Analysis of heterosis based on barley genome

2022, September

Areej Sakkour

Graduate School of Environmental and Life Sciences
(Doctor's course)

Okayama University

Table of contents

General Abstract

Cultivated barley (*Hordeum vulgare* ssp. *vulgare*) is used for food, animal feed, and alcoholic beverages and is widely grown in temperate regions. Both barley and its wild progenitor (*H. vulgare* ssp. *spontaneum*) have large 5.1-Gb genomes. High-quality chromosome-scale assemblies for several representative barley genotypes, both wild and domesticated, have been constructed recently to populate the nascent barley pan-genome infrastructure which will promote the analysis of genomic and genetic diversity of barley.

Basis of barley breeding activities was started in European countries especially in the United Kingdom that already during the first half of the 19th century. After the rediscovery of Mendelian laws of inheritance, cross-breeding and genetic recombination has been used to develop cultivars in many breeding programs around the world. The idea of hybrid barley seed production system using genetic and cytoplasmic male sterility have not become popular around the world due to the labor of developing lines for hybrid seed production and the lower level of hybrid vigor which attracts the interests of growers and breeding programs. For the genetic basis of heterosis, mainly three models have been presented. However, genetic and genomic mechanisms of heterosis have not yet well understood in barley partly due to the lack of genomic diversity information in barley.

The heterosis is the basis of hybrid barley cultivars. The present studies were conducted to estimate the level of heterosis among selected samples of world barley collection and also focus on the genetic analysis of possible loci which contributes to enhance the productivity of biomass based on the barley genome sequence.

In the genetic analysis, Japanese two row malting cultivar 'Haruna Nijo' and Scottish two row malting cultivar 'Golden Promise' were used as common parent of the crosses. Since genome assembly of 'Golden Promise' was published as a part of pan-genome analysis, the genome sequencing of 'Haruna Nijo' has been conducted with the same technical standards of pan-genome analysis. The chromosome-scale assembly of 'Haruna Nijo' has 4.28-Gb with a scaffold N50 size of 18.9 Mb. The assembly showed high collinearity with the barley reference genome 'Morex' cultivar, with some inversions. The pseudomolecule assembly was characterized using transcript evidence of gene projection derived from the reference genome and *de novo* gene annotation achieved using published full-length cDNA sequences and

RNA-Seq data for 'Haruna Nijo'. We found good concordance between our whole-genome assembly and the publicly available BAC clone sequence of 'Haruna Nijo'. Interesting phenotypes have since been identified in 'Haruna Nijo'; its genome sequence assembly will facilitate the identification of the underlying genes.

To estimate the amount of heterosis in morphological and yield component traits, F1s were developed from top crosses of barley and evaluated the influence of each parental genotype on the amount of heterosis, furthermore identify the promising combination of crosses. A total of 207 and 80 barley accessions were selected from diverse types and geographical origins around the world to be crossed with either 'Haruna Nijo' or 'Golden Promise' as common parent, respectively. Parents and F1s were grown in two replications at the field of Okayama University. Mean total biomass of parents and F1s were 37.5 g and 50.6 g in 'Haruna Nijo' crosses ('Haruna Nijo'=41.4 g), and 39.2 g and 60.0 g in 'Golden Promise' crosses ('Golden Promise'=30.3 g). Mean total biomass of six-row types was higher than that of two-row types in both 'Haruna Nijo' (six-row: 53.7 g and two-row: 47.6 g) and 'Golden Promise' (62.9 g and 53.2 g) crosses, respectively. However, the differences were not observed by hulled/hull-less comparisons. There are also differences in total biomass by geographical origin (in 'Haruna Nijo' crosses: ranged 40.4 g in North Africa to 60.1 g in China). All the parents were genotyped by Illumina iSelect 50K and genotype by sequencing to estimate the relationship between heterosis and genotype. Several minor loci were identified for total weight and spike weight in both top crosses. F2 plants from selected cross combinations from each top cross showed wide range of distribution and QTL analyses showed several significant loci related to total weight. Annotations on these loci will be further investigated.

Two chapters of this study were coordinated to identify genetic factors related to heterosis which intend to contribute to the development of hybrid barley cultivars. The genetic basis of heterosis is important to further improve the amount of heterosis with the basis of molecular/genomic information.

Barley is a winter growing plant in Kurashiki and cannot plant in the summer to evaluate traits related to heterosis. Within limited time and generations, this study identified promising amount of heterosis in biomass in barley germplasm which is almost comparable with the historical study of Okayama University, which was conducted in the same field condition. Hybrid barley is still attracting the interests of breeding companies and may

become popular if reasonable amount of heterosis and seed price are available. The key factor is how to identify and collect molecular information responsible for the heterosis in barley. As some of the genes have been identified in other crops, there may be possible ways to achieve higher amount of heterosis by using responsible genes. The other topics related decreasing seed price will be solved at the same time.

Abbreviations

IBSC	international barley sequencing consortium
fl-cDNA	full length complementary DNA
BAC	bacterial artificial chromosome
RNA-Seq	RNA sequencing
GFF	general feature format
ENA	European nucleotide archive
TREP	Triticeae repeat sequence
Hi-C	high-throughput chromosome conformation capture
AHRD	automatic assignment of Human Readable Descriptions
N50	sequence length of the shortest contig at 50% of the total genome length
QTL	quantitative trait locus
GBS	genotype by sequencing

General Introduction

Barley was domesticated ca. 10000 years ago in Middle East (Zohary et al. 2012) and has been distributed around the world (Bothmer et al. 2003). Both domesticated barley (*Hordeum vulgare* ssp. *vulgare*) and its wild progenitor (*H. vulgare* ssp. *spontaneum*) are self-pollinating diploid in the same species and thus the natural mutations have been occurred with phenotypic changes to promote the generation of variation for genotypes and phenotypes of this species. During the process of distribution and adaptation, barley has been diversified to adapt environments and selected uses by human mainly for food, brewing and feed (Fischbeck 2003).

Basis of barley breeding activities was started in European countries especially in the United Kingdom that already during the first half of the 19th century, farm owners observed their cereal fields for off-types, and multiplied the separately harvested seed to provide better seed stocks for the improvement of barley production (Fischbeck 2003). After the rediscovery of Mendelian laws of inheritance, cross-breeding and genetic recombination has been used to develop cultivars in many breeding programs around the world. After the first treatments to induce mutations in barley by X rays (Stadler 1928), a variety of physical and chemical mutagens have been used to induce mutations (Lundqvist and Franckowiak 2003). These mutations and structural variations in chromosomes have been used in genetic studies of barley. Also, many of these mutations have been used directly as cultivars or indirectly as crossing parents in barley breeding.

Idea of hybrid barley breeding might come from extensive genetic studies of barley chromosomes (Taketa et al. 2003) and mutant collections of genetic male sterility (Lundqvist and Franckowiak 2003). The hybrid barley seed production system using genetic male sterility with balanced tertiary trisomics was an excellent genetic platform (Ramage 1965) but was not become popular around the world. Even after the discovery of cytoplasmic male sterility (CMS) and its restorer gene (Ahokas 1982), CMS based hybrid barley cultivars have not become popular around the world due to the labor of developing lines for hybrid seed production and the lower level of hybrid vigor which did not attract the interests of growers and breeding programs.

Mainly three models have been presented for the genetic basis of heterosis. Garcia

et al. (2008) summarizes the cause of heterosis in to three hypotheses. The dominance hypothesis suggests that dominant alleles cancel the negative effects at multiple loci at the F1 hybrid. The overdominance hypothesis assumes that the loci with heterozygous genotypes are superior to both homozygous parents. Epistasis is also a possible cause of heterosis. However, genetic mechanisms of heterosis are still not well understood in barley partly due to the lack of genomic diversity information in barley.

Barley has been genetically well studied with a large amount of mutant collections and detailed genetic maps by mutant traits has been developed (Franckowiak et al 1997). Cytogenetical studies were also advanced on barley due to its large size of chromosomes and fewer number of diploid chromosomes. However, large size of chromosomes also indicates the large genome size (5 Gbp) which was 12 times larger than the rice genome. The genome sequencing analysis has been delayed compared to diploid species of model plants, i.e., rice and Arabidopsis.

Barley research community has generated a large amount of cDNA sequences to develop high density transcript maps (Close et al. 2009, Sato et al. 2009a) which provided basic genomic structures to order genomic clones to develop physical contig of barley genome. Together with several whole genome shotgun sequencing reads, the draft genome of barley was published (IBSC 2012) which had around 2 Gbp including most of the genic region of the genome from the 5 Gbp of the estimated physical size of the contig. Each of the genomic clone in the physical contig was sequenced and ordered by chromosome conformation capture (Hi-C) and optical mapping techniques and the first pseudomolecule sequences of barley was published as MorexV1 (Mascher et al. 2017). The more efficient assembly pipeline was developed by Monat et al. (2019), barley genome assembly has been improved to next stage MorexV2 and more 19 genotypes from the world were sequenced as barley pan-genome with this assembly pipeline (Jayakodi et al. 2020). The most recent assembly MorexV3 (Mascher et al. 2021) was analyzed using PacBio Hifi reads which were quality controlled long nucleotide sequences of ca. 20kb.

The international barley research community has been using American six row malting cultivar Morex as a key genotype since the high quality molecular genetic map was developed by the cross of American cultivars Morex/Steptoe (Kleinhofs et al. 1993). Morex has a Manchurian origin and is different from other malting barley cultivars with European

origin. According to 20 accessions of pan-genome analysis (Jayakodi et al. 2020), the clustering of orthologous gene models yielded 40,176 orthologous groups. Of these, 13,188 were absent from at least one assembly; and 1,760 were present in only one assembly, indicating that many of the genes in Morex were not shared with other 19 genotypes. The results suggested the good reason to sequence other haplotypes used in the specific studies. In this thesis analysis, Japanese two row malting cultivar 'Haruna Nijo' and Scottish two row malting cultivar 'Golden Promise' were used as common parent of the crosses. Since genome assembly of 'Golden Promise' was published as a part of pan-genome analysis (Jayakodi et al. 2020), the genome sequencing of 'Haruna Nijo' has been conducted with the same technical standards of pan-genome analysis.

Genome assembly information accelerates the use of genome wide DNA markers for genotyping and trait mapping. The prefixed single nucleotide marker system was developed by transcript map (Close et al. 2009) and the exome sequence analysis (Bayer et al. 2017). These are based on the genetic/genomic position of cv. Morex. Recently, partial *de novo* sequencing platform has been used for germplasm fingerprinting and trait mapping. Milner et al. (2019) analyzed ca. 20,000 German gene bank accessions of barley which were categorized into European, Asian and Ethiopian germplasm groups. The geographical distribution or collection regions were well agreed with the molecular maker groups. The partial world collection samples of barley in Okayama University were also analyzed by the transcript-based SNP markers by Close et al (2009) and showed that the collection regions and DNA marker polymorphisms were agreed and the accession were grouped into 5-8 subgroups (Wu et al. 2016, Sato et al. 2018, Sato et al. 2020). However, these groupings were based on the Morex genetic/genomic information and analysis with other genotypes have not yet done.

Based on these technical backgrounds, the present studies were conducted to estimate the level of heterosis among partial samples of world barley collection. The heterosis is the basis of hybrid barley cultivars. In addition, the present studies also focus on the genetic analysis of possible loci which contributes to enhance the productivity of biomass in barley.

Chapter 1

Chromosome-scale assembly of barley cv. 'Haruna Nijo' as a resource for barley genetics

Abstract

Cultivated barley (*Hordeum vulgare* ssp. *vulgare*) is used for food, animal feed, and alcoholic beverages and is widely grown in temperate regions. Both barley and its wild progenitor (*H. vulgare* ssp. *spontaneum*) have large 5.1-Gb genomes. High-quality chromosome-scale assemblies for several representative barley genotypes, both wild and domesticated, have been constructed recently to populate the nascent barley pan-genome infrastructure. Here, we release a chromosome-scale assembly of the Japanese elite malting barley cultivar 'Haruna Nijo' using a similar methodology as in the barley pan-genome project. The 4.28-Gb assembly had a scaffold N50 size of 18.9 Mb. The assembly showed high collinearity with the barley reference genome 'Morex' cultivar, with some inversions. The pseudomolecule assembly was characterized using transcript evidence of gene projection derived from the reference genome and *de novo* gene annotation achieved using published full-length cDNA sequences and RNA-Seq data for 'Haruna Nijo'. We found good concordance between our whole-genome assembly and the publicly available BAC clone sequence of 'Haruna Nijo'. Interesting phenotypes have since been identified in 'Haruna Nijo'; its genome sequence assembly will facilitate the identification of the underlying genes.

Keywords: *Hordeum vulgare*, full-length cDNA, RNA-Seq, genome sequencing, pseudomolecules

Introduction

Cultivated barley is used for many purposes, including animal feed, human food, and malting for brewing. Malting barley has only been cultivated in Japan for ca. 140 years (Seko 1987). The founder cultivars were mainly introduced from Europe and crossed with Japanese landraces, which prior to that had been used for human food. In 1978, the malting barley cultivar 'Haruna Nijo' was released from Sapporo Breweries (Tokyo, Japan) and has since been used as a donor of high-quality profiles in Japanese malting barley breeding programs. At Okayama University (Okayama, Japan), 'Haruna Nijo' is used as a key genotype in genetics and genomics studies. 'Haruna Nijo' was used for the generation of expressed sequence tags (Sato 2020) (see also https://harvest.ucr.edu/). Using these transcript sequences, a high-density genetic map was constructed from a cross between 'Haruna Nijo' and the wild barley (*H. vulgare* ssp. *spontaneum*) accession 'OUH602' (Sato et al. 2009a, Close et al. 2009), and a set of recombinant chromosome substitution lines was developed (Sato and Takeda 2009). 'Haruna Nijo' was used to generate full-length cDNA (fl-cDNA) sequences (Sato et al. 2009b, Matsumoto et al. 2011), which have been used for the annotation of gene models in the reference genome of the cultivar 'Morex' (IBSC 2012, Mascher et al. 2017, Jayakodi et al. 2020, Schreiber et al. 2020, Mascher et al. 2021). Whole-genome shotgun sequencing was performed for 'Haruna Nijo' to enable the estimation of the genic regions of the genome (IBSC 2012, Sato et al. 2016a). A BAC library of 'Haruna Nijo' was also constructed to isolate the genes responsible for major traits (Saisho et al. 2007), such as hull-less caryopsis (Taketa et al. 2008) and seed dormancy (Sato et al. 2016b). The mitochondrial genome of 'Haruna Nijo' was also sequenced and found to be highly similar to that of 'OUH602' (Hisano et al. 2016).

After the release of a high-quality barley genome assembly generated using BAC-by-BAC sequencing and scaffold alignment (Mascher et al. 2017), several whole-genome shotgun assembly techniques were developed for Illumina short reads, such as the DeNovoMAGIC assembly pipeline (NRGene, Ness Ziona, Israel), the TRITEX pipeline (Monat et al. 2019), and w2rap-contigger (Clavijo et al. 2017). Using these assembly methodologies, the global landscape of the barley genome (pan-genome; Jayakodi et al. 2021) was recently analyzed using 20 domesticated and wild accessions (Jayakodi et al. 2020) based

on a selection of 22,000 genomic profiling datasets with genotype by sequencing (GBS) from German gene bank accessions (Milner et al. 2019).

Here, we utilized the TRITEX pipeline to generate a chromosome-scale genome assembly of 'Haruna Nijo'. We aligned the assembly to the most recently updated assembly, 'Morex'V3 (Mascher et al. 2021), to identify genomic differences among the genotypes. We also aligned the assembly to the published BAC sequences used for gene isolation to estimate the quality of the assembly. A similar sequencing methodology was also recently applied to the wild barley accession 'OUH602' (Sato et al. 2021); however, the assembly of the 'Haruna Nijo' genome is desirable for its economic and breeding importance.

The present barley genome annotation, e.g., EnsemblPlants (http://plants.ensembl.org/Hordeum_vulgare/), is based on 'Morex', which is the North American malting cultivar with a Manchurian landrace pedigree, and differs from malting barleys in other areas of the world. In a recent barley pan-genome analysis (Jayakodi et al. 2020), gene projection was performed using informant gene models of 'Morex', the German malting cultivar 'Barke', and an Ethiopian landrace 'HOR10350', which were predicted from transcriptome data and protein homology information using a previously described annotation pipeline (Mascher et al. 2017). In addition to this gene projection analysis, we performed *de novo* gene annotation for 'Haruna Nijo' using published fl-cDNA sequences and RNA-Seq data. These procedures may provide alternative gene annotation information on the barley genome by characterizing different sources of transcript and protein information from fl-cDNA sequences and RNA-Seq data.

Materials and Methods

DNA extraction, library construction, and sequencing

High-molecular-weight DNA was isolated from leaf material of seedlings of 'Haruna Nijo' (Dvorak et al. 1988)) and size selected for a molecule size of 40 kb or higher. The 440-bp paired-end (PE) libraries were prepared with the Hyper Kapa Library Preparation kit (Kapa Biosystems) with no PCR amplification. The 8- to 10-kb mate-pair (MP) libraries were constructed with the Nextera Mate Pair library Sample Prep kit (Illumina, San Diego, California, USA) followed by the TruSeq DNA Sample Prep kit. The 10X libraries were

constructed with the Chromium Genome Library Kit & Gel Bead Kit v2 (10X Genomics). Sequencing was performed following Sato et al. (2021). In brief, the 440-bp PE libraries were sequenced for 251 cycles using a NovaSeq 6000 system (Illumina). The 10X and MP libraries were sequenced for 151 cycles from each end of the fragments on the NovaSeq 6000 system. All libraries were prepared and sequenced at the University of Illinois Roy J. Carver Biotechnology Center (Urbana, Illinois, USA). *In situ* Hi-C libraries were prepared as described by Padmarasu et al. (2019). Sequencing data generated from each of the libraries are listed in Table 1-1. The Hi-C data were used to prepare chromosome-scale assemblies using the TRITEX pipeline (Mascher et al. 2019), which was also used for the contig assembly and scaffolding with the PE, MP, and 10X data (Table 1-1).

Transcript sequencing

Published RNA-Seq reads from the seedling root, shoot, spike at flowering, and seeds of 'Haruna Nijo' (Sato et al. 2016a) were used for the transcript sequencing. An additional RNA sample of a young spike (3 cm in length) from 'Haruna Nijo' was also extracted and subjected to an RNA-Seq analysis, as described by Sato et al. (2016a). These RNA-Seq libraries were sequenced with the MiSeq Reagent Kit V3 (2× 300 bp cycles) on a MiSeq system (Illumina).

Gene projection

To derive the projected gene structures for 'Haruna Nijo', informant gene models of 'Morex', 'Barke', and 'HOR10350' were employed, which were predicted from transcriptome data and protein homology information (Jayakodi et al. 2020) using a previously described annotation pipeline (Mascher et al. 2017). The projection was based on a stepwise procedure, as previously described (Jayakodi et al. 2020, Sato et al. 2021). Briefly, BLASTN (Altschul et al. 1990) and Exonerate alignments (Slater and Birney 2005) of the coding sequences (CDSs) of each of the barley sources of the 'Haruna Nijo' genome sequence were computed. The matches were clustered by their genomic loci, and the top-scoring match was selected using a stepwise integration approach. In addition to protein-coding genes, 'pseudogene'-type mappings were previously projected and included in the CDSs and GFF files but were obviously missing from the protein sequence files.

