

BarPlex v1.0: a multiplex PCR-based enrichment of genome-wide short segments that enable genetic studies in barley

Guangqi Gao

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Luxi Yan

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Yu Cai

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Congcong Jiang

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Qiang He

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Sarah Tasnim

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Zongyun Feng

Sichuan Agricultural University

Jun Liu

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Jing Zhang

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Ping Yang (✉ yangping@caas.cn)

Institute of Crop Sciences, CAAS <https://orcid.org/0000-0002-6977-8449>

Research Article

Keywords: Barley, Tibetan semi-wild barley, *H. agriocrithon*, genotyping, multiplexed enrichment

Posted Date: October 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-979583/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Over the past 15 years sequencing methodologies have advanced greatly enabling high-throughput sequencing based genotyping of crop plants. In this study, we developed BarPlex v1.0, a robust and cost-efficient genotyping approach in barley (*Hordeum vulgare* L.). In this multiplex PCR-based amplification of five-hundred genome-wide segments, followed by high-throughput sequencing of barcoded PCR products, we obtained hundreds to thousands of polymorphic markers. Comparison of genotyping with BarPlex v1.0 to genotyping-by-sequencing (GBS) revealed a similar genetic diversity. The polymorphic markers revealed by BarPlex v1.0 were highly accurate, with an average sequencing depth >700x and a data missing rate <0.5%. By analyzing 1,068 genotypes of wild barley (brittle rachis; *Hordeum vulgare* ssp. *spontaneum* L.), Tibetan semi-wild barley (brittle rachis), landraces, cultivars, as well as an F₂ population, this assay has been robust in studies of population diversity, variety pedigree, heterozygosity discrimination, linkage mapping, as well as genome-wide association study (GWAS). Notably, a diversity analysis in a population of Tibetan semi-wild barley suggested a close relationship with Chinese landraces, but a dramatic decrease in its genetic diversity, inferring that Tibet was not a center of domestication for the native wild barley.

Introduction

Cultivated barley (*Hordeum vulgare* ssp. *vulgare* L.) is among the four most important cereal crops for both animal and human consumption (including brewing). In 2019, the global yield production reached 3.1 ton per hectare, and was 2.3-fold higher than that of 1961 (FAO dataset, 2021). Factors contributing to this increase include a significant improvement in irrigation supply, fertilizer, and fungicide. Another important factor is that the use of elite varieties with improved yield production and resistance to biotic/abiotic stresses, relying on exploration and optimal use of agriculturally-important genes (e.g., barley Green Evolution genes *sdw1* and *uzu/BRI1*, and the powdery mildew resistance gene *mlo*) (Buschges et al. 1997; Chono et al. 2003; Kuczynska et al. 2013). Marker-assisted incorporation of resistance genes *rym4* and *rym5* conferred reliable protection from barley yellow mosaic virus disease, which had been severe in Europe since the 1980s (Jiang et al. 2020; Stein et al. 2005).

Over the past few decades different marker technologies, from hybridization-based restriction fragment length polymorphism (RFLP) markers to PCR-based simple sequence repeat (SSR) markers, have been well established in barley (Graner et al. 2010). These markers have contributed to theoretical and applied genetic studies including assessment of population diversity, genetic mapping, map-based gene isolation (Komatsuda et al. 2007; Taketa et al. 2008), and marker-assisted breeding (Ordon et al. 2005). The availability of genomic resources has enabled the development of high-throughput genotyping assays such as SNP arrays (BOPA-1 and BOPA-2; Illumina 9K SNP chip; 50K Illumina Infinium iSelect array) (Bayer et al. 2017; Close et al. 2009; Comadran et al. 2012). In contrast to SNP arrays requiring prior identification of polymorphic sites, a complexity-reduced sequencing approach (genotyping-by-sequencing, GBS) offers a flexible choice for genotyping a previously unclassified species and detecting minor frequency polymorphisms (Darrier et al. 2019; Poland et al. 2012). The relatively low cost and

higher marker density allowed genotyping of >22,000 Genebank-hosted barley accessions (Milner et al. 2019). Along with the dramatic drop in the cost of high-throughput sequencing, exome-sequencing of hundreds of geographically diversified barley genotypes and whole-genome re-sequencing (WGR) of wheat accessions became affordable, thus generating millions of SNPs and indels (Russell et al. 2016; Zhou et al. 2020). The tremendous increase in the number of markers is allowing studies in linkage disequilibrium (LD) based genome-wide association study (GWAS), bulked segregant analysis (BSA), and genomic selection (GS).

Although the sequencing cost significantly decreased in the past few years, genotyping by GBS or WGR of every Genebank accession or all the lines in a breeding program remains costly. Gigabyte or terabyte scale of sequences generated from high-throughput sequencing techniques are difficult to handle by the majority of biologists and geneticists (especially breeders). For most genetic studies, a moderate marker density is often sufficient (Abed et al. 2018; Casas et al. 2021; Moragues et al. 2010). For example, deployment of 384 SNPs was effective in detecting the genetic diversity within barley germplasms (Moragues et al. 2010), and to delimit the ALBOSTRIANS gene *HvAST* in a map-based approach (Li et al. 2019). Use of 38 SNP markers has successfully clarified most of barley cultivars that are valuable on seed certificate (Owen et al. 2019).

In this study, we deployed multiplexed PCR amplification, followed by high-throughput sequencing, to establish a robust genotyping assay (designated BarPlex v1.0) that targets five-hundred genome-wide short fragments with an accumulative size of 91.1 kb, producing hundreds to thousands of polymorphic SNPs. Since this assay is cost-efficient and flexible to incorporate additional segments/genetic markers which associate with particular phenotypic traits, it should gain great interest in genetic studies of barley, and provide a reference for other species.

Materials And Methods

Plant materials and phenotyping

1,068 genotypes consisting of 51 wild barley (*H. vulgare* ssp. *spontaneum* L., brittle rachis), 248 Tibetan semi-wild barley (including *H. agriocrithon*, the six-rowed barley with brittle rachis), 345 barley landraces, 329 barley cultivars, as well as an F₂ population with 95 segregants were analyzed. Passport information of each genotype is given (Supplementary Table 1). The F₂ population was generated by a cross between the barley accessions “Bin Hai Yu Da Mai” (National Crop Genebank of China: ZDM02420, two-rowed/hulled) and “Bai Qing Ke” (ZDM04577, six-rowed/naked). Two-week-old plants were moved into a vernalization chamber for 35 days growth (4 °C with day/10 h, night/14 h), followed by cultivation under normal glasshouse conditions (day 22 °C/14 h and night 18 °C/10 h) until full maturity.

The adherence of caryopsis (naked or hulled) of each F₂ plant was scored after harvest and hand-threshing. Spike brittleness for *H. spontaneum* and Tibetan semi-wild barley was determined as described in an earlier study (Pourkheirandish et al. 2015).

DNA extraction and quantification

Two-leaf-old seedlings were harvested for DNA extraction following a previously described method (Shi et al. 2019). DNA samples were qualified by agarose gel electrophoresis and quantified using Qubit 3.0 (Thermo Fisher, USA).

BarPlex v1.0: Primer design, PCR amplification, library construction, high-throughput sequencing, and bioinformatics analysis

Whole-genome re-sequencing (0.6x) was conducted following a previously-established workflow (Zeng et al. 2018). Short-read mapping and calling for SNPs were performed in the identification of polymorphic sites between two landraces WDM06177 and ZDM00809 (NCBI accession IDs: SRR15661624 and SRR15661625). Flanking sequences of the SNP sites were extracted from the barley reference genome (Morex v2) (Monat et al. 2019), and subjected to primer pickup (size of PCR product: 140 to 350 bp) using BatchPrimer 3 (You et al. 2008). Out of 587 primers, the primers producing multiple fragments (un-specific amplification) in test experiments were discarded. The first round of PCR amplification using the GenoPlexs multiple PCR amplification kit was performed in a volume of 30 μ L containing 10 ng genomic template DNA, 10 μ L Genoplex Master Mix (3x, including high-fidelity polymorphism and PCR buffers) as well as mixed primers of an equal molar. The PCR reaction included denaturation at 95 °C for 5 min, followed by six cycles of 30 sec at 95 °C, 4 min at 60 °C, and a final extension at 72 °C for 5 min. PCR products were purified by adding 15 μ L GenoPrep DNA Clean Beads solution, followed by two steps of washing with 75% ethanol. The purified PCR products, which serve as PCR templates, were subjected for the second round of PCR amplification by adding 1 μ L of barcode solution and 10 μ L 3x Genoplex Master Mix, with the same PCR reaction procedure as above. PCR products were purified as described and eluted with 30 μ L Tris-HCl solution (pH = 8). PCR products were qualified by agarose gel electrophoresis and quantified using Qubit 3.0 (Thermo Fisher, USA). The barcoded PCR products were combined with an equal molar ratio, and were sequenced using Illumina Hiseq X ten or DNBSEQ-T7.

