene flows to Cb-C, respectively (Fig. 4C-D). In IBD analysis, the average length of shared IBD blocks was shorter in comparison between Wb-NE and Wb-T, while that was longer between Wb-T and Cb-C (Supplement Data 3). The results suggested that the gene flow in evolutionary stage from Wb-NE to Wb-T was earlier than domesticated stage from Wb-T to Cb-C. We found that more accessions pairs shared IBD blocks within Wb-T and Cb-C, while only 3 pairs were detected within Wb-NE (Fig. 4B; Table 3). In addition, number of accessions shared IBD block was the minimum in Wb-NE (Fig. 4B; Table 3). The results implied that more gene flow events occurred in the later stage of domestication. Comparing with Wb-T and Cb-C, we found number of accessions in Wb-NE that shared IBD blocks within or between populations was much lesser (Fig. 4B; Table 3). It has been reported that there are large areas of natural habitat regions for wild barley in the Near East Fertile Crescent, which are isolated from barley fields [1, 5]. Moreover, in some micro-ecological environments in the Near East Fertile Crescent, such as Evolution Canyon (EC) and Evolution Plateau (EP) models, due to the ecological or geological-edaphic divergence, there is low-frequency gene flow or no gene flow between the two wild barley or emmer wheat populations separated by only a few hundred meters [45]. Therefore, we suspected that low-frequency gene flow events within Wb-NE might be caused by some divergent micro-ecological environments. Nevertheless, since the wild barley in Tibet, as a weed, grows together with cultivated barley and other crops [5], and doesn't formed independent community. So high frequency of gene flow more likely occur within Wb-T accessions or between Wb-T and Cb-C.

Metabolic divergence provided new insights into the domestication process of cultivated barley of China

The diffusion process of barley from the Near East Fertile Crescent to far beyond the original range of its ancestors is the process of constantly adapting to the new environment [46]. The adaptation to different environments leads to different evolutionary directions of barley, resulting in different ecotypes of wild barley and widely distributed landraces, and finally forming various modern cultivated barley [46–48]. Throughout the domestication process, significant changes have taken place in the morphology of barley, such as seed shattering (Fig. 1B), seed dormancy, and seed size and number [28]. Moreover, morphological changes during domestication were frequently accompanied by coordinated metabolite alterations [29–30]. During the domestication process from cultivated barley of China, we found metabolic divergence among the three

populations (Fig. 7B-C). Moreover, Wb-T resided between Wb-NE and Cb-C (Fig. 7B-C), further suggesting two stages of the domestication process in metabolic level, first from Wb-NE to Wb-T, then from Wb-T to Cb-C. Interestingly, similar results were obtained in a metabolic study of the domestication process from teosinte to temperate maize [33].

Tibetan wild barley was a rich reservoir of many novel alleles for better adapting to severe ecological environment of Tibet, such as high levels of UV-B radiation (about 65 kJ m^{-2} in the summer) [49], low temperatures (average yearly temperature of 7.6°C) [50], and low barometric pressure (about 650 mbar) [51]. Although Wb-T was not originated locally, they significantly differentiated from Wb-NE at genetic and metabolic level (Fig. 6; Fig. 8C-D). The genetic difference between Wb-T and Wb-NE has been reported [43]. However, the metabolic differences between the two populations have not been studied. Two specific alkaloid metabolites (Hordenine and its precursor N-Methyltyramine) were found to be strongly selected in the diffusion stage of wild barley from the Near East Fertile Crescent to Tibet (Fig. 9A). N-Methyltyramine, Hordenine and Gramine are alkaloids which play a key role in allelopathy of barley [52], and the mixture of Hordenine and Gramine has the function of phytotoxin, that impact weeds, insects and pathogens [53]. Lovett and Hoult found that the continuous breeding and domestication process of cultivated barley reduced the synthesis of Gramine, while increased the synthesis of Hordenine [54]. The similar results was found in our study that the content of Hordenine and N-Methyltyramine in Wb-T is higher than that in Wb-NE (Fig. 9E). We speculated that the difference in hordenine and N-methyltyramine might be the result of the long-term adaptation of the two populations to different environments. Our results showed that mutation of a selective gene, *HORVU6Hr1G027650*, which regulating the two metabolites, might be the cause of metabolic difference. However, further functional studies of the gene should be carried out to verify the hypothesis.

In the domestication process of cultivated barley in China, from Wb-T to Cb-C, three specific phenylpropanoids compounds (4-Hydroxycimamic, L-Phenylalanine and Apiin) were found to be strongly selected (Fig. 10A). The metabolites of phenylpropanoids metabolism pathway, such as flavonoids, phenols and lignin, play an important role in plant life activities, and are widely involved in biological and abiotic stress, and help plants resist external diseases and insect pests [55]. A recent study pointed out that the excessive accumulation of phenylpropanoids metabolites not only protect against UV-B damage, but also may help to resist freezing injury of Tibet hulless barley [56]. Our results showed that compared with Cb-C, phenylpropanoid metabolites showed higher contents in Wb-T (Fig. 10E-F). We speculated that the difference in 4-Hydroxycimamic, L-Phenylalanine might be caused by different ecological environment. A selective gene, *HORVU4Hr1G072150* regulates the two metabolites. Our results showed that the missense mutation of the key gene might be the cause of metabolic difference. Phenylpropanoid metabolites including the three compounds have the function on regulating the absorption of UV by plant leaves and preventing plants from being burned [57]. The accumulation of these metabolites provided a guarantee for the survival of wild barley in Tibet under strong UV-B radiation.