De novo gene annotation using RNA-Seq and fl-cDNA sequences

A structural gene annotation was performed by combining *de novo* gene calling and homology-based approaches with RNA-Seq, protein, isoseq, and fl-cDNA datasets. Using evidence derived from expression data, RNA-Seq sequences were first mapped against the 'Haruna Nijo' genome assembly using STAR (Dobin et al. 2013) (version 2.7.8a) and subsequently assembled into transcripts using StringTie (Kovaka et al. 2019) (version 2.1.5; parameters -m 150-t-f 0.3). Triticeae protein sequences obtained from publicly available datasets (UniProt; https://www.uniprot.org; accessed 10 December 2021) were aligned against the genome sequence using GenomeThreader (Gremme et al. 2005) (version 1.7.1; arguments -startcodon-finalstopcodon -species rice -gcmincoverage 70 -prseedlength 7 -prhdist 4). The fl-cDNAs and isoseq were aligned to the genome assembly using GMAP (Wu and Watanabe 2005) (version 2018-07-04). All RNA-Seq, fl-cDNA, and aligned protein sequences were combined using Cuffcompare (Ghosh and Chan 2016) (version 2.2.1) and subsequently merged with StringTie (version 2.1.5; parameters --merge -m150) into a pool of candidate transcripts. TransDecoder (version 5.5.0; http://transdecoder.github.io) was used to find potential open reading frames (ORFs) and to predict protein sequences within the candidate transcript set. An *ab initio* annotation was performed using Augustus[31] (version 3.3.3). GeneMark (Ter-Hovhannisyan et al. 2008) (version 4.35) was additionally used to further improve the structural gene annotation. To avoid potential over-prediction, guiding hints were generated using the above-described RNA-Seq, isoseq, protein, and fl-cDNA datasets and were then trained and optimized using a specific Augustus model for barley, as described by Hoff and Stanke (2019). Structural gene annotations from different prediction methods were combined using EVidenceModeler (Haas et al. 2008) (version 1.1.1), and the weights were adjusted according to the input source: *ab initio* (Augustus: 5, GeneMark: 2) and homology based (10). Additionally, two rounds of PASA (Haas et al. 2003) (version 2.4.1) were run to identify untranslated regions and isoforms using the above-described fl-cDNA dataset.

BLASTP (Altschul et al. 1990) (ncbi-blast-2.3.0+, parameters -max_target_seqs 1 -evalue 1e-05) was used to compare potential protein sequences with a trusted set of reference proteins (Uniprot Magnoliophyta, reviewed/Swissprot; downloaded on 3 August 2016; https://www.uniprot.org). This differentiated candidates into complete and valid genes, non-

coding transcripts, pseudogenes, and transposable elements. In addition, the PTREP database (Release 19; http://botserv2.uzh.ch/kelldata/trep-db/index.html) was used in the BLASTP analysis; this database of hypothetical proteins contains deduced amino acid sequences in which internal frameshifts have been removed in many cases. This step is particularly useful for the identification of divergent transposable elements with no significant similarity at the DNA level. The best hits were selected for each predicted protein to each of the three databases: UniProt, SwissProt, and PTREP. Only hits with an e-value below $10^{e\text{-}10}$ were considered. Furthermore, the functional annotation of all predicted protein sequences was performed using the AHRD pipeline (https://github.com/groupschoof/AHRD).

The proteins were further classified into two confidence classes: high and low. Hits with subject coverage (for protein references) or query coverage (transposon database) above 80% were considered significant. The proteins were classified as high confidence if the sequence was complete and had a subject and query coverage above the threshold in the UniMag database or no BLAST hit in UniMag or PTREP but present in UniPoa. A protein sequence was defined as being low confidence if it was incomplete and had a hit in the UniMag or UniPoa database but not in PTREP. Alternatively, complete protein sequences with no hit in UniMag, UniPoa, or PTREP were also classified as low confidence. In a second refinement step, low-confidence proteins with an AHRD-score of 3* were promoted to high-confidence.

Repeat and transcript annotation

The final assembly was analyzed for repetitive regions using RepeatMasker (Smit et al. 2013, 2014, 2015) (version 4.0.9) with the TREP repeat library (Wicker et al. 2002) (trep-db_complete_Rel-19; downloaded from http://botserv2.uzh.ch/kelldata/trep-db/downloadFiles.html on 13 September 2020). The repetitive regions were changed to lowercase (-xsmall parameter). The output of RepeatMask was condensed using the perl script "one-code-to-find-them-all" (Bailly-Bechet et al. 2014) with the parameters –strict and –unknown.

Data validation and quality control

Benchmarking Universal Single-Copy Orthologs (Waterhouse et al. 2018) (BUSCO; version 3.0.2) was used with the plant dataset (embryophyta_odb10) to validate the assembly and

gene models. For gene prediction, BUSCO uses Augustus (Stanke et al. 2004; König et al. 2016) (version 3.3). For the gene-finding parameters in Augustus, the species was set to wheat and BUSCO was run in genome mode (-m geno -sp wheat).

Alignment of published BAC sequences

Published 'Haruna Nijo' BAC clone sequences of kernel row type *Vrs1* (Komatsuda et al. 2007), brittle rachis *Btr1* and *Btr2* (Pourkheirandish et al. 2015), and quantitative locus seed dormancy 1 *Qsd1* (Sato et al. 2016b) were downloaded from NCBI. Each clone sequence was aligned to the pseudomolecule sequences of 'Haruna Nijo' and 'Morex'V3 using minimap2 (Li 2018).

Genome browser

Pseudomolecule assembly, gene models of the CDS, and amino acid models were visualized in JBrowse genome browser (version 1.16.9). The BLAST (version 2.2.18) and BLAT (version 34) servers were also installed to search for target sequences in the pseudomolecules and gene models.

Data availability

Raw reads have been deposited in the ENA sequence read archive. Bioproject: PRJEB44504 [ERS_ID: paired-end reads: ERS6294308; mate-pair reads: ERS6294309; 10X reads: ERS6294307; Hi-C reads: ERS6294313; assembly: ERS6294316] (Table 2-2).
The reference assembly is available for download or BLAST search from
http://viewer.shigen.info/harunanijo/index.php

Results and Discussion

Genome assembly

We generated the genome assembly from PE and MP short reads and 10X reads. Approximately 868 Gb of raw data was generated, providing an estimated 170× coverage of the genome (Table 1-1). An assembly generated using the TRITEX pipeline (Monat et al. 2019) resulted in a scaffold N50 value of 18.9 Mb (Table 1-3). We integrated Hi-C data into

the assembly, which uses a genomic distance matrix inferred from native chromatin folding to increase the scaffold-level contiguity to full chromosome size (Fig. 1-1). The final pseudomolecule size was 4.28 Gb, comprising 552 scaffolds and a cumulative size of unanchored scaffolds of 154.3 Mb. The pseudomolecule size of 'Haruna Nijo' is comparable with that of the pan-genome assemblies of 'Morex'V2 obtained using similar sequencing platforms but with a smaller scaffold N50 value. The datasets for 'Morex'V3 showed improved statistics compared with our assemblies due to the use of accurate long-read sequencing by circular consensus sequencing on the PacBio platform in the generation of this assembly (Mascher et al. 2021). The alignment of the pseudomolecules of 'Haruna Nijo' to 'Morex'V3 individual chromosomes revealed some small inversions (Fig. 1-2); however, the overall contiguity of entire chromosomes was retained between 'Haruna Nijo' and 'Morex'V3.

For easy access, the reference sequence is available in BLAST-searchable form at http://viewer.shigen.info/harunanijo/index.php.

Quality of assemblies

We used the spectra-cn function from the Kmer Analysis Toolkit (KAT) (Mapleson et al. 2017) to compare *k*-mer contents in the scaffolds and pseudomolecules. KAT generates a *k*-mer frequency distribution from the PE and MP reads and identifies how many times *k*-mers from each part of the distribution appear in the assemblies being compared (Sato et al. 2021). The spectra-cn plot in Fig. 1-3 generated from the contigs shows sequencing errors (*k*-mer multiplicity <20) in black, as these are not included in the assembly. Most of the content appears in a single red peak, indicating sequences that appear once in the assembly. The black region under the main peak is small, indicating that most of this content from the reads is present in the assembly. The content that appears to the right of the main peak and is present two or three times in the assembly represents repeats. Pseudomolecules may contain more miss-assemblies than scaffolds; this is not obvious in the spectra-cn plot in Fig. 1-3b.

We evaluated the quality of the 'Haruna Nijo' assembly using BUSCO (Simão et al. 2015; Waterhouse et al. 2018). This program assesses the completeness of a genome by identifying conserved single-copy orthologous genes. The scaffold and pseudomolecule stages had complete single-copy genes at a rate of 96.0% and 95.7%, respectively (Table 1-4). These values are very close to those recently published for the 'Morex'V2 assembly, which had 97.2% single-copy genes (Schreiber et al. 2020). The differences are mainly due to the greater

number of duplicated genes in the scaffolds (1.3%) than the pseudomolecules (1.2%). Only 1.0% of the fragmented sequences were present in both the scaffolds and pseudomolecules.

Repeat masking

We analyzed each chromosome of the 'Haruna Nijo' assembly for repetitive regions using RepeatMasker with the TREP repeat library. This analysis identified 72.8% (3.23 Gb) of the 'Haruna Nijo' assembly as transposable elements (Table 1-5), almost all of which were retroelements. The same analysis was performed for 'Morex'V2 and 'Morex'V3, producing similar results (Table 1-5). The differences from the published results for the 'Morex'V2 (Monat et al. 2019) and 'Morex'V3 (Mascher et al. 2021) assemblies were due to the different repeat libraries used.

Gene projection

We assessed the gene content of 'Haruna Nijo' using a gene projection approach, as described by Jayakodi et al. (2020) for the 20 barley pan-genome assemblies. The total number of loci was 47,367, which is within the range of 42,464 to 47,588 reported for the 20 pan-genome assemblies. Of the 44,579 protein-coding genes, between 42,800 and 43,211 loci had a BLAST match with an e-value of <1–30, and 34,427 and 38,005 were one-to-one reciprocal BLAST orthologs between 'Haruna Nijo' and 'B1K-04-12' or 'Morex'V2, respectively. The overall and orthologous gene content of 'Haruna Nijo' is therefore highly conserved in comparison with other barley lines. Likewise, 15.9% (7,109) of the tandem-repeated genes in 'Haruna Nijo' had similar ranges as were detected for the 20 barley pan-genome assemblies and were located in 2,735 clusters. The gene content statistics above indicate that the 'Haruna Nijo' assembly contains a gene set with highly similar characteristics to those reported for the 20 barley pan-genome assemblies.

De novo gene annotation using RNA-Seq and fl-cDNA sequences

A final structural gene annotation yielded 161,721 gene models, including 49,524 high- and 112,197 low-confidence gene models (Table 1-6). The high number of total gene models is likely due to the *ab initio* prediction step, which was run without the use of transposable elements hints; the high number of low-confidence gene models supports this rationale. The

BUSCO score of the high-confidence genes was 98.4 (Fig. 1-4). The average number of transcripts per gene was 1.39 for the high-confidence gene models, which was much higher than 1.01 for the low-confidence gene models.

We next compared our sequences with the fl-cDNA dataset, which consisted of 22,651 sequences generated from 'Haruna Nijo' (Sato et al. 2009b; Matsumoto et al. 2012). These sequences were created from plants grown in 12 different conditions and thus represent a good snapshot of the barley transcriptome. The average insert size of these fl-cDNA sequences was 1,711 bp, which was close to the cDNA length of the high-confidence gene models. Sequence similarities among our data and the fl-cDNA sequences, gene models of gene projection, and *de novo* gene annotations were compared using a BLASTN analysis with a threshold e-value of <−100 (Table 1-6). The 22,651 fl-cDNA query sequences showed high similarity with the sequences from the gene projection (19,771) and *de novo* annotation (19,636) (Table 1- 7). These numbers are consistent with the number of fl-cDNA sequences with complete ORFs (19,335) reported by Matsumoto et al. (2012); other fl-cDNA sequences had truncated ORF or non-protein coding sequences. The results also indicated that almost 10% of each gene model did not overlap each other. The amino acid sequences showed a lower level of overlapping than the nucleotide sequences (0.707–0.731; Table 1-8).

An example multiple alignment of fl-cDNA sequence AK371992 (BAK03190.1 for predicted protein) is shown in Fig. 1-5. Other than the completely identical sequences to BAK03190.1 (Horvu_HAR_1H01G006500.1 for gene projection and HarunaNijo.r1.1HG00001200.2 for *de novo* annotation), multiple other sequences showed high sequence similarity (Fig. 1-5A). Even under the higher similarity level of e-value=0, multiple sequences were identified partial agreement among amino acid sequences (Fig. 1-5B, C). The location and direction of nucleotide sequences are indicated on the genome browser (Fig. 1-6). Thus, some gene models with vicinity gene nomenclature numbers are suggested to be resulted from multi gene family or pseudo gene sequences.

Alignment with BAC clone sequences

We aligned 'Haruna Nijo' BAC clone sequences to pseudomolecules of 'Haruna Nijo' to estimate the contiguity of both sequences (Fig. 1-4). The BAC clones were analyzed using shotgun Sanger sequencing and assembled on an individual clone basis. The BAC clone

sequences of *Btr1*/*Btr2* were composed of several clones and showed apparent discontinuity with the pseudomolecule sequence of 'Haruna Nijo'. The alignment of these BAC sequences with the 'Morex'V3 pseudomolecule sequence revealed fragmentation at the 3' region, but the 5' region showed higher contiguity. Another BAC clone sequence, *Qsd1*, which was derived from a single clone, showed more contiguity with the pseudomolecules of 'Haruna Nijo'; however, there was a significant gap between the BAC sequence and the pseudomolecule sequence of 'Morex'V3. The quality of the BAC sequences was comparable with that of 'Morex'V3, but with some structural disorders.

The observed mismatches between BAC clones and pseudomolecules indicate that the pseudomolecules of 'Haruna Nijo' do not have as high of a sequencing quality as those of 'Morex'V3; however, they are useful for examining contiguity in the genome for gene identification.

Genome browser

The high-performance and user-friendly graphical interface genome browser JBrowse was used to visualize the pseudomolecule sequence and the gene models. Tracks of *de novo* annotations and gene projections each display the result of the associated annotation (e.g., exon structure, protein names, and transposable elements) to allow a comparison of each gene model. The fl-cDNA sequence track based on the BLAST search result against the pseudomolecule sequence was also provided, showing strict similarity to clones only. In addition to the browser, the user interface of the sequence similarity search programs BLAST and BLAT was also provided. The BLAST search results are directly linked to JBrowse as a user track, which allows the mapping of query sequences against the reference genome and their comparison with the gene models. The assembled sequence and annotation files can be downloaded from the website (http://viewer.shigen.info/harunanijo/index.php), so that our data can be used in the local user's environment.

Conclusion

Here, we present an assembly of 'Haruna Nijo' that is of similar quality to the 'Morex'V2 reference (Monat et al. 2019). Importantly, it is a European-style Japanese two-row cultivar, expanding barley genomic resources to Japanese and European breeding materials in contrast

to the American six-row cultivar ‘Morex’. Interesting phenotypes have since been identified in “Haruna Nijo”; its genome sequence assembly will facilitate the identification of the underlying genes.

Table 1-1 Number of reads in each platform

Library	Lane	No. of reads
PE440 (251 cycles)	1R1	382,415,190
	1R2	382,415,190
	2R1	247,652,536
	2R2	247,652,536
	3R1	286,023,953
	3R2	286,023,953
	Total	1,832,183,358
MP9K (151 cycles)	1R1	299,273,260
	1R2	299,273,260
	2R1	375,971,859
	2R2	375,971,859
	Total	1,350,490,238
10X (151 cycles)	1I1	349,118,027
	1R1	349,118,027
	1R2	349,118,027
	2I1	344,415,895
	2R1	344,415,895
	2R2	344,415,895
	Total	1,387,067,844
Hi-C (102 cycles)	R1	330,191,592
	R2	330,191,592
	Total	660,383,184

Table 1-2 Accession numbers for database submission

Study	Assembly	Sample
PRJEB44504	ERZ2088793*	ERS6294307-ERS6294316 (shared)

Experiment/Contigs	Run/Chromosome	Type
ERX5471660- ERX5471667	ERR5762904- ERR5762911	WGS, Genome

Name
Chromosome-scale assembly of barley (Hordeum vulgare) cultivar Haruna Nijo.

*GCA accession number will be available only once the data is public

ERS_ID	SAMEA_ID	Description
ERS6294307	SAMEA8609692	Hordeum vulgare cultivar Haruna Nijo - 10X
ERS6294308	SAMEA8609693	Hordeum vulgare cultivar Haruna Nijo - PE450
ERS6294309	SAMEA8609694	Hordeum vulgare cultivar Haruna Nijo - MP9
ERS6294313	SAMEA8609698	Hordeum vulgare cultivar Haruna Nijo - Hi-C
ERS6294316	SAMEA8609701	Hordeum vulgare cultivar Haruna Nijo - assembly

Table 1-3 Statistics of 'Haruna Nijo' and two versions of 'Morex' assemblies

Parameter	'Haruna Nijo'	'Morex'V2	'Morex'V3
Number of scaffolds in pseudomolecules	552	273	103
Pseudomolecule size (Gb)	4.28	4.34	4.20
Scaffold N50* [Mb]	18.9	43.7	118.9
Scaffold N90 [Mb]	2.6	5.9	21.8
Cumulative size of unanchored scaffold (Mb)	154.3	82.9	29.1

* "Scaffold" refers to top-level entities that constitute the pseudomolecules. In 'Morex'V3, these are Bionano scaffolds of PacBio HiFi contigs; in the other assemblies, superscaffolds were constructed from PE, MP, and 10X data.

Table 1-4 BUSCO statistics of 'Haruna Nijo'

Factor	Scaffolds	Pseudomolecule
Complete BUSCOs	1,403 (97.5%)	1,396 (96.9%)
Complete BUSCOs: single copy	1,382 (96.0%)	1,378 (95.7%)
Complete BUSCOS: duplicated	21 (1.3%)	18 (1.2%)
Fragmented BUSCOs	14 (1.0%)	14 (1.0%)
Missing BUSCOs	23 (1.5%)	30 (2.1%)
Total BUSCO groups searched	1,440	1,440

Table 1-5 Condensed output of RepeatMask for each chromosome of 'Haruna Nijo' in comparison with 'Morex' assemblies

category	1H	2H	3H	4H	5H	6H	7H	Un	Total_bp
DHH	1313509	1883031	1419595	1276444	1559800	1282331	1706571	306580	10747861
DTA	67598	73657	75149	60587	65938	48796	79625	9500	480850
DTB		0	0	81					81
DTC	37095769	47043958	43962982	45166305	43209929	43879459	50621122	36972607	347952131
DTH	879286	1164454	1068315	792965	1019951	797710	1058062	157799	6938542
DTM	1308897	1323471	1333615	1010121	1285027	1516951	1395035	311506	9484623
DTP			0		0			0	0
DTT	1083464	1381257	1305213	1002148	1286628	985410	1345735	281136	8670991
DTX	130416	168266	152412	122355	146351	134220	173963	31592	1059575
DXX	40309	58836	71399	40271	46528	43472	46207	3758	350780
NULL	114	206	0	177	456	0	352	0	1305
RIC				0					0
RII	0	89	314	0	115	0	0	0	518
RIJ	883	3029	349	720	298	7784	1065	884	15012
RIL	348	89	0	89	89	136	340	0	1091
RIP	0	0	0	0			82		82
RIR	0	0	0	0	0	0	0	0	0
RIX	1618273	1945288	1734292	1540078	1851025	1501594	1716219	342905	12249674
RLA	0	0		0	0				0
RLC	114425950	147864255	137564201	128716827	130846467	120300496	132900626	29895575	942514397
RLG	229535363	285597714	271500986	271826189	255248480	242211598	266728091	31736335	1854384756
RLH	0	0	0	0	0	0	0	0	0
RLX	1888833	2403922	2482920	2612049	2191854	2008159	2340725	320549	16249011
RSX	134703	188933	180719	122385	171684	114235	202494	66741	1181894
XXX	1014510	859245	1409929	2210778	937906	1935384	732818	15589153	24689723
Total_Bp	390538225	491959700	464262390	456500569	439868526	416767735	461049132	116026620	3236972897
Total length	528447123	678170541	639558213	629672760	608467472	565695744	634886329	158722699	4443620881
Total_Bp/Total length	0.739029901	0.725421808	0.725911075	0.724980653	0.722912146	0.73673479	0.72619162	0.73100206	0.728453886

category	'Morex'V2 tota	'Morex'V3 total
DHH	11065641	11080334
DTA	482089	477270
DTB	81	81
DTC	370506483	453854270
DTH	6889135	6881395
DTM	9221039	9516805
DTP	0	0
DTT	8618385	8709108
DTX	1105004	1093280
DXX	357660	354947
NULL	1212	1212
RIC	0	0
RII	329	410
RIJ	12308	11133
RIL	1112	1013
RIP	0	0
RIR	0	0
RIX	12118688	12091222
RLA	0	0
RLC	963087870	968561840
RLG	1900096742	1872636075
RLH	0	0
RLX	16474636	16109744
RSX	1191710	1204480
XXX	32030815	13009774
Total_Bp	3333260939	3375594393
Total length	4342738950	4196495466
Total_Bp/Total length	0.767548079	0.804384139

Table 1-6 *De novo* gene annotation statistics

Statistics	Complete sequences	High confidence	Low confidence
Number of genes	161,721	49,524	112,197
Number of monoexonic genes	67,724	12,645	55,079
Number of transcripts	181,980	68,751	113,229
Transcripts per gene	1.13	1.39	1.01
cDNA lengths (mRNAs)	1,294	1,696	1,050
CDS lengths (mRNAs)	1,154	1,377	1,018
Exons per transcript (mRNAs)	3.45	5.21	2.38
Exon lengths (mRNAs)	375	326	441
Intron lengths (mRNAs)	675	623	770
CDS exons per transcript (mRNAs)	3.33	4.95	2.35
CDS exon lengths	346	278	434
5' UTR exon number	54,193	48,584	5,609
3' UTR exon number	52,989	44,690	8,299

Table 1-7 BLASTN hits (< e-100) among nucleotide sequences of fl-cDNA, gene projection, and *de novo* gene model sequences.

