SUPPLEMENTARY INFORMATION APPENDIX

The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau

Xingquan Zeng[a,b,1], Hai Long[c,1], Zhuo Wang[d,1], Shancen Zhao[d,1], Yawei Tang[a,b,1], Zhiyong Huang[d,1], Yulin Wang[a,b,1], Qijun Xu[a,b], Likai Mao[d], Guangbing Deng[c], Xiaoming Yao[d], Xiangfeng Li[d,e], Lijun Bai[d], Hongjun Yuan[a,b], Zhifen Pan[c], Renjian Liu[a,b], Xin Chen[c], QiMei WangMu[a,b], Ming Chen[d], Lili Yu[d], Junjun Liang[c], DaWa DunZhu[a,b], Yuan Zheng[d], Shuiyang Yu[c], ZhaXi LuoBu[a,b], Xuanmin Guang[d], Jiang Li[d], Cao Deng[d], Wushu Hu[d], Chunhai Chen[d], XiongNu TaBa[a,b], Liyun Gao[a,b], Xiaodan Lv[d], Yuval Ben Abu[f], Xiaodong Fang[d], Eviatar Nevo[g,2], Maoqun Yu[c,2], Jun Wang[h,i,j,2], Nyima Tashi[a,b,2]

[a]Tibet Academy of Agricultural and Animal Husbandry Sciences, Lhasa, Tibet 850002, China; [b]Barley Improvement and Yak Breeding Key Laboratory of Tibet Autonomous Region, Lhasa 850002, China; [c]Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, P. R. China; [d]BGI-Tech, BGI-Shenzhen, Shenzhen 518083, China; [e]College of Life Science, University of Chinese Academy of Sciences, Beijing 100049, China; [f]Projects and Physics Section, Sapir Academic College, D.N. Hof Ashkelon 79165, Israel; [g]Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel; [h]BGI-Shenzhen, Shenzhen 518083, China; [i]Department of Biology, University of Copenhagen, Copenhagen 2200, Denmark; [j]Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21441, Saudi Arabia

[1]These authors contributed equally to this work.

[2]To whom correspondence should be addressed.

Email: Nyima Tashi, nima_zhaxi@sina.com; Jun Wang, wangj@genomics.org.cn; Maoqun Yu, yumaoqun@cib.ac.cn; or Eviatar Nevo, nevo@research.haifa.ac.il.

Supplementary Information

CONTENTS

I. Note

1 Genome sequencing, assembly and quality assessment

1.1 Whole genome shot-gun sequencing using Illumina technology

Genomic DNA was isolated from an individual of *H. vulgare* L. var. *nudum*, also known as "Qingke" in Chinese and "Ne" in Tibetan, and a landrace of Tibetan hulless barley *Lasa Goumang* (**Fig. S1**). For each short-insert size library, about 5 μg of DNA was fragmented, end-repaired, ligated to Illumina paired-end adapters, size selected at 250,500 and 800 base pairs (bp) on agarose gels, PCR amplified, linked to vector to yield short-insert size libraries. For each large-insert size mate-pair library, 20–60 μg of genomic DNA was sheared to the desired insert size using nebulization for 2 kilobases (kb) or HydroShear for 5 kb, 10 kb, 20 kb and 40 kb. DNA fragments were biotin-labeled, size selected, and circularized. Circular DNA molecules were sheared with Adaptive Focused Acoustic (Covaris) to an average size of 500 bp. Biotinylated fragments were purified on magnetic beads (Invitrogen) and used to construct libraries. All libraries were sequenced on the Illumina sequencing platform. Details of the data amount are shown in **Table S1**.

We performed a stringent filtering process on raw reads before the subsequent analysis as follows: (1) remove reads containing more than 2% Ns or with poly-A structure; (2) remove reads containing 40% or more low-quality bases for short-insert size libraries, and 60% for large-insert size libraries; (3) remove adapter-polluted reads; (4) remove reads with overlap between read1 and read2; (5) remove PCR-duplicated reads. The statistics of high-quality sequencing data are shown in **Table S2**.

1.2 De novo assembly of the Tibetan hulless barley genome

The 17-mer analysis using 133 Gb of high-quality sequencing data indicated that the peak frequency of K_depth was about 26. Therefore, the estimated genome size of Tibetan hulless barley was calculated as 4.48 Gb (**Fig. S2**). Additional K-mer analysis (from 19-mer to 31-mer)

showed similar estimated genome size from 4.33 to 4.5 Gb (**Table S3**). A K-mer refers to an artificial oligo nucleotide sequence with length K. A sequencing read with L bases contains (L-K+1) K-mers. The frequency of each K-mer follows a Poisson distribution in a given data set except for a high proportion of low frequency K-mers due to sequencing errors. The genome size G is estimated as G = K_num/K_depth, where K_num is the total number of K-mers, and K_depth is the expected K-mer frequency, which is the peak value of Poisson distribution of K-mer frequencies.

Reads generated by Solexa pipeline have a low proportion of error, while *SOAPdenovo* (1, 2) is sensitive to sequencing error. Clean data of short-insert sizes (<1 kb) need to be corrected; however, there is no need for large-insert size sequencing data. Sequencing error will lead to low frequency K-mers, so we built a K-mer frequency table and set a cutoff for dividing low and high frequency K-mers. We corrected the error bases by changing the site of the error base to the other base in order to form high frequency K-mers.

The WGS assembly was conducted by *SOAPdenovo* (1, 2) with the K-mer parameter set to 67. The genome assembly procedure was as follows: (1) Constructed contig: split the short-insert size corrected reads into K-mers, constructed de Bruijn graph, simplified the graph and solved the K-mer path to get the contigs; (2) Constructed scaffold: aligned both short-insert and large-insert size reads onto the contigs, used paired-end relationships to construct the scaffolds. *SSPACE-V1.1* (3) was used to further construct scaffold by PE relationships; (3) Filled gap: retrieved the read pairs mapped to the local region around gaps then carried out a local assembly to fill the gaps.

The resulting assembly totaled 4.59 Gb with a contig N50 of 14 kb and a scaffold N50 of 190 kb. The distribution of scaffold lengths identified many short scaffolds less than 200 bp. We assumed that the sequences <200 bp were caused by sequencing errors, or were difficult regions to assemble. We discarded these sequences to obtain the final genome assembly, of which total

scaffold length was 3.89 Gb, total contig length was 3.64 Gb, contig N50 size was 18.07 kb, and scaffold N50 size was 242 kb (**Table S4**).

1.3 Evaluation of sequencing depth and GC-depth distribution on the quality of genome assemblies

We calculated the sequencing depth distribution for bases, which shows no more than 2% of the bases have a depth of less than 10 (**Fig. S3a**). Also, we used 10 kb non-overlapping sliding windows and calculated the GC content and average depth among the windows. The filtered reads were aligned onto the assembly genome sequence using *SOAP*. The percentages of bases with different depth frequencies in genomes were calculated (**Fig. S4a**). The GC content distribution of *H. vulgare* is also similar to that of *O. sativa* (**Fig. S4b**).

1.4 Unassembled genome evaluation

We discarded sequences with lengths <200 bp in the assembly results. For the assembled scaffolds in genome sequences with lengths <200 bp, we found that many were repeats and most had a sequencing depth much less than the average sequencing depth, and the peak depth ranged from 20 to 29 (**Fig. S3b**), which suggests that these short assembly sequences with lengths <200 bp are low quality.

We investigated the unmapped reads by aligning the clean data to the assembled genome sequences with *SOAPaligner*. Many of the unmapped reads could be aligned to the genome with megablast and were located in repeat regions of the genome. This may indicate that the unmapped reads exist partly due to the high repeat content in the genome.

1.5 Quality evaluation of the genome assembly

1.5.1 BAC evaluation

The published BAC clone sequences of *H. vulgare* L. cv. Morex were considered reference data. We mapped the assembled genome sequence back to the BAC sequences (*BlastN* -e 1e-5, nucleotide identity >0.97) to check the coverage rate and quality for the assembled genome

sequences. The aligning segments not in synteny blocks were detected and filtered by our in-house pipeline (**Fig. S5 and Table S5**). *H. vulgare* scaffolds covered 90.63–97.80% of the BAC sequences with an average identity of 98.83%. The regions that could not be covered were full of repetitive sequences.

1.5.2 Gene coverage evaluation

The clean reads of transcriptome data from three samples were de novo assembled with Trinity (4) (**Table S6**), and the resulting unigene (EST) sequences became the query file, mapped to the assembled genome sequence with a threshold of e-value <1e-5 and identity >0.99 to check the gene region coverage rate of the assembly (**Table S7**). The number of RNA sequences that can be covered by genome scaffolds, and the number of RNA sequences with >90% (or >50%) in one scaffold were calculated.

1.6 Reconstruct chromosomes based on the barley genetic map

Integrated anchoring data sets generated by The International Barley Genome Sequencing Consortium were obtained via FTP download from: ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/anchoring/.

This released version of the barley genetic/physical map is composed of various anchoring strategies, and the first two strategies (AC1: FPC_PSEUDO_ANCHORED_280512_AC1.FA and AC2: WGS_ANCHORED_280512_AC2.FA) were used as a reference to assign the assembled scaffolds to chromosomes. From these data sets, which created the contig sequences with genetic markers, we can easily assign the assembled scaffold to a particular chromosome according to the sequence homology to the aforementioned anchoring contigs (AC). Sequence alignments of the WGS scaffold against AC data were performed using *BlastN* (minimal sequence hit length of ≥200 bp, e-value $<1.0 \times 10^{-5}$, at least 99% identity to AC data) and the best-scoring match was chosen in cases of multiple matches. We anchored 28,374 scaffolds onto seven chromosomes, with a total size of 3.48 G (about 89.34% of the total assembled genome) (**Tables S8–9**).

2 Genome annotation

2.1 Repeat annotation

We identified repetitive elements using a combination of *Repbase*-based and de novo approaches. We used Tandem Repeats Finder (TRF) (5) to search for tandem repeats (**Tables S10–12**).

2.1.1 De novo identification of repeat sequences

First, de novo prediction programs *RepeatModeler* (6) and *LTR-FINDER* (7) were employed to build the de novo repeat library based on the genome, then contamination and multi-copy genes in the library were removed. *LTR_FINDER* was used to search the whole genome for the characteristic structure of the full-length long terminal repeat retrotransposons (LTR) (its ~18 bp sequence was complementary to the 3' tail of some tRNA), and then our in-house pipeline was used to filter the low-quality and falsely predicted LTRs. Using this library as a database, *RepeatMasker* was run to find and classify the repeats.

2.1.2 Employment of Repbase for repeat identification

The homology-based approach involves alignment with *Repbase* (8) which contains many known repeats, identified from corresponding software such as: *RepeatMasker* and *RepeatProteinMask*, which identify TEs at the DNA and protein level, respectively.

2.1.3 Sequence divergence of TEs

The transposable elements were identified and the sequence divergence was computed by *RepeatMasker* compared to TE libraries. The extent of divergence shows whether the TE repeats were recently produced or anciently produced by transposition. We found that the de novo method could identify more recently-active TE repeats than ancient TEs (**Fig. S6**).

2.2 Gene model prediction and functional annotation

Protein-coding genes were predicted by homolog-based and de novo methods, in combination with the transcriptome sequencing data.

1) De novo prediction

De novo gene prediction was performed on the repeat-masked genome. *AUGUSTUS* (9) and *GENSCAN* (10) were applied.

2) Homolog-based prediction

Homologous proteins of other species (*A. thaliana*, *B. distachyon*, *O. sativa*, *S. bicolor*, and *Z. mays*) were mapped to the *H. vulgare* genome using *TBlastN* (11) with an e-value cutoff of $1e^{-5}$ and a protein similarity cutoff of 50%. The aligned genome sequences and the query proteins were then passed to *GeneWise* (12) to search for accurate spliced gene model structures.

3) RNA-Seq based prediction

Three samples of hulless barley landrace *Lasa Goumang*, namely, G2, S1, and Y2 as well as the full length cDNA of cultivated barleys (HvuFLcDNA_rep and HvuFLcDNA23614), were subjected to transcriptome sequencing for evaluation of genome assembly quality and complement of gene prediction, of which, G2 comprised mixed root tissues from 10-day seedlings and mature plants; Y2 comprised mixed leaf samples and stems from same plants of G2; Y2 were young spikes of about 2 cm and 5 cm (**Table S13**).

RNA-Seq method can solve the problem of alternative splicing at one gene locus (13). Transcriptome reads were aligned against the genome using *TopHat* (14) to identify candidate exon regions and splicing sites. *Cufflinks* (14) was performed to assemble the alignments into transcripts. ORFs were predicted on the transcripts by using an HMM-based training parameter.

4) Integration evidence

Gene model evidence generated from the above three methods were integrated by *GLEAN* (15) to produce a consensus gene-set (**Table S14–15**). The final gene set had similar gene length, CDS length, exon length and intron length distribution compared to other plant species (**Fig. S7; Table S16**). 93.9% (33,928) genes, which could be linked to chromosomes (**Table S17**).

2.3 Gene function annotation

Gene functions were assigned according to the best match of the alignments using *BlastP* to *SwissProt* and *TrEMBL* databases (16). The motifs and domains of genes were determined by *InterProScan* (17) against protein databases such as *ProDom, PRINTS, Pfam, SMART, PANTHER,* and *PROSITE*. Gene Ontology (18) IDs for each gene were obtained from the corresponding *InterPro* entries. All genes were aligned against *KEGG* (19) proteins, and the pathway in which the gene might be involved was derived from the matched genes in *KEGG* (**Table S18**).

3 Comparison with the cultivated barley genome

3.1 Orthologous identification

We aligned the gene set of Tibetan hulless barley and 26,159 high-confidence gene sets of the published Merox genome (20) using *BlastP*. Reciprocal best hits were identified with the threshold of e-value<1e-5 and protein identity >60%. A total of 22,673 orthologous genes were identified (**Fig. S8**).

3.2 Mapping cultivars and wild barleys to the Tibetan hulless barley genome

Ten accessions of wild and cultivated Tibetan hulless barleys as well as six published barleys (20) were selected. The DNA sequencing data were aligned to the reference genome of Tibetan hulless barley. To reduce the complexity of the Tibetan hulless barley genome and to acquire a high-quality mapping result, we only retained the scaffolds and contigs that were anchored to the seven linkage groups and the contigs with lengths >200 bp as the reference genome (**Table S19**).

3.3 Aligning the genome sequences of the Tibetan hulless barley and Morex barley

BlastN was used to align the genome sequences of the Morex barley to the Tibetan hulless barley genome with the e-value cut off at 1e-5. Next, the in-house pipeline was used to further filter the aligning result. The homologous sequences and specific sequences of Tibetan hulless barley and Morex barley could then be detected (**Table S20-21**).

4 Genome evolution

4.1 Identification of orthologous genes

We identified gene families with OrthoMCL (21) methods based on the all-versus-all *BlastP* alignment of protein sequences from 13 species (*H. vulgare, B. distachyon, O. sativa, S. bicolor, Z. mays, A. thaliana, C. papaya, S. italic ,V. vinifera, P. heterocycla, T. urartu, Ae. Tauschii,* and *T. aestivum*) with e-values less than 1e-5. Barley-specific genes were determined by combining genes in barley-specific families and unclustered barley genes (**Fig. S9; Tables S22–23**).

B. distachyon, ftp://ftpmips.helmholtz-muenchen.de/plants/brachypodium/v1.2
O. sativa, IRGSP1.0, http://rapdb.dna.affrc.go.jp/download/irgsp1.html
S. bicolor, ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Sbicolor/annotation/
Z. mays, ZmB73_5a, http://ftp.maizesequence.org/release-5a/working-set/
A. thaliana, ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/
C. papaya, ftp://asgpb.mhpcc.hawaii.edu/papaya/annotation
S. italic, ftp://ftp.genomics.org.cn/pub/Foxtail_millet/gene_annotation/*V. vinifera*,www.genoscope.cns.fr/externe/Download/Projets/Projet_ML/data/12X/annotation/
T. aestivum,
https://urgi.versailles.inra.fr/download/iwgsc/Gene_models/survey_sequence_gene_models_MIPS_v2.2_Jul2014.zip

4.2 Phylogenetic analysis

4.2.1 Species tree construction

We identified 153 single-copy orthologuous genes in the previous step and used them to build a phylogentic tree for species including *H. vulgare, B. distachyon, O. sativa, S. bicolor, Z. mays, A. thaliana, C. papaya, S. italic, V. vinifera, P. heterocycla, T. urartu, Ae. Tauschii,* and *T. aestivum*.

We aligned CDS sequences from each single-copy family guided by *MUSCLE* (22) alignments of protein sequences, and then concatenated the aligned CDS to one super gene for each species. We extracted 4D sites from the CDS sequences and reconstructed the phylogenetic tree using *PhyML* (23, 24) under the GTR+gamma model. We used aLRT values to assess branch reliability (**Fig. S10a**).

4.2.2 Divergence time estimation

We took the same set of 4D sequences to estimate divergence times. Fossil calibration times were set, as presented in **Table S24**. We used the PAML mcmctree program (PAML version 4.5) (25) to determine split times using the approximate likelihood calculation method and the "Correlated molecular clock" and "REV" substitution models. The shape and scale parameters were set according to the substitution rate per time unit computed by PAML basem (25). The alpha parameter was also set, as computed by PAML baseml. The MCMC process of PAML mcmctree was set to sample 200,000 times with the sample frequency set to 2 after a burn-in of 40,000 iterations. The fine-tuned parameter r was set to make the acceptance proportions fall in the interval (20%, 40%). The other parameters were set at the default values.

When the multidivtime program was used to calculate split time, the MCMC chain was run for 40,000 generations as burn-in and approximately 400,000 generations to calculate posterior distributions. Other parameters were set as suggested in the manual. For r8s, the maximum likelihood trees inferred by PhyML (with branch lengths) were used as input to calculate split times in the global molecular clock with default settings. Tracer (v1.5.0) (26) was applied to check convergence, and two independent runs were performed to confirm convergence **(Fig. S10b).**

4.3 Genome synteny and whole genome duplication analysis

Genome synteny between Tibetan hulless barley and other species were analyzed based on syntenic blocks. All-versus-all *BlastP* (e-value less than 1e-5) was used to detect orthologous

genes between species. Then, syntenic blocks were detected using Mcscan (27, 28) (**Table S25; Figs. S11–15**). The 4DTv values (29) (deviation at the 4-fold degenerate third codon position) of the blocks were calculated and revised in the HKY model. 4DTv distribution was used to analyze whether whole genome duplication (WGD) events occurred.

Among the *H. vulgare* L. var. *nudum*, *B. distachyon*, and *O. sativa* genomes, chromosomes 1H, 3H of *H. vulgare*, Bd2 of *B. distachyon* and Os5, Os10, Os1 of *O. sativa* have collinear relationships, while 2H is collinear to Bd5, Os7, Os4; 4H is collinear to Bd1, Os3; 5H is collinear to Bd4, Os8, Os9, Os12; 6H and 7H is collinear to Bd3, Os6, Os11 (**Figs. S11, S14**). Similar collinearity was found between *H. vulgare–S. bicolor* and *H. vulgare–S. italica* genomes (**Figs. S12, S14**). These results are similar to a previous report (30), which provides a more detailed view of genome collinearity among Poaceae species.