The raw reads were filtered using fastp v0.20.0 (Chen et al. 2018) with parameters of $-q = 20$, $-u = 40$, and $-n = 10$, to remove low quality reads and adapter sequences. Clean reads were mapped to the barley reference genome (Morex v2) (Monat et al. 2019) by using BWA-MEM v0.7.17 (Li and Durbin 2009). SNP calling was performed using the UnifiedGenotype function of GATK v3.5.0 (McKenna et al. 2010) with parameters of $-dcov 1000000$, $-minIndelFrac 0.15$, $-glm BOTH$, and $-I INFO$, followed by a filtration using the VariantFiltration function with parameters of $MQ0 \geq 4 \ \&\& \ (MQ0 / (1.0 * DP) > 0.1)$ and $DP < 5 \ || \ QD < 2$. SNP/InDel sites with a depth $< 5x$ were marked "NA". The homozygosity or heterozygosity of SNPs/Indels were identified based on the allele frequency where homozygous $AF \geq 0.8$ or $AF \leq 0.2$, and heterozygous $0.2 < AF < 0.8$. Each allele was read ≥ 5 times. Polymorphic sites were filtered (with a miss rate $< 20\%$, minor allele frequency (MAF) > 0.01 , and heterozygosity $< 10\%$) using Vcftools v0.1.17 (Danecek et al. 2011).

Genotyping-by-sequencing (GBS)

100 ng samples of genomic DNA were digested for 2 h at 37 °C with 10 U each of *Pst*I and *Msp*I (New England Biolabs, Ipswich, MA) in a 50 µL reaction volume containing 1× NEB Buffer 2.1. In-house pre-treated adapters were then ligated to sticky ends by adding 30 µL of a master mix solution containing 4 U of T4 ligase (New England Biolabs) in each reaction. Samples were incubated at 22 °C for 2 h and heated at 65 °C for 10 min to inactivate the T4 ligase. Ligated samples were purified using DNA clean beads (Automag, USA) according to the manufacturer's instructions. Index PCR reaction was then performed in a 50 µL volume using HotStart ReadyMix (Kapabiosystems, USA). These amplified products were purified as described above and quantified using Qubit 3.0. Equal amounts of purified products were pooled for agarose gel electrophoresis. Fragments with sizes of 300-500 bp were extracted from the gel, purified by column tube, and quantified using Qubit 3.0 before sequencing on Illumina HiSeq X Ten.

Quality control was conducted by using fastp v0.20.0 (Chen et al. 2018) with the default parameters. After trimming low quality sequences using Trimmomatic v0.39 (Bolger et al. 2014) with the parameters LEADING:3, TRAILING:3, SLIDINGWINDOW:10:15, and MINLEN:150, the qualified paired reads were mapped to the reference genome using BWA-MEM v0.7.17 with the default parameters (Li and Durbin 2009). Once the binary alignment map (BAM) was generated, alignment reads were sorted by reference position, filtered by -q 30 and -F 3332 with SAMtools v1.10 (Li et al. 2009). Duplicates derived from PCR amplification were marked with GATK4 (McKenna et al. 2010) and removed with SAMtools. SNPs/Indels were called with GATK HaplotypeCaller and filtered with GATK VariantFiltration. The polymorphic sites were filtered (with miss rate < 20%, minor allele frequency > 0.01, heterozygosity < 10%, and minDP = 2) using Vcftools v0.1.17.

Principal component analysis, genetic diversity analysis, genome-wide association study (GWAS), and linkage mapping

Principal component analysis (PCA) was performed with EIGENSOFT v7.2.1 with the smartpca function (Patterson et al. 2006). The kinship matrix was calculated by Tassel version 5 (Bradbury et al. 2007). Nucleotide diversity (π), and the strength of selection (Fu and Li's D^* and F^* , and Tajima's D) were calculated by using DNASP v6.12.01 (Librado and Rozas 2009). A genome-wide association study was performed with the rMVP package with general linear model (GLM), mixed-linear model (MLM), as well as fixed and random model circulating probability unification (FarmCPU) models (Yin et al. 2021). Linkage mapping was performed using JoinMap v4 (Ooijen and Van 2006) with regression algorithm and the Kosambi function.

Results

Assembling a panel of genome-wide short segments for genotyping

Target fragments harboring polymorphic SNPs were derived from three sources, including (1) whole genome re-sequencing of two landraces, (2) GBS-derived polymorphisms in a global barley diversity panel (Milner et al. 2019), and (3) SNPs reported in a 50K Illumina Infinium iSelect SNP array (Bayer et al. 2017). Flanking sequences were extracted from the barley reference genome (Morex v2) and subjected to

primer pickup. Out of 587 primer pairs tested for multiplex PCR reaction followed by high-throughput sequencing, 87 were discarded due to either a lower capacity of fragment capture, production of multiple fragments, or extensive/low PCR amplification efficiency. As a result, five hundred primer pairs were qualified applicable for barley multiplex PCR amplification (BarPlex v1.0) (Supplementary Table 2). The number of target fragments distributed on each chromosome varied from 54 to 84 (Fig. 1a), with higher density towards the telomeres (where a higher recombination rate is generally present) (Mascher et al. 2017).

We conducted four independent experiments with analysis of the 1,068 genotypes that included 51 wild barley (*Hs*), 248 Tibetan semi-wild barley, 345 landraces, 329 cultivars, as well as 95 F₂ segregants (Table 1). The detection of target fragments in each experiment varied from 99.4% to 99.8% (Fig. 1b), with the average sequencing depth (x) between 467 and 1010 (Fig. 1c). For each of the 1,068 genotypes, the detection rate ranged from 96.4% to 100% with a mean of 99.7% (Fig. 1d), and the average sequencing depth across all samples was 757 (Fig. 1e). In addition, we analyzed the detection rate and sequencing depth of the 500 targets across all samples. Of these, 466 targets (93.2%) were detected in >99.5% samples, whereas only four target fragments were detected in less than 90% samples (Fig. 1f-g). We did not observe a significant difference in the number or size of detected fragments in the populations of wild, Tibetan semi-wild, or cultivated barley (Table 1). In comparison to barley landraces and cultivars, fewer polymorphic target SNPs were observed in the populations of *H. spontaneum* and Tibetan semi-wild barley. This is likely due to the origin of target SNPs which were identified from the population of cultivated barley. By identifying the polymorphic sites that were derived from the 91.1 kb of target sequences, a higher number of SNPs was observed in *H. spontaneum*, but not among Tibetan semi-wild barley. Considering the detection rate and sequencing depth, as well as the number of polymorphisms, we would like to conclude that BarPlex v1.0 is a robust and complexity-reduction assay applicable for genotyping in barley plants.