Target	Query		
	Full-length cDNA	Gene projection	*De novo* annotation
Full-length cDNA	22,651	25,977	28,415
Gene projection	19,711	47,367	43,087
De novo annotation	19,636	42,336	49,524
Total hits	19,937	42,753	44,387
Ratio (total hits/number of queries)	0.880	0.903	0.896

Table 1-8 BLAST hits (< e-100) among sequences of full-length cDNA, gene projection and de novo gene model

Target	Query		
	Nucleotide	Protein	
	full length cDNA	gene projection	*de novo* annotation
full-length cDNA	22,651	20,379	22,985
gene projection	16,403	44,579	35,559
de novo annotation	16,377	31,324	49,524
total hits	16,562	31,500	36,217
ratio	0.731	0.707	0.731

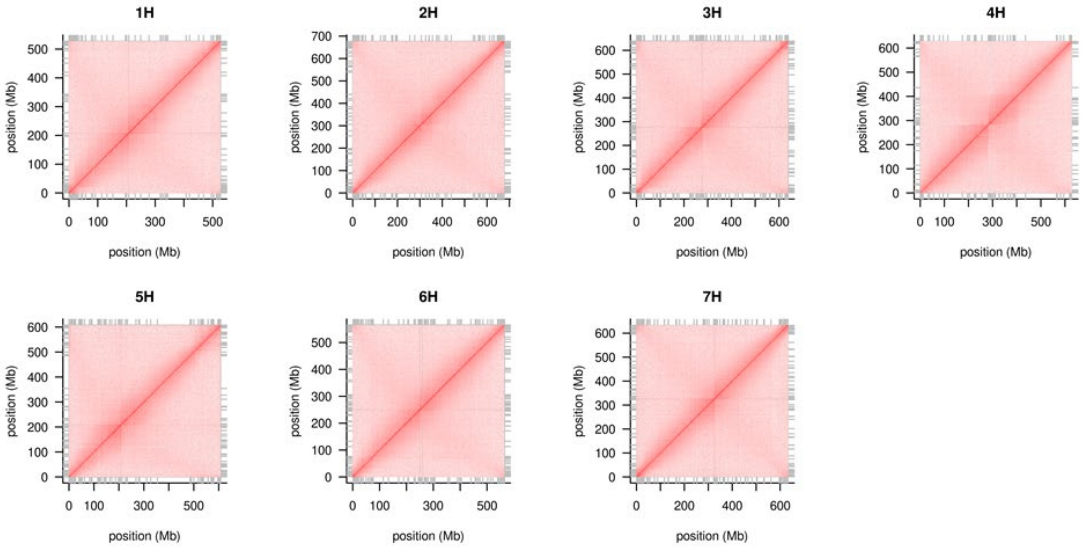


Fig. 1-1. Intra-chromosomal Hi-C contact matrices. Gray lines mark contig boundaries. Some centromeres are spanned by a single contig. The absence of off-(anti)-diagonal signals supports the accuracy of the assembly.

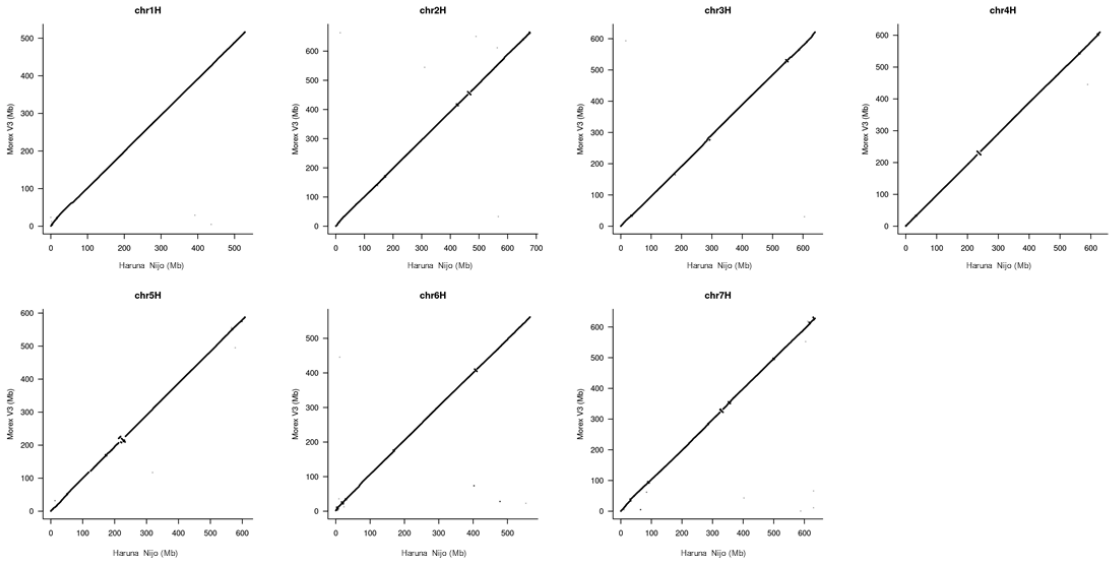


Fig. 1-2. Alignment of pseudomolecules of 'Haruna Nijo' to 'Morex'V3 individual chromosomes.

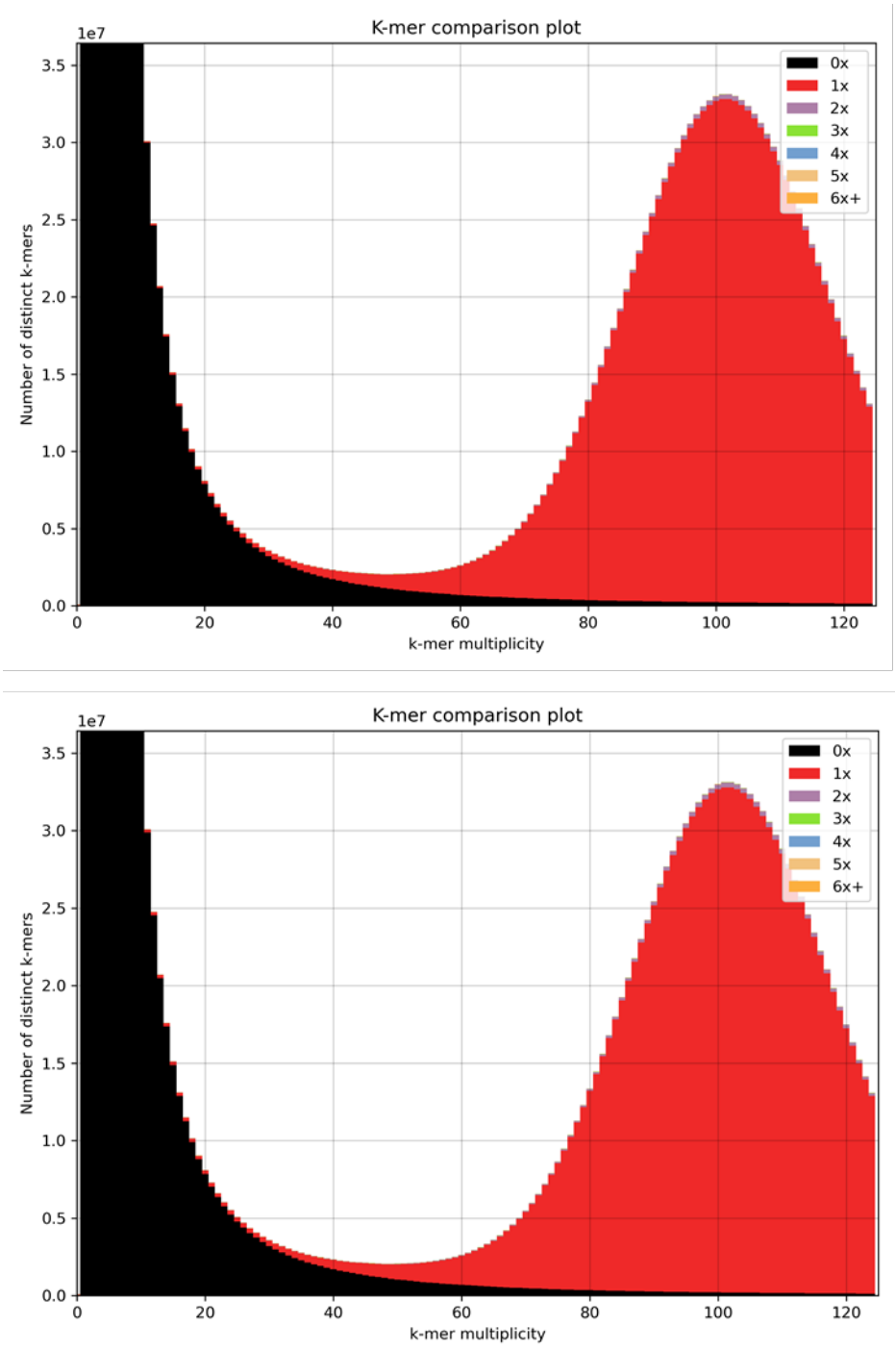


Fig. 1-3. Spectra cn plots comparing *k*-mers from the paired end and mate pair reads to *k*-mers in pseudomolecule assembly (upper) and scaffold assembly (lower).

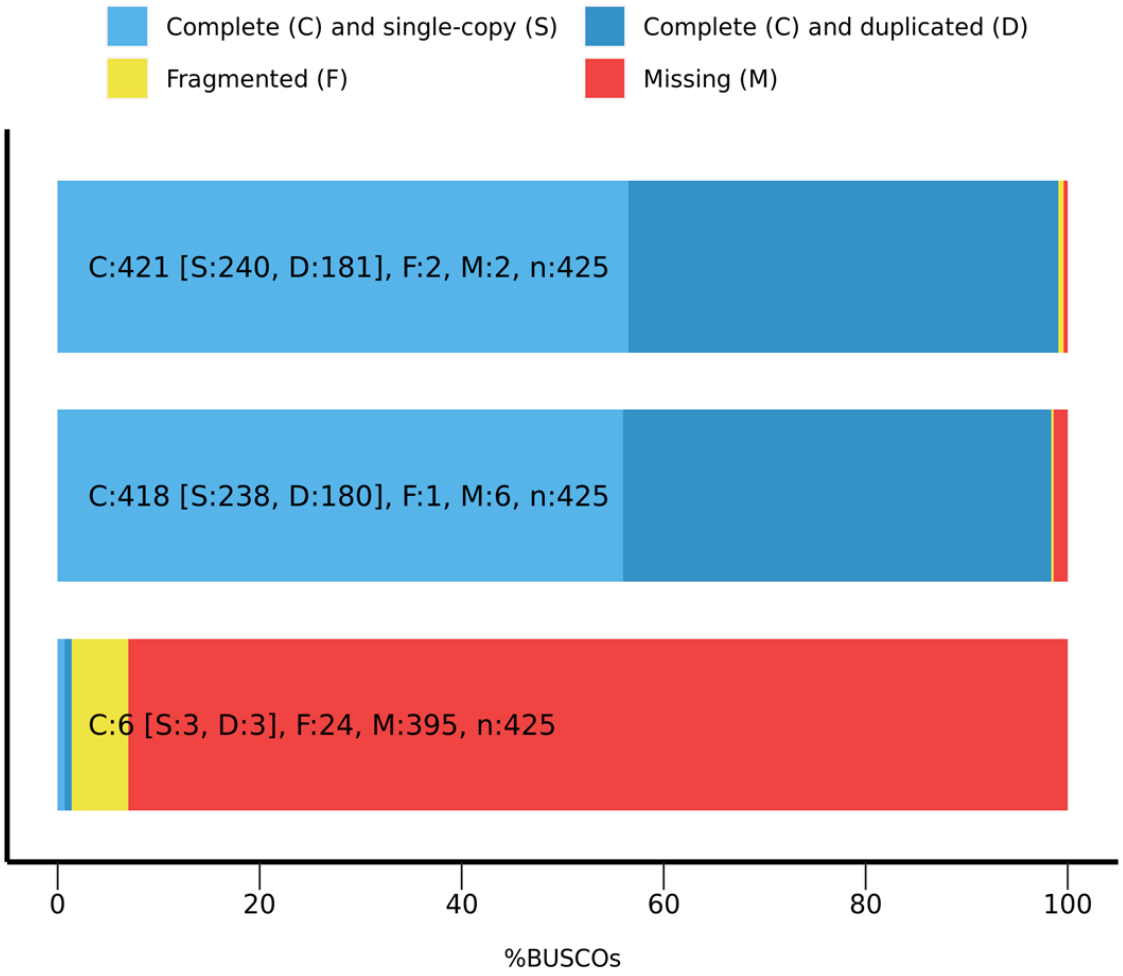


Fig. 1-4. BUSCO assessment results of 'Haruna Nijo' fl-cDNA sequences (upper), high-confidence genes (middle), and low-confidence genes (lower).

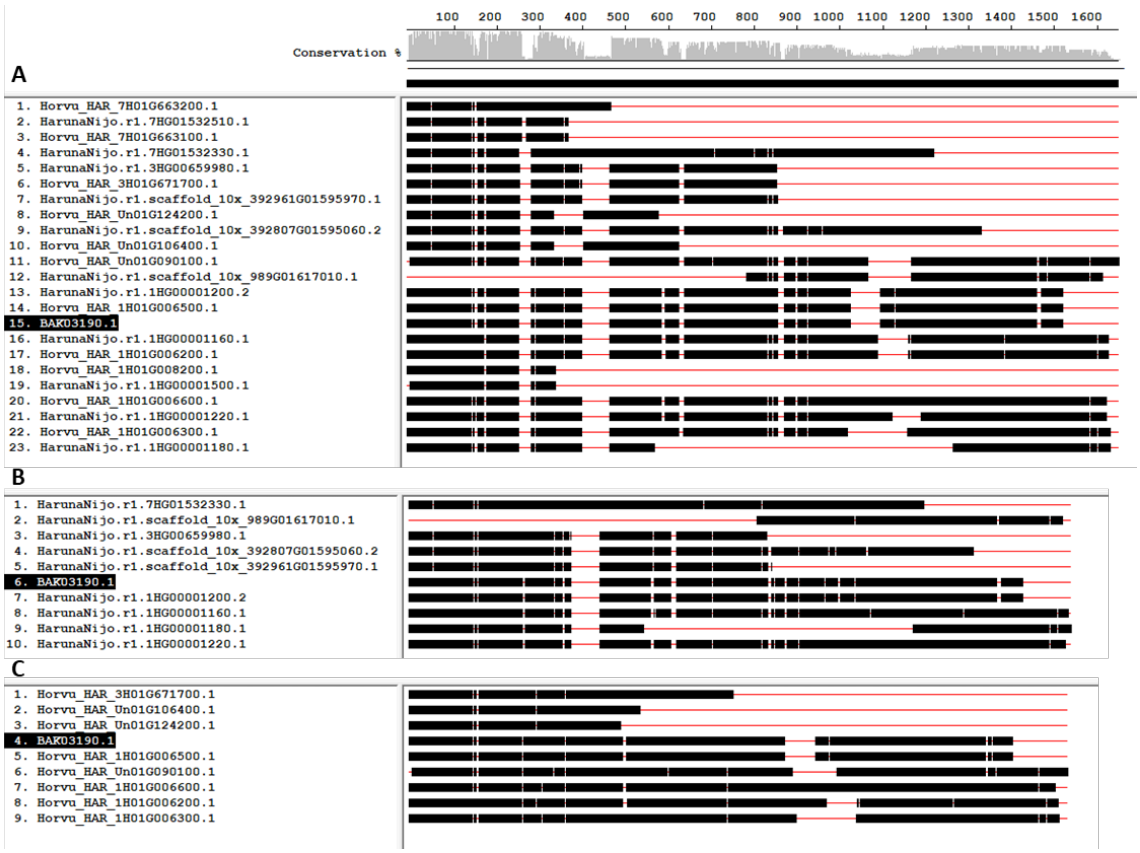


Fig. 1-5. An example multiple alignment of query protein sequences queried by BLASTX showing similarity to predicted protein BAK03190.1 from full-length sequence AK371992. An alignment of protein sequences showing e-value < -100 among protein sequences from gene projection (n=11) and *de novo* annotation (n=11) with BAK03190.1. B an alignment of protein sequences showing e-value =0 among protein sequences from *de novo* annotation (n=9). C an alignment of protein sequences showing e-value=0 among protein sequences from gene projection (n=8).

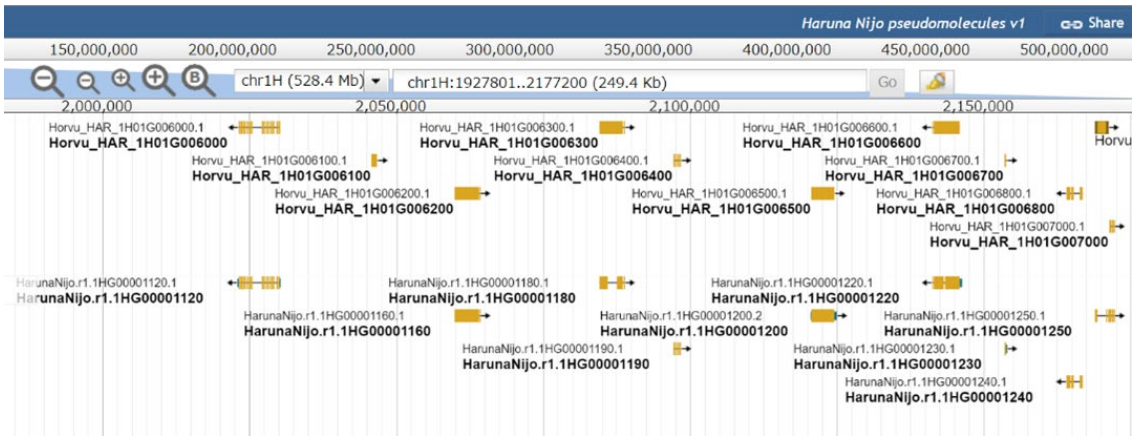
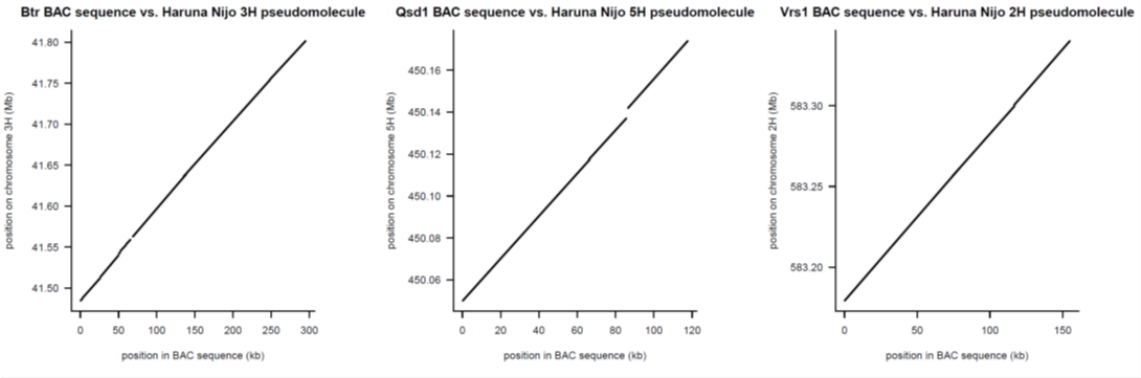


Fig. 1-6. An example snapshot of JBrowse genome browser showing similarity to predicted protein BAK03190.1 from full-length sequence AK371992.