Conclusions

Domestication process of cultivated barley in China has been under debate because of the controversial origin centers of barley. Transcriptomic and non-targeted metabolic data on Wb-NE, Wb-T, and Cb-C

populations allowed us to propose two stages of the domestication process of Cb-C, first from Wb-NE to Wb-T, and then from Wb-T to Cb-C. We also found that Wb-T played an important intermediate role in domestication process of Cb-C, and had made more genetic contribution than Wb-NE to Cb-C. Meanwhile, continuous gene flow, a large number of selective genes and metabolites during domestication process were detected by ABBA-BABA analysis, selective sweep analysis and “ $Q_{st} - F_{st}$ ” comparison, respectively. Two evolutionary stages of Cb-C showed different evolutionary characteristics. Divergent metabolites of alkaloids and phenylpropanoids and related genes were specific targeted in stages from Wb-NE to Wb-T and from Wb-T to Cb-C, respectively. Our study provided new insights into the domestication process of Cb-C from genetic and metabolic perspectives. Moreover, it significantly enriches the domestication and evolutionary theories of barley.

Methods

Determination of eight grain morphology traits

The 29 barley accessions used in this study, included 9 wild barleys (*Hordeum vulgare* ssp. *spontaneum*) of the Near East Fertile Crescent (Wb-NE), 10 two-row wild barley (*Hordeum vulgare* ssp. *spontaneum*) of Tibetan Plateau (Wb-T) and 10 cultivated barley (*Hordeum vulgare* ssp. *vulgare*) of China (Wb-C) (Fig. 1A), which were provided by United States Department of Agriculture (USDA) and Huazhong Agricultural University barley germplasm collection (See Supplementary Data 1 for details). Eight grain morphology traits (grain area, grain perimeter, grain length, grain width, length-width ratio, roundness of grain, density factor and grain diameter) were measured by a camera-assisted phenotyping system (SC-G, Wanshen Detection Technology Co., Ltd., Hangzhou, China) [58]. SPSS 24.0 software was performed to one way ANOVA analysis of all measured traits in the three populations.

Plant Growth Conditions

Each barley accession was seeded in plots (3 individuals per plot) for germination with three replications in a controlled greenhouse environment with 16 h of light (at 5000-lux illumination intensity and mixed wavelength ranging from 450 to 660 nm) and 8 h of dark. The third fully unfolded leaves of each accession were collected for RNA sequencing (one replication per accession) and metabolite profiling (three replications per accession).

RNA Sequencing and Single Nucleotide Polymorphisms (SNPs) Calling

Total RNA of each barley accession was extracted from mixed third fully unfolded leaves using TRIzol® reagent following the manufacturer's instructions (Invitrogen). 29 RNA-seq libraries were constructed following the TruSeq™ RNA sample preparation kit (Illumina, San Diego, CA) and paired-end libraries were sequenced by an Illumina X Ten platform (2 × 151-bp read length). The raw paired-end reads (fastq format) were quality controlled by Trimmomatic (0.38 version) with default parameters [59]. Then, index of reference genome of Morex (*Hordeum vulgare* ssp. *vulgare*) was built and clean reads of each accession were separately aligned to the reference genome using hisat2 (2.2.1 version) with default parameters [60]. The

SAM files generated by aligning were sorted by “SortSam” command by SAMtools [61] and further were marked duplicates by picard. The generated BAM file after marking duplicates was used to detect SNPs and short insertions/deletions (Indels) by the SAMtools [61]. The “mpileup” command was used to identify SNPs and Indels with the parameters as “-q 10 -Q 20 -C 50”. After excluding SNP calling errors caused by incorrect mapping and all Indels, only high-quality SNPs (with coverage depth ≥ 3 , mapping quality ≥ 30 and missing ratio = 0) were retained for subsequent analysis.

Genetic Diversity Analysis

Based on 176,741 high-quality SNPs from three populations, the nucleotide diversity (π) and Watterson estimator (θ_w) were calculated by Dnasp5.0 software [62]. The minor allele frequency (MAF) was calculated by combined VCFtools v0.1.16 [63] and Plink (1.90b6.18 version) [64]. The polymorphism information content (PIC) was calculated by followed formula:

$$PIC = 1 - \left(\sum_{i=1}^n p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2 p_i^2 p_j^2$$

p_i and p_j are the frequencies of the i and j allele at a locus within the population, respectively. n is the number of alleles at the corresponding locus.

Phylogenetic Tree, PCA and Population Structure inference based on SNPs of RNA-Seq

To infer the phylogenetic relationship from a genome-wide perspective, an unrooted neighbor-joining tree was constructed based on the 176,741 high-quality SNPs by the software TreeBest (1.9.2 version), with 1000 times bootstraps [65]. A principal component analysis (PCA) was also performed to evaluate the genetic relationship for samples by the software plink (1.90b6.18 version) [64]. The population genetic structure was inferred using the software ADMIXTURE (1.23 version) [66]. A total of 29 accessions were used to estimate the genetic ancestry, specifying a K ranging from 2 to 4.

In order to infer the origin of cultivated barley of China, rooted neighbor-joining trees was constructed based on the sets of 165,848 high-quality SNPs using the software MEGA with Tajima-Nei model and 1000 times bootstraps [67]. *Hordeum bulbosum* was used as outgroup [68]. The genotype of *Hordeum bulbosum* in each corresponding SNP site were identified by the following two steps. First, the reference sequences consisted with 200bp with the SNP site in the 100th bp were mapped to contigs of *Hordeum bulbosum* by the command of “blastn” in blast software packages [68]. Second, the genotype of *Hordeum bulbosum* in each corresponding SNP site was identified by our shell script. In conclusion, we obtained 165,848 SNPs for inference of rooted phylogenetic trees.