BarPlex v1.0 vs. genotyping-by-sequencing (GBS)

GBS is a complexity-reduced and cost-efficient genotyping approach that has been widely applied in genetic studies of crop plants with complex genomes (Milner et al. 2019; Poland et al. 2012). To test if BarPlex v1.0 works as efficient as GBS in barley, genotyping of 96 barley landraces or cultivars using GBS was conducted, yielding 0.7-Gb of clean sequences per accession (8-fold more than that in the study of sequencing >22,000 Genebank barley accessions) (Milner et al. 2019). GBS unlocked 24,195 qualified polymorphic sites, while 1,372 polymorphisms were detected using BarPlex v1.0 (Fig. 2a). In each sample, 99.2%-100.0% (mean = 99.7%) of the target fragments were detectable using BarPlex v1.0, whereas the detection rate using GBS after quality control (removing missing rate >20%) was 81.5% - 97.6% (Fig. 2b). The sequencing depth using BarPlex v1.0 was 310-fold higher than that of GBS (Fig. 2c). Each of the SNPs in BarPlex v1.0 was detected in 83.3% - 100% (mean = 99.9%) of samples, whereas this number in GBS arranged from 80.2% to 100% with a mean of 91.5% (Fig. 2d-e). Based on the principal component analysis (PCA) with the dataset from either BarPlex v1.0 or GBS, a similarity of unlocking the

population structure diversity was observed (Fig. 2f-g). With these results, we would like to conclude that BarPlex v1.0 is a reliable and highly accurate assay informative for genotyping in barley.

Applications for heterozygosity discrimination, variety pedigree, and linkage mapping and GWAS

We further checked if polymorphisms revealed by BarPlex v1.0 were applicable in genetic studies of barley. Here, 495 target SNPs, as well as their residing fragments (90.3 Kb in total) representing 3,220 polymorphic sites were included. First, the heterozygosity revealed by quantifying the allele frequency on polymorphic sites was analyzed (Fig. 3a). All the individuals in the F₂ population remained heterozygous at a proportion of polymorphic sites (mean = 6.5%). In comparison, only ten samples of 973 in natural materials were found to remain heterozygous (cutoff: $\geq 4\%$). Moreover, we analyzed the heterozygosity revealed by target SNPs. An overestimate in comparison to that revealed by polymorphic sites, was observed (Fig. 3a).

Second, comparisons of various landraces/cultivars values using both target SNPs and polymorphic sites revealed interesting results. “Morex”, a six-rowed US malting variety released in 1978, has been used as the reference genome sequence in barley (Mascher et al. 2017). “Morex” seeds provided by two breeding units in the downstream valley of Yangtze River in China differed from the original “Morex” (Fig. 3b). In contrast to this, three Chinese landraces/historic cultivars which have been cultivated widely during last century (“Chi Ba Da Mai”, “Xiu Ning Ai Jiao Da Mai”, and “Chi Ba Huang”) were found to have a higher identity (Fig. 3c). Comparison of the two-rowed malting barley cultivar “Zao Shu 3 Hao” (Kanto Nijo 3), which was introduced from Japan in 1960s, to its Co⁶⁰-radiation variant “Yan Fu Ai Zao 3” revealed a semi-dwarf mutation. Both “Zao Shu 3 Hao” and “Yan Fu Ai Zao 3” are founder varieties that were widely cultivated in 1970s and 1980s in China, and their genetic similarity was traceable using BarPlex v1.0 (Fig. 3d). The historic cultivar “Yu Da Mai 1 Hao” was reported to be developed from a mutation of “Zao Shu 3 Hao”, whereas the genotyping result suggested it may be derived from an out-pollination rather than a mutation (Fig. 3d). Notably, the polymorphic sites from the 90.3 kb of sequences showed a better resolution on variety pedigree discrimination than 495 target SNPs.

Third, we further examined if BarPlex v1.0 is applicable for preliminary genetic mapping. Cultivated barley is traditionally classified into hulled or naked barley, according to the adherence of the caryopsis in fully mature grains (Taketa et al. 2008). The naked trait (*NUD*) is inherited qualitatively and controlled by a loss-of-function allele of the ethylene response factor (*ERF*) gene that resides on chromosome 7HL (Lei et al. 2020; Taketa et al. 2008). We conducted a genome-wide association study (GWAS) in 973 barley accessions (excluding 95 F₂ lines out of 1068 samples) with 3,220 polymorphic sites, and a single peak associating with the *NUD* locus was revealed (Fig. 4a). In addition, linkage mapping using a bi-parental F₂ population was conducted by genotyping with BarPlex v1.0 (Fig. 4b-c). Here, 108 target SNPs and 442 polymorphic sites between both parents were revealed on the seven chromosomes (Table 1). The naked trait was delimited to a 40.2 cM genetic interval spanning approximately 20 Mb (Fig. 4d) where the previously-identified *NUD* locus was present (Taketa et al. 2008).

Collectively, these results indicated that BarPlex v1.0 is applicable in multiple barley genetic studies.

Tibetan semi-wild barley represented genetic similarity with Chinese landraces but lower diversity

Tibetan semi-wild barley, also referred as Tibetan weedy barley (Zeng et al. 2018), has a characteristic brittle rachis that is also seen in wild barley *H. spontaneum* of the Near East. This has raised a hypothesis that Tibet may be an independent domestication center of native wild barley (Aberg, 1938; (Dai et al. 2014). Although this hypothesis was questioned by re-sequencing and phylogenetic analyses of the brittle rachis domestication genes *Btr1* and *Btr2* (Pourkheirandish et al. 2018), the population diversity of Tibetan semi-wild barley (mainly including *H. agriocrithon*), if compared to wild or cultivated barley population, remained ambiguous. Using the polymorphic sites revealed from BarPlex v1.0, we investigated the genetic diversity in a large panel of barley accessions including 51 *H. spontaneum*, 248 representative Tibetan semi-wild barley, 345 Chinese barley landraces, and 329 cultivars. A higher number of polymorphic sites were detected in the *H. spontaneum* population than in Tibetan semi-wild barley (2,652 vs. 1,496), even though the analyzed population size of *H. spontaneum* was considerably smaller (51 vs. 248) (Table 1). With the polymorphic sites, principal component analysis (PCA) revealed that *H. spontaneum* constitutes a group isolating from domesticated barley, while the Tibetan semi-wild barley was grouped with Chinese barley landraces (Fig. 5a). A rapid decrease in genetic diversity (specifically, nucleotide diversity, π) from Chinese barley landraces to Tibetan semi-wild barley was also observed (Table 2). Selections represented by the estimated significant negative values of Tajima's D , as well as Fu and Li's D^* and F^* suggested a founder effect occurring in Tibetan semi-wild barley population. Statistical analysis using pre-selected target SNPs derived from cultivated barley may give an underestimation of the genetic diversity within the *H. spontaneum* population (Table 1), thus resulting in a deviation of the genetic relationships in principal component analysis (Fig. 5b). The genetic diversity and the PCA analysis in a large panel of Tibetan semi-wild barley, Chinese barley landraces, and *H. spontaneum* inferred that Tibet was not an independent center of domestication for the native wild barley.

Discussion

High-throughput sequencing technology enabled establishment of genotyping approaches that have revolutionized genetic studies in crop plants (Scheben et al. 2017; Torkamaneh et al. 2018). Targeted sequencing of a subset of the genomic regions by either hybridization or multiplexed PCR amplification provides a considerably simplified approach to genotyping (Gasc et al. 2016; Shirasawa et al. 2016). Using this strategy, we established a high-throughput genotyping assay (BarPlex v1.0) that combines multiplexed PCR (that amplifies 500 genome-wide targeted segments of barley) with high-throughput sequencing of a pool of barcoded PCR products. By genotyping thousand samples from multiple experimental runs, this assay has identified target SNPs (with 54 to 84 SNPs on each chromosome), as well as additional polymorphic sites on 91.1 kb of combined sequences. Importantly, this method has worked well with several types of barley (*H. spontaneum*, Tibetan semi-wild barley, landraces, and cultivars, as well as a segregation population) in genetic studies with multiple objectives (i.e., population

genetic diversity, variety pedigree, heterozygosity discrimination, preliminary mapping, and GWAS). In contrast to the popular complexity-reduced GBS methodology, fewer polymorphisms were revealed by BarPlex v1.0. Here, only 91.1 kb of sequences were captured. Still, this assay is most attractive due to (1) accuracy and reproducibility, considering a sequencing depth of over 700x and a missing rate of less than 0.5%, (2) a customized design that enables incorporation of additional loci/markers especially associated with particular targeted traits, and (3) a lower cost for both data producing and handling (ca. 20% and 5%, if compared to the commercial price charged by GBS and KASP, respectively). Thus, we believe BarPlex v1.0 to be a choice high-throughput and robust method to genotype barley and the genotyping strategy that also works for other species.