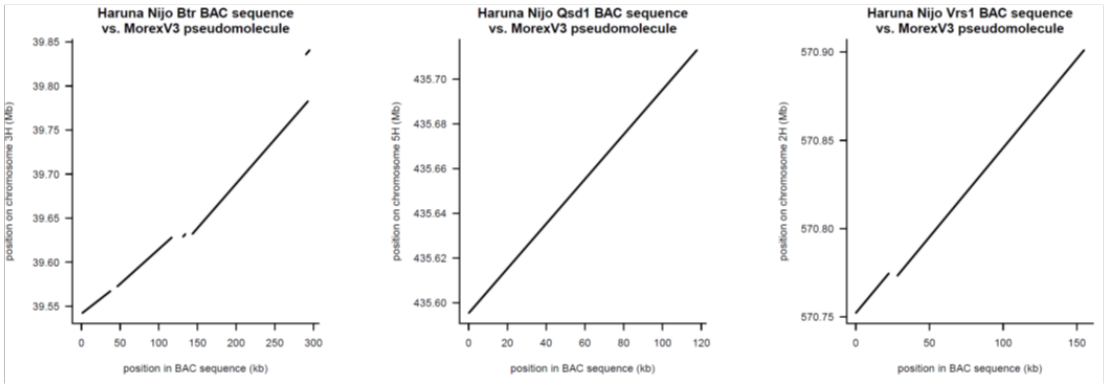


Fig. 1-7. Alignment of 'Haruna Nijo' BAC sequences of *Btr*, *Qsd1*, and *Vrs1* regions to pseudomolecules of 'Haruna Nijo' and 'Morex'V3.

Chapter 2

Field evaluation of heterosis in the top crosses of barley

Abstract

To estimate the amount of heterosis in morphological and yield component traits, F1s were developed from top crosses of barley and evaluated the influence of each parental genotype on the amount of heterosis, furthermore identify the promising combination of crosses. A total of 207 and 80 barley accessions were selected from diverse types and geographical origins around the world to be crossed with either 'Haruna Nijo' (HN) or 'Golden Promise' (GP) as common parent, respectively. Parents and F1s were grown in two replications at the field of Okayama University. Mean total biomass of parents and F1s were 37.5 g and 50.6 g in HN crosses (HN=41.4 g), and 39.2 g and 60.0 g in GP crosses (GP=30.3 g). Mean total biomass of six-row types was higher than that of two-row types in both HN (six-row: 53.7 g and two-row: 47.6 g) and GP (62.9 g and 53.2 g) crosses, respectively. However, the differences were not observed by hulled/hull-less comparisons. There are also differences in total biomass by geographical origin (in HN crosses: ranged 40.4 g in North Africa to 60.1 g in China). All the parents were genotyped by Illumina iSelect 50K and genotype by sequencing to estimate the relationship between heterosis and genotype. Several minor loci were identified for total weight and spike weight in both top crosses. F2 plants from selected cross combinations from each top cross showed wide range of distribution and QTL analyses showed several significant loci related to total weight. Annotations on these loci will be further investigated.

Key words: Hybrid vigor; Biomass; *Hordeum vulgare*; Genome wide association study

Introduction

F1 hybrid production is the most popular breeding technology especially in outbreeding species due to its hybrid vigor in field performance. Hybrid cultivar also has an advantage for breeding companies for one-time use of seeds in the seed market. Most of the cases outbreeding species show growth depression by self-pollination during the development of inbred lines (Charlesworth and Willis 2009) which are necessary to achieve higher level of hybrid performance. This inbreeding depression may reduce the amount of seed production thus the seed price of hybrid cultivar is more expensive than non-hybrid cultivars. Higher price of hybrid seeds has been attracted the interest of breeding companies but enough levels of hybrid vigor in yield or biomass are required for hybrid cultivars to pay the additional seed cost.

Major small grain cereals including rice (*Oryza sativa*), wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*) are inbreeding species and no depressions are existed by self-pollination. On the other hand, the level of hybrid heterosis in these crops is generally lower than outbreeding species. In case of rice, hybrid cultivars from the cross between japonica (*O. sativa* ssp. *japonica*) and indica (*O. sativa* ssp. *indica*) show higher level of hybrid vigor and they are widely cultivated in China (Ma and Yuan 2015). There are smaller markets of hybrid wheat cultivars mainly in EU countries where private companies are major providers for seed samples of new cultivars (Schwarzwälder et al. 2022). The situation of hybrid barley is similar in EU countries (Friedt and Ordon 2013).

Historically, there was a commercial hybrid barley breeding system in the United States. They used the system of genetic male sterility and balanced tertially trisomics to select hybrid seed samples (Ramage 1965), which was sophisticated but complicated to develop lines for hybrid seed production. Cultivars 'Hembar', 'Amy' and 'Belle' are spring type hybrid barley produced by commercial seed companies and grown in Arizona, USA in 1960-1970. They exhibited 15-20% yield advantage over the highest yielding cultivar in that time (Ramage 1975). However, the release of semi-dwarf cultivars has replaced commercial hybrids because they showed tolerance to lodging and ergot with the same yield advantage of hybrid barley. For that reason, hybrid barley has been omitted from the industry in the USA (Ramage 1983).

The present hybrid barley seed production mainly uses cytoplasmic male sterility (CMS) which was found in the germplasm of wild barley (*H. vulgare* ssp. *spontaneum*) (Ahokas 1982). Due to the use of hybrid seed production system based on CMS line, its maintainer and fertility restorer, cost for seed production is more expensive than the non-hybrid seed production. The increase in yield or biomass of barley crop should cover the price difference between hybrid and non-hybrid cultivars. There have been several reports on the hybrid vigor or heterosis in barley (Ramage 1983; Zhang et al. 2015). The degree of heterosis in barley F1 hybrid was reported high especially between different sources of parental combinations, e.g., between Japanese and Ethiopian accessions (Yasuda et al. 1994).

There is a barrier for hybrid production among barley germplasm. Domesticated barley has one of the genotypes of *Btr1Btr1btr2btr2* or *btr1btr1Btr2Btr2* in *BRITTLE RACHIS* (*Btr*) gene (Pourkheirandish et al. 2015) and the hybrid between different genotypes show shattering of rachis and its grain yield could not be estimated. This is also true for hybrid between domesticated barley (one of the above genotypes) and wild barley (*Btr1Btr1Btr2Btr2*). However, due to different evolutionary background for different *Btr* genotypes of domesticated barleys, its degree of heterosis could be high but unknown between *Btr* genotypes by the shattering of rachis. Exploration of unknown heterosis effects between different *Btr* genotypes may give important genetic information for heterosis. Barley cultivars also have different categories for uses which may have different genetic background. The hull-less caryopsis is controlled by a recessive allele of *Nud* gene (Taketa et al. 2008), which is important for food use of barley. Most of the malting barley cultivars have two kernel row type on the spike controlled by *Vrs1* gene (Komatsuda et al. 2007) due to its uniform grain sizes and thus six rowed types are mainly used for feed barleys. There have been no practical hybrid cultivars between parents with different germplasm of alleles between hulled and hull-less types or two-row and six-row types.

Mainly three models have been presented for the genetic basis of heterosis. Garcia et al. (2008) summarizes the cause of heterosis in to three hypotheses. The dominance hypothesis suggests that dominant alleles cancel the negative effects at multiple loci at the F1 hybrid. The overdominance hypothesis assumes that the loci with heterozygous genotypes are superior to both homozygous parents. Epistasis is also a possible cause of heterosis. There have been several attempts to prove these hypotheses. Garcia et al. (2008) develop a platform

to map QTLs to detect their additive and dominance effects, and epistatic interactions in maize and rice. They found that the estimated QTL dominant effects could not explain the observed heterosis but there is evidence that additive × additive epistatic effects of QTL could be the main cause for the heterosis in rice. They also show that heterosis in maize is mainly due to dominant gene action. Recently, Hashimoto et al. (2021) identified five QTLs for culm length in sorghum which were explained using the dominance model. They also cloned one of the uncharacterized genes (*Dw7a*) revealed that it encoded a MYB transcription factor, and suggested combining classic *dw1* or *dw3*, and new (*dw7a*) genes in hybrid breeding of sorghum.

The purpose of our research is to estimate the amount of heterosis in morphological and yield component traits in F1s developed from top crosses of barley. We evaluate the influence of each parental genotype on the amount of heterosis and identify the promising combination of crosses including different allelic combination of major loci differentiating germplasm. We also try to analyze genetic loci responsible for the cause of heterosis.

Materials and Methods

Plant materials and data collection

Each 207 and 80 barley accessions were selected from diverse cultivar types and geographical origins around the world to be crossed with either 'Haruna Nijo' (HN, a Japanese two-row hulled malting cultivar) or 'Golden Promise' (GP, a Scottish two-row hulled malting cultivar) as common parent, respectively (Table 2-1). Five plants of parents and F1s were grown in single rows which were 90 cm apart. Within-row spacing was 8 cm. The experiment has two replications located in different fields of Kurashiki at the Institute of Plant Science and Resources, Okayama University (34° 35′ N and 133° 46′ E). Heading date was scored for a plot and morphological traits (no. of culms, culm length, spike length, no. of kernels on the spike) were measured for three plants in each plot. After heading parental and hybrid plants showing rachis brittleness were bagged with plastic net to collect shattered grains. After maturity, spike weight and straw weight were also measured for each parent and F1 plant. The crosses showing higher performance in F1 generation were further planted F2 generation in

the same condition. Total weight, spike weight and number of spikes for each plant were measured.

To genotype parents and F2 plants, genomic DNA was extracted using a GENE PREP STAR PI-480 (KURABO, Osaka, Japan) according to the manufacturer's protocol (PLANT version 1).

SNP genotyping and population structure analysis

After quality check by gel electrophoresis, NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit (Thermo Fisher Scientific), DNA samples were genotyped via an Illumina iSelect® assay with 44,040 single nucleotide polymorphism (SNP) platform (Bayer et al. 2017). All SNP genotyping data were analyzed using GenomeStudio software (Illumina, USA).

The same DNA samples were analyzed by genotype by sequencing (GBS) according to the methods of Wendler et al. (2014). In brief, Illumina sequencing libraries were constructed as described in Meyer and Kirchner (2010) with some modifications. Subsequently, the indexed samples were pooled in equimolar ratios. The GBS library was eluted from the column in 20 μL EB and analysed with an Agilent 2100 Bioanalyser (Agilent Technology, Santa Clara) using the Agilent High Sensitivity DNA kit (Part# 5067-4626). Finally, the library was quantified using qPCR essentially as described previously (Mascher et al., 2013). The concentration was determined based on a standard curve and the average size of the GBS library. The sample was diluted to 10 nm for cluster formation on an Illumina cBot (Illumina, Inc.). Cluster formation and 2 × 150 bp sequencing-by-synthesis using Illumina's HiSeqX instrument were performed according to protocols provided by the manufacturer (Illumina Inc.).

Variant calling was conducted according to Milner et al. (2019). After adapter trimming, reads were aligned to the reference genome sequence of barley cultivar Morex (Mascher et al. 2020), HN (Sakkour et al. 2022) or GP (Jayakodi et al. 2020). Genotypes at bi-allelic sites with a minimum QUAL (i.e., mapping quality) score of 40 were called based on read depth ratios calculated from the DP (total read depth) and DV (depth of the alternative allele) fields. For the analysis of population structure and genetic similarity of samples, homozygous and heterozygous calls had to be supported by two and four reads,

respectively. SNP sites were retained if they had less than 10% missing data and less than 10% heterozygous calls and the number of heterozygous calls did not the exceed the number of homozygous calls for either allele. The filtered SNP matrix was exported as a VCF file.

Population structure and statistical analysis

Model-based estimation of ancestry coefficient was done with ADMIXTURE. ADMIXTURE's cross-validation procedure was used with the function –cv. The lowest cross-validation errors were selected for estimation of ancestry model. Parameter standard errors were estimated using bootstrapping by including the -B flag for 1000 replicates.

Analysis of variance was used to detect the differences between parents and F1s. Marker-phenotype associations were detected using TASSEL ver. 5 (Bradbury et al. 2007) with a general linear model. Heterozygous alleles identified in the Illumina iSelect® assay were eliminated from the analysis. Markers with less than 5% minor allele frequency and genotyping less than 90% of accessions were also eliminated from the analysis.

Results

Comparisons of traits between parents and F1s

Relative mean values of two replications for F1s and parental lines (P) were compared to the two common parents (1.0 in each top cross) of HN and GP (Fig. 2-1). As measurements for heterosis in barley, spike weight for grain yield and total weight for biomass were used. Analysis of variance (Table 2-2) showed that there were significant interactions for spike weight and total weight in HN top crosses. Factors for P (parent other than HN) vs F1 and accessions were also highly significant indicating that some accessions showed different performances between P and F1. Pairwise comparison of P and F1 in total biomass, spike weight and stem weight (Fig. 2-2A) indicated that most F1s showed higher value from mid-parent (MP: (P+HN)/2). The average heterosis of total biomass (F1/MP) was 1.35. The detailed observation indicated that the highest F1 showed total weight of 90.7 g which were 1.94 times higher than the respective P. Among biomass traits in GP, only stem weight showed significant interactions. The spike weight and total weight showed differences among P vs F1 comparison or accessions. Most pair wise comparison of P and F1 (Fig. 2-2B) showed

higher value from MP with average heterosis 1.53. The highest F1 biomass was 103.1 g and 1.80 times higher than P.

There was a significant difference in heading date between HN (129 days after planting: DAP, March 28) and GP (147 DAP, April 15). Thus, average heading dates of Ps and F1s were later in HN top crosses and earlier in GP top crosses. Analysis of variance (Table 2-2) also showed that there was a significant interaction in heading date between in both HN and GP top crosses. The average heading date in F1s was earlier the Ps in HN top crosses, but those were later in F1s than Ps in GP top cross.

There were also significant differences in interactions and primary factors. Of these traits, large average differences in number of grains per spike might be resulted from the cross between six row parent and two row top cross parent since the F1s show two rowed caryopsis which have smaller number of grains per spike.

Comparisons of heterosis between different main traits and origins

Mean total biomass of parents and F1s were 37.5 g and 50.6 g in HN crosses (HN=28.6 g), and 39.2 g and 60.0 g in GP crosses (GP=30.2 g) (Fig. 2A, 2B). Mean total biomass of six-row types was significantly higher than that of two-row types in both HN (six-row: 53.7 g and two-row: 47.6 g) and GP (62.9 g and 53.2 g) crosses, respectively (Table 2-3). However, the difference was not significant between means of hulled/hull-less (Table 2-3). Total biomass by collected regions was compared in parents and F1s (Table 2-4). There were no significant differences among parental means by collected regions both in HN and GP top crosses. However, those among F1s were significant in both top crosses. The F1 means in HN top cross ranged from 40.4 g (North Africa) to 60.1 g (China). Those in GP top cross ranged from 41.0 g (Turkey; n=2) to 74.4 g (Korean Peninsula; n=1). These results suggested that levels of heterosis between accessions and HN or GP were different among collected regions of accessions.

Genetic diversity of barley accession revealed by iSelect 50K system

HN top cross: We estimated genome-wide diversity of barley accessions using the position of barley MorexV1 reference genome derived from barley iSelect 50K system (Bayer *et al.*, 2017). Ancestry coefficient obtained from ADMIXTURE analysis (Alexander et al. 2009) indicated

the number of clusters as K = 8 for the 210 accessions. Clusters 1 to 3 mostly included East Asian accessions from China, the Korean Peninsula, Nepal, and Japan. Cluster 4 was composed of Southwest Asian and North African accessions. Cluster 5 was a mixture, with majority of accessions from Europe. Cluster 6 was mainly composed of accessions from Europe. Cluster 7 was composed mostly of Ethiopian accessions. A marker dissimilarity matrix showed that genetic distance and clustering partly agreed with that obtained by ADMIXTURE analysis, indicating that the materials used in the analysis have specific haplotypes linked to geographic origins.

The data from GBS was also analyzed using the reference genome assembly of HN (Sakkour et al. 2022). The clustering by ADMIXTURE indicated similar results from the analysis by iSelect 50K with the number of clusters with K=8.

GP top cross: The same sets of population structure analyses was conducted on the topcrosses of GP. The data by iSelect 50K system were analyzed with ADMIXTURE indicated the number of clusters as K = 5 for the 80 accessions (Fig. 2-1B). Clusters 1 to 3 mostly included East Asian accessions from China, the Korean Peninsula, Nepal, and Japan. Cluster 4 was composed of Southwest Asian and North African accessions. Cluster 5 was a mixture, with majority of accessions from Europe. Cluster 6 was mainly composed of accessions from Europe. Cluster 7 was composed mostly of Ethiopian accessions. A marker dissimilarity matrix showed that genetic distance and clustering partly agreed with that obtained by ADMIXTURE analysis, indicating that the materials used in the analysis have specific haplotypes linked to geographic origins.

The data from GBS was also analyzed using the reference genome assembly of GP V1 (Jayakodi et al. 2020). The clustering by ADMIXTURE indicated similar results from the analysis by iSelect 50K with the number of clusters with K=5.

Genome wide association studies on biomass and related traits in P and F1

For a genome-wide association study (GWAS), markers showing less than 5% minor allele frequency were omitted, and three biomass traits (total weight, spike weight) and other related traits were associated with markers in two sets of top crosses.

HN top crosses: A total of 29,080 markers from iSelect 50K were applied for GWAS with the reference assembly of MorexV1. For the total weight, associations below the significance threshold of $P < 0.05$ with Bonferroni correction (1.7e-06) were detected for 16 and 14 regions for P and F1 (Fig. 2-3a, b), respectively. The same analyses detected 24 and 2 regions for spike weight (Fig. 2-3c, d), respectively. The GWAS on total weight with 13,108 GBS markers with the reference assembly of HN detected 6 and 12 regions for P and F1 below the significance threshold of $P < 0.05$ with Bonferroni correction (3.8e-06) (Fig. 2-4a, b). The same analysis detected 25 and 0 regions for spike weight (Fig. 2-4c, d), respectively. The GWAS on total weight with 13,035 GBS markers with the reference assembly of MorexV3 detected 6 and 11 regions for P and F1 below the significance threshold of $P < 0.05$ with Bonferroni correction (3.8e-06) (Fig. 2-5a, b). The same analysis detected 23 and 0 regions for spike weight (Fig. 2-5c, d), respectively. Some of the detected positions agreed between the assemblies but differed in some positions.

GP top crosses: The same sets of analyses were conducted on GP top crosses. The iSelect 50K (29,666 markers) GWAS with the reference assembly of MorexV1 did not detect significant ($P < 0.05$ with Bonferroni correction: 1.7 e-06) associations for P and F1. The analysis with replication 1 detected 0 and 1 significant associations for P and F1, respectively (Fig. 2-6a, b). The same analyses by replication 1 detected 0 and 1 region for spike weight (Fig. 2-6c, d), respectively. The GWAS on total weight with 30,241 GBS markers with GP assembly did not detect both regions for P and F1 below the significance threshold of $P < 0.05$ with Bonferroni correction (1.7e-06) (Fig. 2-7a, b). The same analysis detected 1 and 0 regions for spike weight (Fig. 2-7c, d) respectively. The GWAS on total weight with 30,085 GBS markers with the reference assembly of MorexV3 detected no regions for P and F1 below the significance threshold of $P < 0.05$ with Bonferroni correction (1.7e-06). The analysis with replication 1 detected 0 and 5 significant associations for P and F1, respectively (Fig. 2-8a, b). The same analysis with replication 1 detected 1 and 1 regions for spike weight (Fig. 2-8c, d). Some detected positions agreed between the assemblies but differed in some positions.

F2 analyses of the crosses showing higher biomass in F1

Selected promising top crosses were further analyzed for segregation of biomass related traits in F2 generation. The segregation of total weight in five combinations with HN and five combinations with GP were shown in Fig. 2-9. Some crosses showed wider range of segregations and the marker analysis by iSelect 50K were applied for genetic analysis of total weight. A cross combination HN × SV172 detected three major QTL on the chromosomes 4H and 5H (Fig. 2-10). The cross GP × SV183 detected a QTL on chromosome 4H. The QTL on chromosome 4H showed vicinity position and could be the same QTL with dominance effects of 22.6 g and 11.8 g of total weigh in HN × SV172 and GP × SV183 crosses, respectively.

Discussion

The practical level of heterosis for commercial hybrid barley cultivar could be 5-10% (Friedt and Ordon 2013). The amount of heterosis has been reported in barley (Ramage 1983; Zhang et al. 2015). Since most of the commercial barley hybrid cultivars may use cytoplasmic male sterility facilitated seed production system, the change of cross combinations has some limitations due to the time and labor to develop parents by backcrossing. The exploration to evaluate exotic germplasm which potentially show higher level of heterosis could be avoided to keep practically acceptable quality of grain and straw products from hybrid barleys.