Chromosome rearrangement exists extensively in grass genomes and evolved from the reconstructed Poaceae ancestral genome containing five chromosomes, intermediated by the 12 chromosome ancestor (31). Since *O. sativa* retained 12 chromosomes, we used it as the reference to investigate the evolutionary pattern of *H. vulgare* chromosomes. The collinear relationship between *H. vulgare* and *O. sativa* chromosomes is evidence that at least four major nested chromosome fusions (NCF) occurred in *H. vulgare* from the intermediate ancestor. The 1H of *H. vulgare* originated from ancestral chromosomes 10 and 5, while 2H came from ancestral chromosome 7 and 4. 1H and 2H each contain one nested insertion to the centromeric region to form the *H. vulgare* chromosomes. And the two NCFs in 5H of *H. vulgare* originated from ancestral chromosomes 3, 9, and 12. There are also two minor syntenic disruptions presented in 4H and 7H of the *H. vulgare* genome. (**Fig. S15**)

4.4 Gene family expansion and contraction

We used *CAFÉ* (32) software to study gene gain and loss across the phylogenetic tree of 12 species under a random birth and death model. The λ (lambda) parameter, which described both gene birth

(λ) and death (μ = –λ) rate across all branches, was estimated to be 0.00771358 using maximum likelihood. For families with conditional P-values less than the threshold (0.05), we also computed Viterbi P-values for each branch. The expanded and contracted gene family numbers in *H. vulgare* were 2,185 and 5,631, respectively (**Fig. S16; Table S26**). The *H. vulgare* significantly expanded gene families were enriched in the functions of gene regulation and stress resistance (**Table S27**). Cold-related gene and AP2/TF families were expanded in the Tibetan hulless barley lineage (**Table S28**).

4.5 Positively selected gene identification

We calculated pairwise Ka, Ks between *H. vulgare* L. var. *nudum* and *H. vulgare* L. cv. Morex, *B. distachyon, O. sativa, A. tauschii, T. urartu* by KaKs_Calculator (33). The Ka/Ks dot chart was then drawn for orthologous gene pairs of species. From the Ka-versus-Ks figure, the evolutionary process of Tibetan hulless barley diverging from Morex barley had more positively selected genes (Ka/Ks >1) than the other species (**Fig. S17**). KEGG pathway analysis indicated that many of the positively selected genes are involved in pathways related to environmental responses and adaptation (**Table S29**).

5 Genetic diversity in landrace and wild progenitor

5.1 Re-sequencing ten cultivars and wild barleys

We selected ten accessions to investigate diversity in the barleys. Of these lines, five are considered cultivated and five are wild barleys (**Table S30**). Re-sequencing the ten barley genotypes yielded 7.12 G 90-bp paired-end reads, which comprised 641.12 Gb of high-quality clean data, with a sequencing depth of more than 15X for each line. To reduce the complexity of the Tibetan hulless barley genome and to acquire a high-quality mapping result, we only retained the scaffolds and contigs that were anchored to the seven linkage groups and the contigs with lengths >200 bp as the reference genomes.

5.2 Individual SNP and InDel calling

The clean reads (the reads with deleted adapter and low-quality reads removed (quality value ≤5 (E) is ≥50% of the reads) were mapped in the reference genome using BWA software (34). The detailed parameters were as follows: "bwa aln -m 10000 -o 1 -e 10 -i 15 -L -I -t 4 -n 0.04" and "bwa sampe -a 800". Due to the high content of repeats in the Tibetan hulless barley genome, some short reads (90 bp) had multiple hits in the reference genome. We only kept reads with mapping quality ≥20 for further analysis. The alignment results were merged and indexed as BAM files with potential PCR duplicates removed using the *SAMtools* package (34).

We used the Genome Analysis Toolkit (GATK) (35) that uses a realignment algorithm to minimize the number of mismatched bases across all reads for the SNP and InDel calling. The raw SNP or InDel calling results were filtered based on the following criteria: 1) Confidence score of SNP or InDel calling >50, 2) Minimum number of required reads supporting each SNP or InDel ≥3, 3) Distance of SNPs or InDels must be at least 5 bp from each other, and 4) The threshold of SNPs calling was set to 20 for base quality (**Table S31**).

5.3 Population re-sequencing analysis

To get SNPs and InDels in a population with ten accessions, the individual realignment results were used by GATK. The final SNPs or InDels dataset in the population was selected and then filtered under the following requirements: 1) Confidence score of SNP or InDel calling ≥50; 2) Base quality >20; 3) Total depth between 30X and 272X retained, and 4) Reads supporting each individual SNP or InDel should be ≥3 (**Table S32**). R version 2.15.0 and EIGENSOFT version 3.0 (36) were used in the principal component analysis (PCA). FRAPPE version 1.1 (37) was used in the population structure analysis (**Figs. S18–20**).

5.4 Selective sweep analysis

To investigate the genome regions under selective sweeps due to the plateau environment, re-sequencing samples were classified into two different groups: the plateau group (wild and cultivated Tibetan hulless barleys) and the non-plateau group (cultivated barley). We used Tajima's D (test to distinguish between a DNA sequence evolving randomly and one evolving under a non-random process (38)) and F_{st} (measure of genetic differentiation between two groups (39)) in every 50 kb genome window to measure whether some regions were under selective sweep. Genome regions with F_{st} >0.5 (or 5% top F_{st} windows) between the two groups and Tajima's D <–2 within the plateau group were considered under selective sweeps (**Tables S33-35, Fig. S21**).

5.5 Correlation analysis

The graph in **Fig. 3d** shows correlations between environmental stress variables that appear in groups B1-B10 and randomly selected genes from **Table S34** (**Table S36**). The environmental stress variables include salinity, oxygen, (low and high), solar radiation, CO2, drought, temperature (low and high), day length (short and long), and dormancy. This correlation, which was done by Matlab software(40), used algorithms that take each of the genes found in **Table S34** and correlated them to environmental stress variables, which were described above and were found in wheat and barley. If the effect is direct, the correlation value is high; if the correlation is indirect, the value is lower (41, 42).

6 References

1. Li R, *et al.* (2010) The sequence and *de novo* assembly of the giant panda genome. *Nature* 463(7279):311-317.
2. Li R, *et al.* (2010) *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20(2):265-272.
3. Boetzer M, Henkel CV, Jansen HJ, Butler D, & Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578-579.
4. Grabherr MG, *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644-652.
5. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27(2):573-580.
6. Smit, A.F.A. & Hubley, R. (2004), RepeatModeler. http://www.repeatmasker.org.
7. Xu Z & Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35(Web Server issue):W265-268.
8. Jurka J, *et al.* (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1-4):462-467.
9. Stanke M, *et al.* (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 34(Web Server issue):W435-439.
10. Burge C & Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268(1):78-94.
11. Altschul SF, *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402.
12. Birney E, Clamp M, & Durbin R (2004) GeneWise and Genomewise. *Genome Res* 14(5):988-995.
13. Trapnell C, *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511-515.
14. Trapnell C, *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 28(5):511-515.
15. Elsik CG, *et al.* (2007) Creating a honey bee consensus gene set. *Genome Biol* 8(1):R13.
16. Bairoch A & Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28(1):45-48.
17. Zdobnov EM & Apweiler R (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17(9):847-848.
18. Ashburner M, *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1):25-29.
19. Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27-30.
20. Mayer KF, *et al.* (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491(7426):711-716.
21. Li L, Stoeckert CJ, Jr., & Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178-2189.
22. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797.
23. Guindon S & Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696-704.
24. Guindon S, *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307-321.
25. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586-1591.
26. Rambaut, A. & Drummond, A.J. Tracer v1.4. http://beast.bio.ed.ac.uk/Tracer (2007).
27. Tang H, *et al.* (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res* 18(12):1944-1954.
28. Tang H, *et al.* (2008) Synteny and collinearity in plant genomes. *Science* 320(5875):486-488.

29. Huang S, *et al.* (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat Genet* 41(12):1275-1281.
30. Anonymous (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* 463(7282):763-768.
31. Salse J, *et al.* (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell* 20(1):11-24.
32. De Bie T, Cristianini N, Demuth JP, & Hahn MW (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22(10):1269-1271.
33. Zhang Z, *et al.* (2006) KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics, proteomics & bioinformatics* 4(4):259-263.
34. Li H, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
35. McKenna A, *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297-1303.
36. Price AL, *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904-909.
37. Tang H, Peng J, Wang P, & Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* 28(4):289-301.
38. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585-595.
39. Akey JM, Zhang G, Zhang K, Jin L, & Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Research* 12(12):1805-1814.
40. MATLAB 8.0 and Statistics Toolbox 8.1, The MathWorks, Inc., Natick, Massachusetts, United States.
41. Liu Z, Sun F, Braun J, McGovern D, & Piantadosi S (2014) Multilevel Regularized Regression for Simultaneous Taxa Selection and Network Construction with Metagenomic Count Data. *Bioinformatics*.
42. Nasilowski K, Awrejcewicz J, & Lewandowski D (2014) Kinematic analysis of the finger exoskeleton using MATLAB/Simulink. *Acta of Bioengineering and Biomechanics / Wroclaw University of Technology* 16(3):129-134.
43. Matsumoto T, *et al.* (2011) Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. *Plant Physiol* 156(1):20-28.
44. Wikstrom N, Savolainen V, & Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. *Proc Biol Sci* 268(1482):2211-2220.
45. Arakaki M, *et al.* (2011) Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad Sci U S A* 108(20):8379-8384.

II. Tables

Table S1 Statistics of raw sequencing data

Pair-end libraries	Insert size	Average reads length (bp)	Total data (Gb)	Sequence depth (X)	Physical depth (X)
Solexa reads	250 bp	150	688.73	153.73	128.11
	500 bp	100	125.8	28.08	70.2
	800 bp	100	41.95	9.36	37.46
	2 kb	90	131.39	29.33	325.87
	5 kb	90	121.18	27.05	751.36
	10 kb	90	108.86	24.3	1349.95
	20 kb	49	51.34	11.46	2338.74
	40 kb	49	39.27	8.77	3577.81
Total	–	–	1308.51	292.08	8579.5

Table S2 Statistics of high-quality sequencing data

Pair-end libraries	Insert size	Average reads length (bp)	Total data (Gb)	Sequence depth (X)	Physical depth (X)
Solexa reads	250 bp	150	505.54	112.84	94.04
	500 bp	100	110.76	24.72	61.81
	800 bp	100	35.5	7.92	31.7
	2 kb	90	59.05	13.18	146.45
	5 kb	90	42.67	9.52	264.57
	10 kb	90	28.29	6.31	350.82
	20 kb	49	9.2	2.05	419.1
	40 kb	49	6.37	1.42	580.36
Total	–	–	797.38	177.99	1948.84

Table S3 Genome size estimation by K-mer analysis

K-mer	Correct K-mer number	Peak	Genome size (Gb)	Repeat	Hete	Reads coverage
17	1.16E+11	26	4.48	-	-	29.72
19	1.91E+11	44	4.35	81.10%	0.50%	54.92
21	1.87E+11	43	4.34	78.20%	0.41%	55.25
23	1.82E+11	42	4.34	75.59%	0.36%	55.53
25	1.79E+11	41	4.38	82.18%	0.42%	55.24
27	1.75E+11	39	4.50	82.46%	0.31%	53.97
29	1.71E+11	39	4.39	82.43%	0.37%	55.49
31	1.67E+11	38	4.40	82.53%	0.36%	55.62

We estimated the genome size of Tibetan hulless barley of 4.48 Gbp by 17-mer analysis using about 30X of short-insert size sequencing data. We tested the K-mer analysis with Jellyfish software[80] under different K-mers with more sequencing data to check genome size, repeat content (Repeat) and heterozygous rate (Hete) of the genome. Please note that the K-mer analysis usually gets the most accurate genome size estimates at about 30X of sequencing data.

Table S4 Statistics of de novo assembly of Tibetan hulless barley genome

	Contig		Scaffold	
	Size (bp)	Number	Size (bp)	Number
N90	3,248	231,123	**50,201**	17,626
N80	6,677	154,062	93,463	12,075
N70	10,064	109,752	136,739	8,641
N60	13,859	78,957	184,593	6,189
N50	18,067	55,924	**242,004**	4,344
Longest	276,948	–	3,066,885	–
Total Size	3,643,636,907	–	**3,890,743,166**	–
Total Number(≥200 bp)		660,465	–	169,831
Total Number(≥2 kb)		287,049	–	37,441

Table S5 Comparison of assembled scaffolds with five BAC sequences of cultivated barley

BAC_ID	Length	Coverage	%	Identity
HVVMRXALLmA0257K17_c1	154,523	148,919	96.37	98.94
HVVMRXALLmA0104M01_c1	165,391	161,758	97.80	99.52
HVVMRXALLmA0024C01_c1	174,348	158,009	90.63	98.23
HVVMRXALLeA0103D13_c1	149,246	143,214	95.96	98.62
HVVMRXALLeA0088K01_c1	181,550	171,570	94.50	98.85
Average	165,012	156,694	95.05	98.83

Table S6, Statistics of de novo transcriptome assembly result

	Sample	Total number	Total length (nt)	Mean length (nt)	N50
Contig	G2	105,943	36,702,894	346	583
	S1	101,213	35,759,482	353	577
	Y2	99,185	37,787,189	380	685
Unigene	G2	54,625	36,577,502	669	969
	S1	56,989	36,510,457	640	890
	Y2	56,496	40,922,106	724	1,061

Table S7 Assessment of sequence coverage of the genome assembly by homologous search with de novo assembled transcriptome data

Samples*	Dataset	No.	Total length (Mb)	Covered by assembly (%)	With >90% sequence in one scaffold		With >50% sequence in one scaffold	
					No.	Percent (%)	No.	Percent (%)
G2	>0 bp	52,753	35,523,701	90.60	40,390	76.56	45,674	86.58
	>200 bp	52,753	35,523,701	90.60	40,390	76.56	45,674	86.58
	>500 bp	23,735	26,486,011	94.66	17,850	75.21	21,264	89.59
	>1000 bp	10,842	17,330,445	95.70	7,942	73.25	9,795	90.34
S1	>0 bp	55,799	35,795,745	95.60	44,449	79.66	50,563	90.62
	>200 bp	55,799	35,795,745	95.60	44,449	79.66	50,563	90.62
	>500 bp	24,498	26,068,131	96.47	18,255	74.52	22,110	90.25
	>1000 bp	10,225	15,958,209	96.15	7,188	70.30	9,159	89.57
Y2	>0 bp	56,496	40,922,106	91.78	43,650	77.26	49,567	87.74
	>200 bp	56,496	40,922,106	91.78	43,650	77.26	49,567	87.74
	>500 bp	27,537	31,901,061	92.68	20,205	73.37	24,271	88.14
	>1000 bp	13,135	21,604,344	92.71	9,236	70.32	11,591	88.25

* The three samples G2, S1, Y2 came from root, immature spike, leaf and stem, respectively. Transcriptome was assembled to unigenes by Trinity.

Table S8 Statistics for linkage groups construction

Inform	Total length (Gb)	% Scaf	% Chrom	Scaf no.	Gene no.	% of gene no.
Scaffold	3.89	100	-	169,831	39,197	100
Chrom	3.48	89.41	100	28,374	36,566	93.29
With Order*	3.48	89.34	99.93	28,339	-	-

* If the scaffold direction cannot be deduced, it keeps the original order. Therefore, the percentage of scaffold with order may be over-estimated. Chrom, chromosome; Scaf, Scaffold

Table S9 Statistics for each pseudo-chromosomes

Chromosome	Length (Mb)	Scaffold number	% of genome
1H	425.00	3,601	12.21
2H	562.09	4,776	16.15
3H	525.90	4,384	15.11
4H	501.49	3,347	14.4
5H	476.44	4,063	13.69
6H	468.32	3,739	13.45
7H	522.19	4,464	15

Table S10 General statistics of repeats in genome

Type	Repeat size (bp)	Percentage of genome (%)
TRF	150,343,089	3.86
RepeatMasker	2,637,091,676	67.78
RepeatProteinMask	926,152,304	23.80
De novo	3,040,288,155	78.14
Total	3,166,813,964	81.39

Table S11 TEs content in the assembled genome

	RepBase TEs		TE proteins		De novo		Combined TEs	
	Length (bp)	**Percentage of genome (%)**	**Length (bp)**	**Percentage of genome (%)**	**Length (bp)**	**Percentage of genome**	**Length (bp)**	**Percentage of genome (%)**
DNA	245,017,883	6.30	84,574,833	2.17	249,631,774	6.42	306,362,537	7.87
LINE	48,169,246	1.24	53,296,750	1.37	41,652,506	1.07	74,507,684	1.91
SINE	413,423	0.01	0	0.00	0	0.00	413,423	0.01
LTR	2,353,883,601	60.50	788,297,959	20.26	2,340,407,773	60.15	2,656,980,324	68.29
Other	8,703	0.00	0	0.00	0	0.00	8,703	0.00
Unknown	0	0.00	152,001	0.00	432,152,063	11.11	432,304,064	11.11
Total	2,637,091,676	67.78	926,152,304	23.80	3,016,748,833	77.54	3,121,892,234	80.24

Table S12 Comparison of repeat sequence in four genomes

	H. vulgare L. var. *nudum*		*O. sativa*		*A. thaliana*		*V. vinifera*	
	Length (bp)	**Percent in genome**	**Length (bp)**	**Percent in genome**	**Length (bp)**	**Percent in genome**	**Length (bp)**	**Percent in genome**
Class I: Retro element								
LTR Retrotransposon								
Gypsy	1,669,626,336	42.913	67,225,687	18.011	6,466,390	5.404	70,723,343	14.546
Copia	714,744,841	18.370	12,697,237	3.402	1,916,698	1.602	43,604,253	8.968
Others	12,213,821	0.314	2,455,034	0.658	154,863	0.129	5,436,208	1.118
Non- LTR Retrotransposon								
LINE	68,812,934	1.767	5,245,506	1.405	1,893,355	1.582	19,888,282	4.091
SINE	413,724	0.011	1,124,541	0.301	74,595	0.062	14,914	0.003
Total Class I	2,465,811,656	63.376	88,748,005	23.777	10,505,901	8.779	139,667,000	28.726
Class II: DNA transposon								
DNA Transposon Superfamily								
CMC	263,427,335	6.771	14,041,570	3.762	1,773,022	1.482	5,988,350	1.232
hAT	10,139,525	0.261	3,123,836	0.837	722,296	0.604	6,639,038	1.365
MULE	18,963,341	0.487	12,978,576	3.477	4,101,327	3.427	12,913,084	2.656
Harbinger	5,327,839	0.137	474,133	0.127	122,568	0.102	208,946	0.043
Helitron	1,743,443	0.045	3,484,220	0.933	2,239,304	1.871	477,408	0.098
Others	28,965,349	0.744	26,047,931	6.979	849,072	0.710	8,775,290	1.805
Total Class II	328,566,832	8.445	60,150,266	16.115	9,807,589	8.196	35,002,116	7.199
Others	160,704	0.004	420,430	0.113	57,814	0.048	27,564	0.006

The above data are the combined statistics results from *RepeatMasker* and *RepeatProteinMask* based on the assembled genome sequences with >200 bp scaffolds included.