In using BarPlex v1.0, this study genotyped a larger panel of barley genotypes. This powerful and high-throughput method can be used to determine genetic similarity and heterozygosity using either target SNPs or polymorphic sites. This methodology was able to classify different varieties, but failed in the case of a barley variety and its mutagenized offspring. This could be resultant to the limited resolution of this assay (if compared to GBS or WGR). For population diversity analysis, we found that genotyping using pre-selected target SNPs can underestimate the genetic diversity in wild barley populations. That might be explained by ascertainment bias, similar to an earlier study that genotyped barley landraces by SNP array (Moragues et al. 2010). By applying the polymorphic sites revealed by BarPlex v1.0, we were able to detect minor alleles as identified by GBS (Darrier et al. 2019), thus overcoming the bias of genotyping assays (i.e., SNP array) that only rely on pre-identified SNPs (Eagle et al. 2021; Moragues et al. 2010; Rasheed et al. 2017). We tested BarPlex v1.0 to identify the genetic factor controlling the adherence of caryopsis (hulled *vs.* naked) by both linkage mapping and GWAS. Both methods were able to identify the *nud* locus, the determinant of the naked barley. The higher marker density in most of the cases (i.e., GWAS) enabled delimiting the gene of interest in a higher resolution, whereas the marker density correlated with costs and bioinformatics procedures. Several studies using optimal marker density have answered particular questions (Casas et al. 2021; Li et al. 2019; Moragues et al. 2010). Thus, choosing a genotyping platform (BarPlex, GBS, or WGR) by consideration of laboratory effort, costs, and bioinformatics capability is highly recommended.

Cultivated barley (rough rachis) was domesticated from its wild progenitor *H. spontaneum* (brittle rachis) in the Fertile Crescent about 10,000 years ago. Cultivated barley carries the loss of functional alleles (*btr1* or *btr2*) at either *Btr1* or *Btr2* locus (Pourkheirandish et al. 2015), while Tibetan semi-wild barley has functional *Btr1* and *Btr2* as in wild barley *H. spontaneum*. By genotyping in a large collection of Tibetan semi-wild barley accessions as well as wild barley, landraces, and cultivars, this study uncovered a dramatic decrease in genetic diversity within the Tibetan semi-wild barley population. This is consistent with a previous study involving a population of cultivated hullless barley “qingke” (Zeng et al. 2018). Without enduring the domestication bottleneck, wild progenitors (i.e., *H. spontaneum*) are expected to host higher genetic diversity and abundant minor alleles (Milner et al. 2019). However, neither of these characters were observed in Tibetan semi-wild barley population, thus questioning Tibet as a center of domestication for the native wild barley. Tibetan semi-wild barley accessions were mixed with Chinese barley landraces, but separated from *H. spontaneum*. This implied that Tibetan semi-wild barley was

unlikely derived from mutations of *H. spontaneum* or hybridization between *H. spontaneum* and local cultivated barley, while both models were proposed to explain the origin of *H. agriocrithon*, the six-rowed Tibetan semi-wild barley (Berg 1940; Tanno and Takeda 2004). Unlocking the sequence diversity in dozens of Tibetan semi-wild barley accessions confirmed that *Btr1* and *Btr2* haplotypes were derived from cultivated barley, suggesting Tibetan semi-wild barley originated from hybridization between two independent domestication events (*btr1Btr2* and *Btr1btr2* in barley landraces), followed by combination and reoccurrence of the brittle rachis (*Btr1Btr2*) (Pourkheirandish et al. 2018). This explanation aligns with our results showing that a genetic similarity between Tibetan semi-wild barley and Chinese barley landraces/cultivars was higher than with *H. spontaneum*.

Declarations

Data availability statement

The short reads generated by high-throughput Illumina sequencing have been deposited in the public NCBI database. The accession IDs for each entry are listed in Supplementary Table 1.

Conflict of interest

The authors declare no conflicts of interest.

Author Contribution Statement

PY, JZ, JL and ZF designed research; GG, LY, YC and ST performed experiments; GG, PY and QH analyzed data; PY, GG, and CJ wrote the paper.

Acknowledgements

The authors would like to thank the National Crop Genebank of China, Drs. Rugen Xu (Yangzhou University), Huiquan Shen (Institute of Agricultural Science in Jiangsu Coastal Areas), Junmei Wang (Zhejiang Academy of Agricultural Sciences), Chenhong Liu (Shanghai Academy of Agricultural Sciences) for providing seeds, Molbreeding Biotechnology (Shijiazhuang, China) for technique assistance, Dr. Chunchao Wang for advice on sequence analysis, and Ms. Haiying Guan for technical support. This work was supported by National Key Research and Development Program of China (2019YFD1001704/2019YFD1001700), National Natural Science Foundation of China (32071997), as well as Agricultural Science and Technology Innovation Program of CAAS.

References

1. Abed A, Perez-Rodriguez P, Crossa J, Belzile F (2018) When less can be better: How can we make genomic selection more cost-effective and accurate in barley? *Theor Appl Genet* 131:1873–1890
2. Aberg E (1938) *Hordeum agriocrithon* nova sp., a wild six-rowed barley. *Ann Agric Coll Sweden* 6:159–212

3. Aberg E (1940) The taxonomy and phylogeny of *Hordeum* L. Sect. Cerealina Ands. with special reference to Thibetan barleys. *Symb Bot Upsal* 4:1–156
4. Bayer MM, Rapazote-Flores P, Ganai M, Hedley PE, Macaulay M, Plieske J, Ramsay L, Russell J, Shaw PD, Thomas W, Waugh R (2017) Development and evaluation of a barley 50k iSelect SNP array. *Front Plant Sci* 8:1792
5. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
6. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635
7. Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Topsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell* 88:695–705
8. Casas AM, Gazulla CR, Monteagudo A, Cantalapiedra CP, Moralejo M, Pilar Gracia M, Ciudad FJ, Thomas WTB, Molina-Cano JL, Boden S, Contreras-Moreira B, Igartua E (2021) Candidate genes underlying QTL for flowering time and their interactions in a wide spring barley (*Hordeum vulgare* L.) cross. *Crop J* 9:862–872
9. Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:884–890
10. Chono M, Honda I, Zeniya H, Yoneyama K, Saisho D, Takeda K, Takatsuto S, Hoshino T, Watanabe Y (2003) A semidwarf phenotype of barley *uzu* results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. *Plant Physiol* 133:1209–1219
11. Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC Genom* 10:582
12. Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganai M, Shaw P, Bayer M, Thomas W, Marshall D, Hedley P, Tondelli A, Pecchioni N, Francia E, Korzun V, Walther A, Waugh R (2012) Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat Genet* 44:1388–1392
13. Dai F, Chen ZH, Wang X, Li Z, Jin G, Wu D, Cai S, Wang N, Wu F, Nevo E, Zhang G (2014) Transcriptome profiling reveals mosaic genomic origins of modern cultivated barley. *Proc Natl Acad Sci USA* 111:13403–13408
14. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156-2158
15. Darrier B, Russell J, Milner SG, Hedley PE, Shaw PD, Macaulay M, Ramsay LD, Halpin C, Mascher M, Fleury DL, Langridge P, Stein N, Waugh R (2019) A comparison of mainstream genotyping platforms