Identity-by-descent (IBD) blocks analysis and ABBA-BABA analysis

In order to gain further insight into the relationships between Cb-C and its wild relatives, the Identity-by-descent (IBD) blocks analysis for paired barley accessions were performed by using the software BEAGLE

(4.1 version) with the following parameters: window = 1,000; overlap = 100; ibdtrim = 100; ibdlod = 10 [69]. The IBD blocks analysis were performed based on 165,848 SNPs. We looked forward to discovering IBD blocks between paired accessions, and this was satisfied by a LOD score ≥ 10 . The sum of shared IBD blocks between paired accessions were visualized by the use of pheatmap package in R [70].

The vcf file contains 165,848 high-quality SNPs was employed to perform ABBA-BABA analysis using the software Dsuite (0.4 version) with “Dtrios” command [71]. According to the results of rooted neighbor-joining tree, we selected HS18 (a wild barley accession from the Near East Fertile Crescent) as outgroup.

Genomic Similarity Analysis

Based on the results of Phylogenetic Tree, PCA and population structure inference, 25 barley accessions were divided into 3 pure groups for genomic similarity analysis. The three groups included the wild barley of the Near East Fertile Crescent (Wb-NE, including HS7, HS18, HS23, HS26, HS38, HS39 and HS56), the wild barley of Tibetan Plateau (Wb-T, including HS100, HS101, HS102, HS103, HS104, HS105, HS106, HS107 and HS109) and the cultivated barley of China (Cb-C, all of the cultivated barley excluding Huangqingke).

The set of 114,120 SNPs in the 25 barley accessions was used for genomic similarity analysis. It took four steps to complete the analysis. First, three barley gene pools for the three groups were constructed according to [72]. Briefly, for all sites with two or more variants, the minority variants were treated as errors rather than using the reference base. If the variants in a certain site had the same frequency, they were selected randomly. Second, to maximize the genomic coverage and maintain the high accuracy, we used 150-kb windows and 75-kb overlapping slide windows along the barley genome to study the genomic similarity. The numbers of SNPs were counted for each window, and the windows were removed when the window with ≤ 10 SNPs numbers. Third, the number of identical SNPs in each window was counted between Cb-C and Wb-NE, between Cb-C and Wb-T, and between Wb-NE and Wb-T, respectively. Then, the genomic similarity (number of identical SNPs / total number of SNPs) of each window was calculated. Lastly, the windows with similarity $\geq 95\%$ were retained and were visualized by Circlize package in R [73].

Selective Sweep Analysis and Identification of selective genes

Selective sweep analysis was conducted for two of the three pure groups (including 25 accessions totally) by measuring F_{st} in each 15-kb genomic fragment with a step of 5 kb along all chromosomes, using the 114,120 high-quality SNPs randomly distributed on chromosomes. Genomic regions under selective sweeps were measured by the sliding windowed F_{st} using VCFtools v0.1.16 [63]. The windows with F_{st} values > 0.5 were considered putative selection regions. The selective genes were considered to be in the window of with F_{st} values > 0.5 , and were annotated by KEGG database.

Extraction and detection of untargeted metabolites

Untargeted metabolites were extracted from the 25 accessions, with 3 replicates per accession. Analyses were performed using an UHPLC (1290 Infinity LC, Agilent Technologies) coupled to a quadrupole time-of-flight (AB Sciex TripleTOF 6600). For HILIC separation, samples were analyzed using a 2.1 mm $\times$ 100 mm ACQUIY UPLC BEH 1.7 μ m column (waters, Ireland). In both ESI positive and negative modes, the mobile

phase contained A = 25 mM ammonium acetate and 25 mM ammonium hydroxide in water and B = acetonitrile. The gradient was 85% B for 1 min and was linearly reduced to 65% in 11 min, and then was reduced to 40% in 0.1 min and kept for 4 min, and then increased to 85% in 0.1 min, with a 5 min re-equilibration period employed. The ESI source conditions were set as follows: Ion Source Gas1 (Gas1) as 60, Ion Source Gas2 (Gas2) as 60, curtain gas (CUR) as 30, source temperature: 600°C, IonSpray Voltage Floating (ISVF) $\pm$ 5500 V. In MS only acquisition, the instrument was set to acquire over the m/z range 60-1000 Da, and the accumulation time for TOF MS scan was set at 0.20 s/spectra. In auto MS/MS acquisition, the instrument was set to acquire over the m/z range 25-1000 Da, and the accumulation time for product ion scan was set at 0.05 s/spectra. The product ion scan was acquired using information dependent acquisition (IDA) with high sensitivity mode selected. The collision energy (CE) was fixed at 35 V with $\pm$ 15 eV. Declustering potential (DP) was set as $\pm$ 60 V.

After detection, AB Sciex TripleTOF 6600 was used to collect the primary and secondary spectra. The raw data was converted into mzXML format by Proteowizard, and then peak alignment, retention time correction and peak area extraction were performed by XCMS program. The structure of metabolites was identified by accurate mass number matching (< 25 ppm) and secondary spectrum matching. For each metabolite, the area of mass spectrum peak was calculated and corrected for preliminary quantification. After that, we searched the self-built database (with standard products) for annotating the metabolites.

PCA and hierarchical clustering analysis based on metabolites

After filtering, 8828 non redundant metabolites were retained. Among these non-redundant metabolites, 328 metabolites had annotated structure. All metabolite contents were converted into log2 to enhance data homogeneity for further analysis. PCA and hierarchical cluster analysis were performed using the online tools of metaboanalyst (<https://www.metaboanalyst.ca>) based on all non-redundant metabolites.