Table S13 Statistics of transcriptome sequencing data in different samples

Genotype	Samples	Tissue	Total raw reads (M)	Total clean reads (M)	Total nucleotides (Mb)	Q20 percentage (%)	GC percentage (%)
Lasa Goumang	G2	root	72.47	64.75	5827.44	97.59	56.46
	S1	immature spike	72.79	65.33	5879.64	97.31	57.84
	Y2	leaf and stem	75.72	65.33	5879.64	97.35	56.16
Himalaya 10	A1	leaf	75.61	67.79	6101.54	96.64	55.83
	B2	leaf	75.11	67.95	6115.13	96.82	54.84
	C3	leaf	75.84	64.22	5780.00	95.22	55.57
	D4	leaf	75.15	66.79	6010.96	96.74	55.45
	E5	leaf	73.23	65.40	5886.12	96.69	56.21
	F6	leaf	75.41	69.19	6227.49	97.53	54.47
	G7	leaf	74.46	68.07	6126.09	97.49	55.49
	H8	leaf	73.41	66.20	5958.16	97.37	56.19
	I9	leaf	69.88	64.24	5781.42	97.37	56.63

Table S14 General statistics of predicted protein-coding genes

Gene set		Number	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
De novo	AUGUSTUS	46,539	2013.78	895.53	3.99	224.36	373.81
	GENSCAN	36,023	7523.52	525.62	3.00	174.99	3492.43
Homolog	*A. thaliana*	33,339	2032.28	824.60	3.30	249.61	524.26
	B. distachyon	77,949	1919.41	835.75	2.60	321.46	677.35
	O. sativa	53,452	1488.42	710.92	2.43	292.04	542.07
	S. bicolor	62,169	1887.34	820.92	2.71	303.26	624.73
	Z. mays	69,099	1049.13	569.56	1.85	307.21	561.56
Full length cDNA (HvuFLcDNA_rep)		17,547	5167.48	1679.07	5.59	300.55	760.55
Full length cDNA (HvuFLcDNA23614)		17,027	4774.78	1660.58	5.58	297.62	680.03
Tibetan hulless barley transcriptome		134,416	1699.23	833.92	2.38	349.73	2251.22
Barley transcriptome		122,407	2334.28	1112.11	2.74	405.18	2601.43
***H. vul* Gene-set**		36,151	4825.52	1080.97	4.41	245.10	1098.02

Full length cDNA: 23,614 full length cDNAs, obtained from Matsumoto et al. (43), were taken as query sequences.
Cultivated barley transcriptome: downloaded from NCBI SRA, ERP001573 and ERP001600.
*: 6,892 additional potential gene models could be predicted from flcDNA and cultivated barley transcriptome.
.

Table S15 Summary of evidence for the GLEAN gene models

	≥20% overlap		≥50% overlap		≥80% overlap	
	No.	**Ratio (%)**	**No.**	**Ratio (%)**	**No.**	**Ratio (%)**
H (single)	30	0.08	154	0.43	292	0.81
H (more)	76	0.21	240	0.66	353	0.98
P (single)	0	0.00	1,490	4.12	5,558	15.37
P (more)	0	0.00	1,605	4.44	2,040	5.64
C (single)	5	0.01	45	0.12	142	0.39
C (more)	3	0.01	16	0.04	154	0.43
HC	1,359	3.76	1,826	5.05	2,336	6.46
PC	3,087	8.54	3,047	8.43	3,333	9.22
PH	9,369	25.92	8,315	23.00	5,836	16.14
PHC	22,222	61.47	19,127	52.91	14,462	40.00

P: *ab initio* prediction; H: homology-based prediction; C: cDNA or transcriptome-based prediction; single: with one gene source; more: with two or more gene sources.

Table S16 Statistics of gene structures of five grass genomes

Gene set	**Number**	**Average gene length (bp)**	**Average CDS length (bp)**	**Average exon per gene**	**Average exon length (bp)**	**Average intron length (bp)**
*H. vulgare**	36,151	4825.52	1080.97	4.41	245.10	1098.02
Ae. tauschii	32,645	2865.48	1241.75	5.06	245.53	400.19
T. urartu	34,862	3204.34	1080.58	4.68	230.86	577.71
B. distachyon	26,413	2862.71	1286.47	5.04	255.18	390.02
O. sativa	35,402	2177.38	998.58	3.8	262.69	420.79
S. bicolor	27,159	2942.1	1261.01	4.85	259.9	436.44
S. italica	38,801	2522.24	1087.33	4.25	256.1	442.1

* *H. vulgare* L. var. *nudum*

Table S17 Statistics of gene number for each pseudo-chromosomes

Chromosome	**Gene Number**
1H	4239
2H	5769
3H	5476
4H	3903
5H	5239
6H	4099
7H	5203

Table S18 Statistics of function annotation

		Number	Percent (%)
Total		36,151	---
Annotated	InterPro	24,255	67.09
	GO	18,494	51.16
	KEGG	17,474	48.34
	Swissprot	22,434	62.06
	TrEMBL	24,821	68.66
Annotated (Total)		29,730	82.24
Unannotated		6,421	17,76

Table S19 Genome coverage depth, coverage rate of Non-Tibetan barleys and Tibetan barleys

Samples	Coverage depth	Covered bases	Coverage rate
Non-Tibetan barleys			
Morex	17.84	3,384,433,849	92.89%
Bowman	16.42	3,459,979,441	94.96%
Barke	11.15	3,093,169,493	84.89%
H. spontaneum	8.50	3,375,440,279	92.64%
HarunaNijo	6.41	3,000,452,695	82.35%
Igri	4.17	3,241,897,402	88.97%
Tibetan barleys			
W1	16.89	3,548,508,189	97.39
W2	14.57	3,429,783,574	94.13
W3	12.85	3,530,158,350	96.89
W4	15.13	3,559,398,267	97.69
W5	16.90	3,489,453,407	95.77
C1	15.32	-	96.44
C2	15.41	3,507,826,619	96.27
C3	16.57	-	96.66
C4	16.54	3,525,201,697	96.75
C5	14.46	3,538,395,436	97.11

Non-Tibetan barleys: downloaded from NCBI SRA, Morex (ERP001435), Bowman (ERP001449), Barke (ERP001450), *H. spontaneum* (ERP001434), ERP001451 (HarunaNijo), Igri (ERP001433). Tibetan barleys: W1, W2, W3, W4, W5, Tibetan wild barley; C1, C2, C3, C4, C5, Tibetan hulless barley.

Table S20 Homologous sequences and specific sequences of Tibetan hulless barley and Morex barley

Subject	Genome size (Mb)	Effective size (Mb)*	Alignment length (Mb)	Percent (%)#	Specific length (Mb)	Percent (%)
Tibetan hulless barley	3,891	3,644	3,356	92.10	288	7.90
Morex (AC1)	2,136	2,113	2,000	94.65	113	5.35

* effective size is the genome size excluding gaps

percent mean the proportion of the length compared to effective size

Table S21 Kegg enrichment of genes are involved in Tibetan hulless barley specific sequences

MapID	MapTitle	Gene number	AdjustedPv
map00941	Flavonoid biosynthesis	42	1.28E-03
map00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	51	3.36E-03
map04075	Plant hormone signal transduction	123	3.36E-03

Table S22 Statistics of orthologous gene numbers in 13 species

Species	Genes number	Genes in families	Unclustered genes	Family number	Unique families	Average genes per family
H. vulgare	36,151	28,303	7,848	18,849	789	1.5
A. thaliana	26,637	22,887	3,750	13,240	805	1.73
C. papaya	25,599	18,080	7,519	12,972	577	1.39
O. sativa	35,402	25,369	10,033	18,682	778	1.36
V. vinifera	25,329	19,012	6,317	12,828	720	1.48
S. bicolor	27,159	24,417	2,742	18,475	95	1.32
Z. mays	75,347	49,552	25,795	22,929	4,965	2.16
B. distachyon	26,413	23,524	2,889	17,794	216	1.32
S. italica	38,801	31,389	7,412	18,941	675	1.66
P. heterocycla	25,719	19,375	6,344	13,323	146	1.45
T. urartu	30,023	24,061	5,962	17,804	263	1.35
Ae. tauschii	32,645	26,421	6,224	19,347	188	1.37
T. aestivum	111,982	68,801	43,181	23,353	2,582	2.95

Unclustered genes refer to species-specific genes; unique families refer to species-specific gene families.

Table S23 Statistics for different types of orthologous genes in 13 species

	H.vul	***O.sat***	***S.bic***	***Z.may***	***B.dis***	***S.ita***	***P.het***	***T.ura***	***Ae.tau***	***T.aes***	***A.tha***	***C.pap***	***V.vin***
1:1:1	153	153	153	153	153	153	153	153	153	153	153	153	153
N:N:N	7,382	7,214	7,449	8,818	7,277	7,572	8,061	7,280	7,621	21,225	7,935	6,184	7,035
Poaceae	1,260	1,144	1,160	1,213	1,142	1,248	1,159	1,439	1,536	2,892	0	0	0
Dicotyledoneae	0	0	0	0	0	0	0	0	0	0	1,034	770	875
Other	16,730	14,857	15,425	20,319	14,372	17,266	9,683	14,610	16,692	35,826	10,720	8,620	8,674
SD	2,778	2,001	230	19,049	580	5,150	319	579	419	8,705	3,045	2,353	2,275
ND	7,848	10,033	2,742	25,795	2,889	7,412	6,344	5,962	6,224	43,181	3,750	7,519	6,317
Total	36,151	35,402	27,159	75,347	26,413	38,801	25,719	30,023	32,645	111,982	26,637	25,599	25,329

1:1:1, single-copy orthologs, N:N:N, multi-copy orthologs, Poaceae, Poaceae-specific orthologs, Dicotyledoneae, Dicotyledoneae-specific orthologs, SD, duplicated species-specific genes, ND, species-specific genes. *H. vul, T.ura, T.aes, Ae. tau, B. dis, P.het, O. sat, S. bic, Z. may, S. ita, C. pap, A. tha,* and *V. vin* represent *H. vulgare, T. urartu , T.aestivum, Ae. tauschii , B. distachyon, P. heterocycla, O. sativa, S. bicolor, Z. mays, S. italic, C. papaya, A. thaliana,* and *V. vinifera*, respectively.

Table S24 Calibration time points used in split time estimation

	Species 1	Species 2	Lower bound (Mya)	Upper bound (Mya)	Reference
(1)	*H. vulgare*	*A. thaliana*	139	156	Wikström et al. (44)
(2)	*H. vulgare*	*O. sativa*	34	-	Arakaki et al. (45)

Table S25 Syntenic blocks between Tibetan hulless barley and other Poaceae genomes

Genomes	No. of syntenic Blocks	No. of gene pairs in blocks	Average no. of gene pairs per block	Mean block length in query genome (Mb)	Mean block length in subject genome (Mb)	Percent in query genome (%)	Percent in subject genome (%)
H. vulgare–B. distachyon	596	6,044	10.14	3.79	0.41	58.03	90.88
H. vulgare–O. sativa	462	4,576	9.9	3.56	0.37	42.24	45.69
H. vulgare–S. bicolor	508	5,045	9.9311	3.84	0.62	50.1	42.45
H. vulgare–S. italica	698	6,592	9.4441	3.99	0.46	71.67	80.58
H. vulgare–T. urartu	486	4,346	8.9424	3.84	2.89	47.95	30.19
H. vulgare–Ae. tauschii	637	6,209	9.7473	3.12	3.55	51.10	54.58
hulless – Morex barley*	454	6,569	14.47	8.46	14.52	--	--

* Hulless barley represents *H. vulgare* L. var. *nudum*, Morex barley represents *H. vulgare* L. cv. Morex.

Table S26 Number of expanded and contracted gene families

Branch	Branch Length	n	Expansions			Contractions			No Change	Avg. Exp.
			Families	Genes	Gene Gain/Family	Families	Genes	Gene Loss/Family		
H.vulgare	17	18,577	2,185	4,816	2.20412	5,631	5,810	1.03179	15,978	-0.0255
T.urartu	10	17,232	1,281	1,638	1.27869	4,610	4,747	1.02972	15,289	-0.07977
Ae.tauschii	10	18,576	1,539	2,077	1.34958	2,979	3,025	1.01544	16,662	-0.02432
Ae.tauschii|T.urartu	6	21,180	510	687	1.34706	2,789	2,854	1.02331	20,496	-0.0556
Ae.tauschii,T.urartu|H.vulgare	31	23,795	983	1,500	1.52594	1,976	1,994	1.00911	22,404	-0.01267
B.distachyon	47	17,692	1,699	2,280	1.34197	8,000	8,117	1.01463	15,664	-0.14976
B.distachyon|Ae.tauschii,T.urartu,H.vulgare	16	25,363	298	365	1.22483	1,019	1,019	1.00000	24,966	-0.01678
P.heterocycla	64	13,196	3,060	3,542	1.15752	13,779	14,207	1.03106	9,444	-0.27363
O.sativa	71	18,597	2,591	3,717	1.43458	10,057	10,147	1.00895	15,434	-0.16497
S.bicolor	24	18,399	1,335	1,760	1.31835	7,304	8,101	1.10912	16,622	-0.16269
Z.mays	24	22,815	8,971	19,808	2.208	2,790	2,940	1.05376	13,500	0.432779
Z.mays|S.bicolor	26	25,261	840	1,298	1.54524	1,729	1,837	1.06246	24,308	-0.01383
S.italica	50	18,888	2,446	5,725	2.34056	8,282	8,360	1.00942	16,149	-0.06761
S.italica|Z.mays,S.bicolor	29	26,877	644	833	1.29348	8,411	8,417	1.00071	26,125	-0.19458
C.papaya	45	12,891	1,388	3,075	2.21542	2,986	3,364	1.12659	10,843	-0.00741

A.thaliana	45	13,227	3,805	6,906	1.81498	2,151	2,320	1.07857	9,261	0.117662
A.thaliana|C.papaya	26	15,217	198	319	1.61111	1,421	1,472	1.03589	14,922	-0.02958
V.vinifera	71	12,791	2,295	4,011	1.74771	4,259	4,477	1.05119	9,987	-0.01196
V.vinifera|A.thaliana,C. papaya	77	16,541	236	347	1.47034	23,568	23,743	1.00743	15,172	-0.60027

Note: *T.aestivum*|*T.urartu* means the branch lead to the most recent common ancestor node of *T.aestivum* and *T.urartu*.

Table S27 Gene gain in *H. vulgare* branch (FDR<0.05)

GO ID	GO description	Type	Number of genes	FDR
GO:0043232	intracellular non-membrane-bounded organelle	CC	112	1.89E-118
GO:0008270	zinc ion binding	MF	131	1.17E-69
GO:0032991	macromolecular complex	CC	113	3.22E-56
GO:0003735	structural constituent of ribosome	MF	58	8.89E-51
GO:0005840	ribosome	CC	58	8.89E-51
GO:0000786	nucleosome	CC	54	7.37E-44
GO:0006259	DNA metabolic process	BP	60	1.41E-43
GO:0006334	nucleosome assembly	BP	54	2.72E-42
GO:0046914	transition metal ion binding	MF	142	1.60E-34
GO:0006412	translation	BP	60	1.86E-33
GO:0043229	intracellular organelle	CC	113	3.48E-32
GO:0046872	metal ion binding	MF	143	4.00E-21
GO:0044424	intracellular part	CC	116	4.17E-19
GO:0090304	nucleic acid metabolic process	BP	115	7.99E-19
GO:0044444	cytoplasmic part	CC	59	1.41E-18
GO:0010467	gene expression	BP	115	2.33E-18
GO:0034645	cellular macromolecule biosynthetic process	BP	116	3.66E-18
GO:0043170	macromolecule metabolic process	BP	220	5.42E-18
GO:0003676	nucleic acid binding	MF	131	1.20E-15
GO:0006952	defense response	BP	33	5.27E-15
GO:0006139	nucleobase-containing compound metabolic process	BP	116	1.16E-14
GO:0005622	intracellular	CC	125	1.23E-14
GO:0044260	cellular macromolecule metabolic process	BP	192	2.14E-14
GO:0009058	biosynthetic process	BP	127	1.70E-12
GO:0005737	cytoplasm	CC	61	2.10E-12
GO:0044249	cellular biosynthetic process	BP	119	1.26E-11
GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	MF	13	2.11E-11
GO:0034641	cellular nitrogen compound metabolic process	BP	117	8.13E-11
GO:0006950	response to stress	BP	50	2.73E-10
GO:0009611	response to wounding	BP	14	3.14E-09
GO:0008234	cysteine-type peptidase activity	MF	20	6.11E-09
GO:0003677	DNA binding	MF	77	8.79E-07
GO:0016844	strictosidine synthase activity	MF	8	1.02E-06
GO:0044238	primary metabolic process	BP	225	1.18E-06
GO:0043531	ADP binding	MF	23	2.96E-06
GO:0004867	serine-type endopeptidase inhibitor activity	MF	14	5.33E-06
GO:0044237	cellular metabolic process	BP	196	1.62E-05
GO:2000112	regulation of cellular macromolecule biosynthetic process	BP	57	2.12E-05
GO:0003700	sequence-specific DNA binding transcription factor activity	MF	28	2.20E-05
GO:0010468	regulation of gene expression	BP	57	2.62E-05
GO:0060255	regulation of macromolecule metabolic process	BP	58	3.36E-05
GO:0006355	regulation of transcription, DNA-dependent	BP	55	4.07E-05
GO:0080090	regulation of primary metabolic process	BP	58	4.31E-05

GO:0031323	regulation of cellular metabolic process	BP	58	4.62E-05
GO:0004185	serine-type carboxypeptidase activity	MF	9	1.66E-04
GO:0004527	exonuclease activity	MF	8	3.19E-04
GO:0005634	nucleus	CC	37	1.05E-03
GO:0070011	peptidase activity, acting on L-amino acid peptides	MF	29	2.22E-03
GO:0004612	phosphoenolpyruvate carboxykinase (ATP) activity	MF	2	6.52E-03
GO:0019538	protein metabolic process	BP	105	6.80E-03
GO:0004518	nuclease activity	MF	10	1.26E-02
GO:0006094	gluconeogenesis	BP	2	2.03E-02
GO:0006508	proteolysis	BP	29	2.18E-02

BP: Biological Process; CC: Cellular Component; MF: Molecular Function

Total of 104 families, comprising 1,945 genes, were significantly (P<0.05) expanded in Tibetan hulless barley lineage.