- for the evaluation and use of barley genetic resources. *Front Plant Sci* 10:544
16. Eagle J, Ruff T, Hooker M, Sthapit S, See D (2021) Genotyping by multiplexed sequencing (GMS) protocol in barley. *Euphytica* 217:77
 17. Gasc C, Peyretailade E, Peyret P (2016) Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids Res* 44:4504–4518
 18. Graner A, Kilian A, Kleinhofs A (2010) Barley genome organization, mapping, and synteny. In: Ullrich SE (ed) *Barley: production, improvement, and uses*. Wiley-Blackwell, Oxford, pp 63–84
 19. Jiang CC, Kan JH, Ordon F, Perovic D, Yang P (2020) Bymovirus-induced yellow mosaic diseases in barley and wheat: viruses, genetic resistances and functional aspects. *Theor Appl Genet* 133:1623–1640
 20. Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci USA* 104:1424–1429
 21. Kuczynska A, Surma M, Adamski T, Mikolajczak K, Krystkowiak K, Ogradowicz P (2013) Effects of the semi-dwarfing *sdw1/denso* gene in barley. *J Appl Genet* 54:381–390
 22. Lei MM, Ali M, Jiang CC, Shen ZZ, Cai Y, Yang P, Feng ZY (2020) Marker-assisted selection in a global barley (*Hordeum vulgare* subsp. *vulgare*) collection revealed a unique genetic determinant of the naked barley controlled by the *nud* locus. *Genet Resour Crop Ev* 67:273–280
 23. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
 24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079
 25. Li M, Hensel G, Mascher M, Melzer M, Budhagatapalli N, Rutten T, Himmelbach A, Beier S, Korzun V, Kumlehn J, Borner T, Stein N (2019) Leaf variegation and impaired chloroplast development caused by a truncated CCT domain gene in albstrians barley. *Plant Cell* 31:1430–1445
 26. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
 27. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, Bayer M, Ramsay L, Liu H, Haberer G, Zhang XQ, Zhang Q, Barrero RA, Li L, Taudien S, Groth M, Felder M, Hastie A, Simkova H, Stankova H, Vrana J, Chan S, Munoz-Amatriain M, Ounit R, Wanamaker S, Bolser D, Colmsee C, Schmutzer T, Aliyeva-Schnorr L, Grasso S, Tanskanen J, Chailyan A, Sampath D, Heavens D, Clissold L, Cao S, Chapman B, Dai F, Han Y, Li H, Li X, Lin C, McCooke JK, Tan C, Wang P, Wang S, Yin S, Zhou G, Poland JA, Bellgard MI, Borisjuk L, Houben A, Dolezel J, Ayling S, Lonardi S, Kersey P, Langridge P, Muehlbauer GJ, Clark MD, Caccamo M, Schulman AH, Mayer KFX, Platzer M, Close TJ, Scholz U, Hansson M, Zhang G, Braumann I, Spannagl M, Li C, Waugh R, Stein N (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544:427–433

28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
29. Milner SG, Jost M, Taketa S, Mazon ER, Himmelbach A, Oppermann M, Weise S, Knupffer H, Basterrechea M, Konig P, Schuler D, Sharma R, Pasam RK, Rutten T, Guo G, Xu D, Zhang J, Herren G, Muller T, Krattinger SG, Keller B, Jiang Y, Gonzalez MY, Zhao Y, Habekuss A, Farber S, Ordon F, Lange M, Borner A, Graner A, Reif JC, Scholz U, Mascher M, Stein N (2019) Genebank genomics highlights the diversity of a global barley collection. *Nat Genet* 51:319–326
30. Monat C, Padmarasu S, Lux T, Wicker T, Gundlach H, Himmelbach A, Ens J, Li C, Muehlbauer GJ, Schulman AH, Waugh R, Braumann I, Pozniak C, Scholz U, Mayer KFX, Spannagl M, Stein N, Mascher M (2019) TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol* 20:284
31. Moragues M, Comadran J, Waugh R, Milne I, Flavell AJ, Russell JR (2010) Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *Theor Appl Genet* 120:1525–1534
32. Ooijen J, Van JW (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations
33. Ordon F, Ahlemeyer J, Werner K, Köhler W, Friedt WJE (2005) Molecular assessment of genetic diversity in winter barley and its use in breeding. *Euphytica* 146:21–28
34. Owen H, Pearson K, Roberts AM, Reid A, Russell J (2019) Single nucleotide polymorphism assay to distinguish barley (*Hordeum vulgare* L.) varieties in support of seed certification. *Genet Resour Crop Ev* 66:1243–1256
35. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:e190
36. Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253
37. Pourkheirandish M, Hensel G, Kilian B, Senthil N, Chen G, Sameri M, Azhaguvel P, Sakuma S, Dhanagond S, Sharma RJC (2015) Evolution of the grain dispersal system in barley. *Cell* 162:527–539
38. Pourkheirandish M, Kanamori H, Wu J, Sakuma S, Blattner FR, Komatsuda T (2018) Elucidation of the origin of ‘agriocrithon’ based on domestication genes questions the hypothesis that Tibet is one of the centers of barley domestication. *Plant J* 94:525–534
39. Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, He Z (2017) Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Mol plant* 10:1047–1064
40. Russell J, Mascher M, Dawson IK, Kyriakidis S, Calixto C, Freund F, Bayer M, Milne I, Marshall-Griffiths T, Heinen S, Hofstad A, Sharma R, Himmelbach A, Knauff M, van Zonneveld M, Brown JWS, Schmid K, Kilian B, Muehlbauer GJ, Stein N, Waugh R (2016) Exome sequencing of geographically diverse

- barley landraces and wild relatives gives insights into environmental adaptation. *Nat Genet* 48:1024–1030
41. Scheben A, Batley J, Edwards D (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol J* 15:149–161
 42. Shi LJ, Jiang CC, He Q, Habekuss A, Ordon F, Luan HY, Shen HQ, Liu J, Feng ZY, Zhang J, Yang P (2019) Bulk segregant RNA-sequencing (BSR-seq) identified a novel rare allele of *eIF4E* effective against multiple isolates of BaYMV/BaMMV. *Theor Appl Genet* 132:1777–1788
 43. Shirasawa K, Kuwata C, Watanabe M, Fukami M, Hirakawa H, Isobe S (2016) Target amplicon sequencing for genotyping genome-wide single nucleotide polymorphisms identified by whole-genome resequencing in Peanut. *Plant Genome-US* 9:3835
 44. Stein N, Perovic D, Kumlehn J, Pellio B, Stracke S, Streng S, Ordon F, Graner A (2005) The eukaryotic translation initiation factor 4E confers multiallelic recessive Bymovirus resistance in *Hordeum vulgare* (L.). *Plant J* 42:912–922
 45. Taketa S, Amano S, Tsujino Y, Sato T, Saisho D, Kakeda K, Nomura M, Suzuki T, Matsumoto T, Sato K, Kanamori H, Kawasaki S, Takeda K (2008) Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. *Proc Natl Acad Sci USA* 105:4062
 46. Tanno K, Takeda K (2004) On the origin of six-rowed barley with brittle rachis, agriocrithon [*Hordeum vulgare* ssp. *vulgare* f. *agriocrithon* (Aberg) Bowd.], based on a DNA marker closely linked to the *vrs1* (six-row gene) locus. *Theor Appl Genet* 110:145-150
 47. Torkamaneh D, Boyle B, Belzile F (2018) Efficient genome-wide genotyping strategies and data integration in crop plants. *Theor Appl Genet* 131:499–511
 48. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X (2021) rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteomics Bioinformatics*
 49. You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9:253
 50. Zeng X, Guo Y, Xu Q, Mascher M, Guo G, Li S, Mao L, Liu Q, Xia Z, Zhou J, Yuan H, Tai S, Wang Y, Wei Z, Song L, Zha S, Li S, Tang Y, Bai L, Zhuang Z, He W, Zhao S, Fang X, Gao Q, Yin Y, Wang J, Yang H, Zhang J, Henry RJ, Stein N, Tashi N (2018) Origin and evolution of qingke barley in Tibet. *Nat Commun* 9:5433
 51. Zhou Y, Zhao X, Li Y, Xu J, Bi A, Kang L, Xu D, Chen H, Wang Y, Wang Y-g, Liu S, Jiao C, Lu H, Wang J, Yin C, Jiao Y, Lu F (2020) Triticum population sequencing provides insights into wheat adaptation. *Nat Genet* 52:1412–1422

Tables

Table 1 Information of five-hundred target fragments and derived SNP markers in 1068 genotypes

	<i>H. spontaneum</i>	Semi-wild	Landraces	Cultivars	F ₂ population
Number of samples	51	248	345	329	95
Detected target fragments	500	500	500	500	500
Size of fragments (kb) ¹	91.1	91.1	91.1	91.1	91.1
Average depth (x)	508	458	848	858	992
Target SNPs showing polymorphic	335	351	385	397	108
Polymorphic SNPs	2652	1496	1680	1648	442

¹ Size of fragments are calculated according to the barley reference genome (Morex v2; Monat et al. 2019). The average depth in *H. spontaneum* and semi-wild populations, which were lower than that of other populations, was conferred by lower sequencing coverage in Experiment-2 (as shown in Fig. 1c).