Level of plant based heterosis

As shown by Yasuda (1994), plant level of heterosis in grains could be double of the higher parents in some combinations between Ethiopian or Turkish landraces and Japanese two-row barleys. The similar top cross experiments were used but included diverse germplasm in the present study. We found enough levels of average heterosis in total weight in both HN (1.35) and GP (1.52) top crosses. Many of the crosses showed highest level of heterosis in total weight around 2.0 in both top crosses. We understand that the results of population growth in practical cultivation may different from the results of this study. However, the present analysis may provide important information to estimate the factors involved in genetic system for heterosis and transfer the factors to the parents of hybrid cultivars.

Heterosis among exotic cross combinations

We observed the geographical differences of germplasm showing the heterosis in top crosses. The results may indicate that the genetic factors controlling the heterosis may exist in the different genetic background. We also observe the different levels of heterosis between two-row and six-row germplasms which could be also resulted from the different genetic background since HN and GP are both major malting barley cultivars and may have less related genetic background from food and feed barleys. However, we could not find differences between hulled and hull-less trait for heterosis. The kernel hulls occupy 6-15% of the grain weight (Sato et al. 1989) and thus we measured total weight not by grain but by total spike weight. However, the non-significance between hull types for heterosis were unexpected and may explain their closer genetic background for heterosis.

Genetic factors associated with total weight

We found several minor loci for heterosis in total weight. He applied two different marker systems. Illumina iSelect system a high-density SNP detection platform basing SNP design on exome capture data from 170 lines from a barley diversity panel. However, it is a prefixed marker system based on the polymorphisms of 170 lines mainly derived from European germplasm with the reference annotation by MorexV1 (Mascher et al. 2017). Thus, we also applied partial *de novo* whole genome sequencing of genotype by sequencing with the reference genome of respective top cross parent of HN (Sakkour et al. 2022) and GP (Jayakodi et al. 2020).

Although the evidence of the heterosis by phenotypic data (F1 vs P) is obvious, the clear genetic basis of heterosis was not identified by the GWAS in each of the analysis methods. The results indicate that the genetic factors of heterosis might not be general among germplasm but could be identified in specific crosses.

Genetic model estimated by F2 crosses

From the results of the F2 segregation in some top crosses, we found enough levels of segregation in F2 plants and QTLs segregating in the populations. There was a significant QTL in both crosses of HN × SV172 and GP × SV183 on chromosome 4H. However, we could

not annotate the regions of QTL by any known loci with reasonable function or phenotypic control (e.g. flowering response).

Until the current time, the full magnitude of heterosis in plants remains a mystery. Heterosis is likely to be a response to multiple mechanisms, dominance (Bruce 1910, Jones 1917) single-locus overdominance (Shull 1908, East 1936) and epistasis (Schnell and Cockerham 1992) in which all mutually contribute to the heterosis for a specific trait. Several genes have linked to be contributing to the heterosis in grain yield in plants. In rice the accumulation of positive dominance alleles contributed to improvement in grain number in F1 hybrids. The heterozygous state of *Hd3a*, *Ghd7* and *Ghd8* control flowering time had overdominance effect on plant height and yield per plant and contributed to significant grain yield over parents by improving seed setting rate. However, there have been less information on the genetic factors controlling heterosis of barley. The present results may give opportunities for identifying responsible genes in barley heterosis.

Combining all the above information, we assume that we have enough level of heterosis among barley germplasm in the condition of Kurashiki, Japan. There are differences in levels of heterosis by cultivar groups and geographic origins, however, these differences were not identified by the GWAS type of analysis as genetic factors. There might be more specific factors, which could be detected in each cross-combination basis, however, the identification of these loci as genes requires more focused design for genetic analysis, especially in an inbreeding crop, barley.

The information obtained in this study may contribute to select the genotypes showing heterosis in two major malting barley germplasm and provide useful information for the candidate genetic factors to control heterosis in barley breeding programs.

Table 2-1A Accessions crossed with 'Haruna Nijo' and groups by ADMIXTURE (Q=8)

Accessic	Name	hull type	row type	Origin	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
SV004	Vantage (CI 7324)	hulled	6	Canada	0.10	0.09	0.57	0.00	0.16	0.06	0.02	0.00
SV007	Rabat	hulled	6	Morocco	0.00	0.00	0.00	0.00	0.95	0.00	0.00	0.05
SV009	Giza 117	hulled	6	Egypt	0.00	0.00	0.00	0.07	0.62	0.20	0.02	0.10
SV012	Giza 68	hulled	6	Egypt	0.00	0.00	0.00	0.05	0.43	0.47	0.00	0.05
SV018	Pukou 1	hulled	6	China	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV019	Chiaochuang 1	hulled	6	China	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV022	Tibet Violet 1	hull-less	6	China	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.99
SV025	Chihchou	hulled	6	China	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV027	Tayeh 1	hull-less	6	China	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.06
SV028	Paoanchen 1	hulled	6	China	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV031	Wuhu	hulled	6	China	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV034	Mushinchiang 3	hulled	6	China	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV036	Debre Zeit 1 (1-5-17a)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV037	Addis Ababa 40 (12-24-84)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV039	Mota 1 (1-24-10)	hulled	2	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV043	Debre Zeit 29 (1-5-35)	hulled	2	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV045	Nazareth 1 (1-10-1a)	hulled	2	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV048	Quiha 1 (2-3-73)	hulled	2	Ethiopia	0.00	0.00	0.00	0.98	0.00	0.02	0.00	0.00
SV049	Dembi 1 (2-20-51)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV052	Jijiga 2 (1-13-20)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV053	Kulubi 3 (1-14-38b)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV054	Deder 1 (1-16-27)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV055	Glyorgi 1 (1-24-21a)	hulled	2	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV056	Adi Abun 1 (1-28-2)	hulled	2	Ethiopia	0.00	0.00	0.00	0.98	0.00	0.02	0.00	0.00
SV057	Dessie 1 (2-4-22)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV059	Zeggi 1 (137-1)	hulled	2	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV060	Quesi (471)	hulled	2	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV064	Sheki 2 (260-2)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV065	Sombo 1 (281-1)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV066	Asella 1 (372a-1)	hulled	2	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV071	Meki 2 (481-2)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV072	Adigrat 8 (520c-2)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV075	Iraq Black Barley	hulled	2	Iraq	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
SV079	Khanaqin 2 (KUH 5305)	hulled	2	Iraq	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
SV081	Arbat 1 (KUH 5317)	hulled	2	Iraq	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
SV082	Chamchamal 1 (KUH 5326)	hulled	2	Iraq	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
SV085	Rewari	hulled	6	India	0.00	0.00	0.00	0.03	0.00	0.03	0.00	0.94
SV087	Iraq Barley 1	hulled	6	Iraq	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
SV088	Katana 2 (183)	hulled	2	Syria	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
SV093	Sarab 1 (347 I)	hulled	2	Iran	0.00	0.15	0.00	0.09	0.00	0.46	0.06	0.24
SV094	Dunga Banse (N 5)	hull-less	6	Pakistan	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.82
SV095	Happar 1 (N 14)	hull-less	6	Pakistan	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.82
SV097	Karad	hulled	6	India	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.98
SV101	Katana 1 (182)	hulled	2	Syria	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
SV104	Khanaqin 1 (KUH 5304)	hulled	2	Iraq	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
SV105	Khanaqin 4 (KUH 5307)	hulled	2	Iraq	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
SV111	Hayakiso 2	hulled	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV116	Sazanshu	hulled	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV117	Yatomi Mochi	hull-less	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV118	Sekitori	hulled	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV119	Suifu	hulled	6	Japan	0.00	0.00	0.00	0.00	0.21	0.00	0.79	0.00
SV122	Wase Bozu	hulled	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV126	Zairai 1	hulled	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV128	Shirochinko	hull-less	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV132	Wase Hadaka	hull-less	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV135	Hayatori Hadaka	hull-less	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV136	Zairai 2	hulled	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV141	Masan Naked 1	hull-less	6	Korea	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV147	Hamjong Covered 1	hulled	6	Korea	0.00	0.00	0.58	0.00	0.00	0.00	0.29	0.12
SV151	Buyong Oni Hadaka 2	hull-less	6	Korea	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV152	Dongsan Oni Hadaka 2	hull-less	6	Korea	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
SV159	Sunchang Naked 4	hull-less	6	Korea	0.00	0.00	0.04	0.00	0.00	0.00	0.95	0.00
SV164	Yeongdong Seungmaeg 1	hulled	6	Korea	0.00	0.00	0.59	0.00	0.00	0.00	0.34	0.07
SV168	Sama 1 (1385)	hull-less	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV169	Tilman Camp 1 (1398)	hull-less	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV170	Pisang 1 (1427)	hulled	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV171	Thangja 1 (1433)	hull-less	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV172	Birkna Camp 1 (1487)	hulled	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV175	Sipche 1	hull-less	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV176	Gho 1 (1392)	hull-less	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table 2-1A (continued)

Accessic	Name	hull type	row type	Origin	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
SV177	Katmandu 2 (1438)	hulled	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV179	Ngyak 1 (1524)	hull-less	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV180	Prok 1 (1537)	hull-less	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV181	Lumley 1 (1674)	hulled	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV183	Trisuli Bazar 3 (1449)	hulled	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV186	Dhumpu 1 (1596)	hulled	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
SV187	Sikha 2 (1626)	hull-less	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.96
SV194	Turkey 11	hulled	6	Turkey	0.00	0.00	0.07	0.04	0.77	0.00	0.01	0.10
SV201	Turkey 21	hulled	6	Turkey	0.00	0.00	0.00	0.00	0.85	0.03	0.00	0.11
SV215	Kleinwanz	hulled	6	Germany	0.00	0.00	0.07	0.00	0.11	0.00	0.82	0.00
SV223	Tammi	hulled	6	Finland	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
SV225	Erevan 1 (Cauc.15)	hulled	2	Armenia	0.00	0.25	0.00	0.12	0.00	0.49	0.00	0.14
SV229	KUH 837	hulled	2	Rumania	0.06	0.94	0.00	0.00	0.00	0.00	0.00	0.00
SV234	Hungarian	hulled	2	Hungary	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SV239	Saalegerste	hulled	2	Germany	0.18	0.82	0.00	0.00	0.00	0.00	0.00	0.00
SV240	Dometzkoer Paradies	hull-less	2	Czechoslovakia	0.00	0.34	0.10	0.05	0.01	0.11	0.16	0.23
SV242	Imperial	hulled	2	UK	0.68	0.32	0.00	0.00	0.00	0.00	0.00	0.00
SV245	Vega	hulled	6	Sweden	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
SV246	Olli	hulled	6	Finland	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
SV247	Caucasus	hulled	2	USSR	0.03	0.34	0.12	0.06	0.00	0.36	0.04	0.06
SV248	Geghard 1 (Cauc.44)	hulled	2	Armenia	0.01	0.23	0.04	0.11	0.00	0.43	0.00	0.18
SV250	PLD 5	hulled	2	Poland	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
SV259	Hanna	hulled	2	Germany	0.07	0.93	0.00	0.00	0.00	0.00	0.00	0.00
SV263	Prokupkuy nahy	hull-less	2	Czechoslovakia	0.00	0.00	0.64	0.00	0.00	0.00	0.07	0.29
SV264	Tuxsky nahy	hull-less	6	Czechoslovakia	0.00	0.00	0.63	0.00	0.00	0.00	0.06	0.31
SV265	Binder	hulled	2	Denmark	0.04	0.85	0.05	0.00	0.00	0.00	0.00	0.06
Z301	Adorra	hulled	2	Austria	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Z302	Akashinriki	hull-less	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Z303	Amagi Nijo	hulled	2	Japan	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z304	181652	hulled	2	wild	0.00	0.00	0.01	0.02	0.20	0.30	0.20	0.27
Z305	Betzes	hulled	2	USA	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Z306	Nudijaponicum Type 19	hull-less	6	Tibet	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.78
Z307	Deder 2 (1-16-28a)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Z308	Golden Promise	hulled	2	UK	0.23	0.77	0.00	0.00	0.00	0.00	0.00	0.00
Z309	Hokuiku17	hulled	2	Japan	0.64	0.36	0.00	0.00	0.00	0.00	0.00	0.00
Z310	HOR10350	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Z311	HOR13821	hulled	2	Turkey	0.03	0.32	0.02	0.12	0.04	0.42	0.02	0.03
Z312	HOR13942	hulled	6	S.Europe	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Z313	HOR7552	hull-less	6	Pakistan	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.81
Z314	Aleppo 1 (438)	hulled	2	Syria	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Z315	Katana 2 (183)	hulled	2	Syria	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Z316	Partek 1 (N 114)	hull-less	6	Indiaia	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.83
Z317	Katana 1 (182)	hulled	2	Syria	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Z318	Katana 3 (184)	hulled	2	Syria	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Z319	Igri	hulled	2	Germany	0.00	0.25	0.00	0.00	0.75	0.00	0.00	0.00
Z320	Wasemugi	hulled	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Z321	Ko A	hulled	2	Japan	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z322	Rokujo	hulled	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Z323	Hakata 2	hulled	2	Japan	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z324	Tochigi Goldenmelon	hulled	2	Japan	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z325	Jangseong Covered 2	hulled	6	Korea	0.00	0.00	0.16	0.00	0.00	0.00	0.84	0.00
Z326	Kawasaigoku	hulled	2	Japan	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z327	Murasaki Mochi	hull-less	6	Japan	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Z328	Chame 1 (1493)	hulled	6	Nepal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Z329	OWB DOM	hulled	2	Unknown	0.00	0.00	0.00	0.14	0.23	0.27	0.20	0.16
Z330	Sapporo 2	hulled	2	Japan	0.68	0.32	0.00	0.00	0.00	0.00	0.00	0.00
Z331	H602	hulled	2	wild	0.00	0.00	0.00	0.01	0.23	0.34	0.21	0.20
Z332	Turkey 284	hulled	6	Turkey	0.00	0.00	0.38	0.00	0.59	0.04	0.00	0.00
Z333	Turkey 552	hulled	6	Turkey	0.00	0.00	0.39	0.00	0.58	0.03	0.00	0.00
Z334	Tern	hulled	2	Germany	0.13	0.87	0.00	0.00	0.00	0.00	0.00	0.00
Z335	Turkey 496	hulled	2	Turkey	0.00	0.27	0.00	0.09	0.00	0.53	0.01	0.09
Z336	Turkey 605	hulled	2	Turkey	0.00	0.25	0.00	0.11	0.00	0.53	0.00	0.11
Z337	Aurore	hulled	2	France	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Z338	Hatif de Grignon	hulled	6	France	0.00	0.00	0.11	0.00	0.89	0.00	0.00	0.00
Z339	Yugoslavia 85	hulled	6	Yugoslavia	0.01	0.20	0.07	0.09	0.08	0.22	0.11	0.21
Z340	Baku 3 (Cauc.32)	hulled	6	Azerbaijan	0.00	0.02	0.01	0.00	0.14	0.21	0.32	0.30
Z341	Shanghai 1	hull-less	6	China	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00

Table 2-1A (continued)

Accessic	Name	hull type	row type	Origin	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
BCS098	Advance	hulled	6	USA	0.00	0.14	0.39	0.00	0.34	0.04	0.04	0.05
BCS099	Albany	hulled	2	Mexico	0.00	0.90	0.00	0.00	0.00	0.10	0.00	0.00
BCS102	Antarctica 05	hulled	2	Brazil	0.02	0.98	0.00	0.00	0.00	0.00	0.00	0.00
BCS117	FNC 6-1	hulled	2	Uruguay	0.08	0.76	0.02	0.00	0.12	0.00	0.01	0.01
BCS131	Manchuria	hulled	6	USA	0.00	0.00	0.97	0.00	0.00	0.00	0.02	0.01
BCS141	Teran	hulled	2	Ecuador	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
BCS142	TR 306	hulled	2	Canada	0.01	0.77	0.19	0.00	0.03	0.00	0.00	0.00
BCS148	Igri	hulled	2	Germany	0.00	0.25	0.00	0.00	0.75	0.00	0.00	0.00
BCS150	Pipkin	hulled	2	United Kingdom	0.09	0.76	0.04	0.00	0.10	0.00	0.01	0.00
BCS154	Colonia	hulled	6	Germany	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
BCS164	Elassona	hulled	6	Greece	0.00	0.00	0.26	0.00	0.70	0.04	0.00	0.00
BCS167	Strickhoff	hulled	2	Switzerland	0.00	0.09	0.16	0.00	0.74	0.00	0.00	0.00
BCS170	Blenheim	hulled	2	United Kingdom	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
BCS174	Gryf	hulled	2	Poland	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
BCS176	Hunter	hulled	2	Ireland	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
BCS181	Perun	hulled	2	Germany	0.00	0.47	0.15	0.00	0.38	0.00	0.00	0.00
BCS184	Quantum	hulled	2	Austria	0.00	0.78	0.00	0.02	0.00	0.18	0.01	0.02
BCS185	Särla	hulled	2	Sweden	0.00	0.80	0.20	0.00	0.00	0.00	0.00	0.00
BCS186	Spratt Archer	hulled	2	United Kingdom	0.24	0.69	0.01	0.03	0.00	0.00	0.02	0.01
BCS187	Tyra	hulled	2	Denmark	0.00	0.85	0.00	0.00	0.15	0.00	0.00	0.00
BCS188	Vada	hulled	2	Netherlands	0.00	0.72	0.00	0.00	0.02	0.25	0.00	0.01
BCS189	Valticky	hulled	2	Czech Republic	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
BCS190	Volla	hulled	2	Germany	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
BCS193	Heils Franken	hulled	2	Germany	0.11	0.86	0.00	0.00	0.00	0.03	0.00	0.00
BCS194	HOR 8676	hulled	2	Czech Republic	0.00	0.88	0.00	0.12	0.00	0.00	0.00	0.00
BCS197	Souce 142 Strotzheim	hulled	2	France	0.11	0.80	0.00	0.00	0.00	0.09	0.00	0.00
BCS198	Agneta	hulled	6	Finlandland	0.00	0.00	0.80	0.00	0.20	0.00	0.00	0.00
BCS206	Zhodinskii 5	hulled	2	Belarus	0.01	0.82	0.06	0.02	0.06	0.00	0.03	0.00
BCS208	Auksinjai	hulled	2	Lituania	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
BCS212	K 3222	hulled	2	Russian Federat	0.00	0.80	0.20	0.00	0.00	0.00	0.00	0.00
BCS217	K 5279	hull-less	2	Kazakhstan	0.00	0.10	0.10	0.05	0.02	0.17	0.18	0.37
BCS222	Novosadski 293	hulled	2	Yugoslavia	0.00	0.20	0.00	0.00	0.52	0.17	0.06	0.05
BCS224	Ros	hulled	2	Ukraine	0.00	0.74	0.00	0.04	0.19	0.00	0.02	0.01
BCS225	Novosadski 294	hulled	2	Yugoslavia	0.02	0.45	0.00	0.00	0.52	0.00	0.00	0.00
BCS226	Vetulio	hulled	6	Italy	0.00	0.00	0.00	0.00	0.30	0.00	0.17	0.53
BCS233	Magurele	hulled	2	Romania	0.10	0.82	0.04	0.00	0.00	0.04	0.00	0.00
BCS239	HOR 12721	hulled	2	Italy	0.00	0.27	0.05	0.00	0.68	0.00	0.00	0.00
BCS244	114644	hulled	2	Syria	0.00	0.07	0.00	0.00	0.12	0.82	0.00	0.00
BCS248	115073	hulled	2	Morocco	0.04	0.81	0.00	0.00	0.15	0.00	0.00	0.00
BCS255	115647	hulled	2	Egypt	0.18	0.82	0.00	0.00	0.00	0.00	0.00	0.00
BCS269	121803	hulled	2	Iraq	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
BCS276	123345	hulled	6	Tajikistan	0.00	0.00	0.00	0.00	0.05	0.18	0.55	0.22
BCS277	123395	hulled	2	Georgia	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
BCS282	123706	hulled	2	Turkey	0.05	0.44	0.00	0.08	0.00	0.40	0.00	0.02
BCS287	123723	hulled	6	Afghanistan	0.00	0.00	0.00	0.02	0.00	0.21	0.59	0.19
BCS288	123724	hulled	6	Afghanistan	0.00	0.00	0.00	0.00	0.00	0.10	0.75	0.14
BCS295	123742	hulled	6	Egypt	0.00	0.00	0.00	0.10	0.46	0.33	0.03	0.07
BCS312	123802	hulled	2	Algeria	0.00	0.64	0.00	0.00	0.30	0.05	0.00	0.00
BCS317	123818	hulled	2	Iran	0.11	0.43	0.00	0.11	0.00	0.25	0.00	0.09
BCS318	123823	hulled	2	Afghanistan	0.01	0.99	0.00	0.00	0.00	0.00	0.00	0.00
BCS323	123833	hulled	6	Jordan	0.00	0.00	0.00	0.00	0.29	0.19	0.10	0.42
BCS324	123836	hulled	2	Jordan	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
BCS332	124007	hulled	2	Australia	0.03	0.30	0.00	0.00	0.67	0.00	0.00	0.00
BCS333	124010	hulled	2	Australia	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
BCS335	124016	hulled	2	Australia	0.00	0.92	0.00	0.08	0.00	0.00	0.00	0.00
BCS336	Ethiopia 16 (CI 1225)	hulled	2	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
BCS337	Ethiopia 127 (CI 4373)	hull-less	6	Ethiopia	0.00	0.01	0.00	0.80	0.00	0.06	0.04	0.09
BCS340	Ethiopia 468 (CI 9660)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
BCS344	Ethiopia 107 (CI 3915-1)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
BCS348	Ethiopia 574 (CI 9795)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
BCS355	Ethiopia 524 (CI 9716)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
BCS357	Deder 1 (1-16-27)	hulled	6	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
BCS359	Asella 9 (372e-2)	hulled	2	Ethiopia	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
BCS361	180001	hulled	2	wild	0.00	0.00	0.04	0.02	0.29	0.30	0.18	0.17
BCS362	180007	hulled	2	wild	0.00	0.00	0.02	0.02	0.22	0.40	0.18	0.17
BCS367	180219	hulled	2	wild	0.00	0.00	0.00	0.02	0.21	0.44	0.14	0.18
BCS373	181277	hulled	2	wild	0.01	0.00	0.06	0.02	0.32	0.25	0.16	0.18
BCS387	181590	hulled	2	wild	0.00	0.00	0.00	0.04	0.24	0.29	0.18	0.24
BCS390	181616	hulled	2	wild	0.00	0.00	0.00	0.00	0.17	0.62	0.14	0.08
BCS393	181678	hulled	2	wild	0.00	0.00	0.03	0.01	0.24	0.35	0.17	0.21