Table S28 Number of gene members in cold-related and AP2 TF families

	AP2/ERF					Total	ratio ERF/DREB	Cold-related
	Soloist	RAV	AP2	ERF	DREB			
Tibetan Hulless barley	1	5	21	95	37	159	2.57	230
Morex	1	7	14	64	35	121	1.83	219
Ae. tauschii	1	1	24	48	35	109	1.37	216
T. urartu	1	2	20	36	24	83	1.5	182
B. distachyon	1	4	25	65	58	153	1.12	164
O. sativa	1	5	29	87	58	180	1.5	132
Z. mays	1	3	31	96	90	221	1.07	148
A. thaliana	1	6	17	65	56	145	1.16	209

Table S29 Kegg pathways of genes with Ka/Ks >1 in hulless Morex pairwise kaks calculation

Pathway	Gene number	Pathway ID
Plant-pathogen interaction	31	ko04626
Plant hormone signal transduction	10	ko04075
Purine metabolism	8	ko00230
Aminoacyl-tRNA biosynthesis	6	ko00970
Phosphatidylinositol signaling system	6	ko04070
Pyrimidine metabolism	5	ko00240
Glutathione metabolism	5	ko00480
Metabolism of xenobiotics by cytochrome P450	5	ko00980
Drug metabolism - cytochrome P450	5	ko00982
Oocyte meiosis	5	ko04114
Peroxisome	5	ko04146
Wnt signaling pathway	5	ko04310
Circadian rhythm - plant	5	ko04712
Insulin signaling pathway	5	ko04910
Alzheimer's disease	5	ko05010

Table S30 Information of barley accessions collected on Tibetan Plateau re-sequenced in this study

Group	ID	Accessions	Growth habit	Species
Wild	W1	ZYM00963	spring	*H. vulgare* ssp. *agriocrithon*
	W2	ZYM01262	spring	*H. vulgare* ssp. *spontaneum*
	W3	ZYM01288	spring	*H. vulgare* ssp. *agriocrithon*
	W4	ZYM01375	spring	*H. vulgare* ssp. *agriocrithon*
	W5	ZYM03251	spring	*H. vulgare* ssp. *spontaneum* -
Cultivated	C1	Linzhiheiliuleng	spring	*H. vulgare* L. var. *nudum*
	C2	Z0237	spring	*H. vulgare* L. var. *nudum*
	C3	Z0414	spring	*H. vulgare* L. var. *nudum*
	C4	Z0699	spring	*H. vulgare* L. var. *nudum*
	C5	Ailibai	spring	*H. vulgare* L. var. *nudum*

Table S31 Statistics of SNPs and InDels in each of the re-sequenced genotypes

Genotypes	No. of SNP	Homo	1H–7H (SNP)	Scaffold (SNP)	Percent in 1H–7H (%) (SNP)	No. of InDel	1H–7H (InDel)	Scaffold (InDel)	1H–7H rate (%) (InDel)
W1	9,206,608	6,282,956	8,328,963	877,645	90.47	424,148	384,395	39,753	90.63
W2	15,759,191	15,217,207	14,282,306	1,476,885	90.63	730,330	666,397	63,933	91.25
W3	12,396,204	8,161,457	11,272,379	1,123,825	90.93	458,848	418,369	40,479	91.18
W4	12,560,313	7,041,767	11,431,772	1,128,541	91.02	500,685	456,863	43,822	91.25
W5	9,268,837	8,804,341	8,356,983	911,854	90.16	467,124	422,796	44,328	90.51
C1	5,436,423	5,074,957	4,810,232	626,191	88.48	270,970	241,937	29,033	89.29
C2	6,827,268	6,447,694	6,096,296	730,972	89.29	330,889	296,494	34,395	89.61
C3	6,614,414	6,194,737	5,914,618	699,796	89.42	330,410	295,477	34,933	89.43
C4	6,465,945	6,067,573	5,776,439	689,506	89.34	327,828	294,597	33,231	89.86
C5	5,692,681	5,332,031	5,072,079	620,602	89.10	262,113	234,184	27,929	89.34
Mean value	**9,022,788**	**7,462,472**	**8,134,207**	**888,582**	**90.15**	**410,335**	**371,151**	**39,184**	**90.23**

Homo: No. of homozygous SNPs.

Table S32 Statistics of population SNPs number in wild & cultivated barleys

	Total No.	1H	2H	3H	4H	5H	6H	7H	Scaffold
SNP (all)	36,469,491	3,674,569	5,224,386	5,721,956	3,531,350	4,955,351	4,818,203	4,799,247	3,744,429
Indel (all)	2,281,198	229,319	332,270	353,253	220,394	314,423	286,077	308,706	236,756
SNP (wild)	34,064,490	3,482,017	4,517,625	5,403,444	3,381,968	4,717,311	4,535,832	4,511,621	3,514,672
SNP (Cultivar)	17,872,045	1,706,407	2,640,130	3,145,674	1,184,379	2,410,753	2,667,631	1,944,525	2,172,546

Table S33 Statisics of selective sweeps of Tibetan hulless barleys population

Chr	Total size	Sweep size	%	Gene number
1H	425,002,292	3,850,000	0.91	35
2H	562,091,317	4,600,000	0.82	56
3H	525,897,205	2,750,000	0.52	19
4H	501,487,783	4,275,000	0.85	40
5H	476,442,955	4,225,000	0.89	73
6H	468,321,367	950,000	0.20	16
7H	522,188,531	16,625,000	3.18	179
total	3,481,431,450	37,275,000	1.07	418

Table S34 Genes under selective sweeps of plateau environment

Gene_id	Symbol	Description
Hvulgare_10000619	ECH2	Enoyl-CoA hydratase 2, peroxisomal
Hvulgare_10000640	NA	NA
Hvulgare_10000792	ROC8	Homeobox-leucine zipper protein ROC8
Hvulgare_10000796	NA	NA
Hvulgare_10000803	NA	NA
Hvulgare_10001098	NA	NA
Hvulgare_10001099	GSVIVT00023967001	Peroxidase 4
Hvulgare_10001638	NA	NA
Hvulgare_10001639	Os06g0677700	Zinc finger CCCH domain-containing protein 45
Hvulgare_10001642	BHLH130	Transcription factor bHLH130
Hvulgare_10001786	NA	NA
Hvulgare_10002292	NA	NA
Hvulgare_10002647	NA	NA
Hvulgare_10002648	captC	Uncharacterized CDP-alcohol phosphatidyltransferase class-I family protein 3
Hvulgare_10002649	At2g17670	Pentatricopeptide repeat-containing protein At2g17670
Hvulgare_10003121	DRM2	DNA (cytosine-5)-methyltransferase DRM2
Hvulgare_10003199	NA	NA
Hvulgare_10003247	CML19	Putative calcium-binding protein CML19
Hvulgare_10004106	At1g68400	Probable leucine-rich repeat receptor-like protein kinase At1g68400
Hvulgare_10004107	NA	NA
Hvulgare_10004419	Serp2	Stress-associated endoplasmic reticulum protein 2
Hvulgare_10004568	CYP79A1	Tyrosine N-monooxygenase
Hvulgare_10004767	NA	Pyruvate kinase, cytosolic isozyme
Hvulgare_10004834	NA	NA
Hvulgare_10005219	NA	NA
Hvulgare_10005220	NA	NA
Hvulgare_10005221	NA	NA
Hvulgare_10005421	MRE11	Double-strand break repair protein MRE11
Hvulgare_10005422	LSM7	U6 snRNA-associated Sm-like protein LSm7
Hvulgare_10005423	WIR1A	Protein WIR1A
Hvulgare_10005435	At2g16250	Probable LRR receptor-like serine/threonine-protein kinase At2g16250
Hvulgare_10005745	Os06g0597200	Probable protein phosphatase 2C 57
Hvulgare_10005790	PAIR1	Protein PAIR1

Hvulgare_10005791	NA	Polyubiquitin
Hvulgare_10005830	NA	NA
Hvulgare_10006158	nfrkb	Nuclear factor related to kappa-B-binding protein
Hvulgare_10006211	CPI1	Cycloeucalenol cycloisomerase
Hvulgare_10006621	ALDH2C4	Aldehyde dehydrogenase family 2 member C4
Hvulgare_10006984	NA	NA
Hvulgare_10006985	NA	NA
Hvulgare_10007065	FATB1	Myristoyl-acyl-carrier-protein thioesterase, chloroplastic
Hvulgare_10007159	TT1	Protein TRANSPARENT TESTA 1
Hvulgare_10007632	IAA23	Auxin-responsive protein IAA23
Hvulgare_10007709	NA	NA
Hvulgare_10007716	At5g01610	Uncharacterized protein At5g01610
Hvulgare_10007880	BPM3	BTB/POZ and MATH domain-containing protein 3
Hvulgare_10007984	HT1	Serine/threonine-protein kinase HT1
Hvulgare_10008058	NA	NA
Hvulgare_10008637	NA	NA
Hvulgare_10008790	ROC8	Homeobox-leucine zipper protein ROC8
Hvulgare_10008791	NA	NA
Hvulgare_10008972	CRK	CDPK-related protein kinase
Hvulgare_10008973	NA	NA
Hvulgare_10008978	CBSDUF3	DUF21 domain-containing protein At2g14520
Hvulgare_10008979	SPAC2G11.09	Uncharacterized membrane protein C2G11.09
Hvulgare_10008980	OSH10	Homeobox protein knotted-1-like 4
Hvulgare_10009094	RPL19	50S ribosomal protein L19, chloroplastic
Hvulgare_10009122	PP2A2	Serine/threonine-protein phosphatase PP2A-2 catalytic subunit
Hvulgare_10009704	NA	NA
Hvulgare_10009705	NA	NA
Hvulgare_10009986	ROC8	Homeobox-leucine zipper protein ROC8
Hvulgare_10010139	GALAK	Galacturonokinase
Hvulgare_10010464	Os01g0246100	Probable histone acetyltransferase HAC-like 3
Hvulgare_10010465	NA	NA
Hvulgare_10010466	NA	NA
Hvulgare_10010467	NA	NA
Hvulgare_10010615	mak16-b	Protein MAK16 homolog B
Hvulgare_10010616	NA	NA
Hvulgare_10010658	NA	NA
Hvulgare_10010696	CYP77A4	Cytochrome P450 77A4
Hvulgare_10010697	NA	NA
Hvulgare_10010698	NA	NA

Hvulgare_10011103	RPL10AC	60S ribosomal protein L10a-3
Hvulgare_10011456	NA	NA
Hvulgare_10011934	NA	NA
Hvulgare_10012074	NA	NA
Hvulgare_10012075	NA	Polyubiquitin
Hvulgare_10012076	PHF1	SEC12-like protein 1
Hvulgare_10012186	At1g04390	BTB/POZ domain-containing protein At1g04390
Hvulgare_10012187	DAP	LL-diaminopimelate aminotransferase, chloroplastic
Hvulgare_10012403	NA	NA
Hvulgare_10012609	At4g18375	KH domain-containing protein At4g18375
Hvulgare_10012610	At1g65750	Putative ribonuclease H protein At1g65750
Hvulgare_10012628	GSTT3	Glutathione S-transferase T3
Hvulgare_10012671	NA	NA
Hvulgare_10012672	Slc38a3	Sodium-coupled neutral amino acid transporter 3
Hvulgare_10012947	NA	NA
Hvulgare_10012948	ABCF4	ABC transporter F family member 4
Hvulgare_10013097	RPT6A	26S protease regulatory subunit 8 homolog A
Hvulgare_10013393	NA	NA
Hvulgare_10013466	MBD2	Methyl-CpG-binding domain-containing protein 2
Hvulgare_10013481	ASHH3	Histone-lysine N-methyltransferase ASHH3
Hvulgare_10013482	NA	NA
Hvulgare_10013917	PRPF8	Pre-mRNA-processing-splicing factor 8
Hvulgare_10014700	Os03g0212300	B3 domain-containing protein Os03g0212300
Hvulgare_10014701	NA	NA
Hvulgare_10015198	DRT100	DNA-damage-repair/toleration protein DRT100
Hvulgare_10015199	NA	NA
Hvulgare_10015393	NA	NA
Hvulgare_10015394	Os04g0346900	Putative B3 domain-containing protein Os04g0346900
Hvulgare_10015797	PHF3	PHD finger protein 3
Hvulgare_10015799	St3gal2	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3 -sialyltransferase 2
Hvulgare_10015800	NA	NA
Hvulgare_10015801	NA	NA
Hvulgare_10015802	Gtpbp4	Nucleolar GTP-binding protein 1
Hvulgare_10015834	NA	NA
Hvulgare_10015835	ADH1	Alcohol dehydrogenase 1
Hvulgare_10015836	At3g47200	UPF0481 protein At3g47200
Hvulgare_10015890	NA	NA
Hvulgare_10015891	UKL3	Uridine kinase-like protein 3
Hvulgare_10016088	DAG	DAG protein, chloroplastic

Hvulgare_10016220	Os03g0212300	B3 domain-containing protein Os03g0212300
Hvulgare_10016221	ROC8	Homeobox-leucine zipper protein ROC8
Hvulgare_10016222	APX6	Probable L-ascorbate peroxidase 6, chloroplastic
Hvulgare_10016368	NA	NA
Hvulgare_10016377	LHCA4	Chlorophyll a-b binding protein 4, chloroplastic
Hvulgare_10016378	At1g04910	DUF246 domain-containing protein At1g04910
Hvulgare_10016638	AHK5	Histidine kinase 5
Hvulgare_10016940	MKK4	Mitogen-activated protein kinase kinase 4
Hvulgare_10016945	NA	NA
Hvulgare_10016946	PER1	Peroxidase 1
Hvulgare_10017102	NA	NA
Hvulgare_10017104	At4g30825	Pentatricopeptide repeat-containing protein At4g30825, chloroplastic
Hvulgare_10017348	NA	NA
Hvulgare_10017380	NA	NA
Hvulgare_10017381	MZM1	Mitochondrial zinc maintenance protein 1, mitochondrial
Hvulgare_10017556	FIB1	rRNA 2'-O-methyltransferase fibrillarin 1
Hvulgare_10017668	yakA	Probable serine/threonine-protein kinase yakA
Hvulgare_10017669	alkbh8	Alkylated DNA repair protein alkB homolog 8
Hvulgare_10017670	PCMP-E95	Pentatricopeptide repeat-containing protein At3g22150, chloroplastic
Hvulgare_10017747	PARP1	Poly [ADP-ribose] polymerase 1
Hvulgare_10017840	IAA27	Auxin-responsive protein IAA27
Hvulgare_10017932	NA	NA
Hvulgare_10017933	SCPL40	Serine carboxypeptidase-like 40
Hvulgare_10017934	NA	NA
Hvulgare_10017935	NA	NA
Hvulgare_10018021	PARP2	Poly [ADP-ribose] polymerase 2
Hvulgare_10018022	NA	NA
Hvulgare_10018023	papA	Poly(A) polymerase
Hvulgare_10018205	At4g31140	Glucan endo-1,3-beta-glucosidase 5
Hvulgare_10018206	CXE2	Probable carboxylesterase 2
Hvulgare_10018207	CXE2	Probable carboxylesterase 2
Hvulgare_10018208	UGT73D1	UDP-glycosyltransferase 73D1
Hvulgare_10018490	CYP51G1	Sterol 14-demethylase
Hvulgare_10018826	GTE8	Transcription factor GTE8
Hvulgare_10019125	NA	NA
Hvulgare_10019152	ARF25	Auxin response factor 25
Hvulgare_10019266	At4g09670	Uncharacterized oxidoreductase At4g09670
Hvulgare_10019588	GTE8	Transcription factor GTE8

Hvulgare_10019589	NA	NA
Hvulgare_10019590	NEC3	Bifunctional monodehydroascorbate reductase and carbonic anhydrase nectarin-3
Hvulgare_10020159	NA	NA
Hvulgare_10020455	TIFY9	Protein TIFY 9
Hvulgare_10020620	NA	NA
Hvulgare_10020621	At2g28450	Zinc finger CCCH domain-containing protein 24
Hvulgare_10020673	MAN7	Mannan endo-1,4-beta-mannosidase 7
Hvulgare_10020674	NA	NA
Hvulgare_10020864	Os03g0212300	B3 domain-containing protein Os03g0212300
Hvulgare_10021734	hdhd3	Haloacid dehalogenase-like hydrolase domain-containing protein 3
Hvulgare_10021933	SPS5	Probable sucrose-phosphate synthase 5
Hvulgare_10021990	NA	NA
Hvulgare_10021991	PCMP-E55	Pentatricopeptide repeat-containing protein At1g31430
Hvulgare_10021992	PHN1	Protein argonaute PNH1
Hvulgare_10022334	NA	NA
Hvulgare_10022335	NA	NA
Hvulgare_10022336	Taf7	Transcription initiation factor TFIID subunit 7
Hvulgare_10022337	NA	Polyubiquitin
Hvulgare_10022339	NA	NA
Hvulgare_10022341	NA	NA
Hvulgare_10022342	RPL27A	60S ribosomal protein L27-1
Hvulgare_10022783	PGP	Phosphoglycolate phosphatase
Hvulgare_10022784	NA	NA
Hvulgare_10022869	ADT1	Arogenate dehydratase/prephenate dehydratase 1, chloroplastic
Hvulgare_10022983	MAPRE3	Microtubule-associated protein RP/EB family member 3
Hvulgare_10022985	NA	NA
Hvulgare_10023184	CYP75B2	Flavonoid 3'-monooxygenase
Hvulgare_10023185	DTX1	MATE efflux family protein DTX1
Hvulgare_10023186	At1g20180	UPF0496 protein At1g20180
Hvulgare_10023187	CRK	CDPK-related protein kinase
Hvulgare_10023218	Hsp90b1	Endoplasmin
Hvulgare_10023256	At4g06598	Uncharacterized protein At4g06598
Hvulgare_10023257	polr1d	DNA-directed RNA polymerases I and III subunit RPAC2
Hvulgare_10023301	NA	NA
Hvulgare_10023360	NA	NA
Hvulgare_10023412	spp27	Upstream activation factor subunit spp27

Hvulgare_10023413	NA	NA
Hvulgare_10023612	NA	NA
Hvulgare_10023766	NA	NA
Hvulgare_10024083	NA	NA
Hvulgare_10024084	CML16	Probable calcium-binding protein CML16
Hvulgare_10024085	NA	NA
Hvulgare_10025045	IPP	Soluble inorganic pyrophosphatase
Hvulgare_10025046	At5g52970	Thylakoid lumenal 15.0 kDa protein 2, chloroplastic
Hvulgare_10025051	CLPR2	ATP-dependent Clp protease proteolytic subunit-related protein 2, chloroplastic
Hvulgare_10025052	NA	NA
Hvulgare_10025053	NA	NA
Hvulgare_10025054	UTR3	UDP-galactose/UDP-glucose transporter 3
Hvulgare_10025193	APUM5	Pumilio homolog 5
Hvulgare_10025194	NA	NA
Hvulgare_10025311	NA	NA
Hvulgare_10025429	RPL11	60S ribosomal protein L11
Hvulgare_10025618	TANC2	Protein TANC2
Hvulgare_10025619	PIP5K4	Phosphatidylinositol-4-phosphate 5-kinase 4
Hvulgare_10025785	RPM1	Disease resistance protein RPM1
Hvulgare_10025786	UGT91B1	UDP-glycosyltransferase 91B1
Hvulgare_10025787	RPP13	Disease resistance protein RPP13
Hvulgare_10025815	RGA2	Disease resistance protein RGA2
Hvulgare_10025830	At2g28370	CASP-like protein At2g28370
Hvulgare_10025831	NA	NA
Hvulgare_10025832	gacL	Rho GTPase-activating protein gacL
Hvulgare_10026382	NA	NA
Hvulgare_10026383	Os03g0212300	B3 domain-containing protein Os03g0212300
Hvulgare_10027088	WAK5	Wall-associated receptor kinase 5
Hvulgare_10027090	NA	NA
Hvulgare_10027091	NA	NA
Hvulgare_10027092	NA	NA
Hvulgare_10027377	At4g13360	3-hydroxyisobutyryl-CoA hydrolase-like protein 3, mitochondrial
Hvulgare_10028187	Nfrkb	Nuclear factor related to kappa-B-binding protein
Hvulgare_10028314	NA	Enoyl-[acyl-carrier-protein] reductase [NADH], chloroplastic
Hvulgare_10028315	At4g01730	Probable S-acyltransferase At4g01730
Hvulgare_10028316	XRN4	5'-3' exoribonuclease 4
Hvulgare_10028317	Os02g0794700	Leucine aminopeptidase 2, chloroplastic