Table 2 Statistics of sequence diversity in barley sub-populations

Populations	Number of Samples	Size of fragments (kb)	π	D^*	F^*	<i>Tajima's D</i>
Wild	51	90.3	0.14717	-1.94324	-2.01379	-1.26210
Semi-wild	248	90.3	0.06472	-4.27746**	-3.63251**	-1.91098*
Landrace	345	90.3	0.10068	-1.82826	-1.68786	-1.11200
Cultivar	329	90.3	0.23402	1.58695*	1.79708*	1.48624

Nucleotide diversity (π), Fu and Li's D-test statistics (D^*), Fu and Li's F-test statistics (F^*), *Tajima's D*. The parameters (π , D^* , F^* , *Tajima's D*) are calculated using DnaSP v6.12.01 with the gaps not-considered algorithm. * $P \leq 0.05$; ** $P \leq 0.02$.

Figures

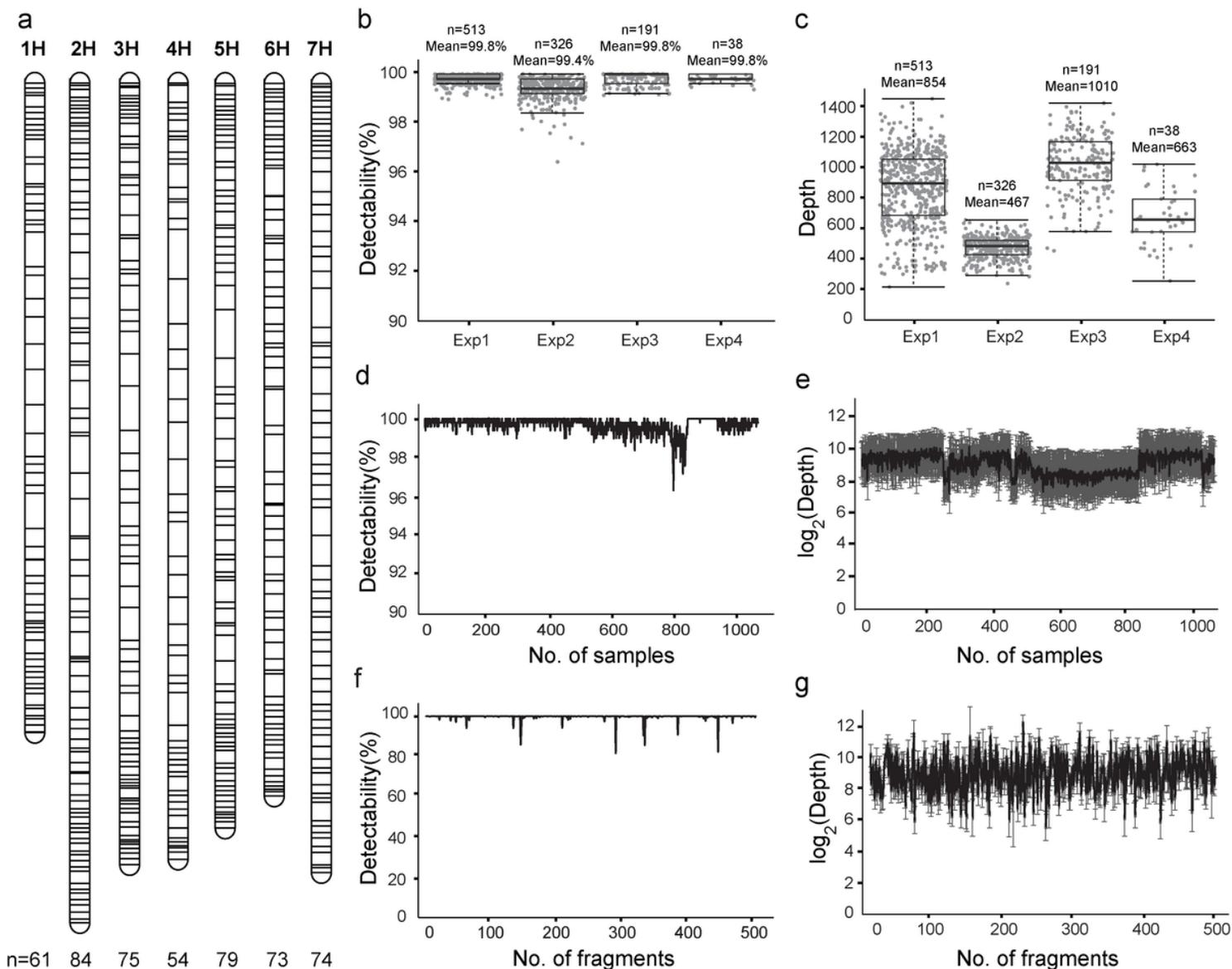


Figure 1

PCR-based target amplification (BarPlex v1.0) of five-hundred genomic fragments. Error bars represent standard deviation (SD) of target fragments. a Chromosomal locations of targeted fragments on the reference genome (v2) of barley cultivar “Morex”. b Percentage of the detection rate of four independent experiments. Detection rates are based on the mean of detected fragments divided by the number of target fragments in each sample. c Mean of sequencing depth of the samples in four experiments. d Percentage of the detection rate in all samples. e The \log_2 value of sequencing depth in each sample. f Percentage of the detection rate in each fragment. g The \log_2 value of the sequencing depth of each fragment.