Q_{st} - F_{st} Comparison

We used a “ Q_{st} - F_{st} Comparison” strategy to identify the divergent metabolites caused by selection rather than neutral processes [74]. The method was described as follows:

Firstly, the F_{st} of SNP between Wb-NE and Wb-T, and between Wb-T and Cb-C were calculated by using Arlequin3.5.1.3 according to 114,120 SNPs sets of 25 samples, respectively [75]. Before calculating F_{st} , monomorphic SNPs were excluded, the retained SNPs were used to calculate F_{st} according to the method of molecular variance (AMOVA) [30] using the following formula:

$$F_{st} = \sigma_B^2 / \sigma_T^2$$

Where σ_B^2 was the variance between populations, σ_T^2 was the total variance. After calculating, SNPs that deviate from neutral expectations using an F_{st} -based outlier test were excluded [30]. The outliers were identified by permuting individuals across populations (the number of permuting is 20000). The outlier SNPs were discarded, and neutral SNPs were retained and used to calculate neutral F_{st} .

Secondly, Q_{st} between Wb-NE and Wb-T, or between Wb-T and Cb-C was calculated based on metabolites data according to the following formula:

$$Q_{st} = \sigma_B^2 / (\sigma_B^2 + 2\sigma_W^2)$$

where σ_B^2 was the variance between populations of converted metabolites content, σ_W^2 was the variance within populations of converted metabolites content. σ_B^2 and σ_W^2 were calculated using R software package of lme4r with the linear mixed model. The 99% confidence interval (CIs) of Q_{st} was calculated by 1000 resampling [76]. CIs of Q_{st} was compared with the F_{st} value of neutral SNP, and those metabolites with CIs greater than the neutral F_{st} value were considered as the selective or divergent metabolites.

Haplotype detection of key genes and “t test” of related metabolite’s contents

HORVU6Hr1G027650 was a key gene in evolutionary stage from Wb-NE to Wb-T and was detected with 8 synonymous SNPs and 5 missense SNPs. Haplotypes of the gene in the Wb-NE accessions and Wb-T accessions (16 accessions totally) were analyzed based on the genotypes of 5 missense SNP sites. The haplotypes with the majority of Wb-NE accessions, and the haplotypes with the majority of Wb-T accessions were performed “t test” for contents of N-Methyltyramine and Hordenine, respectively. *HORVU4Hr1G072150* was a key gene in domesticated stage from Wb-T to Cb-C and was detected with 2 synonymous SNPs and 4 missense SNPs. Haplotypes of the gene in the Wb-T accessions and Cb-C accessions (16 accessions totally) were analyzed based on the genotypes of 4 missense SNP sites. The haplotypes with the majority of Wb-T accessions, and the haplotypes with the majority of Cb-C accessions were performed “t test” for contents of Phenylalanine and 4-Hydroxycimamic (p-coumaric acid), respectively.

Abbreviations

Wb-NE

wild barley of Near East Fertile Crescent; Wb-T:and wild barley of Tibetan Plateau; Cb-C:cultivated barley of China; π :nucleotide diversity; θ_W :Watterson’s estimator; MAF:minor allele frequency; PIC:polymorphism information content; PCA:Principal Component Analysis; SNP:Single nucleotide polymorphisms; IBD:Identical by descent; F_{st} :Fixation index; KEGG:Kyoto encyclopedia of Genes and Genomes; CIs:the 99% confidence interval of Q_{st} .

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication:

Not applicable

Availability of data and materials:

The raw reads (fastq format) of RNA-seq is being submitted to the NCBI SRA database. We promise that once we finish upload the data, we will provide the permanent link on how to access these data in subsequent versions of the manuscript.

Competing interests:

The authors declare no conflict of interest.

Funding:

This project was supported by China Agriculture Research System of MOF and MARA (CARS-5) for the detection of transcriptome and metabolome.

Authors' contributions:

YZ and GL have contributed equally to this work and share first authorship. XFR conceived this study and designed the experiments. YZ, GL and DKS conducted the experiments and phenotyping measurements. YZ and GL performed the statistical analysis and wrote the manuscript. XFR coordinated the experiments and oversaw the data analysis. XFR and GLS modified the manuscript. All authors had read and approved the final version of the manuscript.

Acknowledgements:

Not applicable

References

1. Harlan JR, Zohary D. Distribution of wild wheats and barley. *Science*. 1966;153:1074–1080. doi:10.1126/science.153.3740.1074.
2. Nevo E. Evolution of wild barley and barley improvement. *Advance in Barley Sciences, Proceedings of 11th International Barley Genetics Symposium*. 2012;1–16.
3. Xu TW. On the origin and phylogeny of cultivated barley with preference to the discovery of Ganze wild two-rowed barley *Hordeum spontaneum* c. Koch. *Acta Genet Sin*. 1975;2:129–137.
4. Xu TW. Origin and evolution of cultivated barley in China. *Acta Genet Sin*. 1982;9:440–446.
5. Zohary D, Hopf M, Weiss E. Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. 4th ed. Oxford: Oxford University Press. 2012.
6. Dai F, Nevo E, Wu DZ, Comadran J, Zhou MX, Qiu L, et al. Tibet is one of the centers of domestication of cultivated barley. *Proc Natl Acad Sci*. 2012;109:16969–16973. doi:10.1073/pnas.1215265109.
7. Ren XF, Nevo E, Sun DF, Sun GL. Tibet as a potential domestication center of cultivated barley of China. *PLoS One*. 2013;8:e62700. doi:10.1371/journal.pone.0062700.