Hvulgare_10028318	NA	NA
Hvulgare_10028465	NA	NA
Hvulgare_10028656	NA	NA
Hvulgare_10028658	FLA7	Fasciclin-like arabinogalactan protein 7
Hvulgare_10028669	ACT1	Actin-1
Hvulgare_10028874	At2g28370	CASP-like protein At2g28370
Hvulgare_10028876	NIP1-3	Aquaporin NIP1-3
Hvulgare_10028877	NA	NA
Hvulgare_10028878	NA	NA
Hvulgare_10028881	At5g02060	CASP-like protein At5g02060
Hvulgare_10028882	CML21	Probable calcium-binding protein CML21
Hvulgare_10028910	RPS12	40S ribosomal protein S12
Hvulgare_10028911	Gltp	Glycolipid transfer protein
Hvulgare_10028912	At1g10320	Zinc finger CCCH domain-containing protein 5
Hvulgare_10028913	Os02g0557500	Zinc finger CCCH domain-containing protein 16
Hvulgare_10028914	NA	NA
Hvulgare_10028916	NA	NA
Hvulgare_10029414	NA	NA
Hvulgare_10029833	NA	NA
Hvulgare_10029834	Os03g0650900	Probable E3 ubiquitin-protein ligase BAH1-like 1
Hvulgare_10030134	YKT61	VAMP-like protein YKT61
Hvulgare_10030488	FER	Receptor-like protein kinase FERONIA
Hvulgare_10030489	BHLH79	Transcription factor bHLH79
Hvulgare_10030490	VAMP726	Putative vesicle-associated membrane protein 726
Hvulgare_10030926	ATL40	RING-H2 finger protein ATL40
Hvulgare_10031116	NA	NA
Hvulgare_10031117	ytfP	Uncharacterized protein ytfP
Hvulgare_10031118	VAR3	Zinc finger protein VAR3, chloroplastic
Hvulgare_10031119	MGST3	Microsomal glutathione S-transferase 3
Hvulgare_10031120	At2g02850	Basic blue protein
Hvulgare_10031934	NA	NA
Hvulgare_10032680	yjbI	Group 2 truncated hemoglobin yjbI
Hvulgare_10032821	NA	NA
Hvulgare_10032822	NA	NA
Hvulgare_10032823	NHX2	Sodium/hydrogen exchanger 2
Hvulgare_10033167	NA	NA
Hvulgare_10033514	YSL3	Probable metal-nicotianamine transporter YSL3
Hvulgare_10033515	VPS35B	Vacuolar protein sorting-associated protein 35B
Hvulgare_10033683	NA	NA
Hvulgare_10034207	NA	NA

Hvulgare_10034208	HARBI1	Putative nuclease HARBI1
Hvulgare_10034221	UGT75D1	UDP-glycosyltransferase 75D1
Hvulgare_10034353	G6PDH	Glucose-6-phosphate 1-dehydrogenase, cytoplasmic isoform
Hvulgare_10034490	VTC5	GDP-L-galactose phosphorylase 2
Hvulgare_10034593	Os01g0639100	DEAD-box ATP-dependent RNA helicase 2
Hvulgare_10034595	NA	NA
Hvulgare_10035071	OSH10	Homeobox protein knotted-1-like 4
Hvulgare_10035305	NA	NA
Hvulgare_10035306	NA	NA
Hvulgare_10035419	PCMP-E13	Pentatricopeptide repeat-containing protein At5g56310
Hvulgare_10035420	KCBP	Kinesin-like calmodulin-binding protein
Hvulgare_10036361	ATJ10	Chaperone protein dnaJ 10
Hvulgare_10036362	NA	NA
Hvulgare_10036363	NA	NA
Hvulgare_10036694	KCS4	3-ketoacyl-CoA synthase 4
Hvulgare_10036840	NA	NA
Hvulgare_10036920	PCF1	Transcription factor PCF1
Hvulgare_10037134	STAR2	UPF0014 membrane protein STAR2
Hvulgare_10037135	NA	NA
Hvulgare_10037136	At5g15730	Probable leucine-rich repeat receptor-like serine/threonine-protein kinase At5g15730
Hvulgare_10037138	NA	NA
Hvulgare_10037353	PCMP-E101	Pentatricopeptide repeat-containing protein At4g18840
Hvulgare_10037354	HSP81-3	Heat shock protein 81-3
Hvulgare_10037382	Apeh	Acylamino-acid-releasing enzyme
Hvulgare_10037383	NA	NA
Hvulgare_10037612	AHK5	Histidine kinase 5
Hvulgare_10037846	GAM1	Transcription factor GAMYB
Hvulgare_10037847	NA	NA
Hvulgare_10037848	ISA1	Isoamylase 1, chloroplastic
Hvulgare_10038482	NA	NA
Hvulgare_10038483	NA	NA
Hvulgare_10038484	NA	NA
Hvulgare_10038654	ROMT-17	Tricin synthase 2
Hvulgare_10038655	At1g05000	Probable tyrosine-protein phosphatase At1g05000
Hvulgare_10038657	NA	NA
Hvulgare_10039401	NA	NA
Hvulgare_10039402	FAR1	Protein FAR-RED IMPAIRED RESPONSE 1
Hvulgare_10039403	At1g04910	DUF246 domain-containing protein At1g04910

Hvulgare_10039404	NA	NA
Hvulgare_10039407	CEBIP	Chitin elicitor-binding protein
Hvulgare_10039408	NA	Histone H4
Hvulgare_10039409	NA	Histone H4
Hvulgare_10039410	FMO1	Probable flavin-containing monooxygenase 1
Hvulgare_10039411	FMO1	Probable flavin-containing monooxygenase 1
Hvulgare_10039412	FMO1	Probable flavin-containing monooxygenase 1
Hvulgare_10039480	ASPSCR1	Tether containing UBX domain for GLUT4
Hvulgare_10039510	NA	NA
Hvulgare_10039511	yqxC	Putative rRNA methyltransferase YqxC
Hvulgare_10039617	NA	NA
Hvulgare_10039618	ROMT-15	Tricin synthase 1
Hvulgare_10039619	NA	NA
Hvulgare_10039641	NA	NA
Hvulgare_10039881	NA	NA
Hvulgare_10039882	NA	NA
Hvulgare_10039883	NA	NA
Hvulgare_10039884	PAM18	Mitochondrial import inner membrane translocase subunit TIM14
Hvulgare_10040112	NA	NA
Hvulgare_10040113	NRT2.1	High-affinity nitrate transporter 2.1
Hvulgare_10040114	NRT2.1	High-affinity nitrate transporter 2.1
Hvulgare_10040138	Os08g0520100	RNA pseudourine synthase 3, mitochondrial
Hvulgare_10040139	At3g42630	Pentatricopeptide repeat-containing protein At3g42630
Hvulgare_10040140	TIAL1	Nucleolysin TIAR
Hvulgare_10040220	SCPL40	Serine carboxypeptidase-like 40
Hvulgare_10040565	NA	NA
Hvulgare_10040695	ATB	F-box protein At-B
Hvulgare_10040696	CSN6A	COP9 signalosome complex subunit 6a
Hvulgare_10040755	MRS2-I	Magnesium transporter MRS2-I
Hvulgare_10041234	NA	NA
Hvulgare_10041236	sec14	Sec14 cytosolic factor
Hvulgare_10041331	GLX-I	Lactoylglutathione lyase
Hvulgare_10041336	NA	NA
Hvulgare_10041759	NA	6-hydroxynicotinate 3-monooxygenase
Hvulgare_10041760	NDKR	Nucleoside diphosphate kinase 1
Hvulgare_10041761	NA	Auxin-induced protein 5NG4
Hvulgare_10041970	NA	NA
Hvulgare_10041972	argA	Amino-acid acetyltransferase
Hvulgare_10041973	At1g80150	Pentatricopeptide repeat-containing protein At1g80150,

		mitochondrial
Hvulgare_10042455	At1g05700	Probable LRR receptor-like serine/threonine-protein kinase At1g05700
Hvulgare_10042459	NA	NA
Hvulgare_10042687	HKT7	Probable cation transporter HKT7
Hvulgare_10042819	NA	NA
Hvulgare_10043175	WRAP73	WD repeat-containing protein WRAP73
Hvulgare_10043176	FBX5	Protein ARABIDILLO 1
Hvulgare_10043501	At3g06920	Pentatricopeptide repeat-containing protein At3g06920
Hvulgare_10043505	NA	NA
Hvulgare_10043760	NA	NA
Hvulgare_10043873	SCOA	Succinyl-CoA ligase [ADP-forming] subunit alpha-1, mitochondrial
Hvulgare_10043875	rnf-5	RING finger protein 5
Hvulgare_10043876	THIC	Phosphomethylpyrimidine synthase, chloroplastic
Hvulgare_10044127	Esyt1	Extended synaptotagmin-1
Hvulgare_10044128	Esyt2	Extended synaptotagmin-2
Hvulgare_10044440	NA	NA
Hvulgare_10044442	Os08g0360100	Chloroplastic group IIA intron splicing facilitator CRS1, chloroplastic
Hvulgare_10044467	LECRK91	L-type lectin-domain-containing receptor kinase IX.1
Hvulgare_10044468	NA	NA
Hvulgare_10044527	topA	DNA topoisomerase 1
Hvulgare_10044529	TY5A	Putative transposon Ty5-1 protein YCL074W
Hvulgare_10044748	GPDL1	Probable glycerophosphoryl diester phosphodiesterase 1
Hvulgare_10045485	NA	NA
Hvulgare_10045640	NA	NA
Hvulgare_10045641	NA	NA
Hvulgare_10045994	NA	NA
Hvulgare_10045995	WAK5	Wall-associated receptor kinase 5
Hvulgare_10045996	HP1	Histidine-containing phosphotransfer protein 1
Hvulgare_10045997	NA	NA
Hvulgare_10045998	NA	NA
Hvulgare_10046002	RPS2	Disease resistance protein RPS2
Hvulgare_10046263	IDI2	Methylthioribose-1-phosphate isomerase
Hvulgare_10046882	MBD2	Methyl-CpG-binding domain-containing protein 2
Hvulgare_10046883	NA	NA
Hvulgare_10046885	NA	NA
Hvulgare_10046886	NA	NA
Hvulgare_10047060	NA	NA

Hvulgare_10047061	JKD	Zinc finger protein JACKDAW
Hvulgare_10047397	TAX10	3'-N-debenzoyl-2'-deoxytaxol N-benzoyltransferase
Hvulgare_10047537	MBD2	Methyl-CpG-binding domain-containing protein 2
Hvulgare_10047538	NA	NA
Hvulgare_10047540	MBD2	Methyl-CpG-binding domain-containing protein 2
Hvulgare_10047541	MBD2	Methyl-CpG-binding domain-containing protein 2
Hvulgare_10048289	ARF21	Auxin response factor 21
Hvulgare_10048334	NA	NA
Hvulgare_10048335	NA	NA
Hvulgare_10048337	NA	NA
Hvulgare_10048338	NA	NA
Hvulgare_10048339	NA	NA
Hvulgare_10049312	CYP89A2	Cytochrome P450 89A2
Hvulgare_10049313	OsI_033149	UPF0496 protein 4
Hvulgare_10049332	nep1	Aspartic proteinase nepenthesin-1
Hvulgare_10049479	NA	NA
Hvulgare_10050337	Os04g0650000	Oryzain alpha chain
Hvulgare_10050441	NA	NA
Hvulgare_10050442	NA	NA
Hvulgare_10050969	At1g61330	Putative FBD-associated F-box protein At1g61330
Hvulgare_10050970	PAA1	Putative copper-transporting ATPase PAA1
Hvulgare_10050971	NA	NA
Hvulgare_10051826	PER43	Peroxidase 43
Hvulgare_10052277	MLO1	MLO-like protein 1
Hvulgare_10052292	NA	NA
Hvulgare_10052293	ADT2	Arogenate dehydratase/prephenate dehydratase 2, chloroplastic
Hvulgare_10052294	APX2	L-ascorbate peroxidase 2, cytosolic
Hvulgare_10052295	RMD1	Sporulation protein RMD1
Hvulgare_10052640	SH3GL1	Endophilin-A2
Hvulgare_10052641	yqeI	Probable RNA-binding protein YqeI
Hvulgare_10053406	PSKR2	Phytosulfokine receptor 2
Hvulgare_10053407	TDT	Tonoplast dicarboxylate transporter
Hvulgare_10053409	NA	NA
Hvulgare_10053410	NA	NA
Hvulgare_10054500	RPL31	60S ribosomal protein L31
Hvulgare_10054501	NA	NA
Hvulgare_10054838	Os01g0639100	DEAD-box ATP-dependent RNA helicase 2
Hvulgare_10054839	CYP71D55	Premnaspirodiene oxygenase
Hvulgare_10055789	NA	NA

Hvulgare_10055794	Os08g0500300	Probable protein phosphatase 2C 66
Hvulgare_10055796	Bud13	BUD13 homolog
Hvulgare_10056304	NA	NA
Hvulgare_10056305	yitV	Putative esterase yitV

Table S35 Kegg pathway of genes involved in selective sweeps

Pathway	Gene number	Pathway ID
Plant-pathogen interaction	15	ko04626
Plant hormone signal transduction	14	ko04075
Phenylpropanoid biosynthesis	7	ko00940
mRNA surveillance pathway	7	ko03015
Spliceosome	7	ko03040
Stilbenoid, diarylheptanoid and gingerol biosynthesis	6	ko00945
Ribosome biogenesis in eukaryotes	6	ko03008
Ribosome	6	ko03010
RNA degradation	6	ko03018
Protein processing in endoplasmic reticulum	6	ko04141
Purine metabolism	5	ko00230
Pyrimidine metabolism	5	ko00240
Phenylalanine metabolism	5	ko00360
Glutathione metabolism	5	ko00480
Polycyclic aromatic hydrocarbon degradation	5	ko00624
Aminobenzoate degradation	5	ko00627
Methane metabolism	5	ko00680
Flavonoid biosynthesis	5	ko00941
Ascorbate and aldarate metabolism	4	ko00053
Limonene and pinene degradation	4	ko00903
Drug metabolism - cytochrome P450	4	ko00982
Bisphenol degradation	3	ko00363
Metabolism of xenobiotics by cytochrome P450	3	ko00980
RNA transport	3	ko03013
RNA polymerase	3	ko03020
PPAR signaling pathway	3	ko03320
Endocytosis	3	ko04144
Circadian rhythm - plant	3	ko04712
Chagas disease (American trypanosomiasis)	3	ko05142

Table S36 50 genes randomly collected from Table S34 and 10 stressful environmental variables for correlation analysis

(a)

Name in the graph	Gene
A1	Enoyl-CoA hydratase 2, peroxisomal
A2	Peroxidase 4
A3	DNA (cytosine-5)-methyltransferase DRM2
A4	Putative calcium-binding protein CML19
A5	Probable leucine-rich repeat receptor-like protein kinase At1g68400
A6	Stress-associated endoplasmic reticulum protein 2
A7	Tyrosine N-monooxygenase
A8	Cycloeucalenol cycloisomerase
A9	Protein TRANSPARENT TESTA 1
A10	Auxin-responsive protein IAA23
A11	Serine/threonine-protein kinase HT1
A12	CDPK-related protein kinase
A13	DUF21 domain-containing protein At2g14520
A14	Homeobox protein knotted-1-like 4
A15	Serine/threonine-protein phosphatase PP2A-2 catalytic subunit
A16	Galacturonokinase
A17	Probable histone acetyltransferase HAC-like 3
A18	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 2
A19	Alcohol dehydrogenase 1
A20	Uridine kinase-like protein 3
A21	DAG protein, chloroplastic
A22	Probable L-ascorbate peroxidase 6, chloroplastic
A23	Chlorophyll a-b binding protein 4, chloroplastic
A24	Histidine kinase 5
A25	Mitogen-activated protein kinase kinase 4
A26	Peroxidase 1
A27	Probable serine/threonine-protein kinase yakA
A28	Poly [ADP-ribose] polymerase 1
A29	Serine carboxypeptidase-like 40
A30	Glucan endo-1,3-beta-glucosidase 5
A31	Protein argonaute PNH1

A32	Phosphoglycolate phosphatase
A33	Flavonoid 3'-monooxygenase
A34	Endoplasmin
A35	Phosphatidylinositol-4-phosphate 5-kinase 4
A36	UDP-glycosyltransferase 91B1
A37	Probable S-acyltransferase At4g01730
A38	Aquaporin NIP1-3
A39	Sodium/hydrogen exchanger 2
A40	UDP-glycosyltransferase 75D1
A41	GDP-L-galactose phosphorylase 2
A42	3-ketoacyl-CoA synthase 4
A43	Heat shock protein 81-3
A44	Histidine kinase 5
A45	Magnesium transporter MRS2-I
A46	Lactoylglutathione lyase
A47	6-hydroxynicotinate 3-monooxygenase
A48	Amino-acid acetyltransferase
A49	Probable cation transporter HKT7
A50	Putative transposon Ty5-1 protein YCL074W

(b)

Symbol	Feature
B1	Salinity
B2	Oxygen
B3	Amylase
B4	Radiation
B5	CO2
B6	Drought
B7	Temperature (cold)
B8	Temperature (heat)
B9	Long and short day
B10	Dormancy

III. Figures

Fig. S1, The Tibetan hulless barley is the main crop for people living in Tibet, China. The relative sea level of Tibet is more than 3,100 m.