Table 2-1B Accessions crossed with 'Golden Promise' and groups by ADMIXTURE (Q=5)

Accession	Name	hull type	row type	Origin	Q1	Q2	Q3	Q4	Q5
SV003	Olympia (CI 6107)	1	6	USA	0.00	0.71	0.14	0.15	0.00
SV017	Vladivostok	1	6	China	0.00	1.00	0.00	0.00	0.00
SV018	Pukou 1	1	6	China	0.00	0.00	0.00	0.00	1.00
SV019	Chiaochuang 1	1	6	China	0.00	0.00	0.00	0.00	1.00
SV022	Tibet Violet 1	2	6	China	0.00	0.00	1.00	0.00	0.00
SV023	Manchuria Native 1	1	6	China	0.00	0.33	0.07	0.00	0.60
SV025	Chihchou	1	6	China	0.00	0.00	0.00	0.00	1.00
SV027	Tayeh 1	2	6	China	0.00	0.00	0.05	0.00	0.95
SV031	Wuhu	1	6	China	0.00	0.00	0.00	0.00	1.00
SV034	Mushinchiang 3	1	6	China	0.00	0.00	0.00	0.00	1.00
SV036	Debre Zeit 1 (1-5-17a)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV037	Addis Ababa 40 (12-24-84)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV039	Mota 1 (1-24-10)	1	2	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV043	Debre Zeit 29 (1-5-35)	1	2	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV045	Nazareth 1 (1-10-1a)	1	2	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV046	Asbe Tafari 1 (1-11-19)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV048	Quiha 1 (2-3-73)	1	2	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV049	Dembi 1 (2-20-51)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV051	Addis Ababa 3 (12-24-9)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV052	Jijiga 2 (1-13-20)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV053	Kulubi 3 (1-14-38b)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV055	Glyorgi 1 (1-24-21a)	1	2	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV056	Adi Abun 1 (1-28-2)	1	2	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV059	Zeggi 1 (137-1)	1	2	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV060	Quesi (471)	1	2	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV062	Adigrat 3 (508-2)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV063	Fiche 1 (212a-1)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV064	Sheki 2 (260-2)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV065	Sombo 1 (281-1)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV066	Asella 1 (372a-1)	1	2	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV069	Jimma 2 (265-1)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV071	Meki 2 (481-2)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV072	Adigrat 8 (520c-2)	1	6	Ethiopia	1.00	0.00	0.00	0.00	0.00
SV079	Khanaqin 2 (KUH 5305)	1	2	Iraq	0.00	0.00	0.00	1.00	0.00
SV085	Rewari	1	6	India	0.06	0.00	0.87	0.07	0.00
SV088	Katana 2 (183)	1	2	Syria	0.00	0.00	0.00	1.00	0.00
SV093	Sarab 1 (347 I)	1	2	Iran	0.10	0.00	0.24	0.60	0.06
SV094	Dunga Banse (N 5)	2	6	Pakistan	0.00	0.00	0.79	0.00	0.21
SV095	Happar 1 (N 14)	2	6	Pakistan	0.00	0.00	0.79	0.00	0.21
SV096	Baiji 2 (KUH 5333)	1	2	Iraq	0.00	0.00	0.00	1.00	0.00
SV101	Katana 1 (182)	1	2	Syria	0.00	0.00	0.00	1.00	0.00
SV107	Sinjar 1 (KUH 5337)	1	2	Iraq	0.00	0.00	0.00	1.00	0.00
SV116	Sazanshu	1	6	Japan	0.00	0.00	0.00	0.00	1.00
SV117	Yatomi Mochi	2	6	Japan	0.00	0.00	0.00	0.00	1.00
SV122	Wase Bozu	1	6	Japan	0.00	0.00	0.00	0.00	1.00
SV125	Tainan	1	6	Japan	0.00	0.00	0.00	0.00	1.00
SV126	Zairai 1	1	6	Japan	0.00	0.00	0.00	0.00	1.00
SV132	Wase Hadaka	2	6	Japan	0.00	0.00	0.00	0.00	1.00
SV135	Hayatori Hadaka	2	6	Japan	0.00	0.00	0.00	0.00	1.00
SV136	Zairai 2	1	6	Japan	0.00	0.00	0.00	0.00	1.00

Table 2-1B (continued)

Accession	Name	hull type	row type	Origin	Q1	Q2	Q3	Q4	Q5
SV147	Hamjong Covered 1	1	6	Korea	0.00	0.57	0.13	0.00	0.30
SV169	Tilman Camp 1 (1398)	2	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV170	Pisang 1 (1427)	1	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV171	Thangja 1 (1433)	2	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV172	Birkna Camp 1 (1487)	1	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV174	Ulleri 1 (1633)	1	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV175	Sipche 1	2	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV179	Ngyak 1 (1524)	2	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV180	Prok 1 (1537)	2	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV181	Lumley 1 (1674)	1	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV183	Trisuli Bazar 3 (1449)	1	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV186	Dhumpu 1 (1596)	1	6	Nepal	0.00	0.00	1.00	0.00	0.00
SV187	Sikha 2 (1626)	2	6	Nepal	0.00	0.00	0.99	0.00	0.01
SV223	Tammi	1	6	Finland	0.00	1.00	0.00	0.00	0.00
SV245	Vega	1	6	Sweden	0.00	1.00	0.00	0.00	0.00
SV246	Olli	1	6	Finland	0.00	1.00	0.00	0.00	0.00
SV254	Tripoli	1	6	Italy	0.06	0.40	0.21	0.28	0.05
SV264	Tuxsky nahy	2	6	Czechoslovakia	0.03	0.57	0.28	0.00	0.11
SV267	Vankhuri	1	2	Finland	0.03	0.51	0.01	0.45	0.00
Z101	Katana 1 (182)	1	2	Syria	0.00	0.00	0.00	1.00	0.00
Z102	Harrington	1	2	Canada	0.01	0.59	0.00	0.40	0.00
Z103	Turkey 605 (1986)	1	2	Turkey	0.12	0.00	0.14	0.74	0.00
Z104	Turkey 496	1	2	Turkey	0.12	0.01	0.10	0.76	0.01
Z105	Murasaki Mochi	2	6	Japan	0.00	0.00	0.00	0.00	1.00
Z106	Katana 3 (184)	1	2	Syria	0.00	0.00	0.00	1.00	0.00
Z107	Ghazvin 2 (193 I)	1	2	Iran	0.00	0.00	0.00	1.00	0.00
Z108	Partek 1 (N 114)	2	6	India	0.00	0.00	0.81	0.00	0.19
Z109	Aleppo 1 (438)	1	2	Syria	0.00	0.00	0.00	1.00	0.00
Z110	Adorra	1	2	Austria	0.00	0.57	0.00	0.43	0.00
Z111	Shanghai 1	2	6	China	0.00	0.00	0.00	0.00	1.00

Table 2-2 Analysis of variance for parents and F1s top crossed with 'Haruna Nijo' or 'Golden Promise'

Haruna Nijo top cross		heading date		stem weight		spike weight		stem + spike weight	
factor	d.f.	M.S.	P-value	M.S.	P-value	M.S.	P-value	M.S.	P-value
p vs f1	1	16346.67	2.0403E-145	15720.48	1.27E-59	3887.534	5.15E-30	35243.09	2.84E-49
accessions	206	187.3457	8.1474E-131	156.2989	6.94E-30	102.2804	2.14E-33	461.5739	1.5E-30
interact.	206	48.72127	1.73629E-42	59.52477	0.001839	46.92201	1.01E-07	167.083	0.004628
error	414	10.05918		42.2417		25.52316		122.8368	
total	827								
Golden Promist top cross									
p vs f1	1	2300.512	2.94333E-23	18161.46	4.8E-44	2629.402	1.2E-14	34611.69	9.07E-33
accessions	79	155.1525	2.80823E-32	116.56	1.04E-06	71.35627	0.000165	314.1924	4.44E-05
interact.	79	35.94921	2.48751E-05	81.27861	0.002555	36.9271	0.45772	182.4612	0.153748
error	160	16.7875		47.92807		36.32674		150.493	
total	319								

Haruna Nijo top cross		no. of spikes		stem length		spike length		no. of grains/spike	
factor	d.f.								
p vs f1	1	M.S.	P-value	M.S.	P-value	M.S.	P-value	M.S.	P-value
accessions	206	789.7848	2.70336E-21	23481.65	6.17E-53	15.52174	2.5E-07	60653.76	2.7E-207
interact.	206	27.33936	3.03728E-27	466.9304	1.44E-56	6.164056	9.5E-92	475.9244	1E-164
error	414	10.19387	0.014411008	143.0931	1.07E-08	2.31506	1.8E-34	498.7116	1.6E-168
total	827	7.870102		74.31873		0.564255		16.64687	
Golden Promist top cross									
p vs f1	1								
accessions	79	1205.775	1.37898E-22	4821.806	0.003255	123.132	3.38E-21	46368.45	1.42E-78
interact.	79	35.12732	3.10051E-13	1060.904	0.000166	4.883742	3.13E-17	488.383	1.01E-42
error	160	14.42593	0.008322949	601.8665	0.281041	2.844999	2.4E-08	503.286	1.44E-43
total	319	9.176476		540.2174		1.025587		35.83663	

Table 2-3 Averages of total weight between groups with different kernel traits

Top crosses	Trait	n	Biomass mean (g)	P-value	Trait	n	Biomass mean (g)	P-value
Haruna Nijo	Tow-rowed	106	47.6	0.0006***	Hulled	176	50.5	0.77ns
	Six-rowed	101	53.7		Hull-less	31	51.2	
'Golden Promise'	Tow-rowed	24	53.2	0.0013**	Hulled	63	60.6	0.39ns
	Six-rowed	56	62.9		Hull-less	17	57.6	

***,**,ns: Significant at 0.1 % and 1% levels, and non-significant, respectively

Table 2-4 Analysis of variance of the difference of total weight trait among origin

Area	Haruna Nijo top cross			Golden Promise top cross		
	n	Parent ^{ns} (g)	F1*** (g)	n	Parent ^{ns} (g)	F1*** (g)
Others	10	40.9	55.0	2	47.6	50.6
North Africa	7	34.9	40.4	0	-	-
China	10	40.3	60.1	10	40.8	68.0
Ethiopia	32	39.9	57.8	23	39.4	63.0
Wild	9	32.4	46.9	0	-	-
South Asia	37	36.5	47.4	14	41.5	53.6
Japan	22	41.7	50.6	9	40.8	63.4
Korean	7	41.5	47.6	1	30.8	74.4
Nepal	15	37.5	52.6	12	36.4	60.9
Turkey	8	34.0	48.8	2	31.4	41.0
Europe	50	35.1	47.7	7	34.8	53.6

***,*,ns: Significant at 0.1 %, 5% levels and non-significant, respectively by one way ANOVA among areas of P or F1 in each cross.

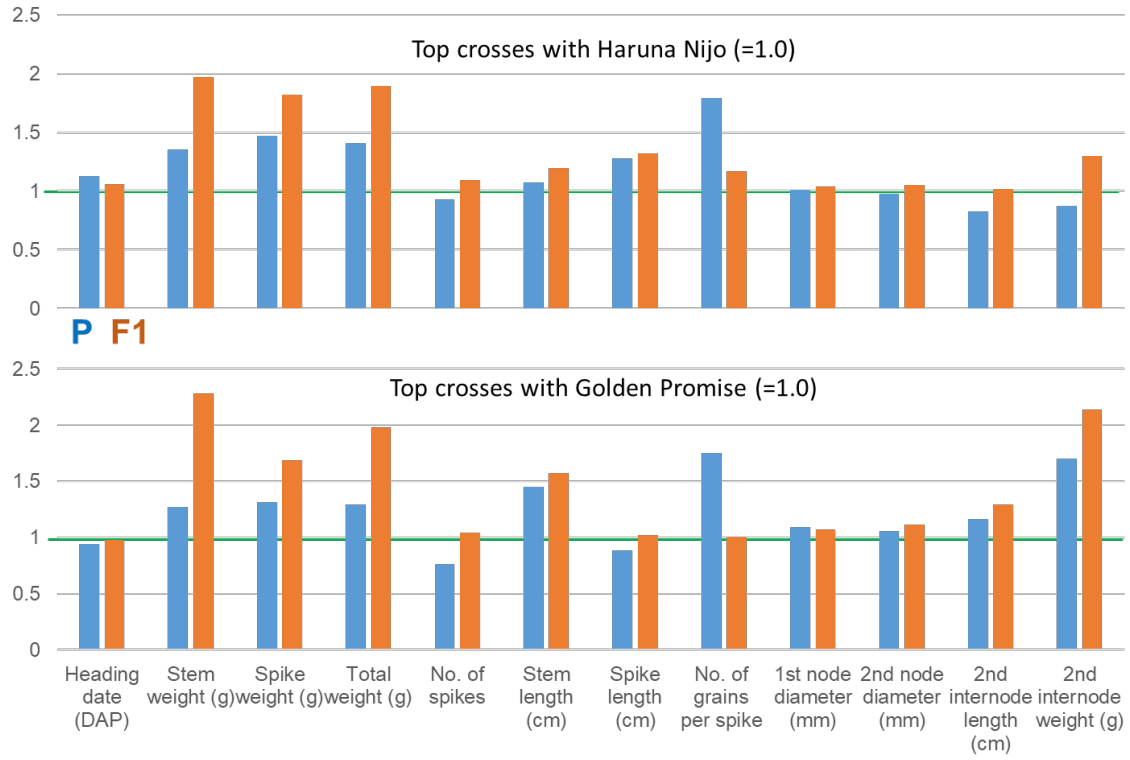


Fig. 2-1. Relative mean performance of Parent (blue) and F1 (orange) from common parent 'Haruna Nijo' and 'Golden Promise'.

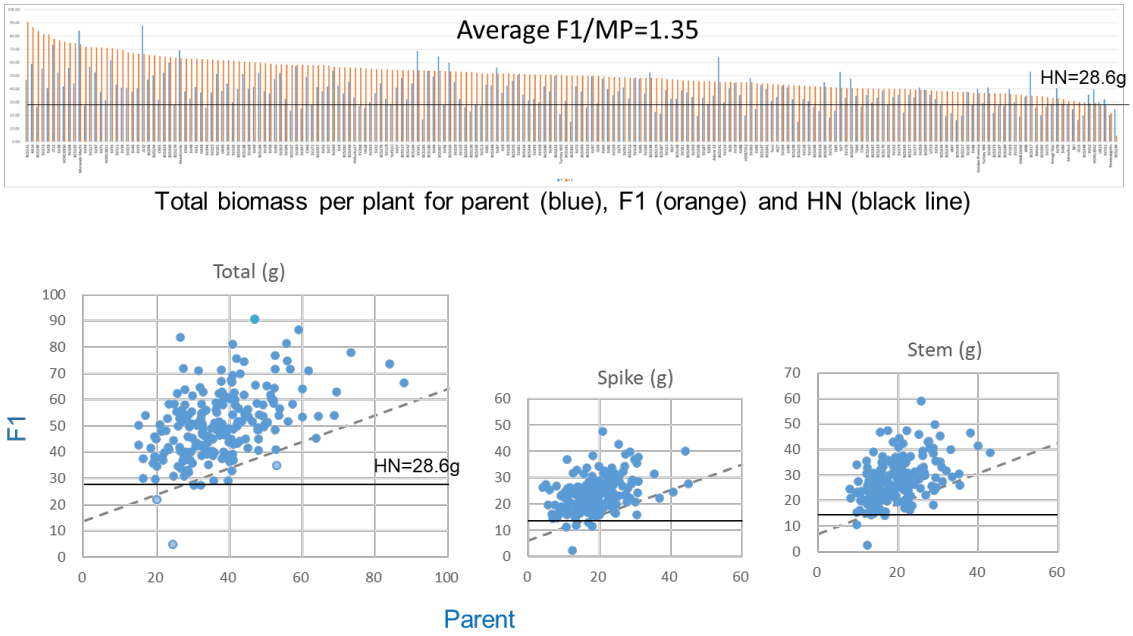


Fig. 2-2A. Total, stem and spike biomass per plant (g) for parent (x-axis), F1 (y-axis), mid-parent (dotted gray line) and HN (black line)

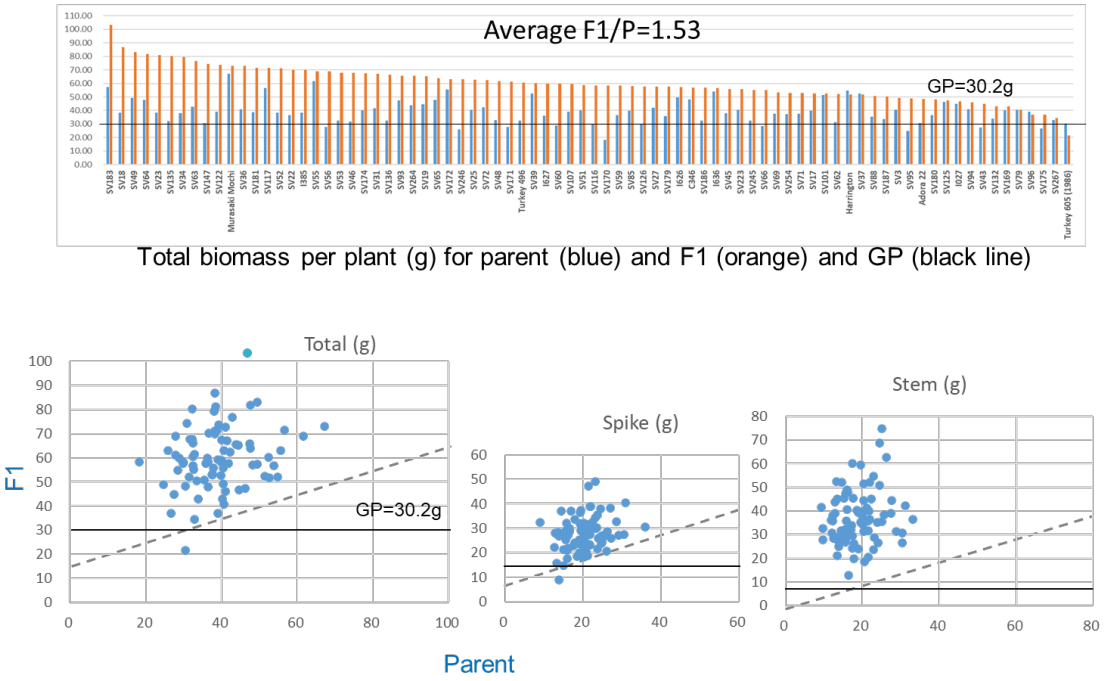


Fig. 2-2B. Total, stem and spike biomass per plant (g) for parent (x-axis), F1 (y-axis), mid-parent (dotted gray line) and GP (black line).