8. Åberg E. *Hordeum agriocrithon* nova sp., a wild six-rowed barley. Ann Agric Coll Swed. 1938;6:159–212.
9. Hsu TW. Origin and phylogeny of cultivated barley with reference to the discovery of Ganze wild two-rowed barley *Hordeum spontaneum* C. koch. Acta Genet Sin. 1975;2:129–137.
10. Ma DQ, Xu TW, Gu MZ, Wu SB, Kang YC. The classification and distribution of wild barley in the Tibet Autonomous Region. Sci Agric Sin. 1987;20:1–6.
11. Ma DQ, Xu TW. The research on classification and origin of cultivated barley in Tibet Autonomous Region. Sci Agric Sin. 1988;21:7–14. doi: 10.14069/j.cnki.32-1769/s.1996.02.003.
12. Zeng XQ, Guo Y, Xu Q, Mascher M, Guo GG, Li SC, et al. Origin and evolution of qingke barley in Tibet. Nat Commun. 2018;9:5433–5445. doi: 10.1038/s41467-018-07920-5.
13. Pourkheirandish M, Kanamori H, Wu J, Sakuma S, Blattner FR, Komatsuda T. Elucidation of the origin of “agriocrithon” based on domestication genes questions the hypothesis that Tibet is one of the centers of barley domestication. Plant J. 2018; 94:525–534. doi: 10.1111/tpj.13876.
14. Civián P, Drosou K, Armisen-Gimenez D, Duchemin W, Brown TA. Episodes of gene flow and selection during the evolutionary history of domesticated barley. BMC Genomics. 2021;22. doi: 10.1186/s12864-021-07511-7.
15. Badr A, Müller K, Schäfer-Pregl R, El Rabey H, Effgen S, Ibrahim H, et al. On the origin and domestication history of barley (*Hordeum vulgare*). Mol Biol Evol. 2000;17:499–510. doi:10.1093/oxfordjournals.molbev.a026330.
16. Hicks T. Genetic diversity of the *DREB1* gene in *Hordeum vulgare* and *Hordeum spontaneum*. (Ph D dissertation). Nova Scotia: Saint Mary's University. 2019.
17. Bonchev G, Dušinský R, Hauptvogel P, Gaplovská-Kyselá K, Švec M. On the diversity and origin of the barley complex agriocrithon inferred by iPBS transposon markers. Genet Resour Crop Evol. 2019;66:1573–1586. doi: 10.1007/s10722-019-00814-5.
18. Konishi T. Genetic diversity in *Hordeum agriocrithon* E. Åberg, six-rowed barley with brittle rachis, from Tibet. Genet Resour Crop Evol. 2001;48:27–34. doi: 10.1023/A:1011254927505.
19. Tanno K, Takeda K. On the origin of six-rowed barley with brittle rachis, agriocrithon [*Hordeum vulgare* ssp. *vulgare* f. *agriocrithon* (Åberg) Bowd.], based on a DNA marker closely linked to the *vrs1* (six-row gene) locus. Theor Appl Genet. 2004;110:145–150. doi:10.1007/s00122-004-1816-4.
20. Badr A, El-Shazly H. Molecular approaches to origin, ancestry and domestication history of crop plants: barley and clover as examples. J Genet Eng Biotechnol. 2012;10:1–12. doi: 10.1016/j.jgeb.2011.08.002.
21. Morrell PL, Clegg MT. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proc Natl Acad Sci. USA. 2007;104:3289–3294. doi:10.1073/pnas.0611377104.
22. Morrell PL, Lundy KE, Clegg MT. Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. Proc Natl Acad Sci USA. 2003;100:10812–10817. doi: 10.1073/pnas.1633708100.
23. Nevo E. Genome evolution of wild cereal diversity and prospects for crop improvement. Plant Genet Resour. 2006;4:36–46. doi: 10.1079/PGR2006108.
24. Hübner S, Hoffken M, Oren E, Haseneyer G, Stein N, Graner A, et al. Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. Mol Ecol.

- 2009;18:1523–1536. doi:10.1111/j.1365-294X.2009.04106.x.
25. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature*. 2017;544: 427–433. doi:10.1038/nature22043.
26. Martin JA, Wang Z. Next-generation transcriptome assembly. *Nat Rev Genet*. 2011;12:671–682. doi:10.1038/nrg3068.
27. Pourkheirandish M, Hensel G, Kilian B, Senthil N, Chen G, Sameri M, et al. Evolution of the Grain Dispersal System in Barley. *Cell*. 2015; 162:527–539. doi:10.1016/j.cell.2015.07.002.
28. Wang YL, Ye H, Liu L, Wu JH, Ru WM, Sun GL. Molecular Insights on the Domestication of Barley (*Hordeum vulgare* L.), *Crit Rev Plant Sci*. 2019;38:280–294. doi:10.1080/07352689.2019.1658856.
29. Shang Y, Ma YS, Zhou Y, Zhang HM, Duan LX, Chen HM, et al. Plant science. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science*. 2014;346:1084–1088. doi:10.1126/science.1259215.
30. Beleggia R, Rau D, Laidò G, Platani C, Nigro F, Fragasso M, et al. Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol Biol Evol*. 2016;33:1740–1753. doi:10.1093/molbev/msw050.
31. Zhou Y, Ma YS, Zeng JG, Duan LX, Xue XF, Wang HS, et al. Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat Plants*. 2016;2:16183. doi:10.1038/nplants.2016.183
32. Zhu G, Wang SC, Wang SC, Huang ZJ, Zhang SB, Liao QG, et al. Rewiring of the fruit metabolome in tomato breeding. *Cell*. 2018;172:249–261. doi:10.1016/j.cell.2017.12.019.
33. Xu GH, Cao JJ, Wang XF, Chen QY, Li Z, Tian F. Evolutionary metabolomics identifies substantial metabolic divergence between maize and its wild ancestor, teosinte. *The Plant Cell*. 2019;31:1990–2009. doi:10.1105/tpc.19.00111.
34. Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. *Mol Biol Evol*. 2011;28:2239–2252. doi:10.1093/molbev/msr048.
35. Purugganan MD, Fuller DQ. The nature of selection during plant domestication. *Nature*. 2009;457:843–848. doi:10.1038/nature07895.
36. Riehl S, Zeidi M, Conard NJ. Emergence of agriculture in the foothills of the Zagros Mountains of Iran. *Science*. 2013;341:65–67. doi:10.1126/science.1236743.
37. Pankin A, Altmüller J, Becker C, von Korff M. Targeted resequencing reveals genomic signatures of barley domestication. *New Phytol*. 2018;218:1247–1259. doi:10.1111/nph.15077.
38. Pankin A, von Korff M. Co-evolution of methods and thoughts in cereal domestication studies: a tale of barley (*Hordeum vulgare*). *Curr Opin Plant Biol*. 2017;36:15–21. doi:10.1016/j.pbi.2016.12.001.
39. Harlan JR. “On the origin of barley: a second look,” in *Barley Genetics II*, eds I. I. Barley Genetics and R. A. Nilan (Pullman, WA: Washington State University Press). 1971;45–49.
40. Harlan JR. “Barley: *Hordeum vulgare*,” in *Evolution of Crop Plants*, ed. N. W. Simmonds (New York, NY: Academic Press). 1976;93–98.