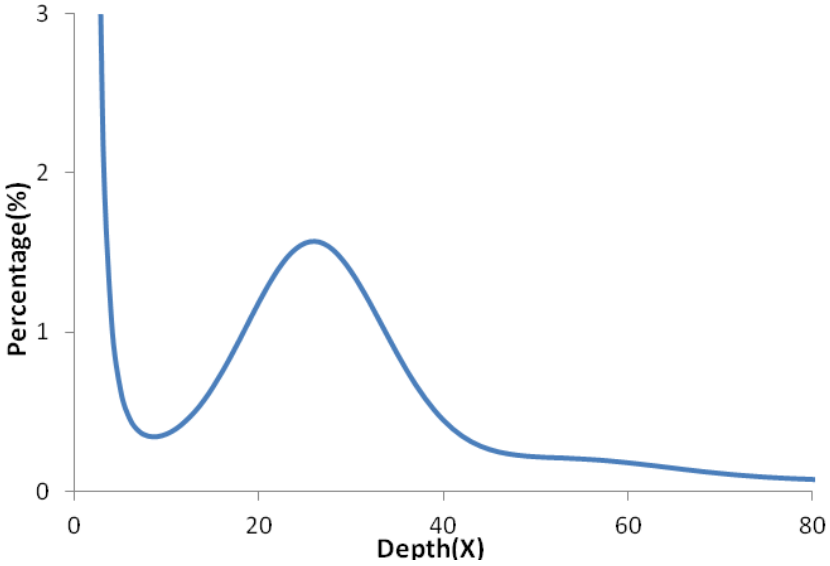


Fig. S2, Distribution of 17-mer frequency in the sequencing reads. 133 Gb data was retained for 17-mer analysis, the peak of distribution is about 26X coverage; the genome size can be estimated as 4.48 Gb (Genome Size=K-mer number/Peak depth).

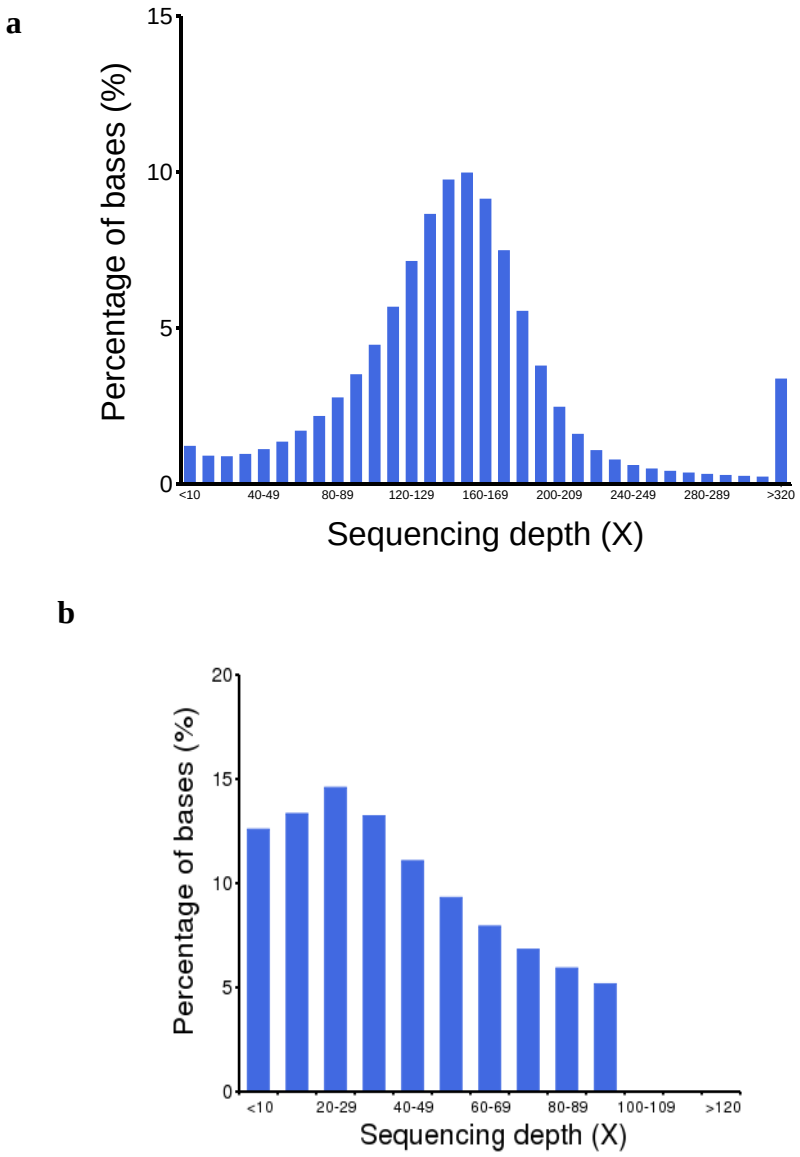


Fig. S3, Sequencing depth distribution for bases of whole genome (a) and scaffolds with length less than 200 bp (b). X-axis is depth and Y-axis is proportional to the base number divided by total bases.

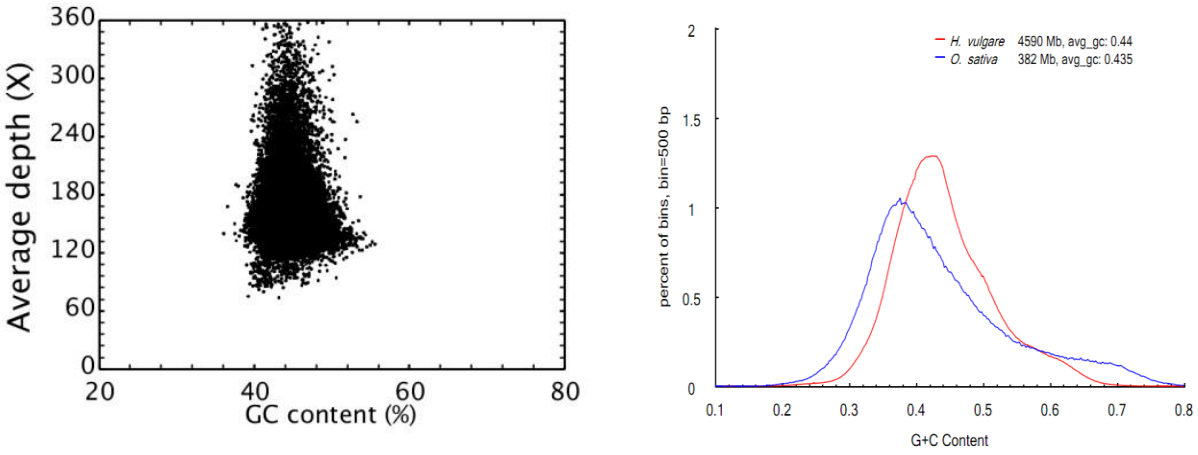
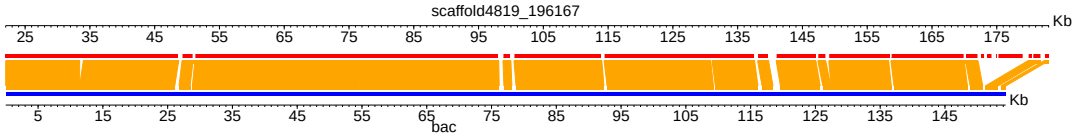
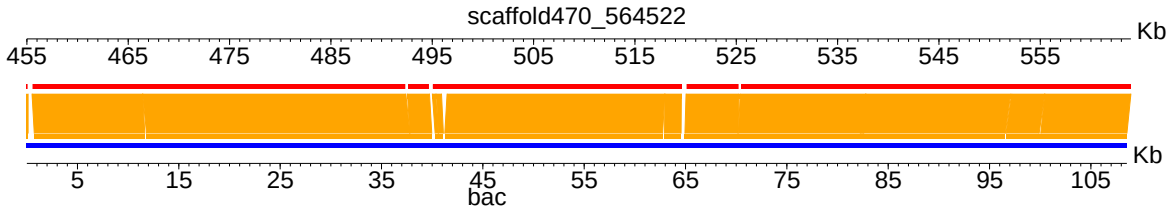


Fig. S4, GC content and sequencing depth. a, GC_depth distribution of Tibetan hulless barley. X-axis represents GC content; Y-axis represents average depth. b, Comparison of GC content distributions in Tibetan hulless barley and rice genomes.

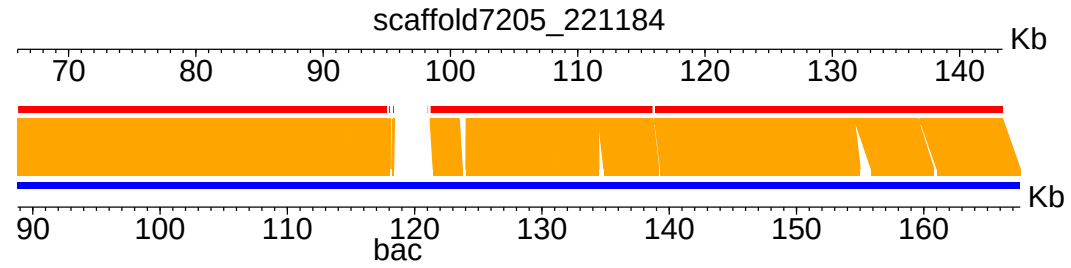
HVVMRXALLmA0257K17_c1 (Genbank ID: AC250041.1)



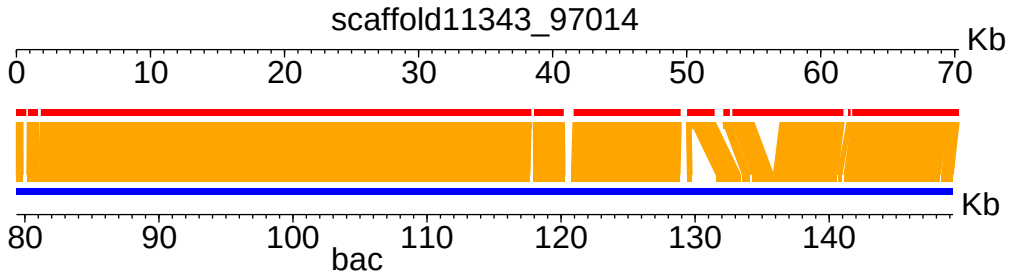
HVVMRXALLmA0104M01_c1 (Genbank ID: AC253422.1)



HVVMRXALLmA0024C01_c1 (Genbank ID: AC253057.1)



HVVMRXALLeA0103D13_c1 (Genbank ID: AC248476.1)



HVVMRXALLeA0088K01_c1 (Genbank ID: AC250041.1)

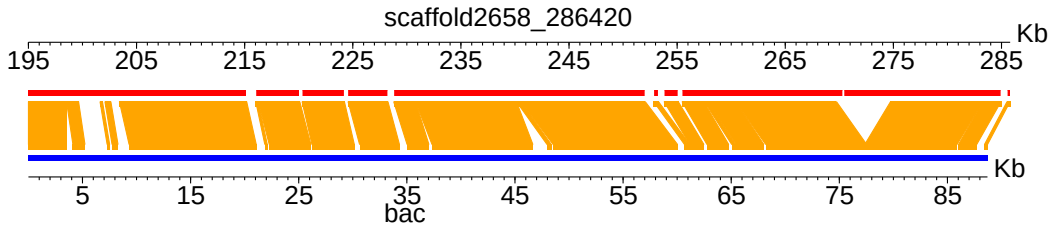


Fig. S5, Comparison of assembled scaffolds with five BAC sequences of cultivated barley

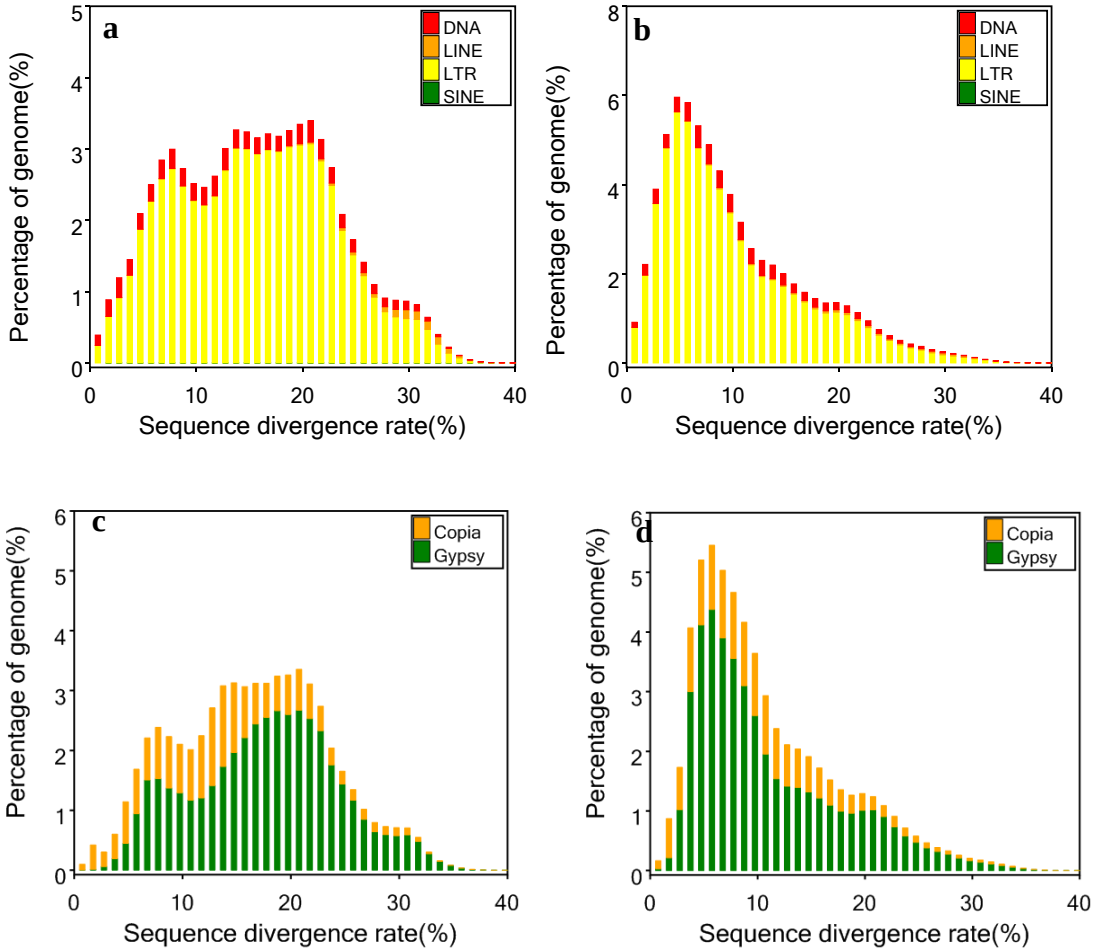


Fig. S6, Divergence rate distribution of different types of TE predicted with *Repbase* (a, c) and de novo (b, d) methods. Divergence rate computed between predicted TE sequence in genome and consensus sequence in the Repbase or de novo predicted TE library. The above figures show that most of the de novo predicted repeats are recently transposed repeats, which are active in the genome. The Copia and Gypsy subfamily of TE repeats are the most enriched in *H. vulgare* genome.

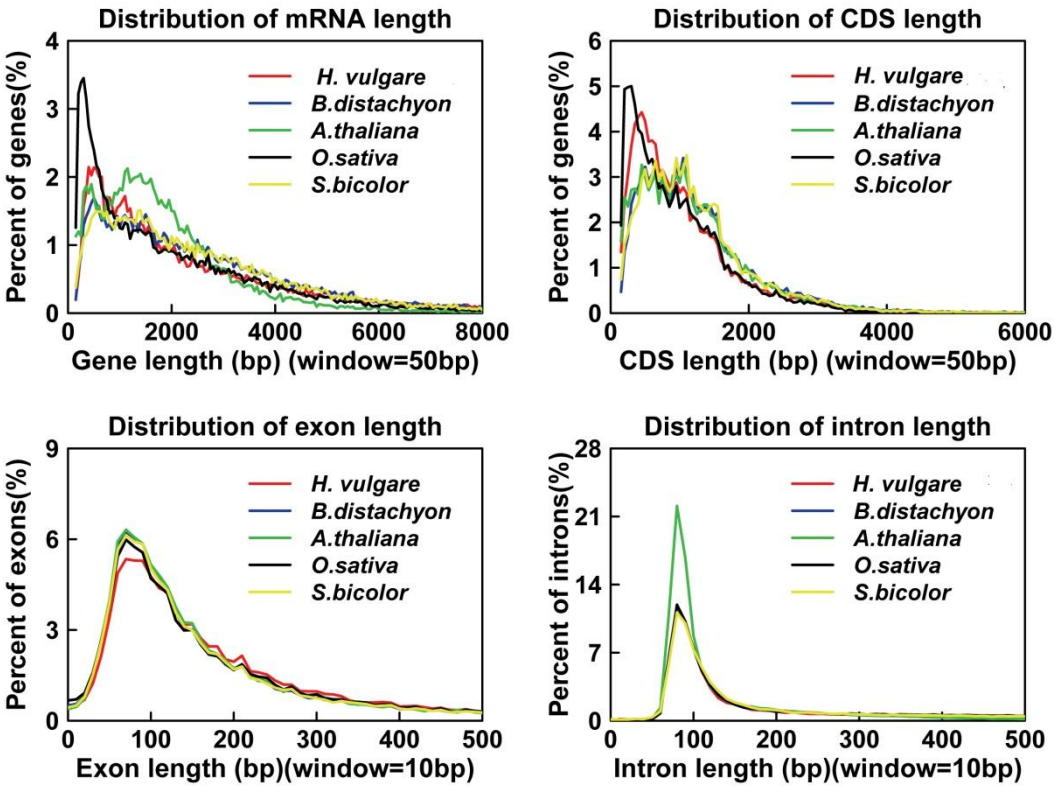


Fig. S7, Comparison of length distribution of gene, exon, intron, and CDS in Tibetan hulless barley (*H. vulgare)* and four other species. Window means the length of every point in the horizontal ordinate.