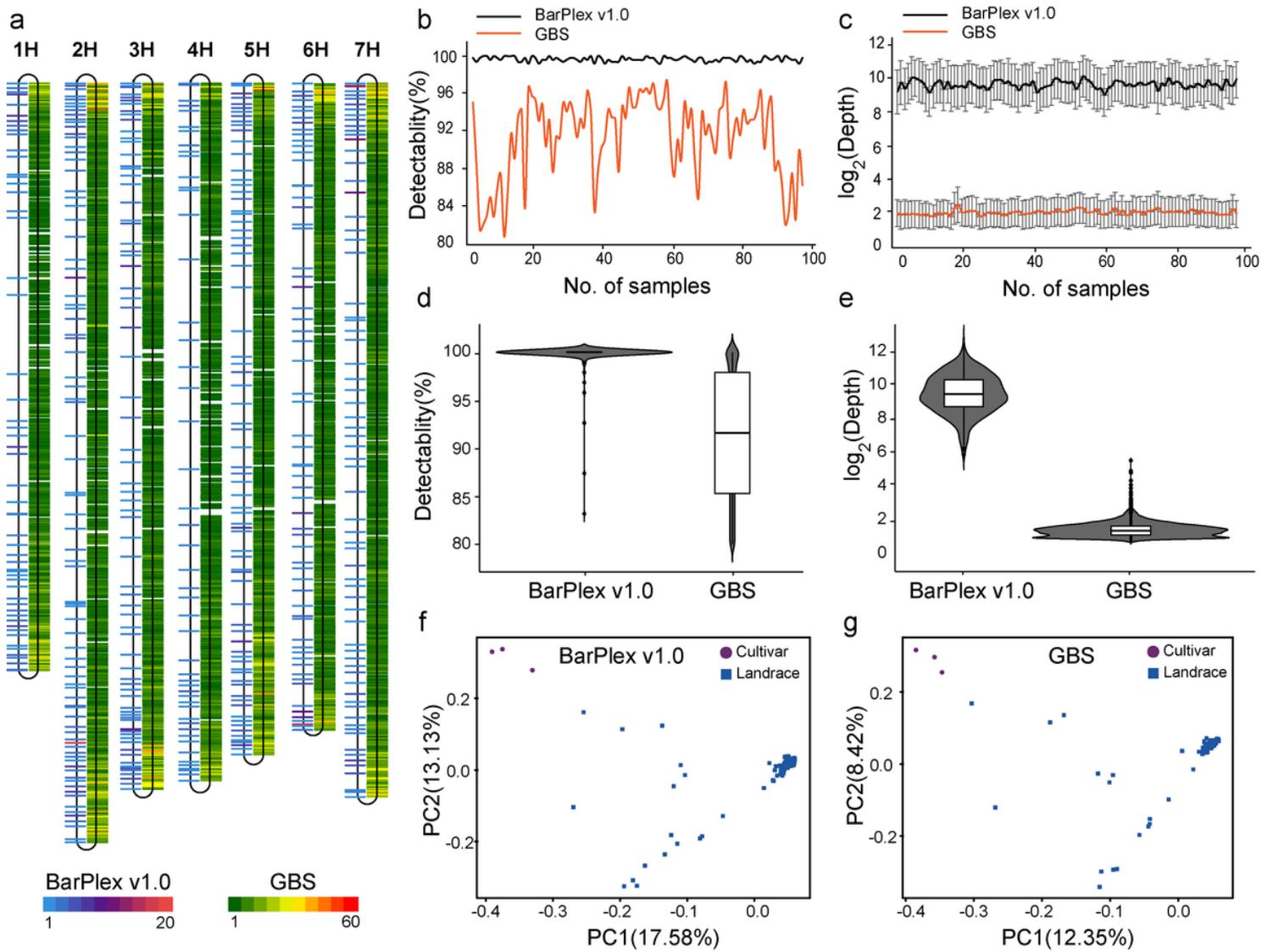


Figure 2

Genotyping of 96 barley accessions using BarPlex v1.0 and GBS. Error bars = SD. a Distribution of SNPs on each chromosome. The horizontal lines on the left and right of each chromosome mark the SNPs derived from BarPlex v1.0 and GBS, respectively. b Percentage of the detection rate in each of 96 samples. c The \log_2 value of sequencing depth in each of 96 samples. d Percentage of the detection rate at each of SNPs. e The \log_2 value of average sequencing depth at each of SNPs. f Principal component analysis of 96 samples using BarPlex-derived 1,372 SNPs. g PCA of 96 samples using GBS-derived 24,195 SNPs.

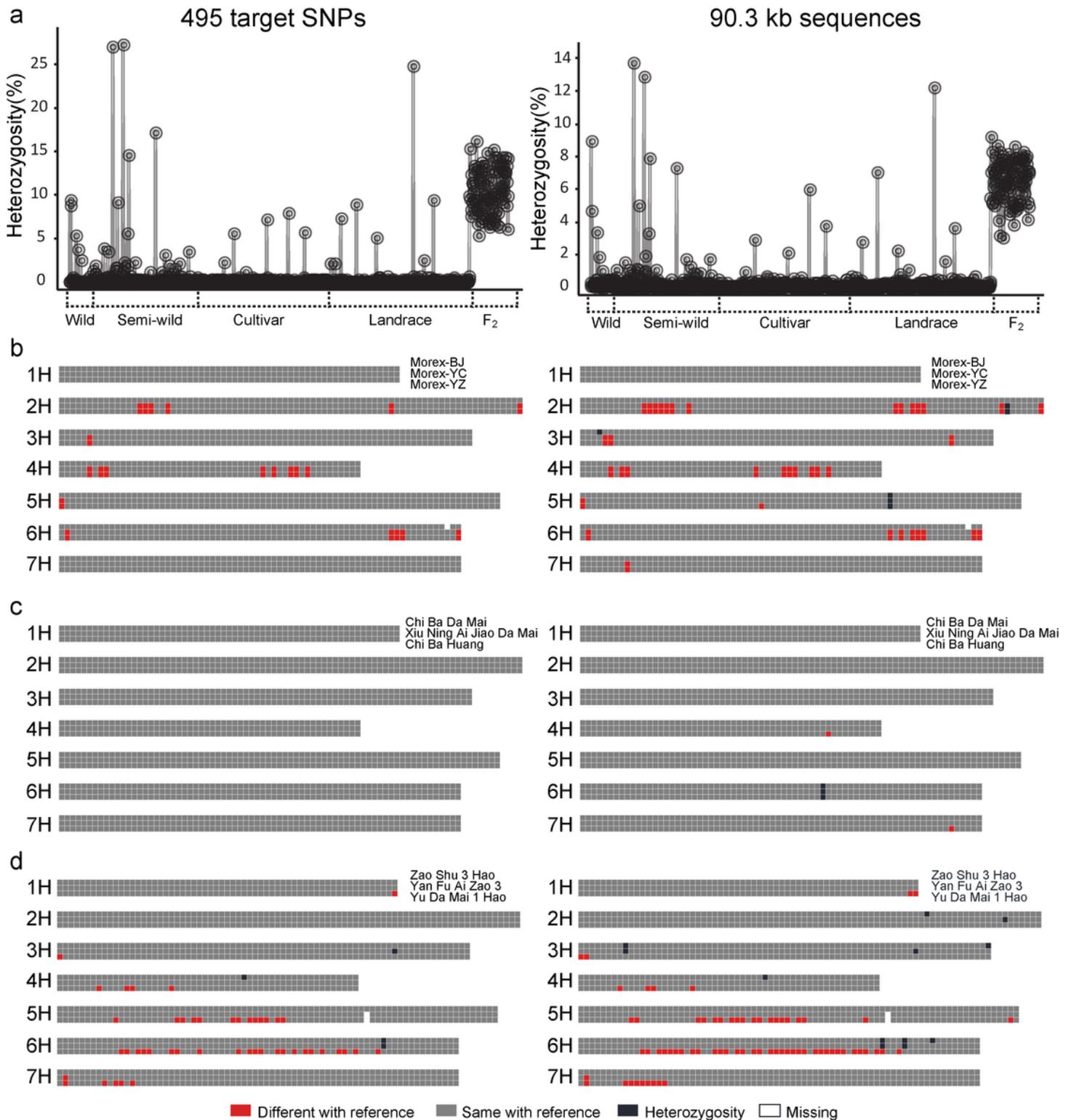


Figure 3

Analysis of heterozygosity and genetic similarity. 495 target SNPs and 3,220 polymorphic sites derived from the 90.3 kb target sequences were applied. a Percentage of heterozygous markers within 1,068 samples. The value of heterozygosity was calculated as the number of heterozygous SNPs/sites divided by total number of SNPs/sites. These samples included 51 wild barley, 248 Tibetan semi-wild barley, 329 cultivated barley, 345 Landrace barley and 95 F₂ individuals (left to right). b Graphical representation of

the polymorphic sites identified in barley cultivar “Morex” provided by different organizations. BJ = Institute of Crop Sciences, Chinese Academy of Agricultural Sciences. YC = Institute of Agricultural Science in Jiangsu Coastal Areas. YZ = Yangzhou University. c Graphical representation of the genetic similarity in three Chinese barley landraces designated by their local names. d Genetic similarity of barley cultivar “Zao Shu 3 Hao” and its progenitor varieties “Yan Fu Ai Zao 3” and “Yu Da Mai 1 Hao”.