41. Zhang Q, Yang GP, Dai X, and Sun JZ. A comparative analysis of genetic polymorphism in wild and cultivated barley from Tibet using isozyme and ribosomal DNA markers. *Genome*. 1994;37: 631–638. doi: 10.1139/g94-090.
42. Feng ZY, Liu XJ, Zhang YZ, Ling HQ. Genetic diversity analysis of Tibetan wild barley using SSR markers. *Acta Genet. Sin.* 2006;33:917–928. doi:10.1016/S0379-4172(06)60126-1.
43. Dai F, Chen ZH, Wang XL, Li ZF, Jin GL, Wu DZ, et al. Transcriptome profiling reveals mosaic genomic origins of modern cultivated barley. *Proc Natl Acad Sci*. 2014;111:13403–13408. doi:10.1073/pnas.1414335111.
44. Wang YG, Ren XF, Sun DF, Sun GL. Molecular evidence of RNA polymerase II gene reveals the origin of worldwide cultivated barley. *Sci Rep*. 2016;6:36122. doi:10.1038/srep36122.
45. Nevo E. Asian, African and European biota meet at “evolution canyon” Israel: Local tests of global biodiversity and genetic diversity patterns. *Proc R Soc Lond B*. 1995;262:149–155. doi:10.1098/rspb.1995.0189.
46. Jones H, Leigh FJ, Mackay I, Bower MA, Smith LMJ, Charles MP, et al. Population based re-sequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. *Mol Biol Evol*. 2008; 25:2211–2219. doi:10.1093/molbev/msn167.
47. Poets AM, Mohammadi M, Seth K, Wang H, Kono TJY, Fang Z, Muehlbauer GJ, Smith KP, Morrell PL. The effects of both recent and long-term selection and genetic drift are readily evident in North American barley breeding populations. *G3-Genes Genom Genet*. 2016;6: 609–622. doi:10.1534/g3.115.024349.
48. Russell J, Mascher M, Dawson IK, Kyriakidis S, Calixto C, Freund F, et al. Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nat Genet*. 2016;48:1024–1030. doi:10.1038/ng.3612.
49. Norsang G, Kocbach L, Tsoja W, Stamnes JJ, Dahlback A, and Nema P. Ground-based measurements and modeling of solar UV-B radiation in Lhasa. *Tibet Atmos Environ*. 2009; 43: 1498–1502. doi:10.1016/j.atmosenv.2008.11.048.
50. Zhang RB, Yuan YJ, Wei WS, Gou XH, Yu SL, Shang HM, et al. Dendroclimatic reconstruction of autumn-winter mean minimum temperature in the eastern Tibetan Plateau since 1600 AD. *Dendrochronologia*. 2015;33:1–7. <https://doi.org/10.1016/j.dendro.2014.09.001>.
51. Ferl RJ, Schuerger AC, Paul AL, Gurley WB, Corey K, Bucklin R. Plant adaptation to low atmospheric pressures: potential molecular responses. *Life Support Biosph Sci*. 2002;8:93–101.
52. Liu DL, Lovett JV. Biologically active secondary metabolites of barley. I. Developing techniques and assessing allelopathy in barley. *J. Chem. Ecol*. 1993;22:17–2230. doi:10.1007/BF00979659.
53. Corcuera LJ, Argandona VH, Zúiga GE. Allelochemicals in wheat and barley: role in plant-insect interactions. *Allelopathy: Basic and applied aspects*. Springer Netherlands: 1992;119–127. doi: 10.1007/978-94-011-2376-1_9.
54. Lovett JV, Hoult AHC. Allelopathy and Self-Defense in Barley. *Washington: Allelopathy*, 1994;170–183.
55. Liu J, Osbourn A, Ma P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol Plant*. 2015;8:689–70. doi:10.1016/j.molp.2015.03.012.