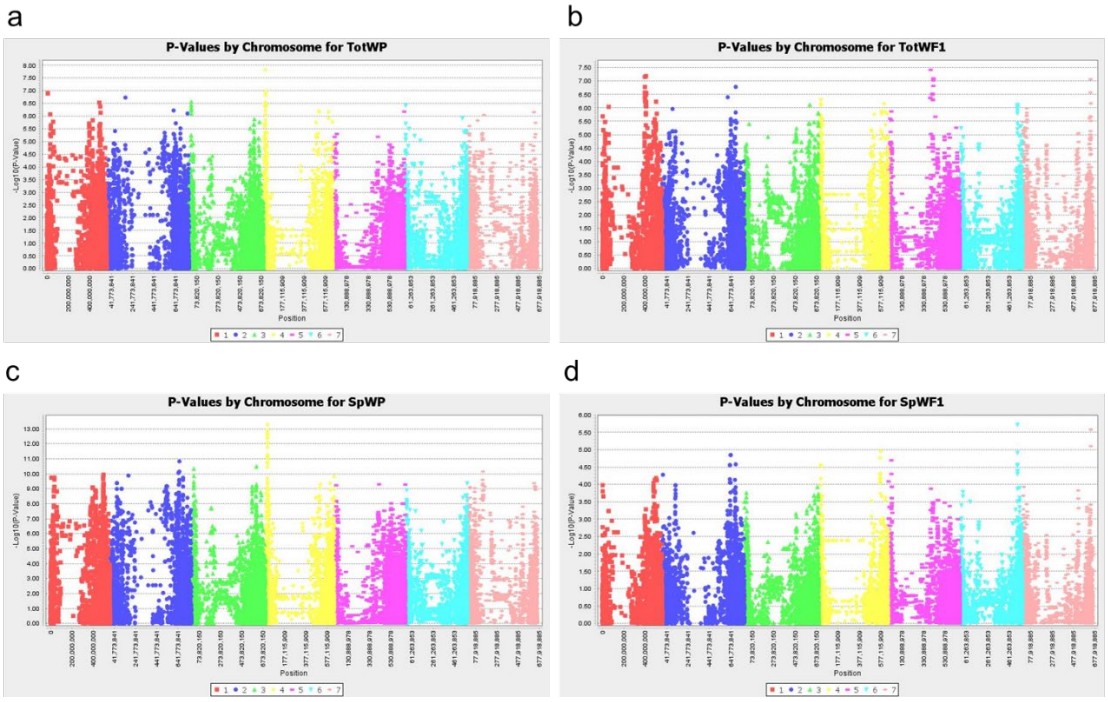


Fig. 2-3. Manhattan plots by GLM procedure for HN (Q=8) top crosses with iSelect 50K markers on reference MorexV1. Significant threshold with Bonferroni correction 5.68 is applied for total weight P (a) and F1 (b) and spike weight P (c) and F1 (d).

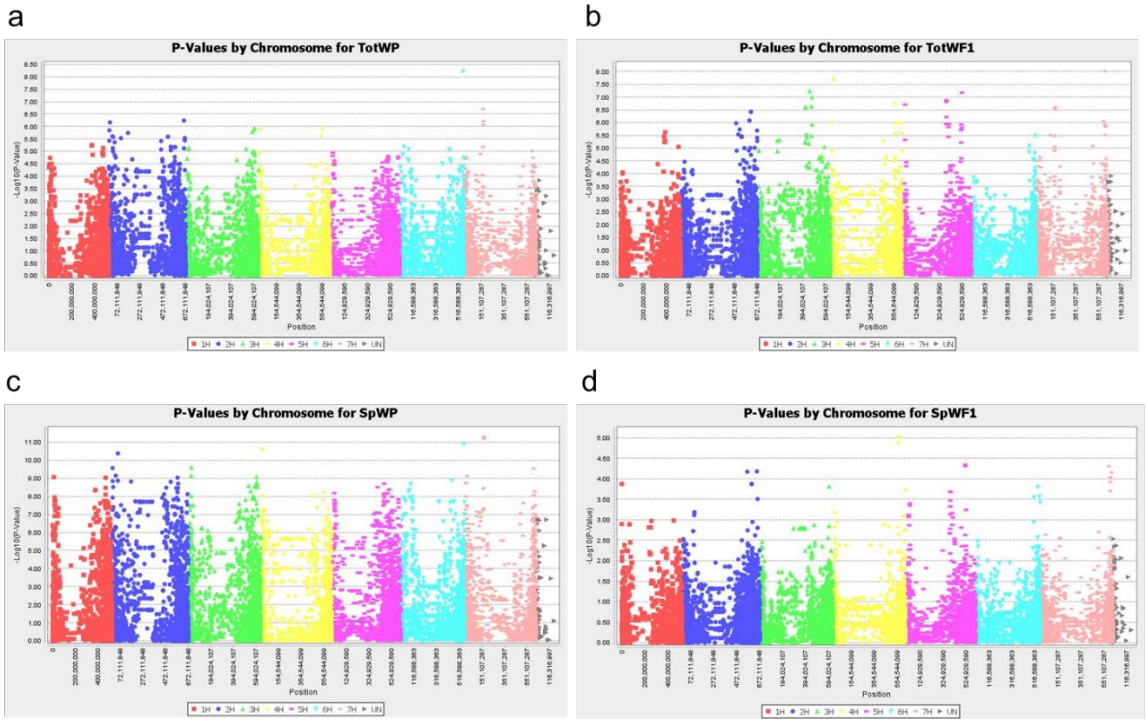


Fig. 2-4. Manhattan plots by GLM procedure for HN (Q=8) top crosses with GBS markers on reference HN. Significant threshold with Bonferroni correction 5.42 is applied for total weight P (a) and F1 (b) and spike weight P (c) and F1 (d).

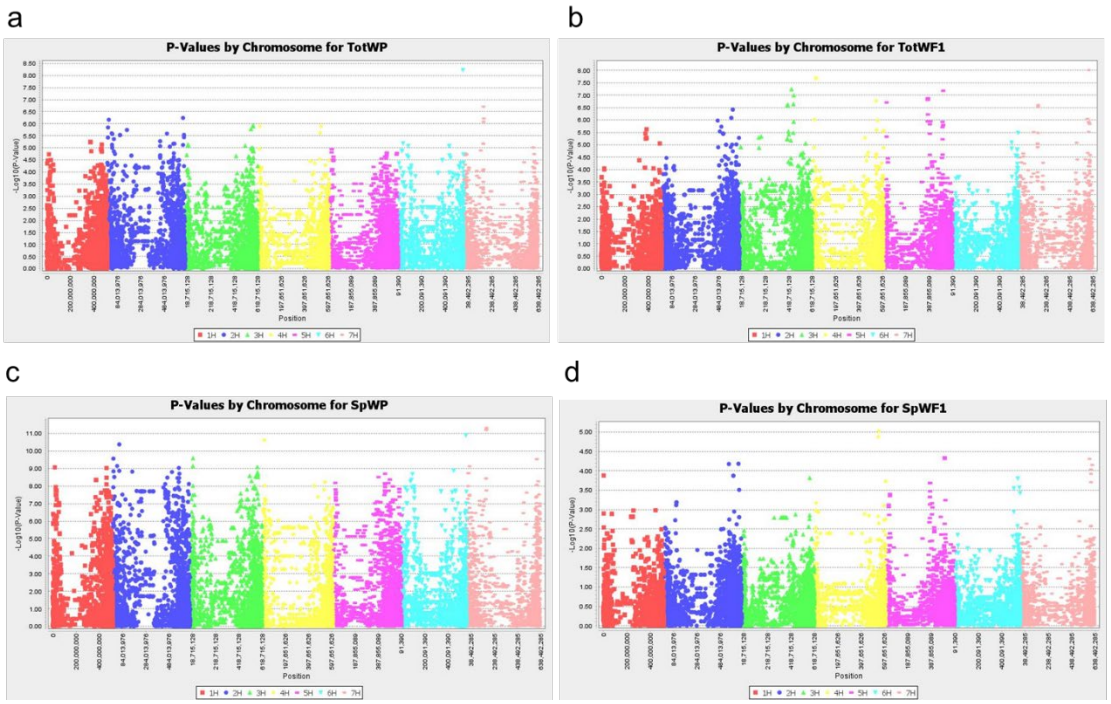


Fig. 2-5. Manhattan plots by GLM procedure for HN (Q=8) top crosses with GBS markers on reference MorexV3. Significant threshold with Bonferroni correction 5.41 is applied for total weight P (a) and F1 (b) and spike weight P (c) and F1 (d).

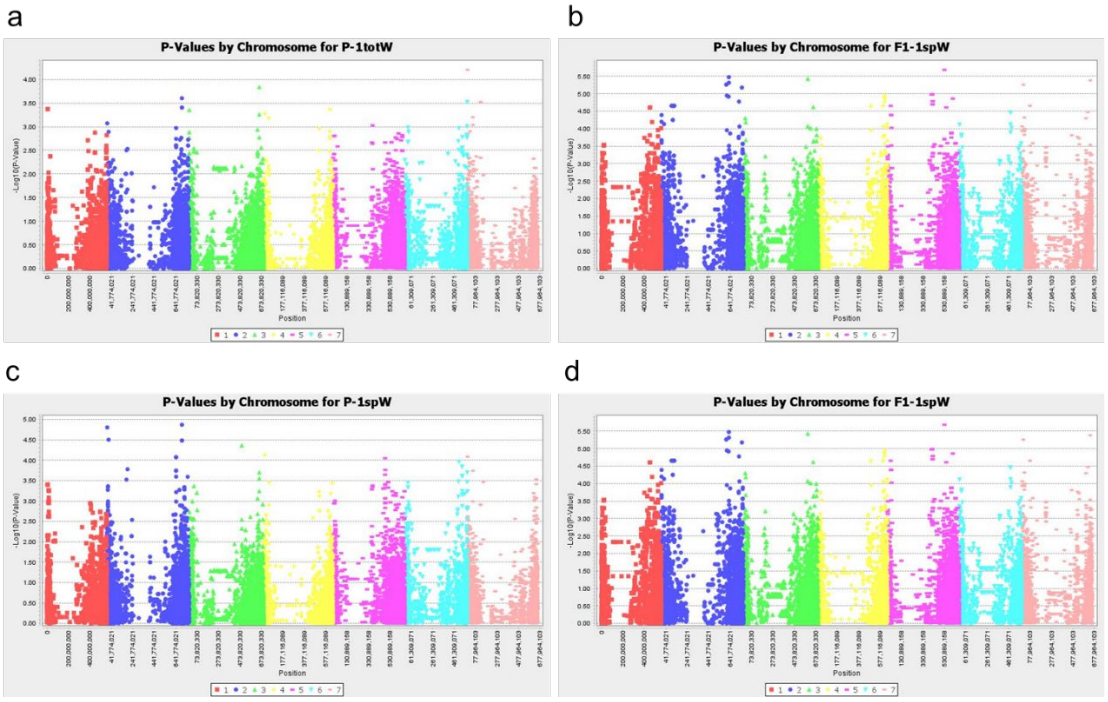


Fig. 2-6. Manhattan plots by GLM procedure for GP rep1 (Q=5) top crosses with iSelect 50K markers on reference MorexV1. Significant threshold with Bonferroni correction 5.77 is applied for total weight P (a) and F1 (b) and spike weight P (c) and F1 (d).

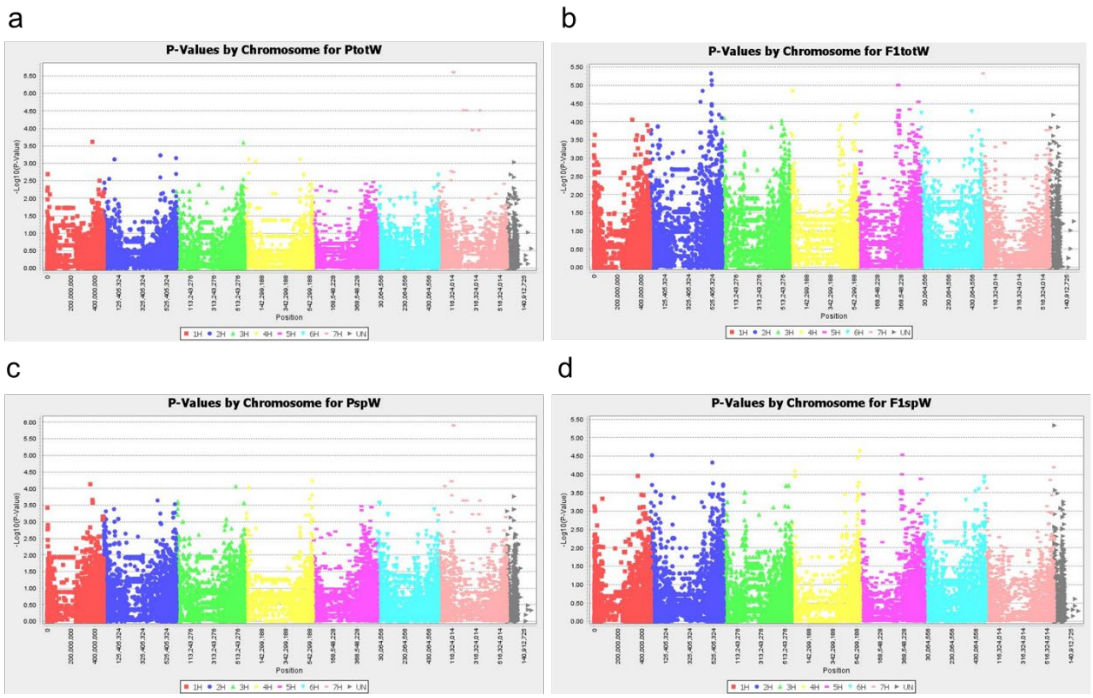


Fig. 2-7. Manhattan plots by GLM procedure for GP (Q=5) top crosses with GBS markers on reference GP. Significant threshold with Bonferroni correction 5.78 is applied for total weight P (a) and F1 (b) and spike weight P (c) and F1 (d).

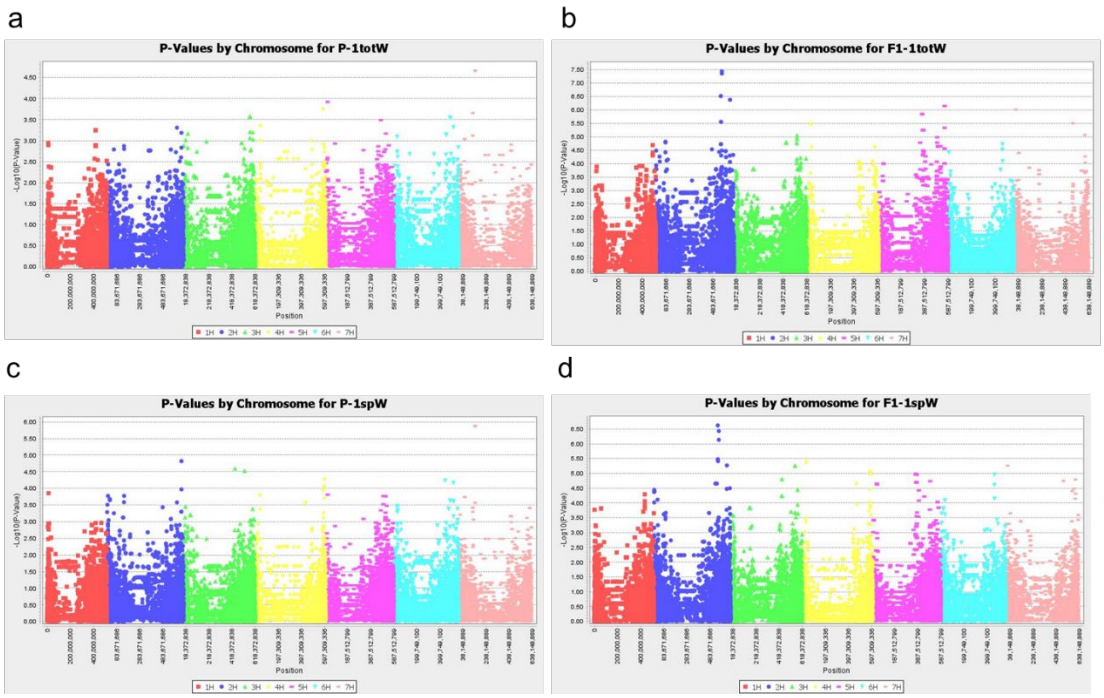
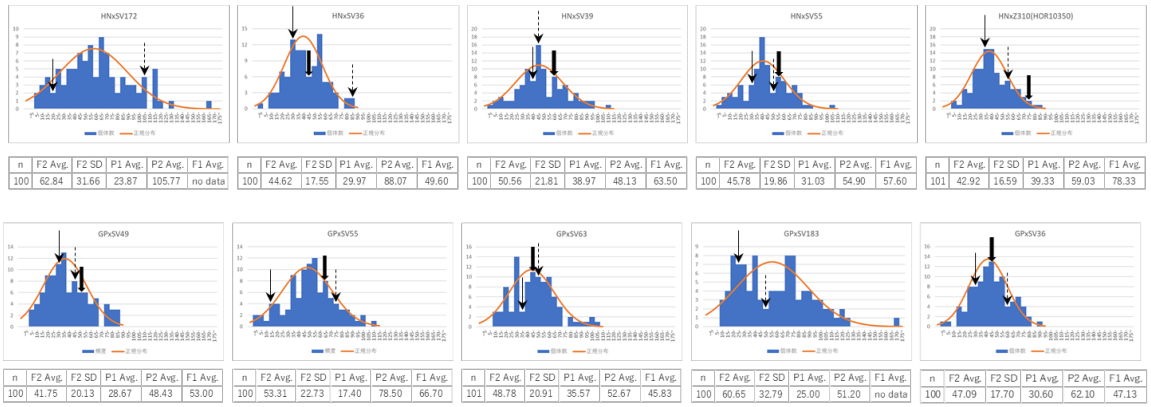


Fig. 2-8. Manhattan plots by GLM procedure for GP rep1 (Q=5) top crosses with GBS markers on reference MorexV3. Significant threshold with Bonferroni correction 5.78 is applied for total weight P (a) and F1 (b) and spike weight P (c) and F1 (d).



n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
100	62.84	31.66	23.87	105.77	no data

n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
100	44.62	17.55	29.97	88.07	49.60

n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
100	50.56	21.81	38.97	48.13	63.50

n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
100	45.78	19.86	31.03	54.90	57.60

n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
101	42.92	16.59	39.33	59.03	78.33

n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
100	41.75	20.13	28.67	48.43	53.00

n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
100	53.31	22.73	17.40	78.50	66.70

n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
101	48.78	20.91	35.57	52.67	45.83

n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
100	60.65	32.79	25.00	51.20	no data

n	F2 Avg.	F2 SD	P1 Avg.	P2 Avg.	F1 Avg.
100	47.09	17.70	30.60	62.10	47.13

Fig. 2-9. Frequency distribution of total weight of F2 plants derived from the crosses with ‘Haruna Nijo’ (upper) and ‘Golden Promise’ (lower).