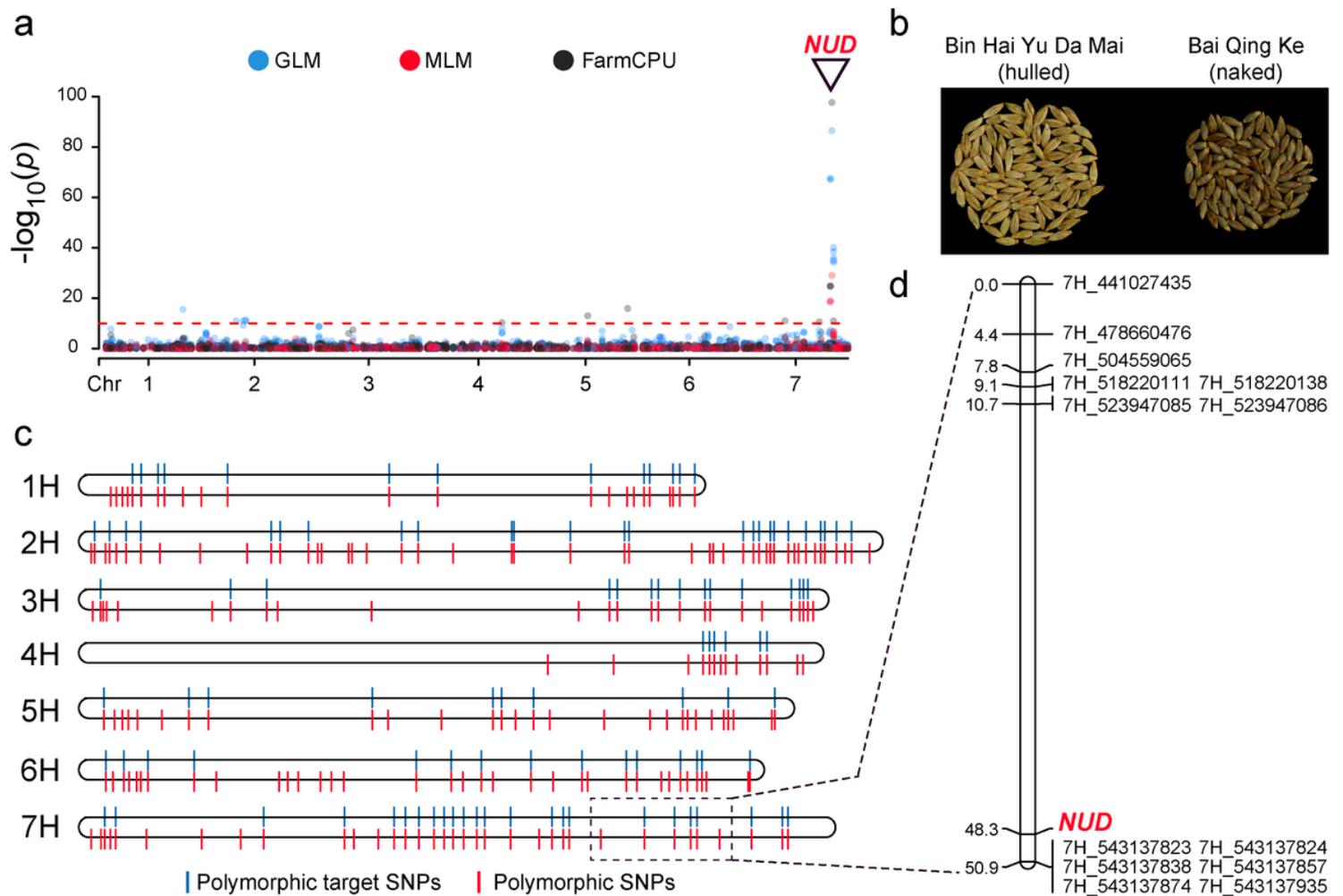


Figure 4

Genetic identification of adherence of caryopsis using linkage mapping and genome-wide association study (GWAS). a GWAS used for determinants that control the adherence of caryopsis. Three algorithms, general linear model (GLM), mixed-linear model (MLM), as well as fixed and random model circulating probability unification (FarmCPU) have been applied. b Hulled and naked grains of the two parental lines. c Polymorphisms detected between the two parents. Blue and red bars indicate polymorphic target SNPs and polymorphic sites detected from 90.3 kb sequences, respectively. d Genetic position of the NUD locus delimited by 95 F2 segregated plants.

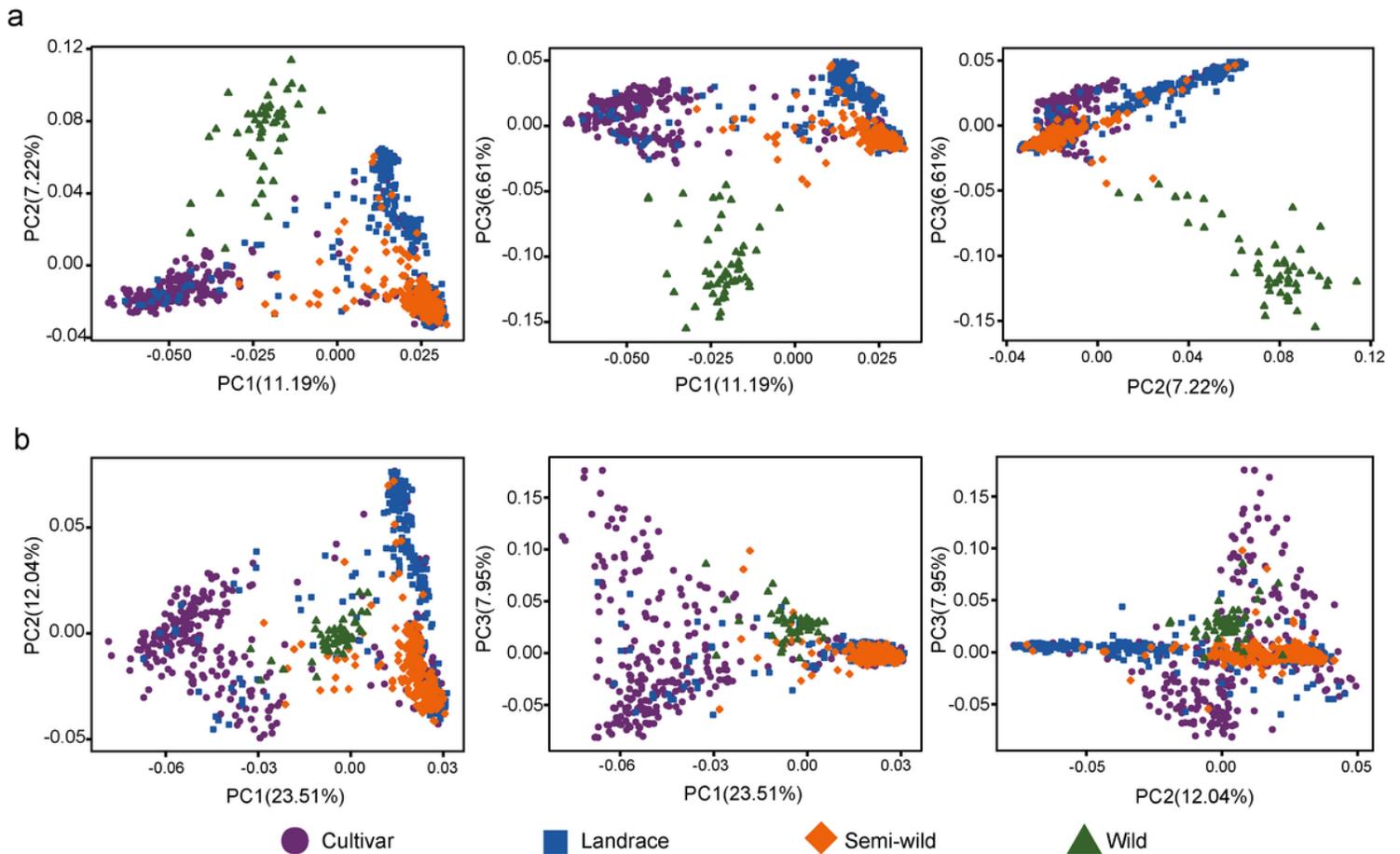


Figure 5

Principle component analysis (PCA) of 973 genotypes including wild barley (*H. spontaneum*), Tibetan semi-wild barley, landraces, and cultivars. a PCA plotting of 3,220 polymorphic sites detected from 90.3 kb sequences. b PCA plotting of 404 target SNPs.

Supplementary Files

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

- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)