56. Zeng X, Yuan H, Dong X, Peng M, Jing X, Xu Q, et al. Genome-wide Dissection of Co-selected UV-B Responsive Pathways in the UV-B Adaptation of Qingke. *Mol Plant*. 2019;13:112–127. doi:10.1016/j.molp.2019.10.009.
57. Markham KR, Ryan KG, Bloor SJ, Mitchell KA. An increase in the luteolin: Apigenin ratio in *Marchantia polymorpha* on UV-B enhancement. *Phytochemistry*. 1998;48:791–804. doi:10.1016/S0031-9422(97)00875-3.
58. Yin CB, Li HH, Li SS, Xu LD, Zhao ZG, and Wang JK. Genetic dissection on rice grain shape by the two-dimensional image analysis in one japonica × indica population consisting of recombinant inbred lines. *Theor. Appl. Genet*. 2015;128:1969–1986. doi:10.1007/s00122-015-2560-7.
59. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–2120. doi:10.1093/bioinformatics/btu170.
60. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol*. 2019;37:907–915. doi:10.1038/s41587-019-0201-4.
61. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009;25:1754–1760. doi:10.1093/bioinformatics/btp324.
62. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25:1451–1452. doi:10.1093/bioinformatics/btp187.
63. Danecek P, Auton A, Abecasis GA, Albers C, Banks E, A DePristo M, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–2158. doi:10.1093/bioinformatics/btr330.
64. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D., et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi:10.1086/519795.
65. Vilella AJ, Severin J, Ureta-Vidal J, Heng L, Durbin R, Birney E. Ensembl Compara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res*. 2009;19:327–335. doi:10.1101/gr.073585.107.
66. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–1664. doi: 10.1101/gr.094052.109.
67. Kumar S, Tamura K, Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput Appl Biosci*. 1994;10:189–191. doi:10.1093/bioinformatics/10.2.189.
68. Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D., et al. Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J*. 2013;76:494–505. doi:10.1111/tpj.12294.
69. Browning BL, and Browning SR. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*. 2013;194:459–471. doi:10.1534/genetics.113.150029.
70. Wu YF, Zhang L, Zhen Y, Liu SY. Bioinformatics analysis to screen for critical genes between survived and non-survived patients with sepsis. *Mol Med Rep*. 2018;18:3737–3743. doi:10.3892/mmr.2018.9408.
71. Malinsky M, Matschiner M, Svoldal H. Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour*. 2020;21:584–595. doi:10.1111/1755-0998.13265.

72. Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*. 2010;464:587–591. doi:10.1038/nature08832.
73. Gu ZG, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30:2811–2012. doi:10.1093/bioinformatics/btu393.
74. Leinonen T, McCairns RJS, O'Hara RB, Merilä J. Qst-Fst comparisons: Evolutionary and ecological insights from genomic heterogeneity. *Nat Rev Genet*. 2013;14:179–190. doi:10.1038/nrg3395
75. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10:564–567. doi:10.1111/j.1755-0998.2010.02847.x.
76. Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, et al. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat Genet*. 2014;46:1089–1096. doi:10.1038/ng.3075.

Figures

Figure 1

The geographic distribution of sampled three barley populations and morphology of grains and spikes(the satellite map is provided by software ShuiJingWeiTu).

Figure 2

Unrooted phylogenetic tree, PCA scatter plot and population structure. Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China. (A) Phylogenetic tree was constructed using the neighbor-joining method, and the percentage of trees from 1,000 bootstrap replications in which the associated group clustered together are shown next to the branches. (B) PCA scatter plot was visualized based on the PC1 (the first principal component) and PC2 (the second principal component), blue hollow ellipse, green hollow ellipse and red hollow ellipse denote 95% confidence interval of Wb-NE, Wb-T and Cb-C. (C) Population structure of three barley populations when K was 2, 3, and 4, respectively.

Figure 3

Rooted phylogenetic tree of three barley populations with *Hordeum bulbosum* as an outgroup. Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China. (A) Bootstrap consensus tree based on Tajima-Nei model. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. (B) Original tree based on Tajima-Nei model. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

Figure 4

Genetic relationship between or within barley populations. Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China. (A) Membership coefficient of each accession When K=3. The small red circles, small green circles and small blue circles denote accessions of Cb-C, Wb-T and Wb-NE, respectively. The colors of the pie charts corresponded to the ADMIXTURE results with K = 3 in Fig. 2C. The satellite map is provided by software ShuiJingWeiTu. (B) Heatmap colors represented the total length of IBD blocks for each pairwise comparison. The blue triangle

with dotted line in the heatmap indicated that paired accessions within Cb-C, while green and yellow triangle with dotted line indicated that paired accessions within Wb-T and within Wb-NE, respectively. (C) The optimal dtrios model when HS18 as the outgroup based on the sets of 165,848 high-quality SNPs by software Dsuite (0.4 version). P1, P2 and P3 denoted the population 1, the population 2 and the population 3, respectively. (D) three trios model when HS18 as the outgroup.

Figure 5

Genomic similarity between populations. Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China. (A) Circos diagram with lines shows genomic similarity regions between Wb-T and Wb-NE (blue lines) and between Wb-T and Cb-C (green), respectively.

The similar regions are connected with lines and each line represents one unique window (150 kb) of the genome with highest similarity between two populations. (B) Circos diagram with lines shows genomic similarity regions between Cb-C and Wb-NE (blue lines) and between Wb-T and Cb-C (green), respectively. The similar regions are connected with lines and each line represents one unique window (150 kb) of the genome with highest similarity between two populations. The green line is consistent with green line of Fig. A. (C) Venn diagram exhibits the numbers of genomic similarity regions among three populations. (D) The numbers of unique genomic similarity regions in 7 chromosomes.

Figure 6

Positive selection signals from Wb-NE to Wb-T and from Wb-T to Cb-C. Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China. Genome-wide distribution of genetic differentiation regions identified by F_{st} analysis between Wb-NE and Wb-T, between Wb-T and Cb-C, and between Wb-NE and Cb-C, respectively. Each point indicated a weighted F_{st} value with a 15-kb window. Red dotted lines indicated the thresholds of $F_{st} = 0.5$. The green vertical line represents the

genomic region from 48Mbs to 49Mbs on the 3H chromosome, which contained the two genes (Btr1 and Btr2) regulating brittle rachis.