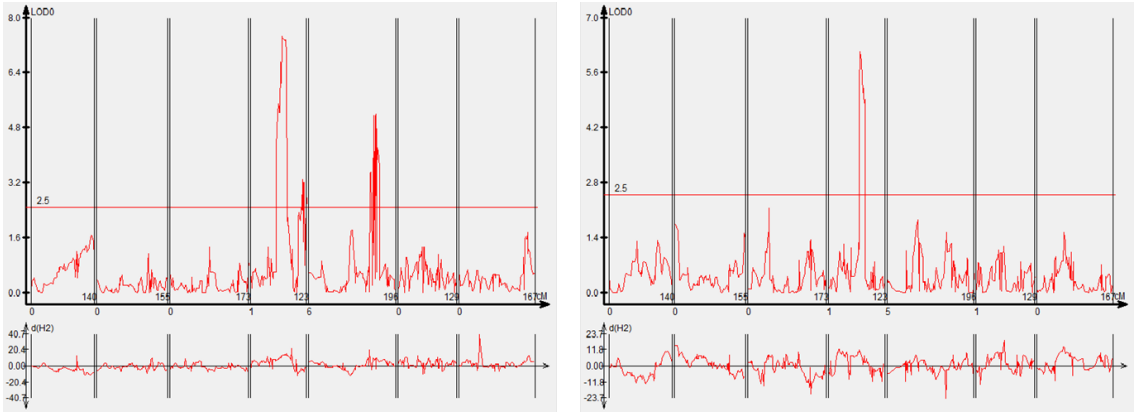


Fig. 2-10. QTL scan for total weight for F2 derived from the cross HN×SV172 (left) and GP×SV183 (right) by iSelect 50K on MorexV1.

General Discussion

Two chapters of this study were coordinated to identify genetic factors related to heterosis which intend to contribute to the development of hybrid barley cultivars. The genetic basis of heterosis is important to further improve the amount of heterosis with the basis of molecular/genomic information.

Genome assembly of 'Haruna Nijo'

After the release of multiple genome assembly in barley as pan-genome analysis (Jayakodi et al. 2020), large structural variations and differences of gene models in barley accessions especially different collected regions were found. Since barley reference assembly has been developed on cv. 'Morex' (Masher et al. 2017, Monat et al. 2019, Mascher et al. 2021), results of genetic analysis may be influenced according to the genome sequence used for the analysis. It is timely that the assembly was published for cv. 'Golden Promise' as a part of pan-genome analysis (Jayakodi et al. 2020) and the methodology was available by the collaboration with German groups who develops pipelines used for pan-genome study for the assembly of cv. 'Haruna Nijo'. The quality of assembly was almost similar to those in pan-genome analysis. It was also confirmed that the large structural variations from Morex assembly was not found. The annotation of full-length cDNA sequences from 'Haruna Nijo' (Sato et al. 2009b, Matsumoto et al. 2012) on the assembly of 'Haruna Nijo' (Sakkour et al. 2022) may give important information for the sequences of genes on this important Japanese malting barley standard.

Sources of germplasm for hybrid barley

The parental genotypes of hybrid barley cultivars may need conflicting genetic components. They both have to have acceptable agronomic traits for cultivation and product uses, however, they may be genetically different in general to achieve the higher level of heterosis. For these reasons, the sources of parents of hybrid barley cultivars could be restricted to the genepool of same cultivar background. The current study aims to expand the crossing parents even to wild or different brittle rachis genotype (Pourkheirandish et al. 2015) which have not been explored due to missing grains by grains after maturity of hybrid plants.

It may be safe to say that the sources of parents for hybrid barley with cvs. 'Harun Nijo' and 'Golden Promise' are abundant if we do not have to consider the shattering or major cultivar types of row-type.

Genomics assisted development of parents for hybrid barley

It was suggested that the factors for controlling heterosis in biomass are mostly controlled by minor effect alleles in general. It is not easy to collect these loci into the parental genome, however, once the collection is made into the common parent, i.e., 'Haruna Nijo', it is an appropriate parent to be crossed with 'Haruna Nijo', since the hybrid plant may show similar phenotype with 'Haruna Nijo' with higher biomass. The genome sequence information may accelerate the collection of effective loci for heterosis for DNA marker generation to backcross alien alleles to 'Haruna Nijo'. There are also effective loci found in F2 genetic analysis from the promising cross combination with 'Haruna Nijo' and 'Golden Promise'. In case of usual map-based cloning, we select heterozygote for the target genomic region to make a large mapping population to narrow down the region of locus. The detection of locus may depend on the heritability of traits. However, the biomass of single plant is influenced by environment and may be difficult to identify locus just by biomass only. We will try to find alternative trait to select biomass which will be less influenced by environmental conditions.

Conclusion

Barley is a winter growing plant in Kurashiki and cannot plant in the summer to evaluate traits related to heterosis. Within limited time and generations, this study identified promising amount of heterosis in biomass in barley germplasm which is almost comparable with the historical study of Yasuda et al. (1994) which was conducted in the same field condition. Hybrid barley is still attracting the interests of breeding companies and may become popular if reasonable amount of heterosis and seed price are available. The key factor is how to identify and collect molecular information responsible for the heterosis in barley. As some of the genes have been identified in other crops, there may be possible ways to achieve higher amount of heterosis by using responsible genes. The other topics related decreasing seed price will be solved at the same time.

Acknowledgments

Barley seeds used in this study were supplied by the National BioResource Project-Barley, Japan (www.nbrp.jp). I would like to thank the lab members of the Group of Genome Diversity for their assistance in carrying out this study. In particular, I would like to thank Mrs. Nami Yamaji for her support not only in research but also in my life at Kurashiki.

References

Ahokas H (1982) Cytoplasmic male sterility in barley: Evidence for the involvement of cytokinins in fertility restoration. Proceedings of the National Academy of Sciences 79:7605-7608

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655-1664

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403-410

Bailly-Bechet M, Haudry A, Lerat E (2014) "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. Mobile DNA 5:13

Bayer MM, Rapazote-Flores P, Ganal 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. Frontiers in Plant Science 8:doi=10.3389/fpls.2017.01792

von Bothmer R, Sato K, Komatsuda T, Yasuda S, Fischbeck G (2003) Chapter 2 - The domestication of cultivated barley. In: von Bothmer R, van Hintum T, Knüpffer H, Sato K (eds) Developments in Plant Genetics and Breeding. Elsevier, pp 9-27

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

Bruce AB (1910) The Mendelian theory of heredity and the augmentation of vigor. Science 32:627-628

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nature Reviews Genetics 10:783-796

Clavijo BJ, Garcia Accinelli G, Wright J, Heavens D, Barr K, Yanes L, Di-Palma F (2017) W2RAP: a pipeline for high quality, robust assemblies of large complex genomes from short read data. bioRxiv:110999

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, Szűcs 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 Genomics 10:582

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2012) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21

Dvorak J, McGuire PE, Cassidy B (1988) Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome 30:680-689

East EM (1936) Heterosis. Genetics 21:375-397

Friedt W, Ordon F (2013) Barley Production and Breeding in Europe: Modern Cultivars Combine Disease Resistance, Malting Quality and High Yield. Springer Netherlands, Dordrecht, pp 389-400

Fischbeck G (2003) Chapter 3 - Diversification through breeding. In: von Bothmer R, van Hintum T, Knüpffer H, Sato K (eds) Developments in Plant Genetics and Breeding. Elsevier, pp 29-52

Franckowiak JD, Lundqvist U, Konishi T (1997) New and revised names for barley genes. Barley Genetics Newsletter 26:4-8.

Garcia AAF, Wang S, Melchinger AE, Zeng Z-B (2008) Quantitative Trait Loci Mapping and The Genetic Basis of Heterosis in Maize and Rice. Genetics 180:1707-1724

Ghosh S, Chan C-KK (2016) Analysis of RNA-Seq Data Using TopHat and Cufflinks. In: Edwards D (ed) Plant Bioinformatics: Methods and Protocols. Springer New York, New York, NY, pp 339-361

Gong X, Westcott S, Zhang X-Q, Yan G, Lance R, Zhang G, Sun D, Li C (2013) Discovery of Novel Bmy1 Alleles Increasing β -Amylase Activity in Chinese Landraces and Tibetan Wild Barley for Improvement of Malting Quality via MAS. PLOS ONE 8:e72875

Gremme G, Brendel V, Sparks ME, Kurtz S (2005) Engineering a software tool for gene structure prediction in higher organisms. Information and Software Technology 47:965-978

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL, White O (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Research 31:5654-5666

Guo M, Rupe MA, Wei J, Winkler C, Goncalves-Butruille M, Weers BP, Cerwick SF, Dieter JA, Duncan KE, Howard RJ, Hou Z, Löffler CM, Cooper M, Simmons CR (2013) Maize ARGOS1 (ZAR1) transgenic alleles increase hybrid maize yield. Journal of Experimental Botany 65:249-260

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biology 9:R7

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL, White O (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Research 31:5654-5666

Hashimoto S, Wake T, Nakamura H, Minamiyama M, Araki-Nakamura S, Ohmae-Shinohara K, Koketsu E, Okamura S, Miura K, Kawaguchi H, Kasuga S, Sazuka T (2021) The dominance model for heterosis explains culm length genetics in a hybrid sorghum variety. Scientific Reports 11:4532

Hisano H, Tsujimura M, Yoshida H, Terachi T, Sato K (2016) Mitochondrial genome sequences from wild and cultivated barley (Hordeum vulgare). BMC Genomics 17:824

Hochholdinger F, Hoecker N (2007) Towards the molecular basis of heterosis. Trends in Plant Science 12:427-432

Hoff KJ, Stanke M (2019) Predicting Genes in Single Genomes with AUGUSTUS. Current Protocols in Bioinformatics 65:e57

Huang X, Yang S, Gong J, Zhao Q, Feng Q, Zhan Q, Zhao Y, Li W, Cheng B, Xia J, Chen N, Huang T, Zhang L, Fan D, Chen J, Zhou C, Lu Y, Weng Q, Han B (2016) Genomic architecture of heterosis for yield traits in rice. Nature 537:629-633

Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, Li W, Zhan Q, Cheng B, Xia J, Chen N, Hao Z, Liu K, Zhu C, Huang T, Zhao Q, Zhang L, Fan D, Zhou C, Lu Y, Weng Q, Wang Z-X, Li J, Han B (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nature Communications 6:6258

International Barley Sequencing Consortium (IBSC) (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711-716

Jayakodi M, Padmarasu S, Haberer G, Bonthala VS, Gundlach H, Monat C, Lux T, Kamal N, Lang D, Himmelbach A, Ens J, Zhang X-Q, Angessa TT, Zhou G, Tan C, Hill C, Wang P, Schreiber M, Boston LB, Plott C, Jenkins J, Guo Y, Fiebig A, Budak H, Xu D, Zhang J, Wang C, Grimwood J, Schmutz J, Guo G, Zhang G, Mochida K, Hirayama T, Sato K, Chalmers KJ, Langridge P, Waugh R, Pozniak CJ, Scholz U, Mayer KFX, Spannagl M, Li C, Mascher M, Stein N (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. Nature 588:284-289

Jayakodi M, Schreiber M, Stein N, Mascher M (2021) Building pan-genome infrastructures for crop plants and their use in association genetics. DNA Research 28: dsaa030

Jones DF (1917) Dominance of Linked Factors as a Means of Accounting for Heterosis. Proceedings of the National Academy of Sciences 3:310-312

Kleinhofs A, Kilian A, Saghai Maroof MA, Biyashev RM, Hayes P, Chen FQ, Lapitan N, Fenwick A, Blake TK, Kanazin V, Ananiev E, Dahleen L, Kudrna D, Bollinger J, Knapp SJ, Liu B, Sorrells M, Heun M, Franckowiak JD, Hoffman D, Skadsen R, Steffenson BJ (1993) A molecular, isozyme and morphological map of the barley (Hordeum vulgare) genome. Theoretical and Applied Genetics 86:705-712

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. Proceedings of the National Academy of Sciences 104:1424-1429

König S, Romoth LW, Gerischer L, Stanke M (2016) Simultaneous gene finding in multiple genomes. Bioinformatics 32:3388-3395

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M (2019) Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biology 20:278

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100

Lundqvist U, Franckowiak JD (2003) Chapter 5 - Diversity of barley mutants. In: von Bothmer R, van Hintum T, Knüpffer H, Sato K (eds) Developments in Plant Genetics and Breeding. Elsevier, pp 77-96

Ma G-h, Yuan L-p (2015) Hybrid rice achievements, development and prospect in China. Journal of Integrative Agriculture 14:197-205

Mapleson D, Venturini L, Kaithakottil G, Swarbreck D (2018) Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. GigaScience 7: giy131

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 X-Q, Zhang Q, Barrero RA, Li L, Taudien S, Groth M, Felder M, Hastie A, Šimková H, Staňková H, Vrána J, Chan S, Muñoz-Amatriaín 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, Doležel 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

Mascher M, Wicker T, Jenkins J, Plott C, Lux T, Koh CS, Ens J, Gundlach H, Boston LB, Tulpová Z, Holden S, Hernández-Pinzón I, Scholz U, Mayer KFX, Spannagl M, Pozniak CJ, Sharpe AG, Šimková H, Moscou MJ, Grimwood J, Schmutz J, Stein N (2021) Long-read sequence assembly: a technical evaluation in barley. The Plant Cell 33:1888-1906

Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K, Munoz-Amatriain M, Close TJ, Wise RP, Schulman AH, Himmelbach A, Mayer KFX, Scholz U, Poland JA, Stein N, Waugh R (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant Journal 76:718-727

Matsumoto T, Tanaka T, Sakai H, Amano N, Kanamori H, Kurita K, Kikuta A, Kamiya K, Yamamoto M, Ikawa H, Fujii N, Hori K, Itoh T, Sato K (2011) Comprehensive Sequence Analysis of 24,783 Barley Full-Length cDNAs Derived from 12 Clone Libraries. Plant Physiology 156:20-28

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols doi:10.1101/pdb.prot5448

Milner SG, Jost M, Taketa S, Mazón ER, Himmelbach A, Oppermann M, Weise S, Knüpffer H, Basterrechea M, König P, Schüler D, Sharma R, Pasam RK, Rutten T, Guo G, Xu D, Zhang J, Herren G, Müller T, Krattinger SG, Keller B, Jiang Y, González MY, Zhao Y, Habekuß A, Färber S, Ordon F, Lange M, Börner A, Graner A, Reif JC, Scholz U, Mascher

M, Stein N (2019) Genebank genomics highlights the diversity of a global barley collection. Nature Genetics 51:319-326

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 Biology 20:284

Padmarasu S, Himmelbach A, Mascher M, Stein N (2019) In Situ Hi-C for Plants: An Improved Method to Detect Long-Range Chromatin Interactions. In: Chekanova JA, Wang H-LV (eds) Plant Long Non-Coding RNAs: Methods and Protocols. Springer New York, New York, NY, pp 441-472

Pourkheirandish M, Hensel G, Kilian B, Senthil N, Chen G, Sameri M, Azhaguvel P, Sakuma S, Dhanagond S, Sharma R, Mascher M, Himmelbach A, Gottwald S, Nair Sudha K, Tagiri A, Yukuhiro F, Nagamura Y, Kanamori H, Matsumoto T, Willcox G, Middleton Christopher P, Wicker T, Walther A, Waugh R, Fincher Geoffrey B, Stein N, Kumlehn J, Sato K, Komatsuda T (2015) Evolution of the Grain Dispersal System in Barley. Cell 162:527-539

Ramage RT (1965) Balanced Tertiary Trisomics for Use in Hybrid Seed Production1. Crop Science 5:177-178

Ramage RT (1975) Techniques for producing hybrid barley. Barley Genetics Newsletter 18:62-65.

Ramage RT (1983) Heterosis and Hybrid Seed Production in Barley. In: Frankel R (ed) Heterosis: Reappraisal of Theory and Practice. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 71-93

Saisho D, Myoraku E, Kawasaki S, Sato K, Takeda K (2007) Construction and Characterization of a Bacterial Artificial Chromosome (BAC) Library from the Japanese Malting Barley variety 'Haruna Nijo'. Breeding Science 57:29-38

Sakkour A, Mascher M, Himmelbach A, Haberer G, Lux T, Spannagl M, Stein N, Kawamoto S, Sato K (2022) Chromosome-scale assembly of barley cv. 'Haruna Nijo' as a resource for barley genetics. DNA Research 29: dsac001

Sato K (2020) History and future perspectives of barley genomics. DNA Research 27: dsaa023

Sato K, Kira K, Ochi H, Narita H. (1989) Diallel Analysis of Husk Content in Malting Barley. Japanese Journal of Breeding 39:471-480

Sato K, Mascher M, Himmelbach A, Haberer G, Spannagl M, Stein N (2021) Chromosome-scale assembly of wild barley accession "OUH602". G3 Genes|Genomes|Genetics 11:jkab244

Sato K, Nankaku N, Takeda K (2009a) A high-density transcript linkage map of barley derived from a single population. Heredity 103:110-117

Sato K, Hisano H, Matsumoto S, Zhou T-S, Kihara M (2018) Detection of QTLs controlling alpha-amylase activity in a diversity panel of 343 barley accessions. Molecular Breeding 38:14

Sato K, Ishii M, Takahagi K, Inoue K, Shimizu M, Uehara-Yamaguchi Y, Nishii R, Mochida K (2020) Genetic Factors Associated with Heading Responses Revealed by Field Evaluation of 274 Barley Accessions for 20 Seasons. iScience 23:101146

Sato K, Shin-I T, Seki M, Shinozaki K, Yoshida H, Takeda K, Yamazaki Y, Conte M, Kohara Y (2009b) Development of 5006 Full-Length CDNAs in Barley: A Tool for Accessing Cereal Genomics Resources. DNA Research 16:81-89

Sato K, Takeda K (2009) An application of high-throughput SNP genotyping for barley genome mapping and characterization of recombinant chromosome substitution lines. Theoretical and Applied Genetics 119:613-619

Sato K, Tanaka T, Shigenobu S, Motoi Y, Wu J, Itoh T (2016a) Improvement of barley genome annotations by deciphering the Haruna Nijo genome. DNA Research 23:21-28

Sato K, Yamane M, Yamaji N, Kanamori H, Tagiri A, Schwerdt JG, Fincher GB, Matsumoto T, Takeda K, Komatsuda T (2016b) Alanine aminotransferase controls seed dormancy in barley. Nature Communications 7:11625

Schnable PS, Springer NM (2013) Progress toward understanding heterosis in crop plants. Annual Review of Plant Biology 64:71-88

Schnell FW, Cockerham CC (1992) Multiplicative vs. arbitrary gene action in heterosis. Genetics 131:461-469

Schreiber M, Mascher M, Wright J, Padmarasu S, Himmelbach A, Heavens D, Milne L, Clavijo BJ, Stein N, Waugh R (2020) A genome assembly of the barley 'transformation reference' cultivar Golden Promise. G3 Genes|Genomes|Genetics 10:1823-1827

Schwarzwälder L, Thorwarth P, Zhao Y, Reif JC, Longin CFH (2022) Hybrid wheat: quantitative genetic parameters and heterosis for quality and rheological traits as well as baking volume. Theoretical and Applied Genetics 135:1131-1141

Seko H. (1987) History of barley breeding in Japan. In Barley Genetics V. Proc. 5th Barley Genetics Symposium, Okayama Japan, pp. 915-922.

Shull GH (1908) The composition of a field of maize. Journal of Heredity 4:296-301

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212

Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31

Smit A, Hubley R, Green P 2013–2015 RepeatMasker Open-4.0. http://www.repeatmasker.org/faq.html.

Stadler L (1928) Mutations in barley induced by X-rays and radium. Science 68: 186-187

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

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. Proceedings of the National Academy of Sciences 105:4062-4067

Taketa S, Linde-Laursen I, Künzel G (2003) Chapter 6 - Cytogenetic diversity. In: von Bothmer R, van Hintum T, Knüpffer H, Sato K (eds) Developments in Plant Genetics and Breeding. Elsevier, pp 97-119

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Research 18:1979-1990

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM (2018) BUSCO: applications from quality assessments to gene prediction and phylogenomics. Molecular Biology and Evolution 35:543-548

Wendler N, Mascher M, Nöh C, Himmelbach A, Scholz U, Ruge-Wehling B, Stein N (2014) Unlocking the secondary gene-pool of barley with next-generation sequencing. Plant Biotechnology Journal 12:1122-1131

Wicker T, Matthews DE, Keller B (2002) TREP: a database for Triticeae repetitive elements. Trends in Plant Science 7:561-562

Wu D, Yamaji N, Yamane M, Kashino-Fujii M, Sato K, Feng Ma J (2016) The HvNramp5 transporter mediates uptake of cadmium and manganese, but not iron. Plant Physiology 172:1899-1910

Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859-1875

Yasuda S, Hayashi J, Moriya I (1994) Studies on heterosis in barley cultivars. Bulletin Research Institute of Bioresources, Okayama University. 2:7-22.

Zhang X, Lv L, Lv C, Guo B, Xu R (2015) Combining Ability of Different Agronomic Traits and Yield Components in Hybrid Barley. PLOS ONE 10:e0126828

Zohary D, Hoph M, and Weiss E (2012) Domestication of plants in the Old World (4th edition), Oxford University Press, New York, pp. 20-74.