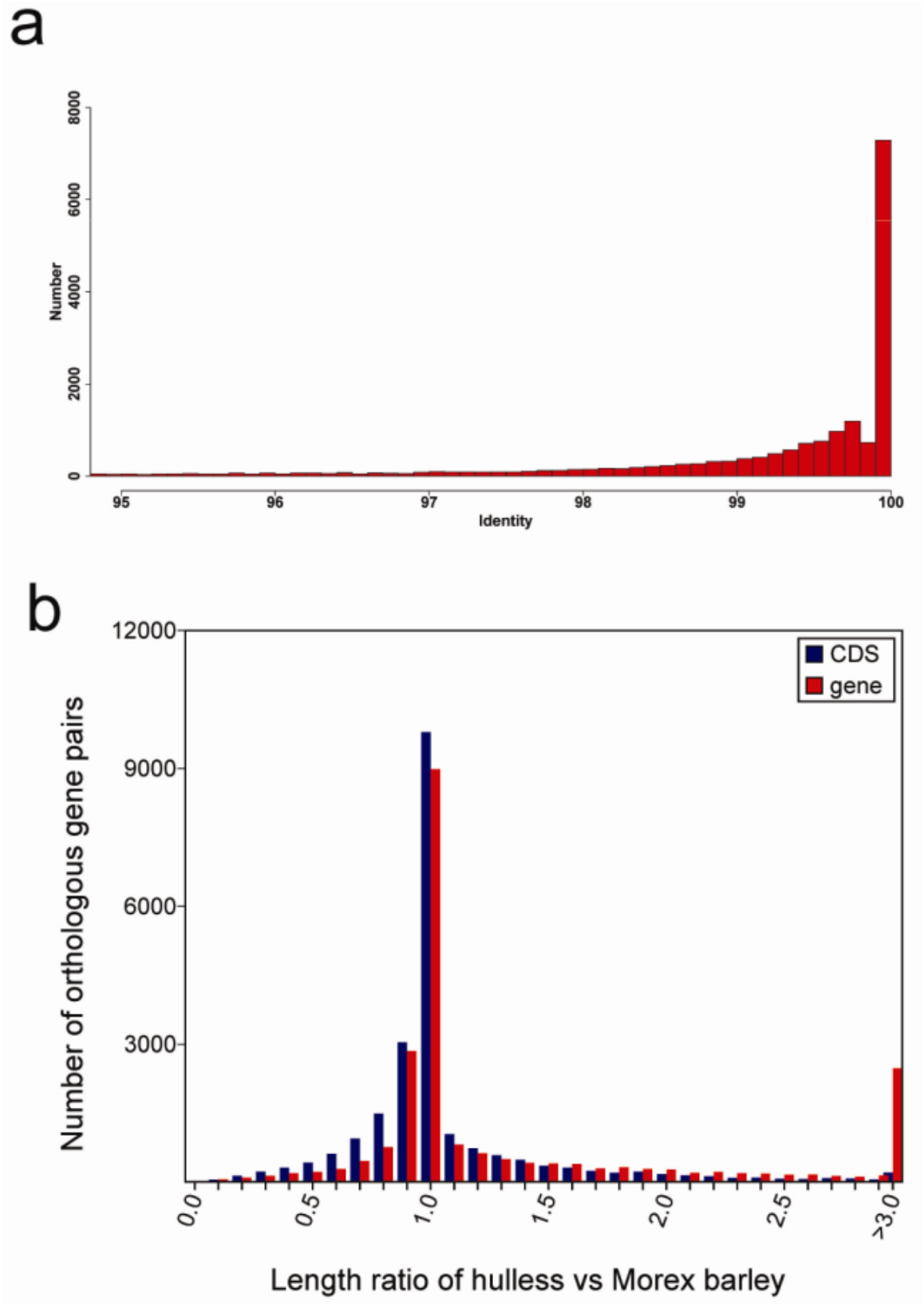


Fig. S8, Comparison between the genes of Tibetan hulless barley and Morex. a. Protein similarity distribution of 22,673 Tibetan hulless barley and Morex orthologous. 7,224

(31.86%) gene pairs had identical aligned sequences and 17,840 (78.68%) had protein similarity higher than 95%. b, The CDS and gene body (CDS + intron) length ratio of Tibetan hulless barley versus Morex barley for their common 22,673 orthologous gene pairs. Orthologous gene pairs were identified by reciprocal best hit of *BlastP* alignments.

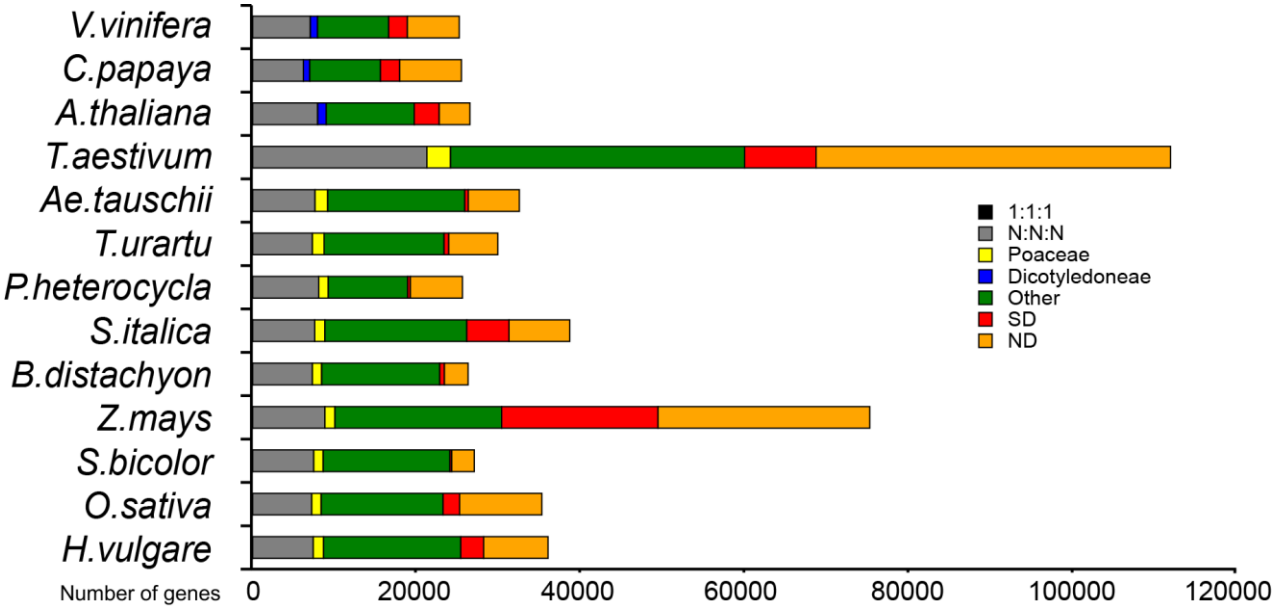


Fig. S9, Stacked bar chart showing orthologous gene numbers among 13 plant genomes. 1:1:1, single-copy orthologs, N:N:N, multi-copy orthologs, Poaceae, Poaceae-specific orthologs, Dicotyledoneae, Dicotyledoneae-specific orthologs, SD, duplicated species-specific genes, ND, species-specific genes.

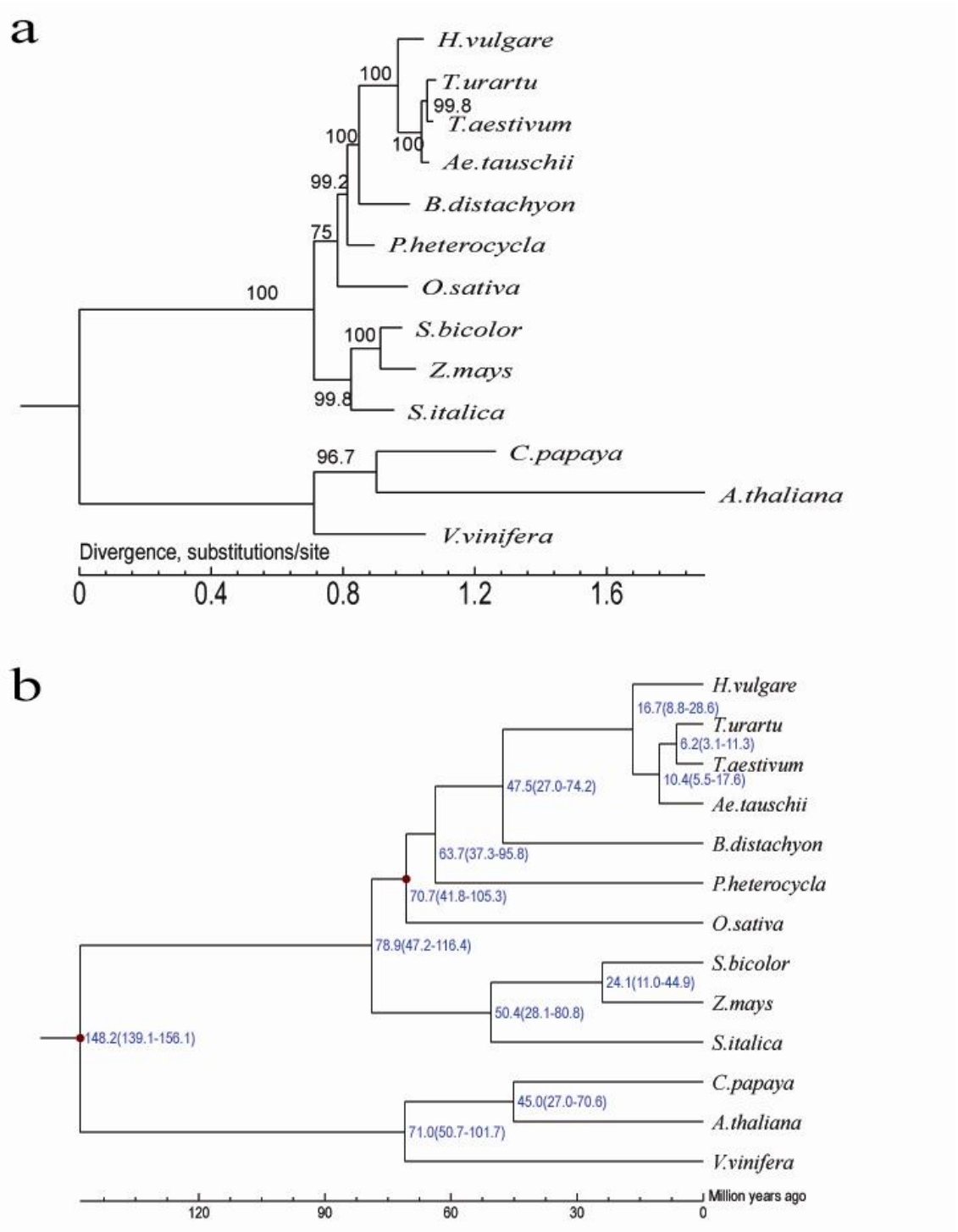


Fig. S10, Phylogenetic tree (a) and divergence time (b) for 13 plant species. a, Phylogenetic tree constructed with orthologous genes on 4-fold degenerate sites by maximum likelihood method. Branch length represents the neutral divergence rate. b, Estimation of divergence time and substitution rate. Blue numbers on the nodes are the divergence time from present (million years ago, Mya). The calibration time *H. vulgare – A. thaliana* divergence (139~156 million years ago), and *H. vulgare – O. sativa* divergence (at least 34 million years ago) is derived from previously published papers (44, 45).

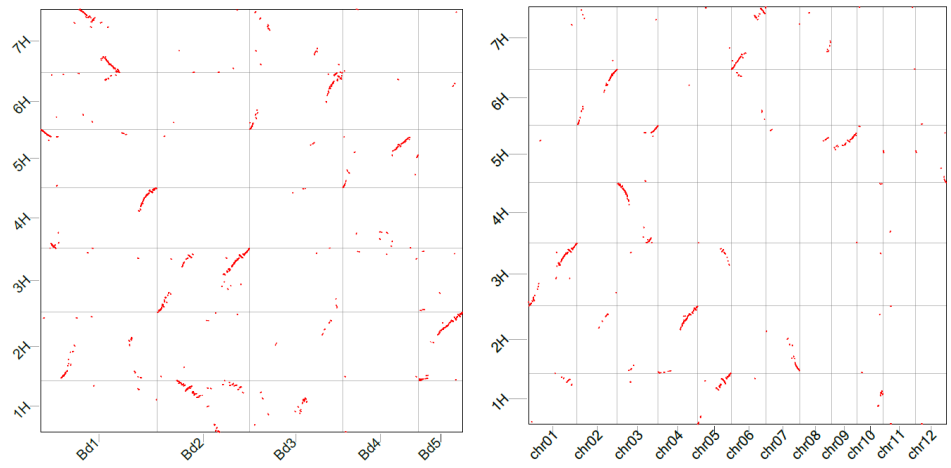


Fig. S11, Syntenic blocks for *H. vulgare–B. distachyon* (left) and *H. vulgare–O. sativa* (right) chromosomes

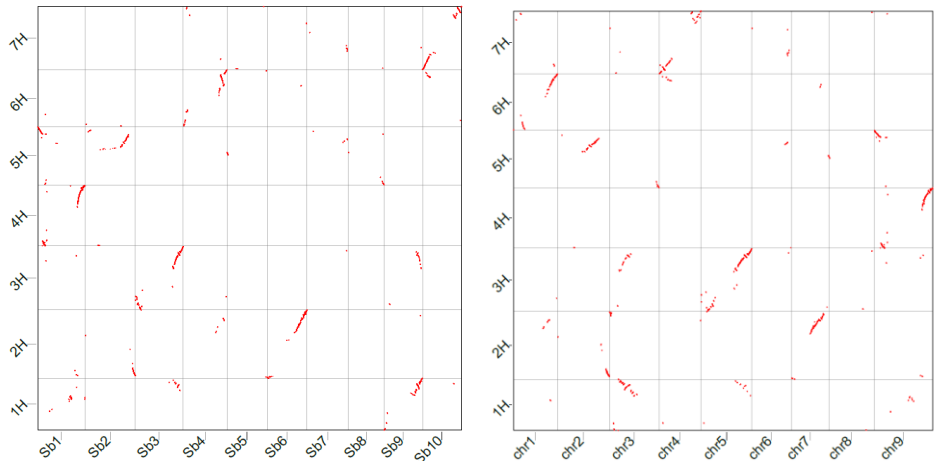


Fig. S12, Syntenic blocks for *H. vulgare–S. bicolor* (left) and *H. vulgare–S. italic* (right) chromosomes

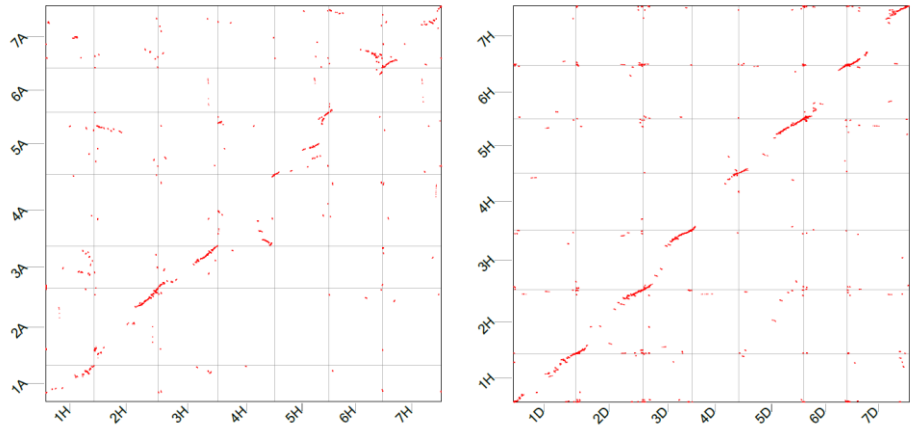


Fig. S13, Syntenic blocks for *T. urartu–H. vulgare* (left) and *H. vulgare–A. tauschii* (right) chromosomes. Chromosomes for *H. vulgare* are named as 1H, 2H, etc. Chromosomes for *T. urartu* are named as 1A, 2A, etc. Chromosomes for *A. tauschii* are named as 1D, 2D, etc.

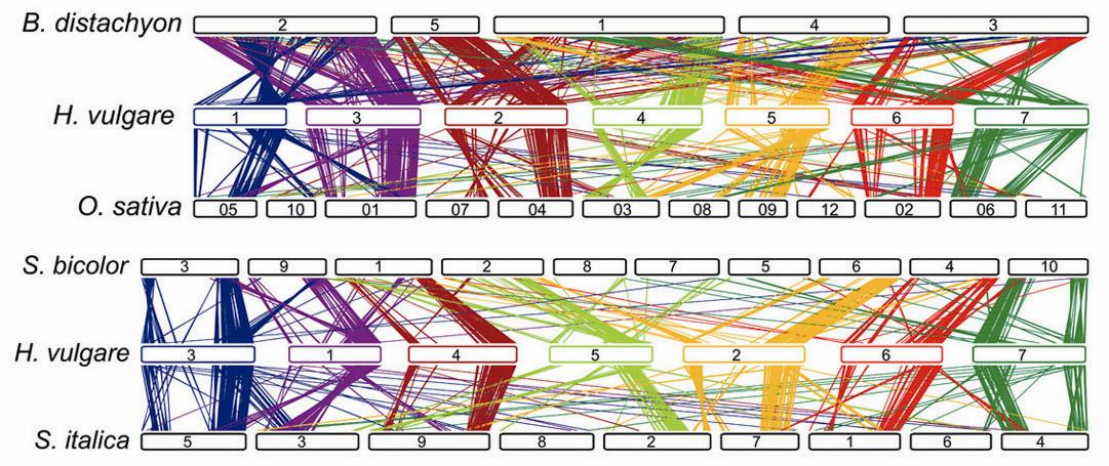


Fig. S14, Syntenic blocks for *B. distachyon–H. vulgare–O. sativa* (up) and *S. bicolor–H. vulgare–S. italic* (down) chromosomes

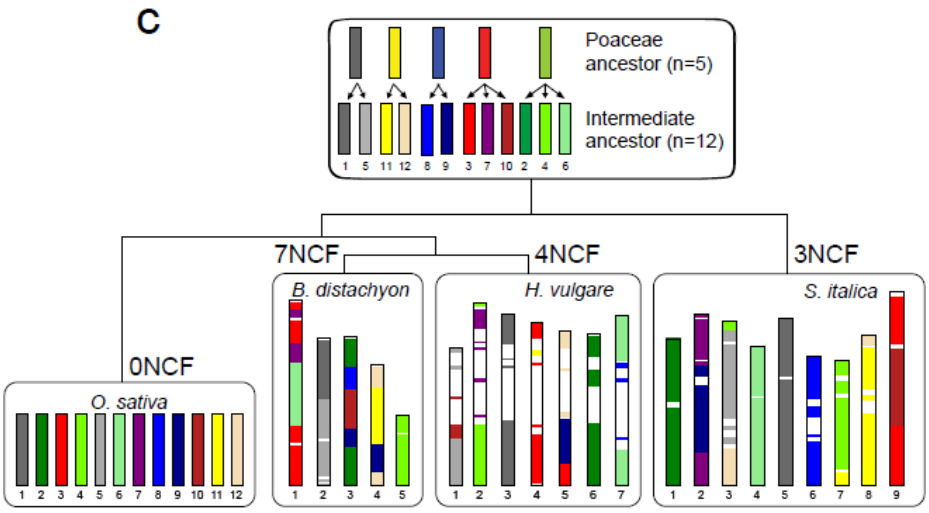


Fig. S15, Chromosome fusion and evolution among four Poaceae species. It shows the Poaceae ancestor contains five chromosomes, then duplicates to 12 chromosomes in the intermediate ancestor. We found four main NCF for *H. vulgare* chromosomes compared to the intermediate ancestor, while 0 NCF, 7 NCF, and 3 NCF events occurred from the intermediate ancestor to generate the *O. sativa*, *B. distachyon*, and *S. italica* chromosomes. NCF represents the nested chromosome fusion event.