Figure 7

Metabolome Divergence between Cb-C and its wild relatives. Wb-NE, wild barley of the Near East Fertile Crescent. Wb-T, wild barley of Tibetan Plateau. Cb-C, cultivated barley of China. (A) Classification of metabolites with annotated structures. (B) PCA of the Cb-C, Wb-T and Wb-NE accessions with all metabolites. PC1, the first principal component. PC2, the second principal component. (C) Hierarchical cluster of Cb-C, Wb-T and Wb-NE accessions. The clustering tree was drawn based on Euclidean distance.

Figure 8

Identification of divergent Metabolites in two domesticated stage. Wb-NE vs Wb-T, evolutionary stage from Wb-NE to Wb-T. Wb-T vs Cb-C, domesticated stage from Wb-T to Cb-C. (A) The box plots of Fst in Wb-NE vs Wb-T and in Wb-T vs Cb-C. Boxes represented the inter-quartile ranges. The lines across the boxes, the whiskers and the black circles represented the median values, the 10th and 90th percentiles and outliers, respectively. (B) Density diagrams of Fst in Wb-NE vs Wb-T and in Wb-T vs Cb-C. Fst values were higher than 0.326 (the purple dotted line) and higher than 0.347 (the yellow dotted line) indicated that SNPs were under natural or artificial selection in Wb-NE vs Wb-T and Wb-T vs Cb-C, respectively. (C) Qst distribution associated

with Wb-NE vs Wb-T and Wb-T vs Cb-C. For each metabolite, Qst was calculated separately for Wb-NE vs Wb-T and Wb-T vs Cb-C. The metabolites that exhibited divergence in Wb-NE vs Wb-T, Wb-T vs Cb-C, and both were represented in purple circles, red circles, and orange circles, respectively. (D) Venn diagram showing the number of divergent metabolites in Wb-NE vs Wb-T, Wb-T vs Cb-C and common. (E) Different metabolite classes exhibited distinct divergence patterns. The bar chart showed the number of divergent metabolites in the corresponding evolutionary process. (F) and (G) Illustration of the divergent pattern using alkaloids (F) and phenylpropanoids (G). Bars and dots represented 99% CIs of Qst and observed Qst, respectively. Red bars with above 0.326 indicated corresponding metabolites were selective in Wb-NE vs Wb-T, and blue bars with above 0.347 indicated corresponding metabolites were selective in Wb-T vs Cb-C.

Figure 9

HORVU6Hr1G027650, a key gene regulating 4-hydroxyphenylpyruvate dioxygenase (HPD), and its haplotype distribution. Wb-NE, wild barley from the Near East Fertile Crescent. Wb-T, wild barley from Tibetan Plateau. Cb-C, cultivated barley from China. Wb-NE vs Wb-T, evolutionary stage from Wb-NE to Wb-T. Wb-T vs Cb-C, evolutionary stage from Wb-T to Cb-C. (A) biosynthesis of N-Methyltyramine and Hordenine. HPD, 4-hydroxyphenylpyruvate dioxygenase. GOT, aspartate aminotransferase. DDC, aromatic-L-amino-acid/L-tryptophan decarboxylase. Bars and dots represented 99% CIs of Qst and observed Qst, respectively. (B) Gene structure of HORVU6Hr1G027650, The red regions represent the coding region of the gene. The vertical lines in the gene indicated the positions of SNPs. The ruler under gene indicated the position of gene on 6H chromosome. (C) Haplotype diversity of Wb-NE and Wb-T in HORVU6Hr1G027650. (D) Four detected haplotypes in accessions of Wb-NE and Wb-T. haplotypes were composed by five missense SNPs. (E) and (F) Comparison of contents of Hordenine (E) and N-Methyltyramine (F) between Wb-NE with haplotype 1 (4 accessions) and Wb-T with haplotype 4 (9 accessions).

Figure 10

HORVU4Hr1G072150, a key gene regulating phenylpropanoid metabolites, and its haplotype distribution. Wb-NE, wild barley from the Near East Fertile Crescent. Wb-T, wild barley from Tibetan Plateau. Cb-C, cultivated barley from China. Wb-NE vs Wb-T, evolutionary stage from Wb-NE to Wb-T. Wb-T vs Cb-C, domesticated stage from Wb-T to Cb-C. (A) Phenylpropanoid metabolic pathway. PAL, phenylalanine ammonia-lyase. C4H, trans-cinnamate 4-monooxygenase. 4CL, 4-coumarate-CoA ligase. Bars and dots represent 99% CIs of Qst and observed Qst, respectively. (B) Gene structure of *HORVU4Hr1G072150*, The red regions represented the

coding region of the gene. The vertical lines in the gene indicated the positions of SNPs. The ruler above gene indicated the position of gene on 4H chromosome. (C) 16 haplotypes composed by four missense SNPs in HORVU6Hr1G027650. (D) Three detected haplotypes in accessions of Cb-C and Wb-T. (E) and (F) Comparison of contents of Phenylalanine (E) and 4-Hydroxycimamic (F) between Cb-C with haplotype 1 (9 accessions) and Wb-T with haplotype 15 (5 accessions).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFig.1.jpg](#)
- [SupplementaryData1.xls](#)
- [SupplementaryData2.xls](#)
- [SupplementaryData3.xls](#)
- [SupplementaryData4.xls](#)
- [SupplementaryData5.xls](#)
- [SupplementaryData7.xls](#)
- [SupplementaryData8.xls](#)
- [SupplementaryData9.xls](#)
- [SupplementaryData10.xls](#)
- [SupplementaryData11.xls](#)
- [SupplementaryData13.xls](#)