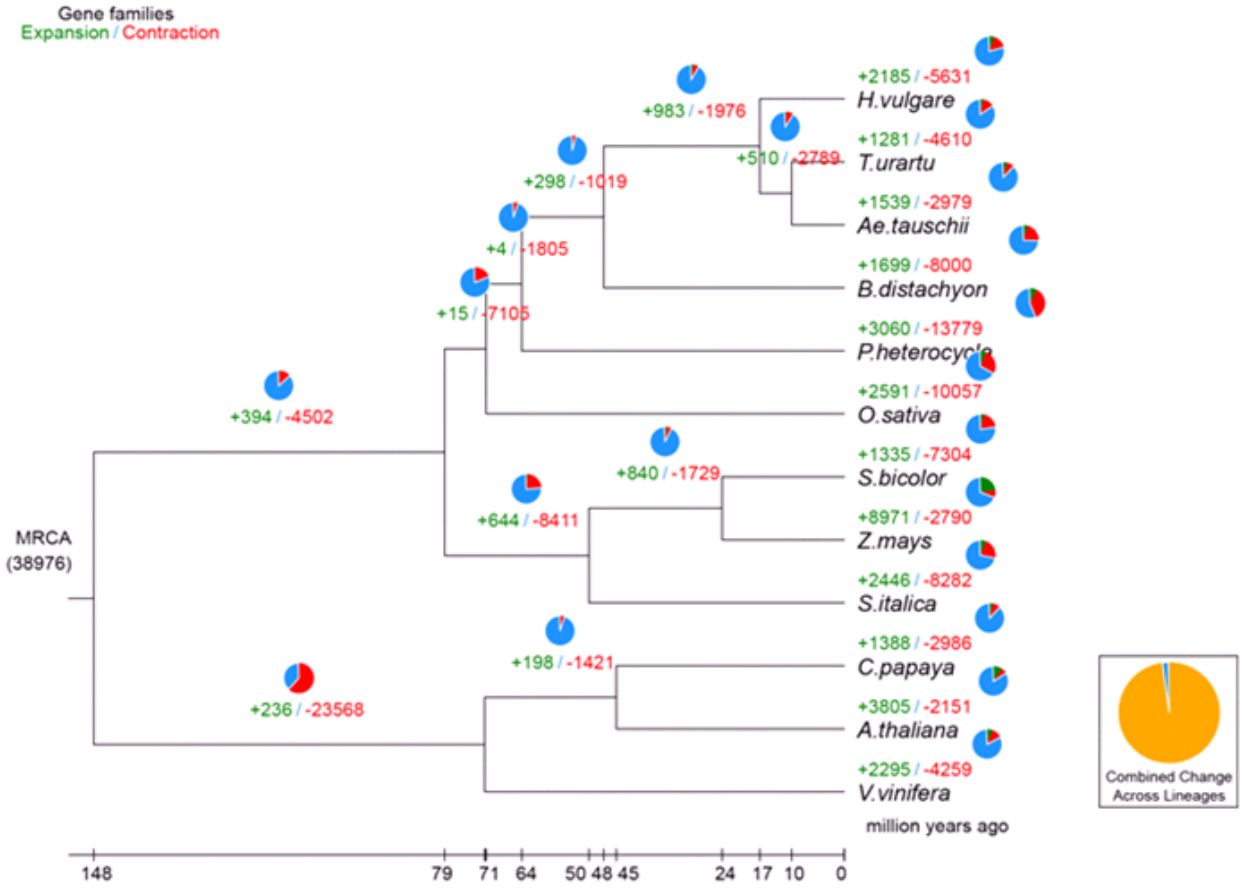


Fig. S16, Expansion and contraction of gene families of barley compared to other plant species. Numbers below branches indicate the number of expanded (green) and contracted (red) gene families. There were 38,976 gene families involved in the analysis, which is shown as MRCA (most recent common ancestor).

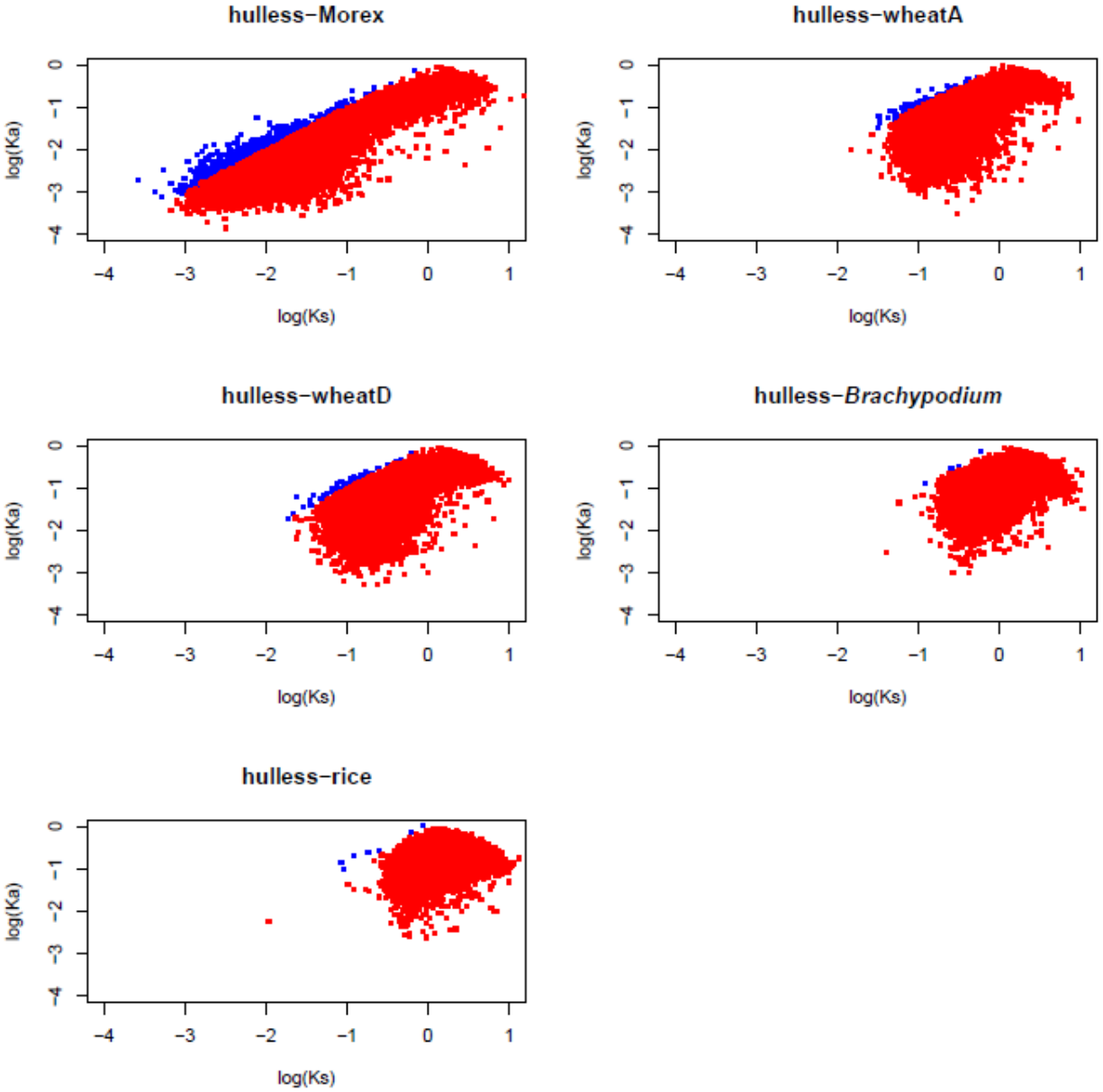


Fig. S17, Ka-vs-Ks dot plot distribution for *H. vulgare* L. var. *nudum* – *H. vulgare* L. cv. Morex gene pairs and *H. vulgare* L. var. *nudum* – *B. distachyon, T. urartu, A. tauschii, O. sativa* orthologous gene pairs. hulless, Morex, wheatA, wheatD, *Brachypodium*, rice represents *H. vulgare* L. var. *nudum, H. vulgare* L. cv. Morex, *T. urartu, A. tauschii, B. distachyon*,and *O. sativa,* respectively.

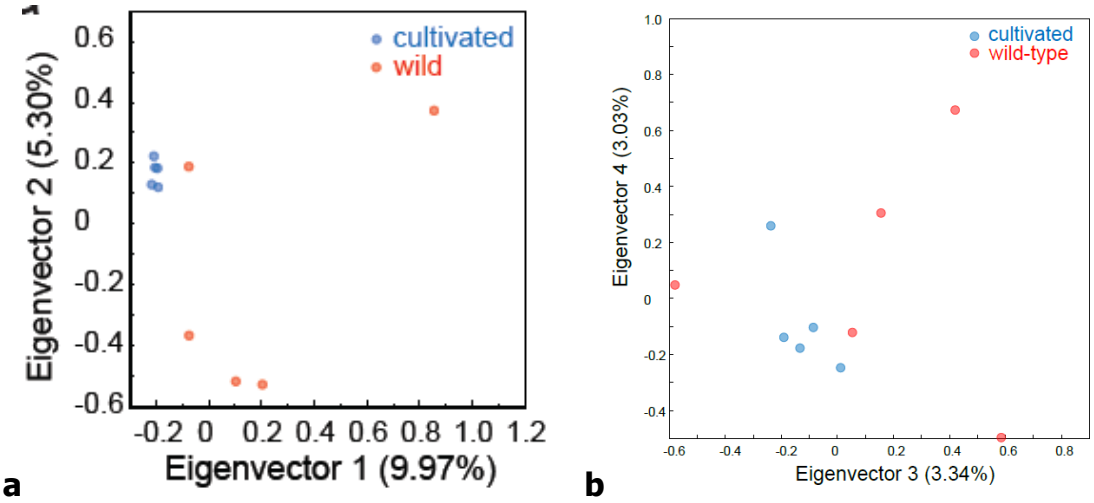


Fig. S18, The 1st, 2nd (a), 3rd, and 4th (b) elements of PCA analysis for 10 Tibetan hulless barley individuals including five wild type and five cultivated barleys.

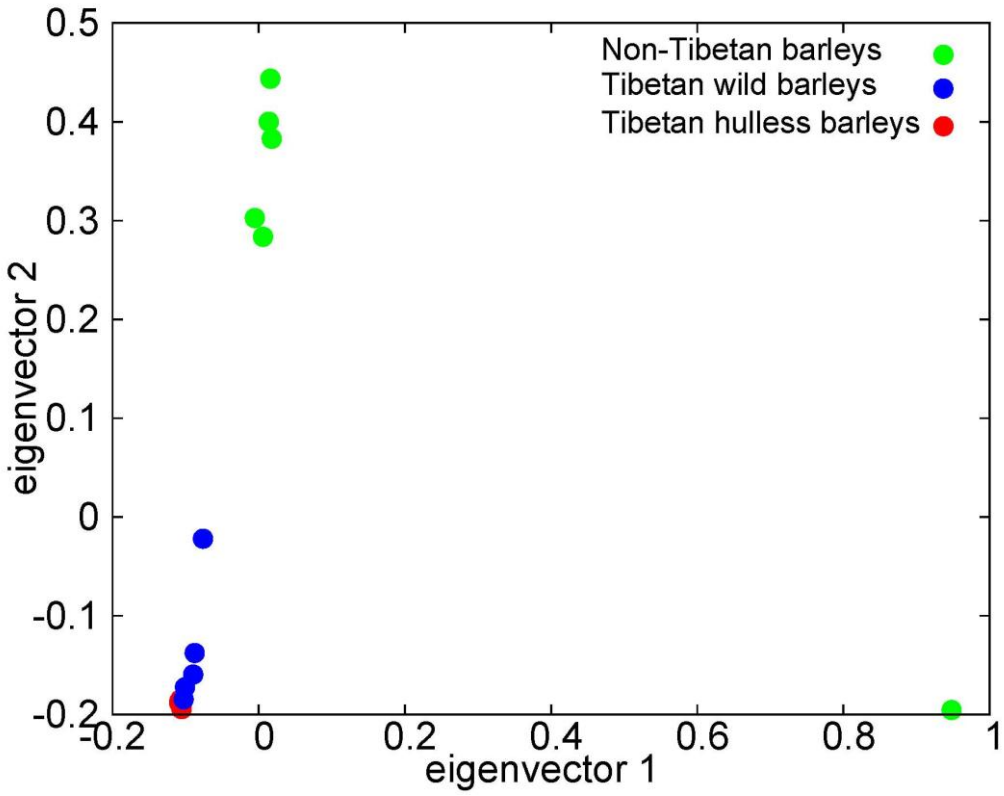


Fig. S19. The first and second eigenvectors of the PCA analysis of barleys. The green circles indicate Non-Tibetan barleys. The blue circles indicate Tibetan wild barleys. The red circles indicate Tibetan hulless barleys.

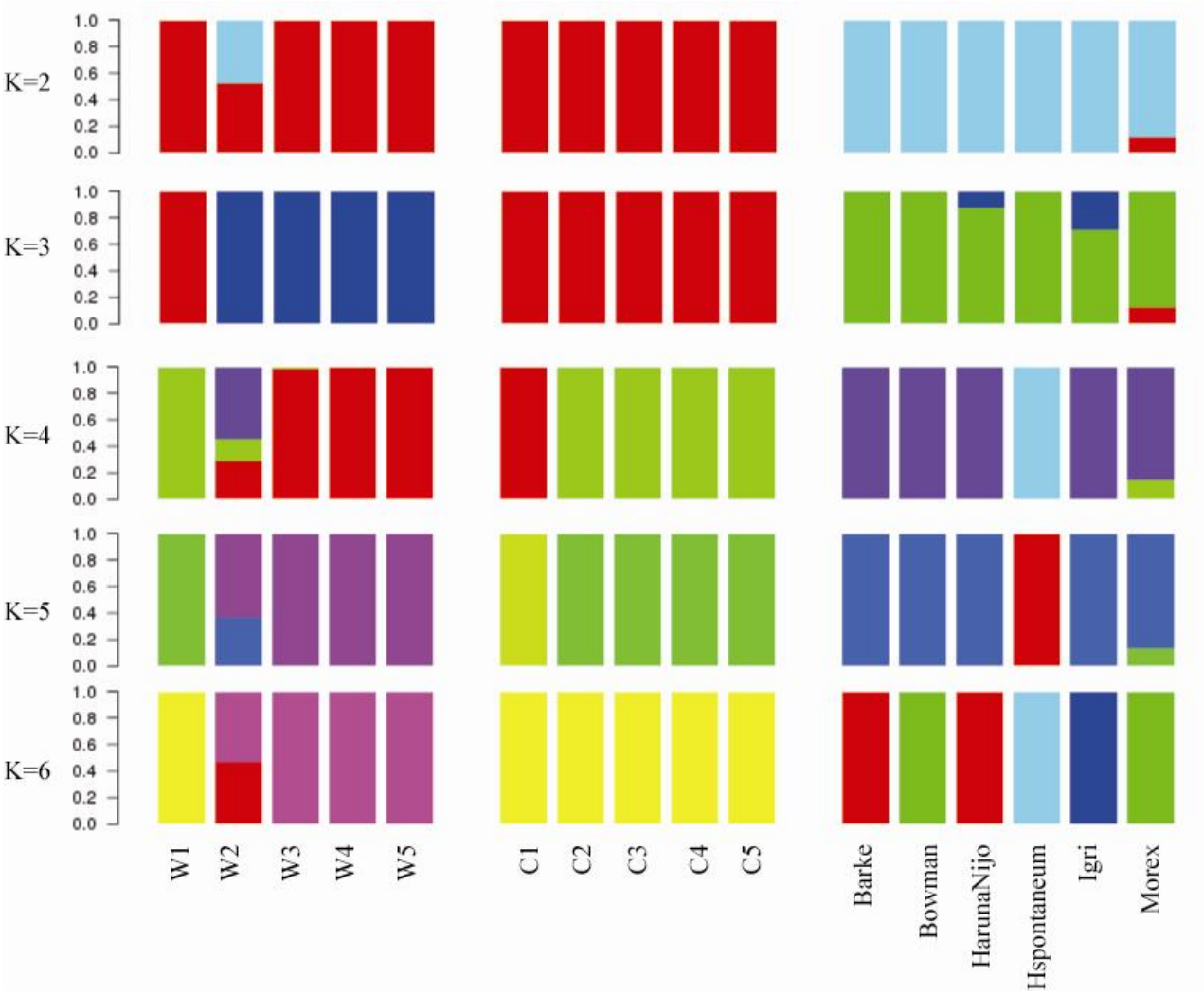


Fig. S20. The population structure analysis of barleys. Each color denotes one population in the population structure analysis. Each vertical bar represents one accession in which the percentages of contribution from the ancestral populations are indicated by the lengths of colored segments. The number of clusters (K) was set from 2 to 6.

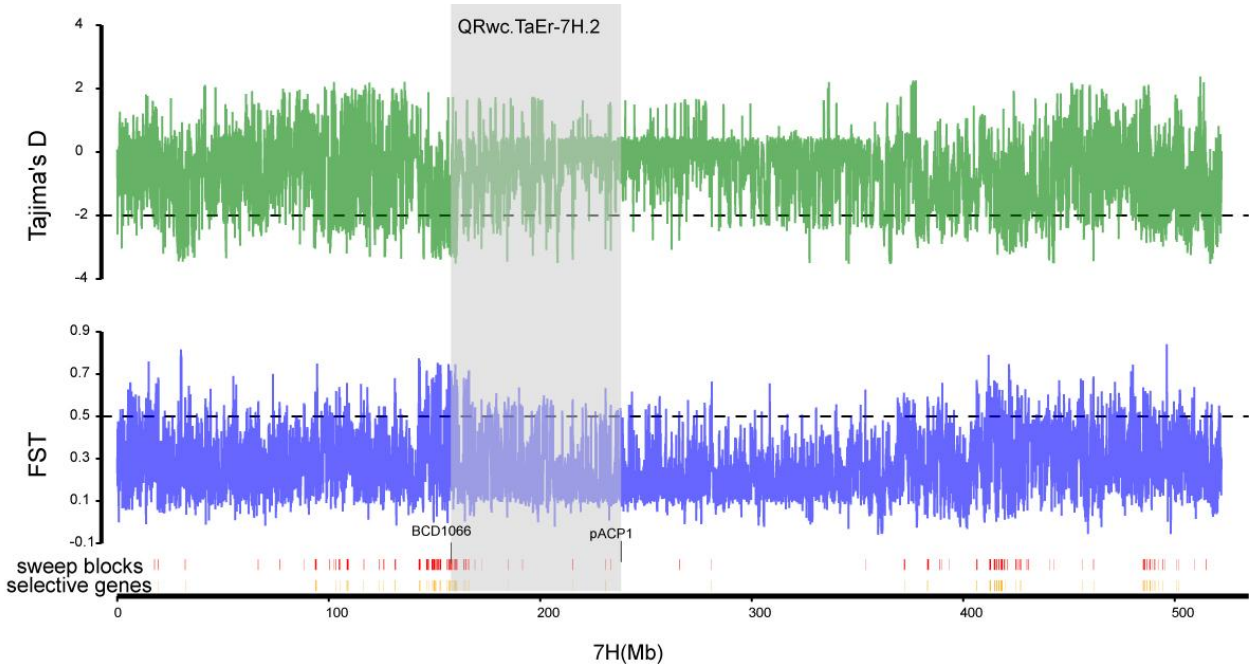


Fig. S21. The selective sweep of 7H chromosome. Tajima's D, Fst, sweep blocks and selective genes were shown from top to bottom. The QTL of QRwc.TaEr-7H.2 was shown in grey background with the highlighted franking marker BCD1066 and pACP1.